

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería
Ingeniería de Sistemas y Computación
Proyecto de Grado

Desarrollo de un Modelo de Inteligencia Artificial para la Predicción de Demanda en Productos de Alta Rotación en PYMEs de Ibagué

William Andres Nova Rojas

Director
MSc. Mario Julián Mora Cardona

Febrero, 2026



Pontificia Universidad
JAVERIANA
Cali

Abstract

Retail small and medium-sized enterprises (SMEs) often replenish inventory using empirical criteria, which increases the likelihood of **stockouts** (inventory shortages and lost sales) and **overstock** (tied-up capital, higher holding costs, and shrinkage). In this context, this undergraduate thesis develops and evaluates an artificial intelligence model for **daily demand forecasting** of **high-rotation products** in two retail stores located in Ibagué (Mercaya and Villa Café), aiming to provide a quantitatively evaluated tool applicable to operations to support inventory management.

Methodologically, a reproducible process was implemented to transform operational data into a consistent *product-day* structure, perform cleaning and standardization, and incorporate temporal signal through *lags* and rolling aggregates. In addition, calendar variables (day, payday period, month, and working-day indicator) were integrated and, when relevant, weather context was included. Linear regression and tree-based ensemble models (Random Forest, XGBoost, and LightGBM) were trained and compared under a **temporal validation** scheme for a one-day-ahead forecast horizon ($h = 1$). Performance was assessed using operationally interpretable error metrics (MAE and WAPE %) and complementary measures of fit (R^2 and R^2_{adj}) and large-error penalization (MSE), as well as training time.

Results show that the best performance is achieved with **compact feature sets** (approximately 9–12 base variables), indicating that adding more variables does not necessarily improve generalization. In both stores, the dominant signal comes from **recent temporal memory** and **calendar effects**; in Villa Café, the categorical weather description provides relevant contextual signal. Random Forest presented the best overall balance, reducing error compared to training with all variables: in Mercaya, MAE decreased from 1075.68 to 933.04 (WAPE % from 35.85 to 31.10), and in Villa Café from 1355.22 to 1099.53 (WAPE % from 43.41 to 35.22), along with improvements in R^2 and lower feature-set complexity. Overall, the thesis delivers a sustainable and replicable procedure to support purchasing decisions, reduce stockouts and overstock, and standardize replenishment through an evidence-based continuous improvement cycle.

Keywords: Demand forecasting, SMEs, High-rotation products, Machine learning, Time series, Feature selection, Temporal validation, Operational efficiency.

Resumen

Las pequeñas y medianas empresas (PYMEs) minoristas suelen reponer inventario con criterios empíricos, lo que aumenta la probabilidad de **desabastecimiento** (ruptura de inventario y ventas perdidas) y de **sobreinventario** (capital inmovilizado, mayores costos de almacenamiento y merma). En este contexto, el presente trabajo de grado desarrolla y evalúa un modelo de inteligencia artificial para la **predicción de demanda diaria de productos de alta rotación** en dos puntos de venta minoristas ubicados en Ibagué (Mercaya y Villa Café), con el propósito de aportar una herramienta evaluada cuantitativamente y aplicable a la operación para apoyar la gestión de inventarios.

Metodológicamente, se construyó un proceso reproducible que transforma datos operativos en una estructura consistente *producto-día*, ejecuta limpieza y estandarización, e incorpora señal temporal mediante *lags* y agregaciones móviles (*rolling*). Adicionalmente, se integran variables de calendario (día, quincena, mes y condición de día hábil) y, cuando es pertinente, contexto climático. Se entrenaron y compararon modelos de regresión lineal y ensambles basados en árboles (Random Forest, XG-Boost y LightGBM) bajo un esquema de **validación temporal** para un horizonte de predicción de un día ($h = 1$). El desempeño se evaluó con métricas orientadas a una lectura operativa del error (MAE y WAPE %) y con medidas complementarias de ajuste (R^2 y R^2_{adj}) y de penalización de errores grandes (MSE), además del tiempo de entrenamiento.

Los resultados muestran que el mejor desempeño se alcanza con **conjuntos compactos de variables** (aprox. 9–12 variables base), lo que evidencia que aumentar el número de variables no garantiza una mejor generalización. En ambos negocios, la señal dominante proviene de la **memoria temporal reciente** y del **calendario**; en Villa Café, además, la descripción categórica del clima aporta señal contextual relevante. Random Forest presentó el mejor equilibrio global, reduciendo el error frente al entrenamiento con todas las variables: en Mercaya, el MAE disminuyó de 1075.68 a 933.04 (WAPE % de 35.85 a 31.10), y en Villa Café de 1355.22 a 1099.53 (WAPE % de 43.41 a 35.22), junto con mejoras en R^2 y una menor complejidad del conjunto de variables. En conjunto, el trabajo entrega un procedimiento sostenible y replicable para apoyar decisiones de compra, reducir el desabastecimiento y el sobreinventario, y estandarizar la reposición mediante un ciclo de mejora continua basado en evidencia.

Palabras clave: Predicción de demanda, PYMEs, Productos de alta rotación, Machine learning, Series temporales, Selección de características, Validación temporal, Eficiencia operativa.

Índice general

1. Introducción	16
2. Descripción del Problema	17
2.1. Planteamiento del Problema	17
2.1.1. Formulación	18
2.1.2. Sistematización	18
2.2. Objetivos	19
2.2.1. Objetivo General	19
2.2.2. Objetivos Específicos	19
2.3. Justificación	19
2.4. Delimitaciones y Alcances	20
2.4.1. Alcances	20
2.4.2. Limitaciones	21
3. Marco Teórico y Trabajos Relacionados	22
3.1. Marco Teórico	22
3.1.1. Marco de referencia	22
3.1.2. Áreas temáticas	23
3.1.3. Términos y conceptos	23
3.1.4. Formulación del problema de pronóstico diario	26

3.1.5.	Modelos supervisados para predicción de demanda	27
3.1.6.	Pipelines de modelado y reproducibilidad	29
3.1.7.	Ingeniería y diagnóstico de variables	29
3.1.8.	Selección de variables y control de redundancias	31
3.1.9.	Optimización de hiperparámetros y búsqueda eficiente	32
3.1.10.	Costo computacional, tiempo de entrenamiento y criterio de selección práctico	34
3.1.11.	Evaluación del desempeño y validación temporal	35
3.1.12.	Trabajos relacionados y diferenciadores del proyecto	37
4.	Adquisición, integración y preprocesamiento de datos	39
4.1.	Adquisición de datos con las empresas	40
4.1.1.	Recolección de datos con dos negocios minoristas en Ibagué . .	40
4.2.	Almacenamiento y estructura de la base de datos	41
4.2.1.	Elección de PostgreSQL y justificación técnica	41
4.2.2.	Descripción funcional del esquema relacional	42
4.3.	Descripción de los conjuntos de datos crudos	44
4.3.1.	Granularidad y unidad de análisis	44
4.3.2.	Periodo de observación y tamaño muestral	45
4.3.3.	Estructura de variables disponibles en el crudo	45
4.3.4.	Consideraciones sobre calidad y particularidades del dato crudo	47
4.4.	Enriquecimiento del dataset: variables temporales, calendario y clima	47
4.4.1.	Librerías empleadas y propósito	47
4.4.2.	Variables agregadas por calendario	48
4.4.3.	Integración de variables meteorológicas: temperatura y estado del tiempo	48
4.4.4.	Justificación del enriquecimiento	49

4.5.	Limpieza y alistamiento de los datos para modelado	50
4.5.1.	Control de faltantes y estabilidad del esquema	50
4.5.2.	Conversión numérica y consistencia contable a nivel ítem . . .	51
4.5.3.	Depuración de transacciones no informativas y no interpretables	51
4.5.4.	Homogeneización de variables categóricas: clima	52
4.5.5.	Definición formal de demanda diaria	52
4.5.6.	Construcción del dataset producto-día y control de integridad	53
4.5.7.	Simplificación del esquema base para modelado	53
4.5.8.	Completado del panel producto $\times$ día	54
4.5.9.	Construcción de variables históricas: <i>lags</i> y <i>rolling</i>	55
4.5.10.	Reducción de redundancias en <i>features</i>	56
4.5.11.	Descomposición de la fecha en componentes calendario	56
4.5.12.	Resultado del alistamiento	57
5.	Análisis Exploratorio y Selección de Características y Productos	59
5.1.	EDA exploratorio orientado a pronóstico: selección de Top-10 y diagnóstico de variables	59
5.1.1.	Selección de productos objetivo: rotación, regularidad e índice compuesto	60
5.1.2.	Diagnóstico temporal agregado de las series completas	67
5.1.3.	Criterio de representatividad: control por familia	70
5.1.4.	Top-10 final por negocio y validación del panel producto-día .	70
5.1.5.	Diagnóstico de variables explicativas: correlaciones como lectura preliminar	71
5.1.6.	MI: dependencia no lineal y relevancia de <i>productoid</i>	72
5.1.7.	Evaluación temporal del aporte de variables exógenas mediante <i>lags</i> ($h=1$) sin fuga temporal	73
5.1.8.	Implicaciones para el pipeline: hipótesis verificables y conjuntos candidatos de variables	73

5.1.9. Salida del EDA: insumos listos para entrenamiento	74
6. Entrenamiento y Evaluación Inicial de los Modelos Predictivos de Demanda	77
6.1. Entrenamiento inicial de modelos con todos los predictores (línea base)	77
6.1.1. Protocolo experimental implementado en el pipeline	78
6.1.2. Predictores incluidos en ORIGINAL_FULL y justificación	79
6.1.3. Modelos comparados y criterio de línea base	79
6.1.4. Métricas globales y tratamiento de MAPE/WAPE	80
6.1.5. Resultados globales en test con ORIGINAL_FULL	80
6.1.6. Inspección visual de predicciones de Villa Café	81
6.1.7. Conclusión de la línea base de Villa Café	88
6.1.8. Inspección visual de predicciones de Mercaya	89
6.1.9. Conclusión de la línea base de Mercaya	95
7. Optimización y Evaluación conjunta: features, hiperparámetros y eficiencia	96
7.1. Pipeline integral de entrenamiento, selección de variables y optimización	96
7.1.1. Propósito y criterios de diseño	96
7.1.2. Validaciones operativas para la búsqueda y la optimización . .	97
7.1.3. Validación cruzada temporal por fecha como unidad indivisible	98
7.1.4. Complejidad efectiva: dimensionalidad posterior al preprocesa- miento	98
7.1.5. Etapa B: selección de variables en dos fases (exploración con- trolada)	99
7.1.6. Etapa C: optimización de modelos en dos fases por conjunto de variables	101
7.1.7. Criterios de selección final: precisión vs. eficiencia	103
7.2. Resultados: Mercaya	104

7.2.1.	Conjunto ganador: variables finales e interpretación	104
7.2.2.	Ganadores por modelo	106
7.2.3.	Alternativas globales cercanas: rapidez y menos variables . . .	107
7.3.	Resultados: Villa Café	114
7.3.1.	Conjunto ganador: variables finales e interpretación	114
7.3.2.	Ganadores por modelo: métricas resumidas	115
7.3.3.	Alternativas globales cercanas: rapidez y menos variables . . .	116
7.4.	Comparación integral entre Mercaya y Villa Café	121
7.4.1.	Comparación global de desempeño	121
7.4.2.	Diferencias en señal dominante (a partir de las variables seleccionadas)	122
7.5.	Efecto de seleccionar conjuntos compactos frente a entrenar con el conjunto completo	123
7.5.1.	Mejor global vs. conjunto completo	123
7.5.2.	Alternativas cercanas vs. conjunto completo base line	124
7.6.	Conclusión integral de resultados	125
8.	Conclusiones	126
9.	Trabajo Futuro	128
9.1.	Extensión del período de observación y construcción de un histórico de mayor cobertura	128
9.2.	Validación prospectiva y evaluación en operación	129
9.2.1.	Incorporación explícita de eventos, intervenciones y cambios estructurales	129
9.2.2.	Pronóstico multi-horizonte alineado con ciclos reales de abastecimiento	130
9.2.3.	De la predicción a la decisión: recomendación de compra integrada con políticas de inventario	130
9.2.4.	Incertidumbre, cuantiles y toma de decisiones basada en riesgo	130

9.2.5. Segmentación por comportamiento y esquemas híbridos de modelado	131
9.2.6. Monitoreo, <i>drift</i> y mantenimiento del modelo en escenarios reales	131
9.2.7. Trazabilidad y gobernanza del dato para reducir fallas operativas	131
9.2.8. Medición ambiental y sostenibilidad con indicadores verificables	132
9.2.9. Escalabilidad y generalización a más PYMEs en el contexto local	132

Índice de cuadros

4.1. Resumen descriptivo de los datasets crudos recolectados (ventana 29/10/2025–21/01/2026).	45
4.2. Diccionario de datos de los conjuntos crudos exportados desde el POS.	46
4.3. Catálogo de categorías para <code>clima_desc</code> tras normalización de texto (espacios, tildes y capitalización).	52
4.4. Esquema final de variables del dataset producto–día y tipos de dato.	58
5.1. Resumen descriptivo del portafolio previo a la selección del Top-10	61
5.2. Resumen descriptivo de la serie agregada diaria por negocio	67
5.3. Top-10 final de productos en Mercaya	70
5.4. Top-10 final de productos en Villa Café	71
6.1. Villa Café: resumen global en test con <code>ORIGINAL_FULL</code> (28→41).	81
6.2. Mercaya: resumen global en test con <code>ORIGINAL_FULL</code> (28→41).	81
7.1. Mercaya: conjunto ganador <code>L4_R3_M0_CAL3_WX0</code> y variables incluidas.	105
7.2. Mercaya: mejor configuración por familia en prueba	106
7.3. Mercaya: mejor global, alternativas dentro del 5 % y mención de una opción de entrenamiento muy rápido	107
7.4. Mercaya: variables en alternativas cercanas e interpretación	108
7.5. Villa Café: conjunto <code>L4_R2_M0_CAL3_WX1</code> y variables incluidas	114
7.6. Villa Café: mejor configuración por familia en prueba	115

7.7. Villa Café: mejor global y alternativas cercanas (5 %)	116
7.8. Villa Café: variables en alternativas cercanas e interpretación	117
7.9. Mejor global por negocio (prueba)	122
7.10. Mejora del mejor global (RF_TUNED) frente al conjunto completo ORIGINAL_FULL	123
7.11. Mejora relativa del mejor global (RF_TUNED) frente a ORIGINAL_FULL	124
7.12. Alternativas cercanas (RF_TUNED) comparadas contra el conjunto completo ORIGINAL_FULL	124

Índice de figuras

4.1. Esquema MRD de la base de datos implementada	44
4.2. Nulos en los DataFrame.	51
5.1. Mercaya: diagnóstico descriptivo previo del portafolio. La figura resume la distribución del volumen acumulado, la frecuencia de venta y la concentración de la demanda.	62
5.2. Villa Café: diagnóstico descriptivo previo del portafolio. La figura resume la distribución del volumen acumulado, la frecuencia de venta y la concentración de la demanda.	63
5.3. Mercaya: relación entre volumen acumulado y continuidad de venta. La figura muestra la relación entre demanda total, número de días con venta y racha máxima consecutiva, además de resaltar el Top-10 preliminar.	65
5.4. Villa Café: relación entre volumen acumulado y continuidad de venta. La figura muestra la relación entre demanda total, número de días con venta y racha máxima consecutiva, además de resaltar el Top-10 preliminar.	66
5.5. Mercaya: evolución temporal de la demanda diaria agregada, con medias móviles de 7 y 14 días.	68
5.6. Villa Café: evolución temporal de la demanda diaria agregada, con medias móviles de 7 y 14 días.	68
5.7. Mercaya: heatmap de demanda agregada por semana y día de la semana.	69
5.8. Villa Café: heatmap de demanda agregada por semana y día de la semana.	69
5.9. Villa Café: muestra del conjunto de datos producto-día con variables de calendario y clima, e ingeniería de características mediante rezagos (<i>lags</i>) y agregaciones móviles (<i>rolling</i>).	76

5.10. Mercaya: muestra del conjunto de datos producto–día con variables de calendario y clima, e ingeniería de características mediante rezagos (<i>lags</i>) y agregaciones móviles (<i>rolling</i>).	76
6.1. Villa Café: comparación global (agregado por fecha) de demanda real vs. predicha para modelos entrenados con el set completo ORIGINAL_FULL	82
6.2. Villa Café (FULL, ORIGINAL_FULL) . Predicción vs. demanda real por producto (Top-10) para Linear	83
6.3. Villa Café (FULL, ORIGINAL_FULL) . Predicción vs. demanda real por producto (Top-10) para RF_TUNED	84
6.4. Villa Café (FULL, ORIGINAL_FULL) . Predicción vs. demanda real por producto (Top-10) para XGB_TUNED	84
6.5. Villa Café (FULL, ORIGINAL_FULL) . Predicción vs. demanda real por producto (Top-10) para LGBM_TUNED	85
6.6. Mercaya: comparación global (agregado por fecha) de demanda real vs. predicha para modelos entrenados con el set completo ORIGINAL_FULL	89
6.7. Mercaya (FULL, ORIGINAL_FULL) . Predicción vs. demanda real por producto (Top-10) para Linear	90
6.8. Mercaya (FULL, ORIGINAL_FULL) . Predicción vs. demanda real por producto (Top-10) para RF_TUNED	91
6.9. Mercaya (FULL, ORIGINAL_FULL) . Predicción vs. demanda real por producto (Top-10) para XGB_TUNED	91
6.10. Mercaya (FULL, ORIGINAL_FULL) . Predicción vs. demanda real por producto (Top-10) para LGBM_TUNED	92
7.1. Mercaya: mejor configuración del modelo Random Forest (RF_TUNED). Comparación global entre demanda real y predicha agregada por fecha.	106
7.2. Mercaya: alternativa global <i>más rápida</i> dentro del 5 % (RF_TUNED). Comparación entre demanda real y predicha agregada por fecha.	109
7.3. Mercaya: alternativa global con <i>menos variables</i> dentro del 5 % (RF_TUNED). Comparación entre demanda real y predicha agregada por fecha.	109

7.4. Mercaya: diagnóstico desagregado por producto para el mejor global (RF_TUNED). Comparación de demanda real vs. predicha (series individuales).	110
7.5. Mercaya: diagnóstico desagregado por producto para la alternativa más rápida dentro del 5 % (RF_TUNED). Comparación de demanda real vs. predicha (series individuales).	111
7.6. Mercaya: diagnóstico desagregado por producto para la alternativa con menos variables dentro del 5 % (RF_TUNED). Comparación de demanda real vs. predicha (series individuales).	112
7.7. Villa Café: desempeño global del mejor modelo en prueba (RF_TUNED). Comparación entre demanda real y predicha agregada por fecha (suma de productos).	115
7.8. Villa Café: alternativa global <i>más rápida</i> (RF_TUNED). Comparación global entre demanda real y predicha agregada por fecha.	117
7.9. Villa Café: alternativa global con <i>menos variables</i> (RF_TUNED). Comparación global entre demanda real y predicha agregada por fecha.	118
7.10. Villa Café: comparación desagregada por producto entre demanda real y predicha en prueba. Escenario <i>mejor global</i> (RF_TUNED).	119
7.11. Villa Café: comparación desagregada por producto entre demanda real y predicha en prueba. Escenario <i>más rápido dentro del 5 %</i> (RF_TUNED).	119
7.12. Villa Café: comparación desagregada por producto entre demanda real y predicha en prueba. Escenario <i>menos variables dentro del 5 %</i> (RF_TUNED).	120

Capítulo 1

Introducción

Las pequeñas y medianas empresas (PYMEs) minoristas cumplen un rol central en la economía local y en el abastecimiento cotidiano de los hogares. Sin embargo, en muchos puntos de venta la **reposición de inventario** aún se guía por criterios empíricos, basados en experiencia, intuición y observación informal del movimiento. Esta práctica se enfrenta a un entorno con **variabilidad diaria**, **dependencia de corto plazo** y presencia de **días atípicos**, lo que aumenta la probabilidad de dos fallas operativas recurrentes: (i) el **desabastecimiento** (rupturas de stock y ventas perdidas) y (ii) el **sobreabastecimiento** (capital inmovilizado, mayores costos de almacenamiento y merma), especialmente en categorías perecederas. En consecuencia, incluso desviaciones pequeñas pero repetidas en las decisiones de compra pueden acumular pérdidas económicas relevantes y generar inestabilidad en el flujo de caja.

En este contexto, el avance de la analítica y la inteligencia artificial (IA) ha abierto la posibilidad de transformar la reposición en un proceso **medible, trazable y ajustable**, soportado por **predicción de demanda diaria** y por métricas interpretables en operación. No obstante, en PYMEs persisten barreras de adopción asociadas a la calidad y disponibilidad de datos, infraestructura limitada y dificultad para sostener procesos de mantenimiento técnico. Por tanto, el reto no consiste únicamente en entrenar un modelo con alta precisión, sino en construir un enfoque **reproducible, justificable y sostenible** que traduzca datos operativos —organizados en una estructura consistente *producto-día* e incorporando señales temporales como *lags* y agregaciones móviles (*rolling*), además de variables de calendario y, cuando aplica, contexto climático— en predicciones útiles para la toma de decisiones de inventario, sin exigir capacidades que excedan las condiciones reales de operación.

Capítulo 2

Descripción del Problema

2.1. Planteamiento del Problema

En Colombia, el tejido empresarial se clasifica en micro, pequeñas, medianas y grandes empresas conforme a la Ley 590 de 2000 y su modificación mediante la Ley 905 de 2004. En este marco, las PYMEs se caracterizan, entre otros criterios, por rangos de activos; específicamente, se consideran dentro de esta categoría aquellas con activos totales entre 500 y 30.000 Salarios Mínimos Mensuales Legales Vigentes (SMMLV) [1].

Dentro de las PYMEs comercializadoras (minoristas o distribuidoras), cuya actividad principal consiste en la compra y venta de bienes sin transformación significativa, la predicción de demanda resulta determinante para planificar la reposición de inventario con criterios verificables. Esta necesidad es especialmente marcada en los productos de alta rotación, dado que concentran una fracción relevante del volumen de ventas y exigen ciclos de reposición frecuentes; en consecuencia, desviaciones sistemáticas entre lo estimado y lo observado se reflejan de manera inmediata en desempeño operativo y resultados económicos [2].

En términos de operación, la subestimación de la demanda tiende a generar compras insuficientes y, por ende, a incrementar la ocurrencia de desabastecimiento y la pérdida de ventas potenciales [2]. De forma complementaria, la sobreestimación favorece compras por encima de lo requerido, elevando niveles de inventario, costos asociados y exposición a obsolescencia o deterioro, particularmente en categorías sensibles al tiempo [2]. Por tanto, más que producir un pronóstico aislado, se requiere un enfoque que conecte la predicción de demanda con decisiones de inventario mediante mecanismos de seguimiento del error e indicadores operativos que permitan ajustar la planificación de manera oportuna [2]. En consecuencia, se justifica considerar enfoques basados en analítica e IA como alternativa para fortalecer la predicción de demanda y su traducción a decisiones de inventario.

Pese a este potencial, la adopción de IA en PYMEs continúa siendo limitada. La evidencia indica que su uso es poco frecuente y que son escasas las organizaciones que desarrollan aplicaciones propias, debido a costos, duración e incertidumbre de estos proyectos; en contraste, se observa una preferencia creciente por soluciones *AI-as-a-service* [3]. Adicionalmente, se reportan barreras recurrentes como la falta de *know-how*, costos percibidos como disuasorios y dificultades para atraer o contratar talento especializado, lo que conduce a postergar la incorporación de enfoques avanzados [3].

Con base en lo anterior, el problema de investigación se formula en torno a cómo aprovechar los registros históricos de ventas disponibles en PYMEs comercializadoras para construir modelos predictivos que estimen la demanda diaria por producto y, con ello, apoyen la reposición de inventario, reduciendo la incidencia de desabastecimiento y sobreabastecimiento.

2.1.1. Formulación

La pregunta ahora es: **¿Cómo predecir la demanda de productos de alta rotación en puntos de venta de PYMEs en Ibagué mediante un modelo de inteligencia artificial?**

2.1.2. Sistematización

Para resolver esta pregunta se deben considerar algunas subpreguntas:

- ¿Cómo garantizar la calidad de los datos para el modelo predictivo?
- ¿Cómo identificar las variables históricas clave en la predicción de demanda?
- ¿Cómo se llevará a cabo el proceso de entrenamiento de los modelos de IA diseñados para la predicción de inventarios en PYMEs?
- ¿Cómo optimizar los modelos para mejorar el desempeño en la predicción de la demanda?
- ¿Cómo evaluar la precisión y eficacia del modelo predictivo?

2.2. Objetivos

2.2.1. Objetivo General

Desarrollar un modelo de inteligencia artificial para predecir la demanda de productos de alta rotación en puntos de venta de PYMEs en Ibagué.

2.2.2. Objetivos Específicos

1. **Realizar la limpieza y preprocesamiento de los datos.** Esta etapa incluirá la eliminación de datos inconsistentes, el tratamiento de valores faltantes y la normalización de los datos, asegurando su calidad para el uso en el modelo predictivo.
2. **Analizar los datos históricos de ventas y seleccionar las características más relevantes para la predicción de demanda.** Se utilizarán técnicas de análisis exploratorio de datos para identificar las variables que más influyen en la demanda de los productos.
3. **Entrenar los modelos de IA seleccionados para la predicción de inventarios en tiendas minoristas y PYMEs.** El entrenamiento se llevará a cabo utilizando los datos preprocesados, con el objetivo de optimizar el modelo para adaptarse a las características específicas del sector.
4. **Optimizar de manera integral los modelos seleccionados, ajustando parámetros, hiperparámetros y estrategias de entrenamiento para maximizar la calidad en la predicción de demanda.** Esta optimización buscará mejorar la precisión y la eficiencia de los modelos.
5. **Evaluar la capacidad de los modelos de IA seleccionados para la predicción de demanda.** Esta evaluación se centrará en medir la precisión y eficacia de los modelos, con el fin de identificar posibles mejoras y ajustar los parámetros del modelo según sea necesario.

2.3. Justificación

Las PYMEs comercializadoras que operan con **productos de alta rotación** suelen enfrentar un reto tan cotidiano como crítico: sostener una disponibilidad estable cuando la **demanda diaria** cambia y no siempre se comporta de forma predecible. En la práctica, esto obliga a que la **reposición de inventario** se mantenga “a ritmo” del consumo; cuando no ocurre, el inventario se vuelve inestable y aparecen, con mayor frecuencia, escenarios de **desabastecimiento** o **sobreabastecimiento** [4]. Esta recurrencia no es marginal: en una encuesta aplicada a personal vinculado

a la operación de una organización minorista, el **67 %** de los encuestados reporta enfrentar estas situaciones **a diario**, lo que sugiere que el problema se manifiesta directamente en la operación y no solo como un resultado contable [4]. Además, se advierte que una gestión inadecuada puede traducirse en perjuicios financieros al afectar la liquidez del negocio [4].

De ahí que, en años recientes, haya crecido el interés por enfoques de transformación digital orientados a fortalecer la toma de decisiones basada en datos. En particular, la inversión en **IA** ha aumentado de forma acelerada (de 17 a 120 millones de dólares entre 2014 y 2021) y se han documentado incrementos de ingresos de hasta un 16 % en ciertas funciones organizacionales en entidades que la adoptan [5]. Sin embargo, esta adopción no avanza al mismo ritmo en todos los tamaños empresariales: mientras en grandes organizaciones la implementación de programas de digitalización alcanza aproximadamente el 70 %, en las pequeñas se reportan niveles cercanos al 14 % [5]. De manera consistente, más del 70 % de las grandes empresas ha incorporado alguna tecnología avanzada, frente a cerca del 30 % de las pequeñas [5].

En el caso de las PYMEs, la evidencia sugiere que la incorporación de **IA** se vuelve más viable cuando existen soluciones accesibles, con menor fricción de adopción, y cuando su implementación se aborda de forma progresiva, priorizando procesos críticos [6]. En particular, se recomienda iniciar por aplicaciones de analítica predictiva orientadas a la **predicción de demanda** y su traducción a decisiones de inventario, para luego escalar hacia otras áreas operativas conforme se consolidan capacidades y prácticas de uso [6].

En este trabajo, dicha utilidad se aborda mediante un enfoque experimental basado en registros reales de ventas, con el propósito de evaluar modelos de aprendizaje automático para la estimación de **demanda diaria por producto** en horizontes de predicción de corto plazo, y analizar su contribución como soporte cuantitativo para la **reposición de inventario**.

2.4. Delimitaciones y Alcances

2.4.1. Alcances

1. **Predicción de demanda para los productos de mayor rotación en negocios minoristas.** El proyecto se centrará en predecir la demanda de los diez productos con mayor rotación en los negocios minoristas Minimercado Villa Café y Autoservicio Mercaya en Ibagué. La selección de estos productos se realizará a partir del análisis de datos históricos de ventas proporcionados por estas empresas. La priorización de estos artículos permitirá optimizar la precisión del modelo y mejorar la eficiencia en la gestión de inventarios.
2. **Pruebas en entornos controlados con datos históricos reales.** El modelo

será implementado y validado en un entorno local, utilizando únicamente datos históricos reales proporcionados por supermercados y tiendas de descuento. Esto permitirá evaluar la precisión del modelo sin requerir pruebas en tiempo real, proporcionando un contexto seguro para la mejora continua.

3. **Evaluación del desempeño basada en métricas de precisión y eficiencia operativa.** Se evaluará el modelo utilizando métricas como el error cuadrático medio (MSE), error absoluto medio (MAE), el porcentaje de error absoluto medio (WAPE), para analizar el rendimiento de los modelos.

2.4.2. Limitaciones

1. **Dependencia de la calidad y cantidad de datos proporcionados por la empresa Minimercado Villa Café y Autoservicio Mercaya.** La precisión del modelo dependerá de la calidad y cantidad de los datos históricos recopilados a través del software de punto de venta (POS) de autoría propia el cual será implementado por estas dos PYMEs para la recolección de los datos de todo el 2025 más los datos que se puedan obtener de los reportes físicos de las PYMEs. Dado que no se considerarán datos en tiempo real ni fuentes externas, el modelo estará limitado a los patrones identificados en la información histórica disponible, lo que podría afectar su capacidad de generalización.
2. **Enfoque en PYMEs de Ibagué del sector comercial.** La investigación se centrará exclusivamente en PYMEs del sector comercial en Ibagué, Colombia. Esto limitará la aplicabilidad de los resultados a otras regiones o sectores con diferentes patrones de demanda y condiciones de mercado.
3. **Evaluación en entornos locales sin acceso a datos en tiempo real.** Las pruebas del modelo se realizarán en un entorno local sin la capacidad de procesar datos en tiempo real. Esta limitación afectará la evaluación del rendimiento del modelo en condiciones operativas continuas, lo que podría influir en la capacidad de respuesta a fluctuaciones inmediatas en la demanda.
4. **Exclusión de factores macroeconómicos y variables externas del modelo.** Para acotar el alcance del proyecto y mantener su viabilidad en el contexto de PYMEs, no se incorporan variables macroeconómicas (p. ej., inflación, tasas de interés o cambios en políticas económicas) ni fuentes externas de alta complejidad o difícil integración (p. ej., indicadores sectoriales, redes sociales o movilidad). En su lugar, el modelo se apoya principalmente en datos internos de ventas e inventario y, de manera complementaria, en variables externas de fácil disponibilidad y alineables temporalmente (calendario y clima). En consecuencia, la ausencia de factores macroeconómicos podría afectar la precisión de las predicciones en escenarios donde estos elementos tengan un efecto significativo sobre la demanda.

se considerarán los lineamientos, recomendaciones y criterios de gobernanza establecidos en el Marco Ético para la Inteligencia Artificial en Colombia[7].

Capítulo 3

Marco Teórico y Trabajos Relacionados

3.1. Marco Teórico

3.1.1. Marco de referencia

Este capítulo reúne los fundamentos conceptuales que sustentan el trabajo y, al mismo tiempo, fija el lenguaje con el que se interpretan los resultados. El problema se aborda desde la convergencia de tres ejes: (i) la **predicción de demanda** como apoyo a la toma de decisiones en el comercio minorista, (ii) la **naturaleza temporal** de los registros históricos, que obliga a modelar dependencias de corto plazo, patrones recurrentes y variaciones a lo largo del tiempo, y (iii) la **analítica predictiva** mediante modelos supervisados de aprendizaje automático. Esta articulación es necesaria porque, en PYMEs comercializadoras, el error del pronóstico no se queda en una cifra: se traduce en decisiones de reposición que pueden materializarse como **desabastecimiento** o **sobreabastecimiento**, con efectos directos sobre rentabilidad y liquidez [2, 4].

Como se discutió en el planteamiento del problema, la adopción de **IA** en PYMEs no depende únicamente de su potencial técnico, sino de condiciones prácticas que determinan si una solución puede sostenerse en el tiempo. Por ello, este marco enfatiza enfoques **realistas**, **auditables** y **reproducibles**, apoyados en los datos disponibles y en procedimientos viables para el contexto organizacional [3, 5, 6]. Finalmente, dado que se trabaja con información operativa, se asumen consideraciones de uso responsable y alineación con marcos éticos aplicables en Colombia [7].

Al cierre del capítulo, se integran **trabajos relacionados** y se explicitan **diferenciadores** del proyecto, con el fin de ubicar el aporte en su contexto y justificar las decisiones metodológicas que se desarrollan posteriormente.

3.1.2. Áreas temáticas

Por su carácter aplicado, el trabajo se apoya en varias áreas que se complementan para responder al problema planteado:

- **Sistemas de información e informática empresarial:** en tanto se parte de información transaccional generada en la operación y se busca que los resultados sean útiles en el día a día.
- **Ciencia de datos y aprendizaje automático:** porque el objetivo central se formula como un problema de regresión supervisada orientado a la **predicción de demanda**, incorporando criterios de selección de variables e interpretación de resultados.
- **Pronóstico con datos temporales:** dado que la evidencia se organiza a nivel *producto-día* y la evaluación debe respetar la secuencia temporal, incluyendo patrones recurrentes (p. ej., comportamiento semanal).
- **Ingeniería de software y de datos:** ya que se requiere asegurar trazabilidad de transacciones, estandarizar datos y preparar insumos consistentes para el modelado.

3.1.3. Términos y conceptos

Con el fin de mantener una nomenclatura consistente a lo largo del documento, a continuación se precisan los conceptos que se emplean con mayor frecuencia. En esta sección se definen ideas del dominio y del enfoque metodológico; las **etiquetas internas de reporte** utilizadas en los experimentos se presentan al final como convención de escritura (no como teoría).

- **PYMEs (Colombia).** De acuerdo con la Ley 590 de 2000 y su modificación (Ley 905 de 2004), la clasificación empresarial en Colombia considera, entre otros criterios, rangos de activos; en particular, las PYMEs se ubican en el intervalo de 500 a 30,000 SMMLV en activos totales [1].
- **Predicción de demanda.** En este trabajo se entiende como la estimación de ventas futuras a partir de datos históricos, con el propósito de apoyar la planificación operativa en contextos minoristas [8]. Metodológicamente, puede abordarse con enfoques cualitativos o cuantitativos; aquí se privilegia el enfoque cuantitativo, en el que el modelo relaciona la demanda con variables explicativas construidas a partir de información disponible [8].
- **Productos de alta rotación.** Son productos con venta frecuente y recurrente, lo que exige reposición continua para sostener disponibilidad. En estos casos, desviaciones pequeñas entre lo esperado y lo observado pueden amplificarse en la operación, especialmente cuando los ciclos de reposición son cortos [4].

- **Desabastecimiento y sobreabastecimiento.** El desabastecimiento corresponde a inventario insuficiente para atender la demanda; el sobreabastecimiento, a inventario por encima de lo requerido. En entornos minoristas, ambos escenarios afectan desempeño operativo y financiero cuando la reposición no logra anticipar adecuadamente el consumo [4].
- **Unidad temporal y horizonte de predicción.** La unidad temporal define el “paso” de observación (por ejemplo, día) y el horizonte h indica cuántos pasos hacia adelante se estima. En este trabajo, el paso temporal es el **día** y el horizonte principal es $h = 1$ [9].
- **Variables endógenas y exógenas.** Las variables *endógenas* provienen de la misma serie de demanda (p. ej., valores pasados). Las variables *exógenas* corresponden a factores externos como calendario o clima. Para que el pronóstico sea operativo (*ex-ante*), estas señales deben alinearse temporalmente sin requerir información futura; si no se dispone de predictores futuros, se emplean versiones rezagadas [9].
- **Variables de calendario (dia_habil y quincena).** Para representar la dimensión temporal sin depender de una marca absoluta de fecha, el trabajo emplea atributos calendario interpretables. En particular, **dia_habil** es un indicador binario que distingue días hábiles (1) de no hábiles (0), considerando fines de semana y festivos; y **quincena** es un indicador binario asociado al ciclo de pago (p. ej., día 15 y fin de mes), ajustado al día hábil inmediatamente anterior cuando la fecha cae en fin de semana o festivo. Estas variables buscan capturar, de forma simple y trazable, patrones operativos recurrentes del consumo minorista.
- **Rezagos (lags) y ventanas móviles (rolling).** Los *lags* incorporan valores pasados de la demanda como predictores (p. ej., $t-1$ o $t-7$), con el fin de capturar **memoria temporal** y patrones recurrentes como el comportamiento semanal [9]. Las variables *rolling* resumen un tramo reciente de la serie: para una ventana de tamaño k , se calcula una estadística (p. ej., promedio, suma o mediana) sobre los últimos k días, representando el nivel y la tendencia reciente del consumo [9].
- **Fuga temporal (data leakage).** Ocurre cuando, durante entrenamiento o evaluación, se incorpora directa o indirectamente información del futuro, produciendo estimaciones optimistas del desempeño. En series temporales se evita respetando el orden cronológico y evaluando con esquemas donde cada pronóstico se emite sin acceso a datos futuros [9].
- **Intermitencia y demanda intermitente.** En series de demanda minorista, se habla de **demanda intermitente** cuando las ventas no ocurren todos los días: se observan **muchos ceros** (días sin venta) alternados con eventos de compra, con intervalos que pueden ser irregulares. En este trabajo, la **intermitencia** se entiende como la presencia frecuente de días con demanda cero dentro del panel producto-día y se usa para (i) caracterizar productos (por ejemplo, en la

selección del Top-10 por negocio junto con la rotación) y (ii) contextualizar la interpretación de métricas porcentuales, dado que valores reales cercanos a cero pueden volver inestable el MAPE.

- **Modelo global multiserie.** Consiste en entrenar un único modelo con múltiples series (por ejemplo, demanda de varios productos) para aprender de forma conjunta y aprovechar patrones compartidos, en contraste con entrenar un modelo independiente por producto [10].
- **Pipeline de modelado.** En este documento se entiende como el procedimiento integral que asegura comparaciones consistentes y trazables entre alternativas. En términos generales, el pipeline fija: (i) preparación de datos y construcción de variables, (ii) esquema de evaluación temporal, y (iii) mecanismos de ajuste (selección de variables e hiperparámetros) aplicados bajo el mismo protocolo. De este modo, los resultados se interpretan como efecto del *proceso completo* y no únicamente del algoritmo.
- **Hiperparámetros.** Son configuraciones definidas antes del entrenamiento (p. ej., profundidad, número de árboles o regularización) que modifican el comportamiento del algoritmo. Su selección suele requerir una fase de búsqueda u optimización y, por ello, incrementa el costo computacional del proceso [10].
- **Estructura *producto-día* (panel diario).** Cada observación representa un **producto** en un **día** específico. En el trabajo se construye un panel diario completo para el subconjunto analizado, de modo que la ausencia de ventas en un día se represente explícitamente (p. ej., con demanda cero), condición necesaria para calcular rezagos coherentes y comparar modelos sobre el mismo soporte temporal.

Convenciones de reporte y nomenclatura experimental

Para facilitar la lectura de resultados, el documento utiliza algunas etiquetas operativas. Estas convenciones no introducen teoría adicional; su objetivo es identificar de forma compacta configuraciones evaluadas y mantener trazabilidad:

- **Top-10 por negocio.** Subconjunto de productos seleccionado por negocio para el análisis y el entrenamiento, definido mediante criterios reproducibles de **rotación** e **intermitencia**.
- **Línea base (*baseline*).** Referencia experimental reproducible que fija el protocolo temporal y el preprocesamiento del pipeline; sirve como punto de comparación para atribuir ganancias posteriores debidas a selección de variables y ajustes adicionales. En este documento, la línea base se implementa como una referencia **exigente** (no necesariamente el modelo más simple), manteniendo fijo: (i) el **set completo de predictores** y (ii) el **protocolo temporal + preprocesamiento**. Operativamente, esta configuración se identifica con la

etiqueta `ORIGINAL_FULL`, que corresponde al conjunto “completo” antes de aplicar reducción/selección de variables.

- **Variables crudas vs. variables efectivas (`raw`→`eff`).** Las variables crudas corresponden al número de señales antes del preprocesamiento. Las variables efectivas corresponden a la dimensionalidad final tras imputación y codificación (p. ej., *one-hot encoding*); por ello, pueden aumentar respecto a las crudas. Esta noción se reporta como aproximación de **complejidad efectiva**.
- **Conjuntos de variables (*feature sets*).** En resultados, cada alternativa de ingeniería de variables se identifica con una etiqueta compacta del tipo `Ll_Rr_Mm_CALc_WXw` (p. ej., `L2_R2_M3_CAL4_WX0`), donde `L` refiere a *lags*, `R` a *rolling*, `M` a variables adicionales de memoria/recencia, `CAL` a calendario y `WX` a clima. Los números son identificadores de configuración (no magnitudes de la serie) y permiten reconstruir qué bloques se activaron. En este documento, el bloque `CAL` incluye variables calendario interpretables como `dia_habil` y `quincena`, entre otras.
- **Modelos reportados.** Se emplean etiquetas operativas para distinguir familias y configuraciones: **Linear** (regresión lineal como referencia), **RF_TUNED** (Random Forest), **XGB_TUNED** (XGBoost) y **LGBM_TUNED** (LightGBM). Aquí, *TUNED* indica que el modelo fue ajustado dentro del pipeline mediante una estrategia de búsqueda en dos fases.

3.1.4. Formulación del problema de pronóstico diario

En la operación minorista, el pronóstico se puede formular como la estimación de la demanda del día siguiente para cada producto. Esta estimación debe construirse exclusivamente con información disponible **hasta el día t** . Por ello, para cada producto se utiliza su historia reciente de ventas (incluyendo *lags* y agregaciones *rolling*) y el contexto del día (calendario y, cuando aplica, clima), con el fin de anticipar la demanda en $t+1$. Bajo esta formulación, la evaluación debe simular el uso real: entrenar con periodos anteriores y evaluar en periodos posteriores, evitando señales que no estarían disponibles al momento de emitir el pronóstico.

Formalmente, sea p un producto y t un día calendario. Si $y_{p,t}$ denota la demanda observada del producto p en el día t , el pronóstico a un día ($h = 1$) se representa como:

$$\hat{y}_{p,t+1} = f(\mathbf{x}_{p,t}), \quad (3.1)$$

donde $\mathbf{x}_{p,t}$ agrupa las variables disponibles hasta t (historia de demanda y variables de contexto).

3.1.5. Modelos supervisados para predicción de demanda

En este trabajo, la **predicción de demanda** se plantea como un problema de **regresión** sobre observaciones *producto-día*. La variable objetivo corresponde a la demanda a estimar y los predictores se construyen exclusivamente con información disponible al momento de decidir, integrando memoria reciente (capturada mediante *lags* y agregaciones *rolling*) con variables de contexto conocidas *ex-ante*, como calendario [9].

Cuando se pronostican múltiples productos, surge una decisión adicional: entrenar modelos **locales** (uno por producto) o un modelo **global** ajustado de forma conjunta, para aprovechar patrones compartidos [11]. En el documento se adopta una aproximación global, coherente con la necesidad de una solución sostenible en operación y con la posibilidad de integrar variables numéricas y categóricas (p. ej., identificador del producto) junto con señales temporales y de contexto.

Regresión lineal como referencia

Descripción formal. La regresión lineal estima la demanda futura como una combinación lineal de predictores disponibles hasta el día t (p. ej., *lags*, *rolling*, calendario y, cuando aplica, clima). Se usa como referencia por su simplicidad, bajo costo computacional e interpretabilidad: cada coeficiente resume el cambio promedio esperado en la predicción ante una variación del predictor, manteniendo constantes las demás variables. En el diseño experimental, funciona como línea base para dimensionar qué parte del patrón puede capturarse mediante relaciones aproximadamente lineales [9].

Formulación matemática. Sea $\mathbf{x}_{p,t} \in \mathbb{R}^d$ el vector de predictores del producto p en el día t y sea $y_{p,t+1}$ la demanda observada al día siguiente. El modelo lineal se expresa como:

$$\hat{y}_{p,t+1} = \beta_0 + \sum_{j=1}^d \beta_j x_{p,t}^{(j)} = \beta_0 + \boldsymbol{\beta}^\top \mathbf{x}_{p,t}. \quad (3.2)$$

Sus parámetros se estiman minimizando el error cuadrático en entrenamiento:

$$(\hat{\beta}_0, \hat{\boldsymbol{\beta}}) = \arg \min_{\beta_0, \boldsymbol{\beta}} \sum_{i=1}^n (y_i - \beta_0 - \boldsymbol{\beta}^\top \mathbf{x}_i)^2. \quad (3.3)$$

Ensamblados basados en árboles

Descripción formal. Los árboles de decisión particionan el espacio de variables en regiones donde la demanda se aproxima por un valor constante, mediante reglas del tipo “si-entonces”. En escenarios con variables tabulares y mezcla de señales

(memoria reciente, calendario y, cuando aplica, clima), su fortaleza es que capturan **no linealidades** e **interacciones** de forma natural [11]. En la práctica, se suelen emplear **ensambles** (conjuntos de árboles) para mejorar estabilidad y generalización [11]. Estas familias resultan adecuadas aquí por el equilibrio que ofrecen entre desempeño predictivo y consideraciones operativas como costo de entrenamiento y necesidad de ajuste [11].

Random Forest. Descripción formal. Random Forest entrena múltiples árboles en paralelo usando muestras *bootstrap* del conjunto de entrenamiento y restringe aleatoriamente el subconjunto de variables candidatas en cada partición. Con ello reduce la correlación entre árboles y la varianza del modelo, produciendo predicciones más estables ante ruido [11].

Formulación matemática. Si $T_b(\cdot)$ es el árbol b y hay B árboles:

$$\hat{y}(\mathbf{x}) = f_{\text{RF}}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B T_b(\mathbf{x}). \quad (3.4)$$

Cada árbol induce regiones (hojas) $\{R_{b,m}\}$ y predice:

$$T_b(\mathbf{x}) = \sum_m c_{b,m} \mathbb{I}(\mathbf{x} \in R_{b,m}), \quad (3.5)$$

donde $c_{b,m}$ suele ser el promedio de y en la hoja m .

Gradient Boosted Trees (XGBoost/LightGBM). Descripción formal. En *boosting* el ensamble se construye secuencialmente: cada nuevo árbol busca corregir los errores acumulados. Implementaciones como XGBoost y LightGBM suelen ser competitivas en datos tabulares; sin embargo, su rendimiento depende de un ajuste cuidadoso y conlleva un costo asociado a la búsqueda y al entrenamiento [11].

Formulación matemática. Modelo aditivo por etapas:

$$F_M(\mathbf{x}) = \sum_{m=0}^M \gamma_m h_m(\mathbf{x}), \quad (3.6)$$

donde $h_m(\cdot)$ son árboles. Para pérdida cuadrática:

$$r_{i,m} = y_i - F_{m-1}(\mathbf{x}_i), \quad (3.7)$$

y la actualización:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \nu \gamma_m h_m(\mathbf{x}), \quad (3.8)$$

donde $\nu \in (0, 1]$ es la tasa de aprendizaje.

Modelo global multiserie y criterios prácticos. Estas familias permiten entrenar un **modelo global multiserie** al incorporar el identificador del producto como variable categórica, favoreciendo el aprendizaje conjunto. La comparación entre alternativas considera, además del error, criterios prácticos como tiempo de entrenamiento e inferencia y esfuerzo de ajuste, porque estos condicionan la viabilidad en escenarios reales de operación [11].

3.1.6. Pipelines de modelado y reproducibilidad

En predicción de demanda, el desempeño observado no depende únicamente del algoritmo, sino del **flujo completo** con el que se preparan los datos, se entrena y se evalúa el modelo. Decisiones como la partición temporal, el preprocesamiento, la selección de variables y la optimización de hiperparámetros deben entenderse como partes de un mismo procedimiento, cuyo propósito es estimar el desempeño de manera realista y reducir el riesgo de sobreajuste [12].

Un problema frecuente aparece cuando entrenamiento y evaluación no son consistentes. Por ejemplo, imputar usando información de todo el dataset o aplicar codificaciones que cambian entre corridas introduce *data leakage* y produce métricas optimistas que luego no se sostienen fuera de muestra [12]. Para controlar este riesgo, el documento adopta un **pipeline** como unidad de trabajo: un procedimiento único que encadena transformaciones y estimador final. La idea central es que cada transformación se ajusta únicamente con los datos de entrenamiento y se aplica de forma idéntica sobre validación y prueba, asegurando consistencia, trazabilidad y comparabilidad entre alternativas [12].

En presencia de variables numéricas y categóricas, el pipeline permite definir tratamientos diferenciados (imputación y codificación consistente, por ejemplo *one-hot encoding*), controlando categorías no vistas y evitando cambios de dimensionalidad entre entrenamiento y prueba [12]. Dado que estas decisiones afectan representación, complejidad y costo, además de las métricas de desempeño se reporta el número de **features efectivas** tras el preprocesamiento como medida de complejidad operativa.

3.1.7. Ingeniería y diagnóstico de variables

En predicción diaria, elegir un algoritmo es solo una parte del problema: una fracción importante del resultado depende de **qué señales** se construyen y **cómo** se entregan al modelo. Por ello, la ingeniería de variables se entiende como el proceso de transformar la información disponible en predictores útiles para describir el comportamiento de la demanda, manteniendo coherencia con el **pipeline** de preparación, entrenamiento y evaluación [12]. En este trabajo, las variables se organizan en dos bloques:

1. **Variables endógenas:** señales derivadas de la serie de demanda, principalmente *lags* y *rolling*, que capturan memoria reciente.
2. **Variables exógenas:** señales externas disponibles *ex-ante*, como calendario y, cuando aplica, clima, que aportan contexto.

A medida que se agregan transformaciones y combinaciones, el número de variables puede crecer rápidamente. Esto trae una tentación natural: probar muchas alternativas hasta encontrar mejoras en métricas. Sin embargo, si esa exploración no se controla, se puede terminar con modelos que “aprenden” particularidades del periodo observado y pierden desempeño al pronosticar días posteriores [12]. En consecuencia, el diagnóstico y la selección se asumen como una depuración: ayudan a distinguir señales que aportan información de aquellas que solo incrementan complejidad. La confirmación final se realiza dentro del mismo protocolo del pipeline, verificando que las mejoras se mantengan en evaluación temporal [12].

Correlaciones como diagnóstico inicial (Pearson, Spearman y Kendall)

Las correlaciones se usan como una lectura inicial por dos razones prácticas: (i) identificar predictores que parecen asociarse con la demanda y (ii) detectar redundancias entre variables antes de invertir costo en entrenar modelos completos [12]. Su interpretación, no obstante, debe ser prudente: son medidas bivariadas y no reemplazan la evaluación predictiva controlada dentro del **pipeline** [12].

- **Pearson.** Aproxima asociación **lineal** entre predictor y demanda [12].
- **Spearman.** Evalúa asociación **monotónica** sobre rangos, sin exigir linealidad [12].
- **Kendall.** Mide concordancia por pares y tiende a ser más conservadora ante ruido [12].

Definiciones matemáticas (síntesis). Sea $\{(x_i, y_i)\}_{i=1}^n$. Pearson se define como covarianza normalizada:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (3.9)$$

Kendall (versión básica) se expresa como:

$$\tau = \frac{C - D}{\binom{n}{2}}, \quad (3.10)$$

donde C y D son pares concordantes y discordantes, respectivamente. En el documento, estas medidas se usan para orientar la exploración inicial; la selección final se valida con desempeño en el pipeline [12].

Información mutua (Mutual Information)

La **información mutua (MI)** complementa las correlaciones cuando se busca responder a una pregunta práctica: *¿cuánta información aporta esta variable para explicar la demanda, incluso si la relación no es lineal?* [12]. A diferencia de las correlaciones clásicas, MI puede reflejar dependencias no lineales o efectos por umbrales. En coherencia con el enfoque del documento, se usa como señal exploratoria para priorizar variables y representaciones, y su utilidad final se confirma mediante evaluación dentro del **pipeline** [12].

Sea X una variable explicativa y Y la demanda (o demanda futura). La información mutua se define como:

$$I(X; Y) = \sum_x \sum_y p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right). \quad (3.11)$$

3.1.8. Selección de variables y control de redundancias

En un problema aplicado de **predicción de demanda**, la cantidad de variables tiende a crecer con rapidez al incorporar *lags*, *rolling*, señales exógenas y codificaciones. Este crecimiento aumenta el riesgo de **sobreajuste**, dificulta la trazabilidad de lo que se está usando y eleva el costo de entrenamiento y validación [12]. Por ello, la selección de variables se justifica como una forma de mejorar generalización, facilitar auditoría y mantener tiempos de cómputo viables [12].

Un punto clave es que la búsqueda debe realizarse de forma **controlada** dentro del **pipeline**. Probar múltiples conjuntos sin un protocolo estable puede producir mejoras aparentes por azar que no se sostienen en periodos posteriores [12]. En este trabajo, la selección se operacionaliza con tres criterios complementarios:

- **Eliminación de redundancias exactas o triviales.** Si dos variables codifican la misma información por construcción, mantener ambas no aporta señal adicional y sí incrementa complejidad [12].
- **Selección por desempeño bajo evaluación consistente.** La utilidad de un subconjunto se define por su desempeño predictivo bajo el mismo protocolo temporal y el mismo pipeline, y no únicamente por diagnósticos bivariados [12].
- **Simplicidad operacional.** Si dos conjuntos rinden de forma similar, se prioriza el conjunto más pequeño por costo, mantenimiento y trazabilidad [12].

Formulación matemática (lectura general). Sea $\mathcal{F} = \{1, \dots, d\}$ el conjunto total de variables y $S \subseteq \mathcal{F}$ un subconjunto candidato. La selección puede verse como

minimizar pérdida de validación y penalizar complejidad:

$$S^* = \arg \min_{S \subseteq \mathcal{F}} \hat{\mathcal{L}}_{\text{val}}(S) + \lambda \Omega(S), \quad (3.12)$$

donde $\hat{\mathcal{L}}_{\text{val}}(S)$ se estima con protocolo temporal y pipeline fijo (p. ej., MAE o WAPE) y $\Omega(S)$ penaliza conjuntos grandes (p. ej., $\Omega(S) = |S|$) [12].

Reducción de dimensionalidad: selección en lugar de transformación. En este trabajo no se emplean transformaciones hacia espacios latentes (p. ej., PCA). La reducción se realiza mediante **selección de subconjuntos** de variables originales, buscando el **menor número de *features*** que mantenga (o mejore) el desempeño bajo evaluación consistente en el pipeline. Esta elección preserva trazabilidad e interpretación, ya que las variables seleccionadas corresponden a señales observables del problema (p. ej., *lags*, *rolling* y variables exógenas) [12].

3.1.9. Optimización de hiperparámetros y búsqueda eficiente

Los **hiperparámetros** son configuraciones externas al ajuste directo del modelo (no se “aprenden” como pesos), pero condicionan el comportamiento del algoritmo (p. ej., profundidad en árboles, tasa de aprendizaje en *boosting*, fracciones de muestreo). Su selección se implementa mediante un **bucle externo** que evalúa desempeño bajo un protocolo de validación, buscando aproximar capacidad de generalización [13].

De manera formal, sea A un algoritmo con espacio de configuraciones Λ y una pérdida $L(\cdot)$. La optimización busca:

$$\lambda^* = \arg \min_{\lambda \in \Lambda} \mathbb{E}_{(D_{\text{train}}, D_{\text{valid}}) \sim \mathcal{D}_V} \left[V(L, A_\lambda, D_{\text{train}}, D_{\text{valid}}) \right], \quad (3.13)$$

donde $\mathcal{D}_V$ representa el mecanismo de partición inducido por V (p. ej., *holdout* o validación cruzada) [13]. En la práctica, la búsqueda opera bajo un **presupuesto computacional** finito; por ello, se debe equilibrar **precisión** y **costo**.

Búsqueda aleatoria en dos fases: exploración amplia y refinamiento

En escenarios aplicados, el ajuste de hiperparámetros se realiza con tiempo limitado. Por esta razón, el trabajo adopta como punto de partida la **búsqueda aleatoria** (*random search*), que muestrea configuraciones λ dentro de rangos razonables y las evalúa con un protocolo fijo hasta agotar el presupuesto [13]. Frente a *grid search*, suele ser más eficiente cuando solo algunos hiperparámetros son realmente influyentes, porque explora combinaciones más diversas con el mismo número de evaluaciones y facilita paralelización [13].

Como el espacio de búsqueda es amplio y no presenta una estructura simple, se emplea un esquema de dos fases. Primero, una **exploración amplia** para ubicar

regiones prometedoras; luego, un **refinamiento** donde se acotan rangos alrededor de las mejores configuraciones y se repite la búsqueda con mayor detalle, concentrando el esfuerzo en alternativas competitivas [13]. En ambas fases se mantiene el mismo protocolo de validación dentro del **pipeline**, de modo que las mejoras observadas se deban a cambios reales de configuración y no a variaciones del procedimiento [13].

Asignación progresiva de presupuesto: Successive Halving y Hyperband

En contextos con recursos limitados, una parte importante del costo del *tuning* proviene de entrenar configuraciones poco prometedoras “hasta el final”. Para reducir este desperdicio, se emplean estrategias de **asignación progresiva de presupuesto** (*multi-fidelity*), donde primero se evalúan muchas configuraciones con un presupuesto pequeño y solo las más competitivas reciben presupuestos mayores [13].

Successive Halving. Idea general. Sea r un presupuesto inicial (por ejemplo, número de estimadores/árboles, iteraciones, o fracción de datos) y sea $\eta > 1$ un factor de reducción. Se evalúan n configuraciones con presupuesto r ; luego se conserva una fracción (típicamente $1/\eta$) con mejor desempeño y se incrementa su presupuesto multiplicándolo por η . El proceso se repite por rondas hasta alcanzar el presupuesto máximo [13].

Esquema simplificado. Si $n_0 = n$ y $r_0 = r$, en la ronda k :

$$n_k = \left\lfloor \frac{n_{k-1}}{\eta} \right\rfloor, \quad r_k = \eta r_{k-1}, \quad (3.14)$$

de modo que se reduce el número de candidatos mientras se incrementa el esfuerzo por candidato. Esta lógica prioriza eficiencia: concentra el cómputo en configuraciones que ya demostraron ser competitivas con presupuestos bajos [13].

Hyperband. Idea general. Successive Halving requiere fijar de antemano cuántas configuraciones evaluar y con qué presupuesto inicial. Hyperband extiende esta idea ejecutando **varios** esquemas de Successive Halving (llamados *brackets*) con distintos balances entre **exploración** y **refinamiento** [13]. En algunos brackets se evalúan muchas configuraciones con presupuestos muy pequeños (exploración agresiva); en otros, pocas configuraciones con presupuestos iniciales mayores (refinamiento temprano). Con ello, Hyperband reduce el riesgo de descartar configuraciones que podrían verse mal con un presupuesto demasiado pequeño y, a la vez, mantiene controlado el costo total [13].

Lectura operativa en este trabajo. En la práctica, estas estrategias se alinean con el criterio del documento: (i) explorar suficiente diversidad de configuraciones sin exceder el presupuesto, y (ii) reservar mayor cómputo para alternativas que ya

muestran desempeño competitivo bajo evaluación temporal consistente dentro del pipeline [13].

3.1.10. Costo computacional, tiempo de entrenamiento y criterio de selección práctico

En un trabajo aplicado, la comparación entre modelos no se sostiene únicamente en una métrica de error. Cada entrenamiento y cada evaluación consumen tiempo y recursos, y ese **presupuesto computacional** termina definiendo cuántas alternativas es posible explorar y con qué profundidad se ajusta cada una. En particular, durante la optimización de hiperparámetros aparece un **trade-off**: repartir presupuesto entre *muchas* configuraciones mejora la cobertura del espacio, pero puede descartar temprano opciones que se beneficiarían de un ajuste más fino; concentrar el presupuesto en *pocas* configuraciones puede llevar a invertir tiempo en alternativas poco competitivas desde el inicio [13]. Por ello, la eficiencia del proceso depende tanto del algoritmo como de la forma de organizar la búsqueda.

Bajo este criterio, además del desempeño predictivo, se consideran tres elementos relevantes para una adopción realista:

- **Tiempo de entrenamiento/ajuste:** se reporta como medida complementaria porque resume el costo de entrenar y de ejecutar la búsqueda de hiperparámetros [13].
- **Complejidad efectiva:** se aproxima mediante el número de *features* resultantes tras el preprocesamiento, por su impacto en memoria, tiempo y riesgo de sobreajuste.
- **Criterio práctico de selección:** además del mejor resultado, se consideran alternativas con desempeño cercano que reduzcan tiempo o complejidad, por ser más sostenibles y auditables dentro del **pipeline** en operación.

Selección multiobjetivo y frontera de Pareto

En un escenario operativo, la selección final raramente se reduce a “minimizar el error”. En cambio, se busca un balance entre objetivos que suelen competir, como **precisión** (error bajo), **tiempo** (entrenamiento/ajuste razonable) y **complejidad** (número de *features* efectivas y mantenimiento). Esta situación puede leerse como una decisión **multiobjetivo** [13].

Sea $\mathbf{g}(\theta) = (g_1(\theta), g_2(\theta), \dots, g_m(\theta))$ un vector de objetivos a minimizar para una alternativa θ (por ejemplo, $g_1 = \text{WAPE}$, $g_2 = \text{tiempo}$, $g_3 = \# \text{features}$). Se dice que θ_a **domina** a θ_b si:

$$g_j(\theta_a) \leq g_j(\theta_b) \quad \forall j, \quad \text{y} \quad \exists k : g_k(\theta_a) < g_k(\theta_b). \quad (3.15)$$

Es decir, θ_a no es peor en ningún objetivo y es estrictamente mejor en al menos uno. Bajo esta definición, la **frontera de Pareto** es el conjunto de alternativas **no dominadas**: configuraciones para las cuales no existe otra alternativa que mejore simultáneamente todos los objetivos [13]. En términos prácticos, la frontera recoge las opciones “razonables” del compromiso precisión–costo: elegir una u otra depende del contexto (por ejemplo, preferir una alternativa con error apenas mayor si reduce de forma marcada el tiempo o la complejidad).

En este documento, este concepto respalda el criterio práctico de selección: además del mejor resultado en error, se reportan y se consideran alternativas cercanas que ofrecen un equilibrio más sostenible en costo y complejidad dentro del pipeline.

En términos conceptuales, la selección se interpreta como un balance entre **precisión** y **costo**. Cuando existen límites operativos explícitos (por ejemplo, tiempo máximo de entrenamiento o memoria disponible), estos se asumen como restricciones que determinan qué configuraciones son viables [13].

3.1.11. Evaluación del desempeño y validación temporal

En pronóstico, la evaluación debe simular el uso real: **entrenar con pasado** y **evaluar en días posteriores**. Por esta razón, la partición de datos y la estimación del desempeño se realizan respetando el **orden temporal**, evitando divisiones aleatorias que introducirían información futura de manera implícita [9, 14].

Holdout temporal. Se reserva un **bloque final** como conjunto de prueba (*out-of-sample*), manteniendo intacta la secuencia cronológica. Este bloque representa el escenario real de uso: días que el modelo no vio durante el entrenamiento. Su tamaño se define de forma coherente con el horizonte considerado (al menos h observaciones futuras) [9, 14].

Validación cruzada temporal. Para complementar el *holdout* y obtener una estimación más estable, se emplea **validación cruzada temporal**. El modelo se re-entrena de manera secuencial con un historial que se actualiza y, en cada corte, se evalúa sobre el bloque inmediatamente posterior (ventana expansiva o *rolling forecasting origin*). Esto permite verificar si el desempeño se mantiene en distintos tramos, sin romper la causalidad temporal [9].

Integridad del día en la estructura producto–día. Dado que la unidad temporal del análisis es el **día**, se evita dividir un mismo día entre entrenamiento y validación/prueba. En una estructura producto–día, partir el día puede mezclar observaciones que comparten el mismo contexto temporal (p. ej., calendario o clima), produciendo una evaluación artificialmente optimista. Por ello, se conserva la integri-

dad del día al definir cortes temporales, como medida para prevenir fuga temporal y mantener una lectura realista del desempeño [9, 14].

Métricas y su interpretación operativa

Las métricas resumen qué tan cerca queda el pronóstico $\hat{y}$ del valor real y . La lectura principal se apoya en métricas de **error**: valores más bajos implican menor desviación y, por tanto, mejor desempeño. Como complemento, se reportan R^2 y R^2_{adj} para contextualizar qué parte de la variación de la demanda captura el modelo; sin embargo, en pronóstico no reemplazan el análisis del error en evaluación temporal [9].

- **MAE (Mean Absolute Error)**. Expresa el **error promedio en unidades**. Permite una lectura directa del desvío típico: en promedio, “me equivoco en x unidades por producto–día”. Un MAE menor indica mejor desempeño.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.16)$$

- **MSE/RMSE**. Penalizan con mayor fuerza los errores grandes, por lo que son útiles cuando los **picos de demanda** tienen impacto operativo. Un valor menor es preferible; el RMSE conserva la escala de y .

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.17)$$

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (3.18)$$

- **WAPE (Weighted Absolute Percentage Error)**. Mide el error absoluto total *en relación con el volumen total* observado. En lectura práctica: “**qué proporción del total demandado** quedó mal pronosticada”. Resulta útil para comparar resultados agregados cuando se evalúan múltiples productos [14].

$$\text{WAPE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i|} \quad (3.19)$$

- **MAPE (limitaciones en demanda intermitente)**. Aunque es intuitiva, puede volverse **inestable** cuando hay valores reales cercanos a cero o ceros, distorsionando comparaciones. Por ello se privilegian MAE y WAPE como lecturas más robustas [9, 14].

$$\text{MAPE}_\epsilon(\%) = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\max(|y_i|, \epsilon)} \quad (3.20)$$

- R^2 . Indica qué tanto el modelo explica la variación de y frente a una referencia basada en el promedio. Se interpreta junto con MAE/WAPE y siempre sobre evaluación temporal [9].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.21)$$

- R^2_{adj} (**coeficiente de determinación ajustado**). Es una variante de R^2 que penaliza la inclusión de muchos predictores, porque ajusta el poder explicativo por el tamaño de la muestra y el número de variables. Se reporta como complemento para contextualizar comparaciones cuando la complejidad (número de *features*) cambia entre configuraciones.

$$R^2_{\text{adj}} = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}, \quad (3.22)$$

donde n es el número de observaciones y p el número de predictores.

3.1.12. Trabajos relacionados y diferenciadores del proyecto

Trabajos relacionados. En Colombia se han consolidado iniciativas orientadas a fortalecer la toma de decisiones mediante pronósticos basados en datos, especialmente en escenarios donde la demanda varía por calendario y condiciones del negocio. En este marco, un trabajo aplicado al sector textil propone y valida un esquema de predicción de ventas diarias que contrasta una referencia lineal con modelos no lineales (por ejemplo, ensambles), incorporando variables como promociones, fines de semana, festivos y quincena, y planteando la integración del modelo en un entorno real con monitoreo y ajustes periódicos [15]. Este antecedente es pertinente porque muestra el valor de modelar señales temporales y de comparar familias de modelos bajo un protocolo consistente.

De forma complementaria, un estudio en consumo masivo reporta mejoras frente a metodologías tradicionales (p. ej., promedios móviles), cuantificadas con métricas de error, y culmina con una herramienta de adopción sencilla que facilita el acceso a pronósticos y apoya la selección de alternativas de modelado [16]. Esta línea resalta que la utilidad del pronóstico depende también de su *operacionalización* en los flujos cotidianos.

Finalmente, un trabajo orientado a tendencias de consumo propone un prototipo predictivo de bajo costo para PYMEs, integrando datos abiertos y presentando resultados en una aplicación interactiva (Streamlit) que prioriza comprensión del usuario y exploración visual [17]. Aunque el dominio difiere, funciona como antecedente de democratización tecnológica: el pronóstico se concibe como un recurso utilizable para decisión, no solo como modelo aislado.

Síntesis de aportes. En conjunto, estos antecedentes convergen en tres ideas: (i) el pronóstico mejora la planeación cuando integra señales temporales (calendario/estacionalidad) y variables del contexto [15]; (ii) las comparaciones requieren métricas consistentes y evaluación temporal para que el desempeño estimado refleje el uso real [16]; y (iii) la adopción se favorece cuando el pronóstico se traduce en procedimientos reproducibles y artefactos interpretables (herramientas y visualizaciones) que faciliten mantenimiento [17].

Diferenciadores del proyecto. El presente proyecto comparte la motivación general de apoyar decisiones mediante pronóstico, pero se diferencia por un énfasis metodológico y operativo alineado con el funcionamiento de PYMEs comercializadoras y con el control explícito del ciclo de modelado:

- **Granularidad producto–día y enfoque multiserie.** El problema se formula sobre observaciones producto–día y se prioriza el aprendizaje conjunto de múltiples productos de alta rotación, incorporando el identificador del producto para capturar diferencias estructurales sin perder consistencia en el preprocesamiento.
- **Evaluación temporal y lectura operativa del error.** La evaluación respeta la lógica temporal (entrenar con pasado, evaluar en días posteriores), evitando sesgos por particiones aleatorias y favoreciendo una interpretación operativa basada en métricas de error.
- **Ingeniería y selección de variables guiada por desempeño.** Se construyen variables endógenas y exógenas y se aplica una reducción controlada orientada a equilibrar precisión, interpretabilidad y sostenibilidad.
- **Pipeline reproducible para comparación justa.** El flujo se encapsula para que preprocesamiento y selección se ajusten solo con entrenamiento y se apliquen igual en validación/prueba, garantizando comparaciones consistentes.
- **Selección práctica precisión–costo.** La elección final considera no solo el mejor desempeño, sino alternativas cercanas que reduzcan complejidad y tiempo, criterio pertinente en contextos con recursos limitados.

Capítulo 4

Adquisición, integración y preprocesamiento de datos

La construcción de un conjunto de datos adecuado para el pronóstico de demanda en entornos minoristas no se limita a “pasar información a una tabla”. En realidad, es un paso metodológico que condiciona todo lo que ocurre después, y por ello debe sostenerse en dos criterios básicos: **trazabilidad** (poder seguir el rastro de cada registro hasta su origen) y **reproducibilidad** (poder obtener los mismos resultados si el procedimiento se repite bajo las mismas condiciones). En la práctica, este trabajo articula tres componentes: (i) la captura de registros transaccionales **trazables** en un contexto donde es común encontrar controles operativos manuales y, por tanto, mayor riesgo de inconsistencias; (ii) una infraestructura de almacenamiento que conserve la **integridad** de la información y facilite consultas **reproducibles**; y (iii) una secuencia de transformaciones que convierta los movimientos de venta en una representación temporal consistente a nivel *producto-día*. Esta coherencia no es un detalle menor: cuando la captura o el procesamiento se realiza de forma poco controlada, pueden aparecer sesgos y riesgos de *data leakage* que, aunque no siempre son evidentes, terminan afectando las métricas y la interpretación de los resultados.

Con base en lo anterior, el flujo implementado incluye la recolección en campo en dos negocios minoristas de Ibagué, el diseño de la arquitectura de almacenamiento en PostgreSQL, la caracterización de los datos exportados desde el POS desarrollado e instalado, y el alistamiento final orientado a aprendizaje supervisado sobre un panel *producto-día*. En consecuencia, los apartados siguientes describen este flujo como un **proceso de preparación de datos** cuidadosamente documentado: se conserva la integridad transaccional en el almacenamiento y se consolida una unidad de análisis estable, que permite soportar posteriormente los experimentos de modelado y ajuste planteados en el documento.

4.1. Adquisición de datos con las empresas

4.1.1. Recolección de datos con dos negocios minoristas en Ibagué

La adquisición de datos se realizó mediante trabajo de campo con dos negocios minoristas ubicados en la ciudad de Ibagué. En la fase de diagnóstico se observó que, a pesar de contar con un flujo de ventas representativo para la categoría de minimercados, los establecimientos no disponían de un sistema POS ni de mecanismos de sistematización que permitieran consolidar un registro estable de ventas, inventarios o pedidos. En consecuencia, una parte importante de la operación cotidiana se apoyaba en anotaciones manuales y en el criterio empírico de los administradores para la reposición.

Con el propósito de reconstruir información histórica relacionada con la demanda y el abastecimiento, se solicitó acceso a los registros existentes de ventas y compras. No obstante, la evidencia disponible correspondía, en su mayoría, a facturas aisladas de algunos productos, con baja consistencia y amplios intervalos temporales entre documentos. Asimismo, no se encontró un historial estructurado de transacciones: en el punto de pago, las ventas se totalizaban de forma manual (por ejemplo, mediante calculadora), lo que impedía contar con un rastro transaccional verificable para caracterizar la demanda diaria y analizar patrones de reposición.

Ante este escenario, se definió como estrategia la implementación de un proceso de sistematización orientado a capturar información pertinente y trazable para los objetivos del proyecto. Para ello, se desarrolló e instaló un software POS de autoría propia en cada negocio, con el fin de registrar las transacciones de manera estructurada y consolidar progresivamente un histórico confiable de ventas.

Durante la implementación se identificó un reto adicional: al no existir sistematización previa, los negocios no contaban con un catálogo digital de productos. Por esta razón, se realizó un inventariado manual, registrando producto a producto los códigos, descripciones y atributos mínimos requeridos para su operación en el POS. Este levantamiento permitió consolidar un catálogo superior a dos mil productos por tienda y exigió actividades de captura, verificación y estandarización directamente en el entorno operativo.

En términos de planificación, la recolección sistemática se proyectó para iniciar desde agosto y fortalecerse de manera incremental con los registros generados por el POS. Sin embargo, el trabajo requerido para construir el inventario desde cero y estabilizar el registro en operación retrasó el inicio efectivo del histórico aprovechable. En consecuencia, el análisis y los experimentos del documento se soportan en un periodo concentrado de aproximadamente tres meses, comprendido entre el **29 de octubre de 2025** y el **21 de enero de 2026**, intervalo en el cual se dispone de un seguimiento estructurado de la **demanda diaria** registrada por el sistema para los

productos comercializados en cada negocio.

4.2. Almacenamiento y estructura de la base de datos

4.2.1. Elección de PostgreSQL y justificación técnica

El software POS desarrollado en este proyecto se integró con una base de datos relacional en **PostgreSQL** para almacenar la información transaccional generada en la operación diaria. Esta decisión se tomó, principalmente, por dos razones que son críticas en el contexto del trabajo: (i) la necesidad de conservar un registro **consistente** de eventos como ventas, movimientos de inventario y actualizaciones asociadas; y (ii) la posibilidad de consultar esa información de forma **estable** a lo largo del tiempo, sin perder trazabilidad entre entidades.

En primer lugar, PostgreSQL ofrece un soporte robusto para **transacciones** y **restricciones de integridad** (por ejemplo, claves primarias y foráneas), lo cual reduce el riesgo de registros incompletos o inconsistentes cuando se registra una venta y, al mismo tiempo, se actualiza inventario y pagos. En un sistema operativo, estos detalles no son secundarios: son los que permiten que, al reconstruir el histórico, una venta siempre tenga sus vínculos correctos (producto, fecha, punto de pago, usuario) y que el estado del inventario pueda auditarse con confianza [18].

En segundo lugar, se optó por un enfoque **relacional** porque el dominio del POS está compuesto por entidades con relaciones claras (productos, ventas, detalle de venta, inventario, usuarios, métodos de pago). En este escenario, un modelo basado en tablas relacionadas facilita tanto la validación de reglas de negocio como la posterior extracción de datos para construir el panel *producto-día*. Si bien motores como **MySQL** también son alternativas viables, PostgreSQL resulta especialmente conveniente cuando se busca mantener esquemas estrictos y consultas analíticas más expresivas sin comprometer la integridad [18].

Por contraste, no se priorizó una alternativa **NoSQL** (p. ej., MongoDB) porque el objetivo del almacenamiento no era la flexibilidad del documento, sino la **consistencia estructural** y la facilidad de reconstruir relaciones entre eventos (por ejemplo, de una venta a su detalle y a sus impactos en inventario). Dado el alcance del proyecto y la naturaleza del dato transaccional, un esquema relacional ofrecía una base más adecuada para sostener el histórico operativo que luego se utiliza en la etapa de preparación de datos [18].

Finalmente, para mantener alineada la estructura de la base de datos con la evolución del software, se gestionó el esquema mediante **migraciones** desde el backend, de forma que los cambios quedaran versionados y fueran reproducibles en distintos despliegues. Con ello, la base de datos no se entiende solo como un

repositorio, sino como un componente controlado de la solución, coherente con los requisitos de consistencia que exige el uso posterior de los datos [18].

4.2.2. Descripción funcional del esquema relacional

Con el fin de soportar la operación diaria y la trazabilidad requerida para el análisis, se diseñó una base de datos organizada en módulos funcionales (catálogo, ventas, inventario, caja y abastecimiento). Las entidades se conectan mediante claves primarias y foráneas, lo cual asegura consistencia y facilita la obtención de variables analíticas mediante consultas reproducibles. A continuación, se describen las tablas principales del esquema.

Catálogo de productos

- **categorías:** almacena la clasificación general del portafolio (p. ej., granos, lácteos, aseo). Se relaciona con **productos** para habilitar segmentación y análisis por categoría.
- **productos:** representa el maestro de productos comercializados (identificador, nombre, precio y atributos operativos como código de barras e información tributaria cuando aplica). Es la entidad central para vincular ventas, inventario y abastecimiento.

Estructura operativa e inventario

- **sucursales:** representa la unidad física del negocio (cuando aplica), utilizada para contextualizar ventas e inventario.
- **puntos pago:** define los puntos de pago o cajas asociadas a una sucursal, aportando trazabilidad a nivel operativo.
- **inventario:** registra existencias por producto y sucursal. Su relación con **productos** y **sucursales** permite control de stock por ubicación y auditoría de niveles de inventario.

Ventas y pagos

- **ventas:** tabla cabecera que registra cada transacción (fecha/hora, total y referencias operativas como cliente, cajero y/o punto de pago cuando corresponde). Permite reconstruir el histórico de la operación.
- **detallesventas:** almacena los ítems de cada venta (producto, cantidad y precio unitario). Su relación con **ventas** y **productos** constituye la fuente primaria para construir la serie de **demanda diaria por producto**.

- **venta_pagos**: registra los pagos asociados a una venta, soportando múltiples métodos de pago en una misma transacción. Se relaciona con **ventas**.

Abastecimiento y proveedores

- **proveedores**: almacena información de proveedores.
- **compras** y **detallescompras**: registran compras a proveedores, separando cabecera e ítems (productos, cantidades), lo cual permite reconstruir entradas de mercancía cuando existe registro.
- **pedidosproveedor** y **detallespedidosproveedor**: gestionan órdenes de compra y su seguimiento (estado, fechas, recibido), separando información general del pedido y sus ítems.
- **preciosproveedor**: conserva referencias de precios por proveedor y producto, útil para análisis de variación y control de costos.

Ajustes, devoluciones y trazabilidad

- **cambiosdevoluciones**: registra eventos de cambios o devoluciones asociados a ventas y productos específicos, preservando trazabilidad sobre movimientos que pueden afectar la demanda observada y el inventario.

En conjunto, el esquema relacional permite mantener consistencia entre entidades (ventas–detalles–productos–inventario), preservar trazabilidad de la operación y soportar la construcción del conjunto de datos utilizado en los modelos de predicción de demanda.

4.3.2. Periodo de observación y tamaño muestral

Con el fin de asegurar comparabilidad entre negocios y trabajar sobre un tramo operativo estable, se definió como ventana de análisis el periodo **29 de octubre de 2025 a 21 de enero de 2026**. En dicho intervalo se dispone de **85 días calendario** con registros en ambos negocios.

Cuadro 4.1: Resumen descriptivo de los datasets crudos recolectados (ventana 29/10/2025–21/01/2026).

Dataset	# Registros (líneas)	# Días	# Productos únicos
Dataset 1 (Villa Café)	69,695	85	2,410
Dataset 2 (Mercaya)	69,694	85	2,399

4.3.3. Estructura de variables disponibles en el crudo

Antes del enriquecimiento con variables externas (Sección 4.4), los archivos crudos contienen variables principalmente transaccionales y del ítem. En particular:

CAPÍTULO 4. ADQUISICIÓN, INTEGRACIÓN Y PREPROCESAMIENTO DE DATOS

Cuadro 4.2: Diccionario de datos de los conjuntos crudos exportados desde el POS.

Columna	Descripción operativa	Tipo en CSV	Tipo en procesamiento
<code>fecha</code>	Fecha calendario de la transacción/venta (referencia temporal para agregación diaria).	Texto (YYYY-MM-DD)	<code>date</code> / <code>datetime</code>
<code>hora</code>	Hora exacta de registro de la transacción (útil para auditoría y trazabilidad; opcionalmente agregable por franjas).	Texto (HH:MM:SS...)	<code>time</code> / <code>datetime</code>
<code>productoid</code>	Identificador numérico único del producto en el POS (clave para consolidación multiserie y diferenciación entre referencias).	Entero	<code>int</code>
<code>producto</code>	Nombre/descriptor textual del producto (etiqueta legible; útil para validaciones y reportes).	Texto	<code>string</code>
<code>precio_lista_actual</code>	Precio de lista vigente del producto al momento del registro (referencia comercial).	Numérico decimal	<code>float</code> / <code>numeric</code>
<code>precio_unitario_venta</code>	Precio unitario efectivamente aplicado en la venta (puede reflejar descuentos/promociones).	Numérico decimal	<code>float</code> / <code>numeric</code>
<code>cantidad</code>	Unidades vendidas en la línea de venta (variable base para construir demanda diaria por producto).	Entero	<code>int</code>
<code>subtotal_linea</code>	Valor subtotal de la línea (consistencia contable: <code>precio_unitario_venta</code> × <code>cantidad</code>).	Numérico decimal	<code>float</code> / <code>numeric</code>

Esta estructura es suficiente para definir la demanda diaria por producto, pero no incorpora explícitamente factores de calendario ni condiciones ambientales que pueden influir en la dinámica de compra.

4.3.4. Consideraciones sobre calidad y particularidades del dato crudo

En el registro transaccional se observaron dos características comunes en contextos minoristas: (i) magnitudes variables en **cantidad** derivadas de productos por fracción o peso, y (ii) registros operativos que pueden incluir etiquetas no interpretables (p. ej., productos sin identificación real), los cuales deben tratarse explícitamente durante el alistamiento. Adicionalmente, las variables monetarias se registran en **pesos colombianos (COP)**, por lo cual es normal observar magnitudes de miles o cientos de miles.

4.4. Enriquecimiento del dataset: variables temporales, calendario y clima

Con el propósito de ampliar la capacidad explicativa del conjunto de datos y capturar condiciones del entorno que pueden incidir en el comportamiento de compra, se realizó un **enriquecimiento experimental** incorporando variables de calendario y variables meteorológicas históricas para Ibagué. Esta decisión se fundamenta en que, en pronóstico de demanda minorista, la inclusión de **variables exógenas** puede aportar señal complementaria a la memoria de ventas y reportar **mejoras en métricas de error** (p. ej., MAE/RMSE) frente a enfoques basados únicamente en el historial [19]. No obstante, este beneficio depende de una **integración temporalmente consistente**: las variables externas deben alinearse correctamente con el instante t y construirse sin introducir información futura [19].

4.4.1. Librerías empleadas y propósito

El proceso se desarrolló utilizando las siguientes librerías:

- **pandas**: carga y transformación tabular, construcción de variables temporales y uniones (*merge*) por llave horaria.
- **requests**: consumo de la API meteorológica histórica (*Open-Meteo Archive API*).
- **holidays**: construcción del calendario oficial de festivos en Colombia para los años presentes en el dataset.

- `urllib.parse.quote`: codificación de la zona horaria al construir la URL de consulta.
- `datetime.timedelta`: desplazamientos controlados para ajustar días de pago a día hábil anterior.

4.4.2. Variables agregadas por calendario

A partir de las columnas `fecha` y `hora` se construyó una marca temporal local y una llave horaria (redondeo por hora) para integrar clima. Se obtuvo:

- `dia`: día de la semana en español (Lunes–Domingo), para capturar estacionalidad semanal.
- `dia_habil`: indicador binario (1 = hábil, 0 = no hábil), donde se considera no hábil a fines de semana y festivos oficiales.
- `quincena`: indicador binario de día de pago, aproximado mediante dos fechas base por mes (día 15 y “fin de mes técnico”), ajustadas al **día hábil inmediatamente anterior** cuando la fecha cae en fin de semana o festivo.

Codificación de dia para el modelado. Aunque `dia` se construyó y reportó como etiqueta (Lunes–Domingo) por trazabilidad e interpretabilidad, en el *pipeline* de entrenamiento se transformó a una codificación **ordinal discreta** (p. ej., 1–7) para su uso en modelos tabulares. Esta decisión se adoptó por (i) **parsimonia**, evitando la expansión dimensional de un *one-hot* (7 indicadores) en un histórico acotado, y (ii) **compatibilidad con modelos no lineales** (ensambles basados en árboles), que capturan efectos semanales mediante particiones por rangos (p. ej., fines de semana vs. días hábiles) sin requerir una representación dispersa por categoría. En contraste, la codificación *one-hot* se reservó para variables **nominales** sin orden natural (véase §6.1.1).

4.4.3. Integración de variables meteorológicas: temperatura y estado del tiempo

Con el fin de incorporar el contexto climático en el análisis, se obtuvo un histórico meteorológico **horario** para coordenadas representativas de la ciudad de Ibagué, manteniendo la misma zona horaria de los registros operativos. De este histórico se conservaron únicamente dos señales, por ser las más directamente interpretables en un entorno minorista: (i) la **temperatura** y (ii) el **estado general del tiempo**.

En el caso de la temperatura, se trabajó en **grados Celsius** y se consolidó como una variable numérica continua. Para el estado del tiempo, el registro meteorológico

original se basa en **códigos estandarizados**; por tanto, estos códigos se tradujeron a una descripción textual en español (por ejemplo, *despejado*, *nublado*, *lluvia ligera/moderada/intensa*, *tormenta*), de manera que la variable fuera más legible y fácil de relacionar con el comportamiento de compra.

La integración con las ventas se realizó utilizando la **hora** como llave de correspondencia, de forma que cada transacción quedara asociada a las condiciones climáticas registradas para ese mismo momento. Para preservar la trazabilidad del histórico de ventas, se adoptó un criterio conservador: se mantuvieron todas las transacciones aun si, de manera puntual, no existía información meteorológica disponible para alguna hora específica.

Como resultado, el conjunto de datos final quedó enriquecido con variables de calendario (día, condición de día hábil y quincena) y con las dos variables meteorológicas incorporadas: **temperatura** y **descripción del estado del tiempo**.

4.4.4. Justificación del enriquecimiento

En un problema de predicción de demanda, la historia de ventas aporta una parte central de la señal, en especial a través de *lags* y agregaciones *rolling*. Sin embargo, en la práctica la demanda también responde a condiciones del entorno y a patrones propios del negocio que no siempre quedan capturados únicamente por la memoria reciente. Por esta razón, distintos trabajos han señalado que incorporar **variables exógenas** puede ayudar a explicar variaciones adicionales y, en algunos casos, mejorar el desempeño del pronóstico en métricas de error [19].

El punto crítico no es solo *qué* variables se agregan, sino *cómo* se integran. Para que el enriquecimiento sea válido desde una perspectiva metodológica, las señales externas deben quedar **alineadas temporalmente** con el instante en que se emite el pronóstico y construirse de manera **ex-ante**, evitando cualquier forma de fuga temporal que produzca estimaciones optimistas del desempeño [19].

En este trabajo, la información climática se integra a partir de registros históricos observados, lo cual permite evaluar, en un entorno offline, el **potencial informativo** del clima sobre la demanda. No obstante, en un escenario operativo el estado meteorológico del día objetivo no se conoce con certeza al momento de decidir inventario; por ello, en un despliegue real esta señal se sustituiría por **pronósticos meteorológicos disponibles ex-ante** o, en su defecto, por versiones rezagadas (p. ej., clima del día anterior). En consecuencia, los resultados asociados al clima se interpretan bajo esta consideración, sin asumir disponibilidad perfecta de la señal en operación.

Con base en lo anterior, el enriquecimiento se trató como una fase **verificable** dentro del alistamiento de datos. En lugar de asumir su utilidad de forma general, las variables exógenas se incorporaron de manera **experimental** para analizar si, en el contexto específico de las dos PYMEs con las que se trabajó, esta información

aportaba valor al momento de predecir la demanda. Esta verificación se realiza bajo el mismo protocolo temporal, con el mismo preprocesamiento y las mismas métricas de evaluación, de modo que la lectura de resultados se mantenga comparable y cualquier variación observada pueda atribuirse razonablemente a la señal contextual incorporada y no a cambios en el procedimiento [19].

4.5. Limpieza y alistamiento de los datos para modelado

La etapa de limpieza y alistamiento tuvo como finalidad convertir los registros transaccionales, capturados a nivel de **línea de venta**, en un conjunto de datos **producto–día** apto para el análisis y el modelado de pronóstico. En un problema temporal, la calidad de esta preparación es determinante: de ella depende que el modelo aprenda patrones del comportamiento real de compra y no irregularidades propias del registro (por ejemplo, fechas mal formateadas, duplicados o inconsistencias de inventario), o incluso señales que no estarían disponibles al momento de predecir.

El flujo se aplicó de manera homogénea en ambos negocios y se orientó a cinco objetivos operativos. Primero, **verificar integridad y estandarizar formatos**, asegurando que fechas, cantidades y códigos de producto fueran interpretables y comparables. Segundo, **depurar registros no útiles para el pronóstico**, eliminando o corrigiendo entradas que no representarían una venta válida o que impedirían una lectura coherente del consumo. Tercero, **consolidar la demanda diaria por producto**, agregando las transacciones del mismo día para obtener una única observación por producto y fecha. Cuarto, **completar el panel temporal**, de forma que los días sin venta quedaran representados explícitamente (demanda cero), condición necesaria para calcular *lags* sin saltos y para evaluar todas las alternativas sobre el mismo soporte temporal. Finalmente, sobre esta estructura ya estabilizada, se construyeron variables históricas basadas en la serie (principalmente *lags* y ventanas *rolling*) respetando el orden cronológico, con el propósito de preservar la consistencia *ex-ante* del pronóstico y reducir el riesgo de fuga temporal.

4.5.1. Control de faltantes y estabilidad del esquema

Como control de calidad, se cuantificaron nulos por columna y se verificó que el enriquecimiento no introdujera faltantes en variables críticas. En los archivos enriquecidos empleados, no se observaron valores nulos en las columnas utilizadas para modelado, lo cual permitió mantener un flujo determinista de transformación.

```
=== Nulos por columna ===  
Series([], dtype: int64)  
  
Filas con al menos un nulo: 0 de 147,713  
  
Filas después de eliminar nulos: 147,713  
¿Quedan nulos?: 0
```

Figura 4.2: Nulos en los DataFrame.

4.5.2. Conversión numérica y consistencia contable a nivel ítem

Se estandarizaron a tipo numérico `cantidad`, `precio_unitario_venta` y `subtotal_linea`. Posteriormente, se ejecutó un control de integridad aritmética verificando que, para cada fila j :

$$\text{subtotal_linea}_j \approx \text{cantidad}_j \times \text{precio_unitario_venta}_j, \quad (4.1)$$

dentro de una tolerancia numérica pequeña. Este control no evidenció inconsistencias, lo cual valida tanto el registro del POS como el proceso de lectura/parseo.

4.5.3. Depuración de transacciones no informativas y no interpretables

Se aplicaron dos reglas de depuración con impacto directo sobre la definición de demanda:

- **Eliminación de cantidad=0:** se removieron registros sin unidades vendidas, dado que no aportan señal de demanda y afectan agregaciones y lags.
- **Eliminación de producto ‘No Ingresado’:** se removieron registros donde el producto no fue identificado, por tratarse de observaciones no interpretables que distorsionan la distribución de demanda y la selección posterior de productos.

Adicionalmente, se verificó la ausencia de duplicados exactos a nivel fila, con el fin de descartar replicaciones artefactuales del proceso de exportación.

4.5.4. Homogeneización de variables categóricas: clima

La variable `clima_desc` proviene como texto libre, por lo que antes de usarla fue necesario **unificar su escritura**. En esta etapa se normalizó el texto (espacios, tildes y capitalización) y se consolidaron las categorías observadas durante el periodo de estudio en un **catálogo cerrado**. Con esto se evita que pequeñas diferencias de escritura (p. ej., mayúsculas o espacios adicionales) generen categorías distintas y, por tanto, se asegura consistencia entre ejecuciones.

La Tabla 4.3 documenta dicho catálogo e incluye un identificador interno por categoría únicamente con fines de registro. Sin embargo, es clave aclarar que `clima_desc` se interpreta como una variable **nominal** (no existe un orden natural entre “despejado” y “lluvia intensa”). Por esta razón, en el entrenamiento de los modelos no se utiliza este identificador como un valor numérico; en su lugar, la variable se transforma mediante **One-Hot Encoding** dentro del pipeline (Sección 7.1), para representar cada categoría sin imponer relaciones artificiales.

Cuadro 4.3: Catálogo de categorías para `clima_desc` tras normalización de texto (espacios, tildes y capitalización).

Categoría <code>clima_desc</code> (normalizada)	ID
Despejado	1
Mayormente despejado	2
Parcialmente nublado	3
Nublado	4
Llovizna ligera	5
Llovizna moderada	6
Llovizna intensa	7
Lluvia ligera	8
Lluvia moderada	9
Lluvia intensa	10

4.5.5. Definición formal de demanda diaria

De acuerdo con la unidad temporal y el horizonte principal ($h = 1$) establecidos en el Marco Teórico, la variable objetivo del pronóstico se define a **granularidad diaria** y por **producto**. En términos operativos, esto implica transformar el registro transaccional (a nivel de *línea de venta*) en una serie *producto-día* que sea consistente para el cálculo de *lags*, la validación temporal y la comparación entre modelos.

Sea p el identificador del producto (**productoid**) y sea t un día calendario (**fecha**). Para cada línea transaccional j , denótese $q_j \equiv \text{cantidad}_j$ como la cantidad vendida y $m_j \equiv \text{subtotal_linea}_j$ como el valor monetario asociado a dicha línea. Se define el conjunto de índices de líneas que corresponden al producto p en el día t como:

$$\mathcal{J}(p, t) = \{ j : \text{productoid}_j = p \wedge \text{fecha}_j = t \}. \quad (4.2)$$

Con base en este conjunto, la **demanda diaria observada** del producto p en el día t se define como la suma de las unidades vendidas en todas las líneas del día:

$$y_{p,t} = \sum_{j \in \mathcal{J}(p,t)} q_j. \quad (4.3)$$

Estas definiciones formalizan la agregación necesaria para pasar del nivel transaccional al nivel *producto–día*. En particular, $y_{p,t}$ representa el total de unidades vendidas del producto p durante el día t , que es la variable utilizada como objetivo en el modelado de demanda.

4.5.6. Construcción del dataset producto–día y control de integridad

A partir de las definiciones anteriores, se consolidó el dataset a nivel **producto–día** agrupando por **fecha** y **productoid**, agregando: suma diaria de **cantidad** (demanda). Con fines de estabilidad, las variables exógenas se mantuvieron constantes por día.

Adicionalmente, se verificó nuevamente la consistencia contable a nivel producto–día, descartando discrepancias atribuibles a errores de conversión numérica o agregación.

4.5.7. Simplificación del esquema base para modelado

Una vez definida la estructura *producto–día*, el siguiente paso consistió en **simplificar** el conjunto base de datos para que fuera más claro, consistente y fácil de usar en los experimentos de pronóstico. En esta fase no se buscó “enriquecer” la información, sino **depurarla**: conservar únicamente los campos que aportan valor directo a la demanda diaria y a su contexto inmediato, y retirar aquellos que pertenecen a otras capas de la operación (caja, contabilidad o detalle transaccional) y que, para el objetivo de este trabajo, no resultaban imprescindibles.

Bajo este criterio, se descartaron variables **horarias** como **hora**. Aunque este tipo de campo es útil cuando se estudia el comportamiento dentro del día, aquí la decisión operativa se formula a nivel diario ($h = 1$), por lo que, después de

consolidar la demanda por fecha, el detalle intradía deja de aportar información adicional y puede introducir variaciones que no son comparables entre jornadas. De manera similar, se eliminaron variables **monetarias** como `subtotal_linea` y `precio_unitario_venta`. Estos valores reflejan dinámicas comerciales (ajustes de precio, promociones, redondeos) que son relevantes en análisis financieros, pero que no son necesarias para representar la demanda como **unidades vendidas**. Mantenerlas en el conjunto base habría mezclado la señal de consumo con una señal distinta (precio), lo cual exigiría un tratamiento metodológico adicional si se quisiera interpretarlas con rigor dentro del mismo experimento.

Como resultado, el esquema diario final que se conservó para modelado quedó compuesto por:

```
fecha, productoid, cantidad, producto, dia, dia_habil, quincena, temp_c,  
clima_desc.
```

En este conjunto, `productoid` mantiene el vínculo unívoco con el catálogo y permite agrupar de forma consistente la información por producto. La columna `producto` se conservó como apoyo de lectura y auditoría (por ejemplo, para validar rápidamente resultados o revisar casos puntuales), aun cuando el identificador operativo del modelo es `productoid`. Finalmente, las variables de calendario (`dia`, `dia_habil`, `quincena`) y las variables climáticas (`temp_c`, `clima_desc`) se retuvieron como contexto disponible para ser evaluado experimentalmente dentro del mismo protocolo de comparación.

4.5.8. Completado del panel producto $\times$ día

Para que los *lags* representaran **días calendario consecutivos** (y no únicamente “filas consecutivas con ventas”), se construyó un **panel completo** a partir del producto cartesiano entre: (i) todos los días del rango observado y (ii) todos los productos presentes en el periodo. Posteriormente, se reindexó el conjunto de datos para garantizar la existencia de **una fila por producto y día**, incluso cuando no hubo transacciones.

En la práctica, esto implicó dos decisiones de imputación coherentes con el significado de cada variable:

- **Demanda diaria.** Se asignó `cantidad = 0` cuando no existieron ventas del producto en la fecha correspondiente, de modo que los días sin movimiento quedaran explícitos en la serie.
- **Contexto diario.** Las variables exógenas se reintegraron por `fecha` para que cada día contara con su contexto completo. En particular, `temp_c` corresponde a la **temperatura promedio del día** y `clima_desc` al **estado modal del día**.

(la condición climática más frecuente), evitando valores faltantes y manteniendo una caracterización diaria única.

Este paso es determinante porque, sin un panel completo, los *lags* quedarían “saltados” por días sin ventas y terminarían representando distancias irregulares en el tiempo. Al imponer la continuidad del calendario, los *lags* y *rollings* se calculan sobre una base temporal consistente, lo que reduce errores sistemáticos y mejora la validez de los experimentos de pronóstico.

4.5.9. Construcción de variables históricas: *lags* y *rolling*

Con el panel *producto-día* ya completo, el siguiente paso fue incorporar información del **historial reciente de ventas** al conjunto de entrada. En la práctica, un pronóstico diario no se apoya únicamente en el contexto del día, sino también en la **memoria inmediata** de la serie: lo que ocurrió ayer y en los últimos días. Por esta razón, se generaron variables históricas basadas en *lags* de la demanda y en *rolling*, con el propósito de representar de forma explícita el nivel reciente y el comportamiento acumulado de cada producto, manteniendo un enfoque trazable y coherente con el objetivo operativo del trabajo [20].

En concreto, se construyeron tres grupos de variables por producto:

- **Lags de demanda:** `lag_1 ... lag_7`, que corresponden a la demanda observada entre 1 y 7 días atrás.
- **Rolling acumulado:** `rolling_1 ... rolling_7`, que resumen la demanda acumulada de los últimos k días.
- **Rolling promedio:** `rolling_mean_1 ... rolling_mean_7`, que describen el promedio de la demanda en los últimos k días.

Se adoptó una ventana máxima de **7 días** por dos razones principales. Primero, porque en el comercio minorista es frecuente observar comportamientos asociados al ciclo semanal (por ejemplo, diferencias sistemáticas entre días hábiles y fines de semana), y una ventana de una semana permite capturar esa dinámica sin introducir supuestos adicionales [20]. Segundo, porque ampliar la ventana exige un histórico más largo para mantener cálculos consistentes; dado el periodo disponible, se priorizó una configuración estable y comparable a lo largo de todo el análisis.

Finalmente, estas variables requieren contar con historia suficiente para ser calculadas. Por ello, para evitar el **arranque frío**, se excluyeron los primeros 7 días del periodo al conformar el conjunto de entrenamiento, garantizando que todas las observaciones utilizadas dispusieran de *lags* y *rolling* completos.

4.5.10. Reducción de redundancias en *features*

Se identificó una redundancia **exacta** cuando $k = 1$: una ventana de un solo día no agrega información, por lo que `rolling_1` y `rolling_mean_1` coinciden con `lag_1`. Para reducir dimensionalidad sin pérdida de señal y evitar colinealidad trivial, se eliminaron `rolling_1` y `rolling_mean_1`, manteniendo `lag_1`.

4.5.11. Descomposición de la fecha en componentes calendario

Con el panel *producto-día* ya consolidado y con las variables históricas (*lags* y *rolling*) construidas, se incorporaron **atributos calendario explícitos** derivados de **fecha**. El objetivo de este paso fue doble: por un lado, facilitar que los modelos tabulares identifiquen patrones estacionales de **mediano plazo** (mensuales y anuales) que no siempre quedan completamente reflejados en el ciclo semanal; y, por otro, mantener un preprocesamiento **simple, interpretable y trazable**.

En particular, se convirtió la columna `fecha` a tipo `datetime` y se extrajeron tres componentes:

- **ano**: año calendario de la observación.
- **mes**: mes calendario (1–12).
- **dia-num**: día del mes (1–31).

Tratamiento de dia-num como entero discreto. El atributo `dia-num` se conservó como **entero** (1–31) en lugar de expandirse mediante *one-hot*. Esta elección responde a criterios de **parsimonia** y control de dimensionalidad (evitando hasta 31 indicadores adicionales), particularmente relevantes en un histórico acotado, y a la naturaleza de los modelos no lineales evaluados (ensambles basados en árboles), los cuales pueden explotar **particiones por rangos** para capturar efectos intrames (p. ej., inicio/cierre de mes) sin requerir una representación dispersa por categoría. En consecuencia, la codificación *one-hot* se reservó para variables **nominales** sin orden natural (véase §6.1.1).

Esta descomposición se justifica por dos razones principales. Primero, aunque el comportamiento semanal se representa con `dia` (día de la semana) y el comportamiento reciente queda capturado por *lags* y *rolling*, en retail pueden aparecer variaciones asociadas al **inicio y cierre de mes** (cambios en hábitos de compra, cortes de presupuesto, reposiciones) y a **temporadas del año** (periodos específicos con dinámica distinta). Incluir `ano`, `mes` y `dia-num` permite que el modelo aprenda estos patrones sin introducir transformaciones complejas ni supuestos adicionales.

Segundo, trabajar con la **fecha** como un dato completo puede inducir lecturas no deseadas en modelos tabulares: dependiendo del tratamiento, puede terminar

actuando como un identificador del tiempo absoluto o forzar relaciones artificiales entre días consecutivos. En cambio, al separar la fecha en componentes numéricos, la señal temporal se entrega de forma más controlada y comprensible, manteniendo consistencia con el resto de variables calendario ya definidas.

Adicionalmente, esta decisión se implementó atendiendo una **sugerencia del director de tesis**, quien recomendó explicitar la dimensión temporal mediante atributos calendario interpretables (en lugar de conservar la marca temporal completa) para favorecer la trazabilidad del preprocesamiento y la compatibilidad con los modelos supervisados evaluados.

Finalmente, una vez extraídos los componentes, se eliminó la columna original **fecha**. Con ello, el conjunto de entrada queda expresado únicamente mediante variables numéricas/categóricas y la historia de demanda, reduciendo el riesgo de que la fecha funcione como una codificación implícita del tiempo absoluto. En consecuencia, la información temporal utilizada por el modelo queda representada por **dia**, **dia_habil**, **quincena**, **ano**, **mes** y **dia-num**, junto con *lags* y *rolling*, mientras que el resto de atributos del panel (producto y clima, cuando aplica) se conserva sin cambios.

4.5.12. Resultado del alistamiento

Como resultado del alistamiento, los registros exportados desde el POS se transformaron en un dataset a nivel **producto-día** con panel temporal completo, demanda diaria y un conjunto consistente de variables de contexto y de memoria (*lags* y *rolling*). Esta estructura deja la información en un formato directamente utilizable para el EDA y, posteriormente, para el entrenamiento y evaluación de modelos de predicción de demanda bajo partición temporal (sin romper el orden cronológico).

En términos del esquema final, ambos negocios quedaron con una estructura homogénea orientada a modelado tabular, compuesta por: (i) identificadores y soporte de auditoría (**productoid**, **producto**), (ii) variable objetivo diaria (**cantidad**), (iii) calendario y condiciones externas (**dia**, **dia_habil**, **quincena**, **temp_c**, **clima_desc**), (iv) memoria de demanda (**lag_1-lag_7**, **rolling_2-rolling_7**, **rolling_mean_2-rolling_mean_7**) y (v) componentes de fecha (**ano**, **mes**, **dia-num**). En consecuencia, cada fila representa un par (producto, día) e integra la demanda observada, su contexto y la historia inmediata necesaria para pronosticar horizontes cortos.

Cuadro 4.4: Esquema final de variables del dataset producto-día y tipos de dato.

Variable(s)	Tipo de dato		Rol en el modelado
productoid	Entero (int)		Identificador de producto (enfoque global multiserie).
producto	Texto (string)		Etiqueta legible para auditoría/reportes.
cantidad	Numérico (float)	real	Variable objetivo (demanda diaria).
dia	Entero (int)		Calendario: día de la semana.
dia_habil	Binario (0/1)		Calendario: día hábil / no hábil.
quincena	Binario (0/1)		Calendario: señal de pago/quincena.
temp_c	Numérico (float)	real	Exógena: temperatura agregada diaria.
clima_desc	Catórica (string/category)		Exógena: condición del tiempo.
lag_1-lag_7	Numérico (float)	real	Memoria: <i>lags</i> de demanda.
rolling_2-rolling_7	Numérico (float)	real	Memoria: <i>rolling</i> (agregación móvil).
rolling_mean_2-rolling_mean_7	Numérico (float)	real	Memoria: promedio móvil (<i>rolling mean</i>).
ano	Entero (int)		Calendario: ciclo anual.
mes	Entero (int)		Calendario: ciclo mensual.
dia-num	Entero (int)		Calendario: patrones intramensuales.

Para **Villa Café**, el dataset resultante contiene **206,975 filas** y **30 columnas**, correspondientes al periodo **29/10/2025–21/01/2026**. Para **Mercaya**, el dataset final contiene **206,550 filas** y **30 columnas** en el mismo intervalo, manteniendo el mismo esquema de variables. Estas matrices constituyen el insumo directo para el análisis exploratorio, la selección de productos de alta rotación y la etapa de modelado.

Capítulo 5

Análisis Exploratorio y Selección de Características y Productos

5.1. EDA exploratorio orientado a pronóstico: selección de Top-10 y diagnóstico de variables

El análisis exploratorio (EDA) se desarrolló como una etapa **orientada a decisión** dentro del protocolo experimental: su función fue verificar que el dataset construido fuera consistente con un escenario de pronóstico diario y, a partir de esa evidencia, dejar definidas reglas reproducibles para el entrenamiento posterior. En particular, el EDA permitió (i) comprobar la **integridad del panel *producto-día*** (incluyendo la representación explícita de días sin venta), (ii) caracterizar la señal en términos de **rotación** e **intermitencia**, y (iii) establecer un subconjunto de productos sobre el cual el proceso de selección de variables y ajuste de modelos fuera viable bajo el presupuesto computacional disponible.

En coherencia con el alcance del trabajo, se adoptó la decisión de entrenar y comparar modelos sobre un subconjunto de **10 productos por negocio**. Esta elección responde a dos criterios complementarios. Primero, un criterio de **viabilidad experimental**: al restringir el universo de productos, se reduce el costo de ejecutar validación temporal, *feature search* y *tuning* de hiperparámetros sin sacrificar rigor metodológico. Segundo, un criterio de **relevancia operativa**: el pronóstico se concentra en referencias que, por su rotación y comportamiento de venta, tienden a impactar con mayor fuerza las decisiones de reposición e inventario. La selección del Top-10 se realizó de forma reproducible y se aplicó de manera análoga en ambos negocios (**Mercaya** y **Villa Café**).

El EDA se ejecutó sobre el dataset *producto-día* obtenido tras el alistamiento (Sección 4.5). En términos de estructura, dicho dataset incluye la variable objetivo diaria **cantidad**, variables exógenas y variables endógenas. Para mantener trazabilidad, el análisis se apoyó en las mismas columnas que luego alimentan el pipe-

line: calendario (dia, dia_habil, quincena, ano, mes, dia_num), clima (temp_c, clima_desc) y memoria temporal mediante *lags* (lag_) y agregaciones *rolling* (rolling_, rolling_mean_).

A partir de esta base, las subsecciones siguientes presentan: (i) el criterio aplicado para seleccionar el Top-10 en cada negocio, (ii) un diagnóstico temporal agregado de las series completas, (iii) el control de representatividad por familia para consolidar el subconjunto final, y (iv) un diagnóstico inicial de variables, enfocado en identificar señales útiles y posibles redundancias antes de ejecutar la etapa de entrenamiento y comparación de modelos.

5.1.1. Selección de productos objetivo: rotación, regularidad e índice compuesto

El portafolio real del negocio incluye miles de referencias, por lo que no es viable —ni metodológicamente conveniente— entrenar y evaluar el pronóstico diario sobre *todo* el surtido. En la práctica, una parte importante de los productos presenta ventas muy esporádicas: muchos días en cero y pocos días con compra. Tal como se definió en el Marco Teórico para demanda intermitente, ese patrón suele producir una señal diaria poco informativa y vuelve más difícil tanto el aprendizaje como la interpretación de resultados [9]. Por esta razón, antes de entrenar modelos, se estableció un criterio para priorizar productos con dos características deseables: **movimiento suficiente** y **regularidad**.

Antes de construir el ranking por producto, se realizó un diagnóstico descriptivo del portafolio en ambos negocios con el fin de verificar si el conjunto completo ofrecía condiciones homogéneas para pronóstico diario o si, por el contrario, presentaba concentración e intermitencia que justificaran una priorización inicial. En ambos casos, el análisis se ejecutó sobre el panel *producto-día* ya consolidado, preservando explícitamente los días con venta cero y manteniendo consistencia con la estructura que posteriormente alimenta el pipeline de entrenamiento y evaluación.

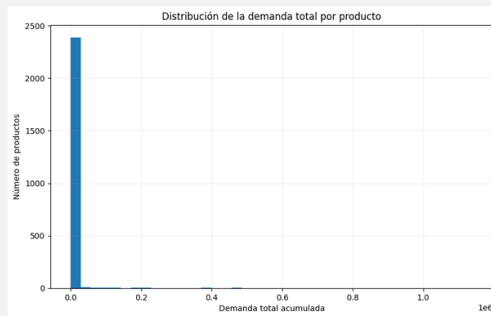
Cuadro 5.1: Resumen descriptivo del portafolio previo a la selección del Top-10

Indicador	Mercaya	Villa Café
Productos únicos	2430	2435
Días del periodo	85	85
Filas del panel producto-día	206550	206975
Demanda total global	10169740	10232900
Mediana de demanda total por producto	9	9
Mediana de días con venta por producto	7	7
Mediana de racha máxima consecutiva	2	2
Participación acumulada del Top-10	63.4 %	61.1 %
Participación acumulada del Top-20	81.5 %	80.9 %
Productos nunca vendidos	1.3 %	1.1 %
Productos con frecuencia de venta ≤ 10 %	54.9 %	55.4 %

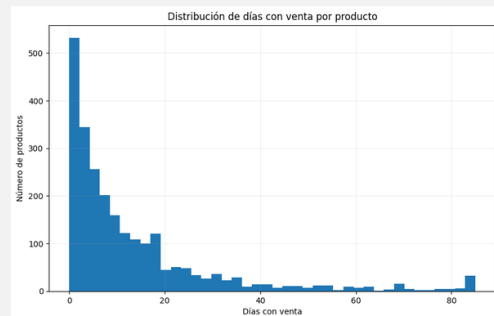
El Cuadro 5.1 muestra un patrón consistente en ambos negocios: una fracción reducida del catálogo concentra gran parte de la demanda, mientras que más de la mitad de las referencias presenta baja frecuencia de venta. En consecuencia, tratar todo el surtido como homogéneo resultaría poco conveniente tanto para el entrenamiento como para la interpretación. Por ello, antes de modelar, se justificó una priorización inicial orientada a referencias con mayor relevancia operativa y menor intermitencia.

Mercaya - Paso 0

(a) Hist. demanda total



(b) Hist. días con venta



(c) Concentración acumulada

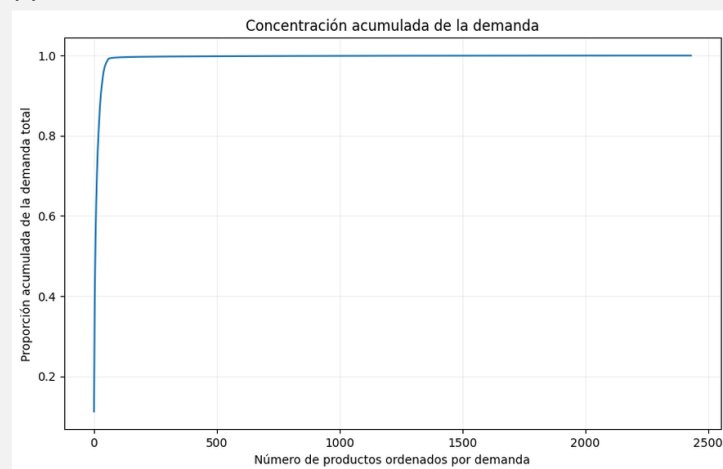


Figura 5.1: Mercaya: diagnóstico descriptivo previo del portafolio. La figura resume la distribución del volumen acumulado, la frecuencia de venta y la concentración de la demanda.

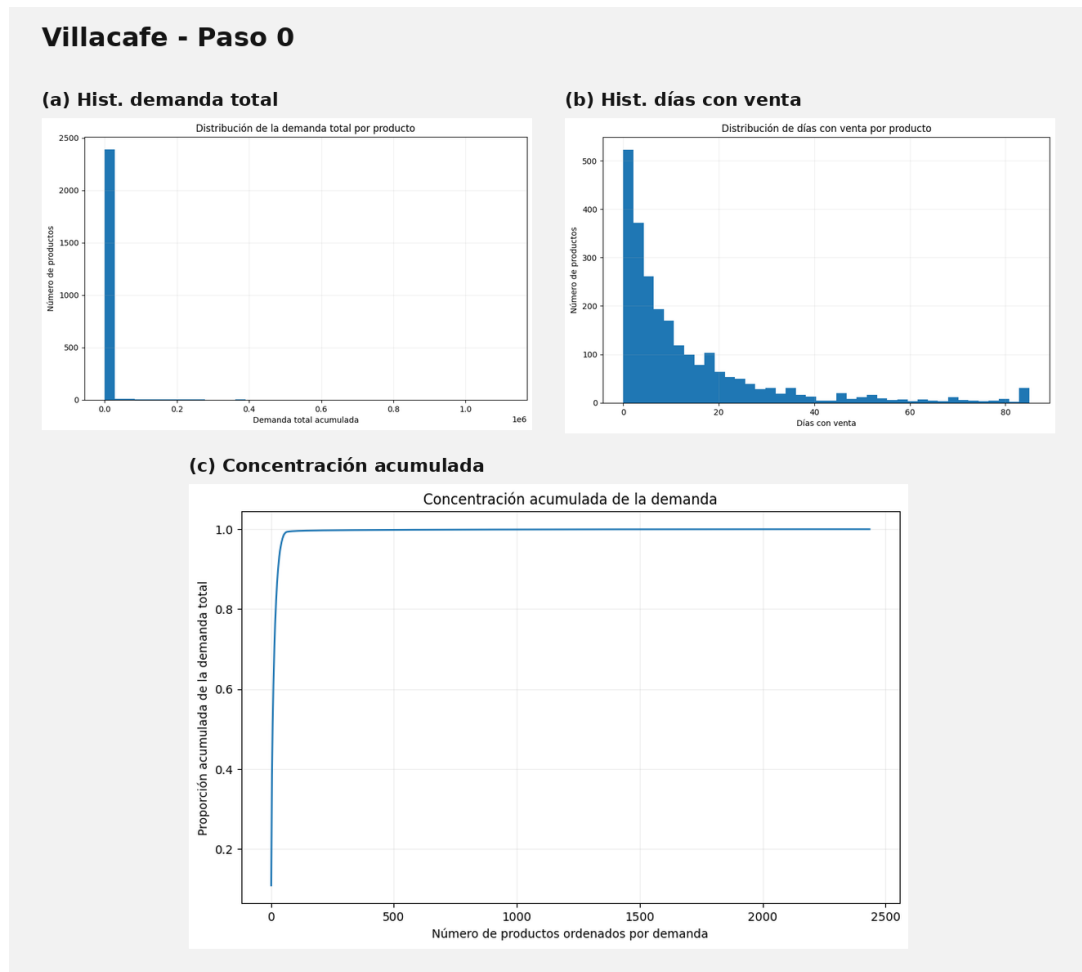


Figura 5.2: Villa Café: diagnóstico descriptivo previo del portafolio. La figura resume la distribución del volumen acumulado, la frecuencia de venta y la concentración de la demanda.

La Figura 5.1 y la Figura 5.2 precisan esta heterogeneidad desde tres ángulos complementarios. En primer lugar, el histograma de demanda acumulada evidencia una distribución fuertemente asimétrica: la mayoría de las referencias presenta volúmenes bajos, mientras un subconjunto reducido concentra valores considerablemente superiores. En segundo lugar, el histograma de días con venta muestra que una porción amplia del catálogo se activa en pocos días del periodo, lo que confirma la presencia de intermitencia. Finalmente, la curva de concentración acumulada indica que una fracción relativamente pequeña de productos explica una parte importante del movimiento total. En conjunto, estos resultados respaldan la necesidad de construir un criterio reproducible de selección inicial antes de pasar al entrenamiento comparativo.

Para cada `productoid` se construyó un resumen con dos medidas operativamente relevantes. La primera fue el **volumen acumulado** (`total_demand`), obtenido como la suma de `cantidad` en toda la ventana de análisis; en términos prácticos, esta medida responde a la pregunta *¿cuánto se mueve este producto en total?*. La segunda

fue la **regularidad** (`zero_rate`), calculada como la proporción de días sin venta, esto es, `days_zero/days_total`; esta medida resume *con qué frecuencia aparece venta en el día a día*. Un `zero_rate` bajo implica menor intermitencia y, por tanto, una señal diaria potencialmente más estable para pronóstico [9].

Con estas dos dimensiones se definió un ranking cuyo propósito no fue introducir complejidad innecesaria, sino formalizar una decisión que en operación suele asumirse de manera intuitiva: priorizar lo que más rota y, al mismo tiempo, lo que se vende con mayor constancia. Para ello, se estabilizó el volumen acumulado mediante una transformación logarítmica suave:

$$\text{sales_log} = \log(1 + \text{total_demand}).$$

Posteriormente, tanto `sales_log` como la continuidad de venta ($1 - \text{zero_rate}$) se llevaron a una escala comparable mediante un **ranking percentil**, donde 1 representa el mejor desempeño relativo dentro del portafolio:

$$\text{score_sales} = \text{rank_pct}(\text{sales_log}), \quad \text{score_nonzero} = \text{rank_pct}(1 - \text{zero_rate}).$$

Finalmente, el puntaje compuesto se definió como el promedio simple de ambos sub-puntajes:

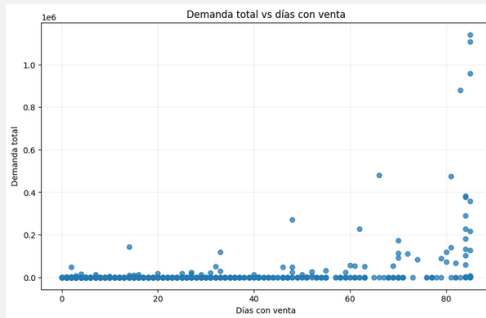
$$\text{score_final} = 0,5 \cdot \text{score_sales} + 0,5 \cdot \text{score_nonzero}.$$

Esta formulación busca un balance deliberado entre **rotación** y **regularidad**, manteniendo un criterio simple, trazable y reproducible para la selección. Asimismo, se excluyeron los productos con `total_demand=0`, dado que no aportan evidencia histórica útil para la priorización ni para el entrenamiento posterior.

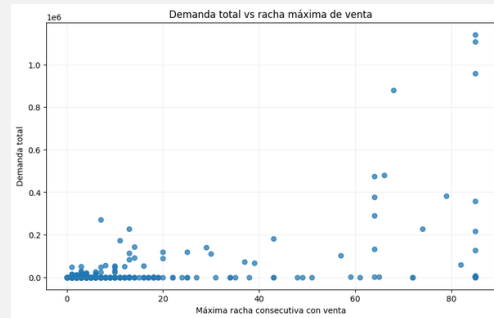
Aunque el índice compuesto se definió formalmente con base en volumen y regularidad, se complementó el análisis con una lectura descriptiva de continuidad temporal para fortalecer su interpretación operativa. En particular, se revisó la continuidad de venta mediante el número de días con venta (`days_sold`) y la racha máxima consecutiva (`max_streak_sold`). Estas medidas no sustituyen el criterio formal basado en `zero_rate`, pero permiten observar con mayor detalle la persistencia temporal de las referencias más representativas.

Mercaya - Paso 2

(a) Demanda vs días



(b) Demanda vs racha



(c) Frecuencia vs continuidad

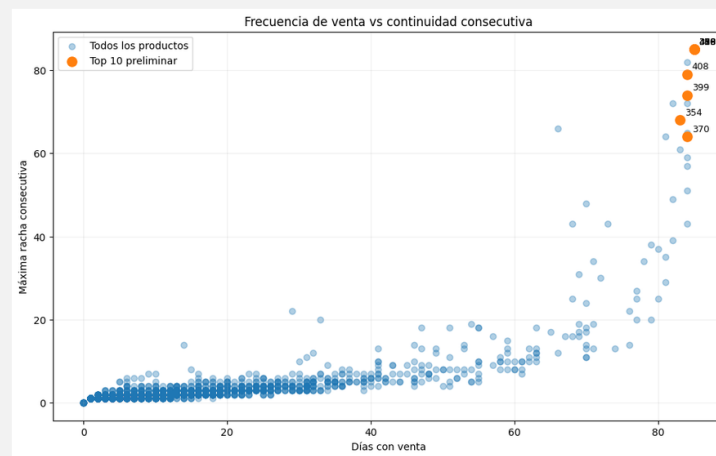


Figura 5.3: Mercaya: relación entre volumen acumulado y continuidad de venta. La figura muestra la relación entre demanda total, número de días con venta y racha máxima consecutiva, además de resaltar el Top-10 preliminar.

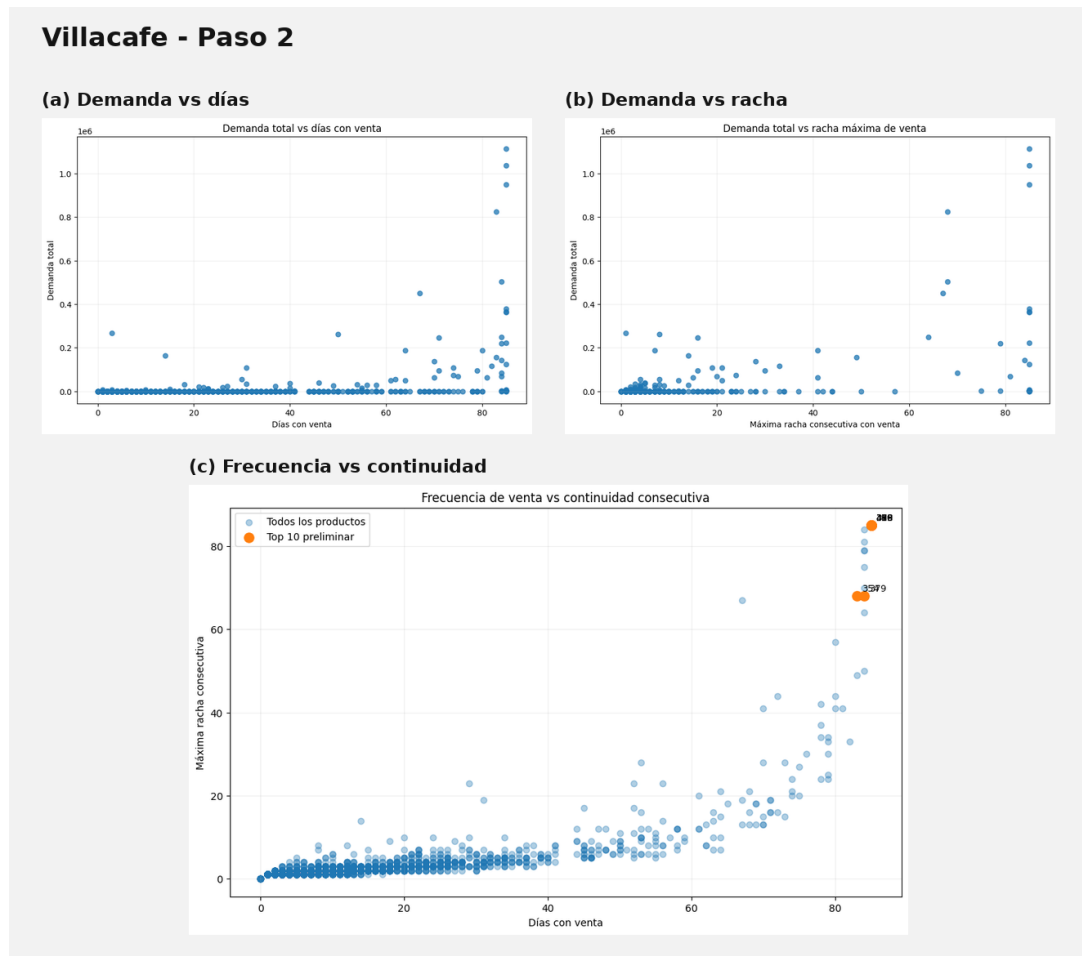


Figura 5.4: Villa Café: relación entre volumen acumulado y continuidad de venta. La figura muestra la relación entre demanda total, número de días con venta y racha máxima consecutiva, además de resaltar el Top-10 preliminar.

La Figura 5.3 y la Figura 5.4 permiten profundizar esta lectura. El panel de demanda total frente a días con venta muestra que, en términos generales, las referencias de mayor volumen tienden también a venderse en una mayor cantidad de días. De forma similar, el panel de demanda total frente a racha máxima evidencia que los productos más relevantes no solo venden más, sino que además tienden a sostener secuencias de venta más largas. Por su parte, el panel de frecuencia frente a continuidad resume la relación entre ambas dimensiones y resalta el Top-10 preliminar, mostrando que las referencias priorizadas se ubican precisamente en la zona donde convergen mayor recurrencia y mayor persistencia temporal.

Esta evidencia sugiere que la rotación y la continuidad están fuertemente asociadas, aunque no sean equivalentes. En términos descriptivos, esta lectura ayuda a evitar dos extremos: seleccionar solo por volumen podría incorporar productos con ventas concentradas en pocos episodios, mientras que seleccionar solo por continuidad podría favorecer referencias constantes pero de menor relevancia operativa. El índice compuesto, por tanto, conserva su utilidad como punto medio entre impacto comercial

y estabilidad temporal. Asimismo, al contrastar el ranking compuesto con criterios aislados, se observó que el Top-10 final preserva el núcleo de productos de mayor impacto sin quedar determinado por una sola dimensión.

En síntesis, las evidencias descriptivas de esta subsección muestran que el ranking inicial no surge de una decisión arbitraria, sino de una estructura de datos donde la demanda está concentrada, la intermitencia es elevada y las referencias de mayor peso operativo tienden también a exhibir mayor continuidad temporal.

5.1.2. Diagnóstico temporal agregado de las series completas

Como complemento al diagnóstico del portafolio y al criterio de selección del Top-10, se incorporó una lectura explícita de la evolución temporal completa de la demanda agregada por negocio. Este análisis se realizó sobre la suma diaria de **cantidad** en cada fecha del periodo de observación, con el propósito de caracterizar el nivel general de la serie, su variabilidad, la presencia de picos puntuales y la posible existencia de patrones semanales antes del entrenamiento de modelos.

En ambos casos, la serie agregada cubre los mismos **85 días** comprendidos entre el **29 de octubre de 2025** y el **21 de enero de 2026**, en coherencia con la ventana de observación definida para el estudio. Esta comparación directa entre negocios resulta relevante porque permite distinguir diferencias de dinámica temporal sin introducir sesgos por longitud o recorte desigual del periodo analizado.

Cuadro 5.2: Resumen descriptivo de la serie agregada diaria por negocio

Indicador	Mercaya	Villa Café
Número de días	85	85
Promedio diario	119643.99	120387.04
Mediana diaria	115524.00	113608.00
Coefficiente de variación	0.251	0.321
Mínimo diario	61292.00	68896.00
Fecha del mínimo	2025-11-08	2026-01-01
Máximo diario	239458.00	354375.00
Fecha del máximo	2025-12-31	2025-12-24
Día de mayor promedio semanal	Martes	Domingo
Día de menor promedio semanal	Jueves	Jueves

El Cuadro 5.2 muestra que ambos negocios operan sobre un nivel medio de demanda diaria similar, pero difieren en la variabilidad relativa de la serie. En particular, Villa Café presenta un coeficiente de variación superior y un pico máximo más pronunciado dentro del mismo horizonte temporal, lo que sugiere una trayectoria agregada más irregular y con episodios de demanda más extremos respecto a su

nivel típico. En contraste, Mercaya exhibe una dinámica algo más contenida, aunque también con picos relevantes hacia el cierre del año.

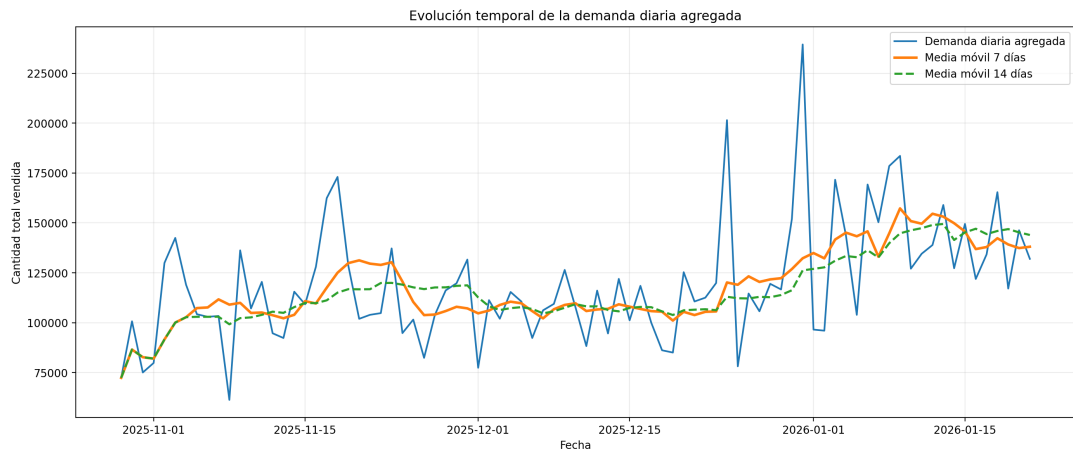


Figura 5.5: Mercaya: evolución temporal de la demanda diaria agregada, con medias móviles de 7 y 14 días.

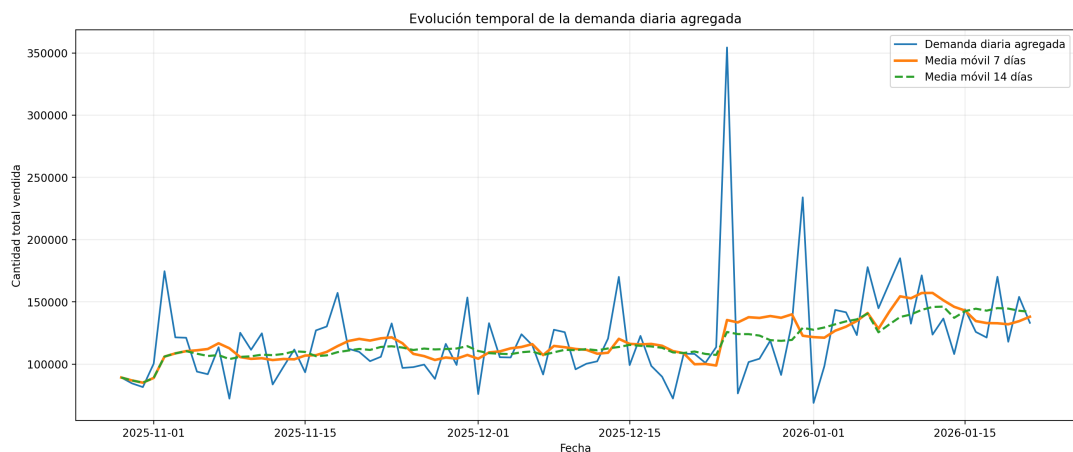


Figura 5.6: Villa Café: evolución temporal de la demanda diaria agregada, con medias móviles de 7 y 14 días.

La Figura 5.5 y la Figura 5.6 permiten observar que la serie agregada no se comporta como una sucesión puramente aleatoria de valores diarios. En ambos negocios se aprecia un nivel general relativamente estable, sobre el cual emergen picos puntuales y fases cortas de incremento o descenso. La inclusión de medias móviles de 7 y 14 días facilita distinguir la variación diaria de alta frecuencia frente a cambios algo más persistentes en el nivel agregado. Desde una lectura operativa, este resultado sugiere que existe estructura temporal útil para pronóstico, pero también episodios abruptos que pueden dificultar el seguimiento de modelos más conservadores.

A nivel comparativo, Mercaya exhibe una trayectoria agregada algo más estable, mientras que Villa Café presenta oscilaciones de mayor amplitud y un pico espe-

cialmente marcado hacia el cierre de diciembre. Esta diferencia es consistente con la lectura posterior del modelado, en la que Mercaya se explica principalmente por memoria reciente y calendario compacto, mientras que en Villa Café la inclusión de `clima_desc` aporta valor adicional al capturar regímenes de demanda menos estables.

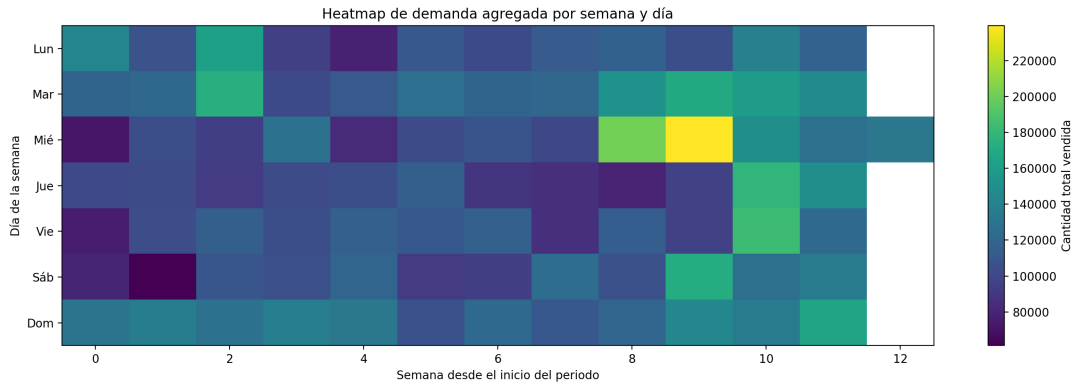


Figura 5.7: Mercaya: heatmap de demanda agregada por semana y día de la semana.

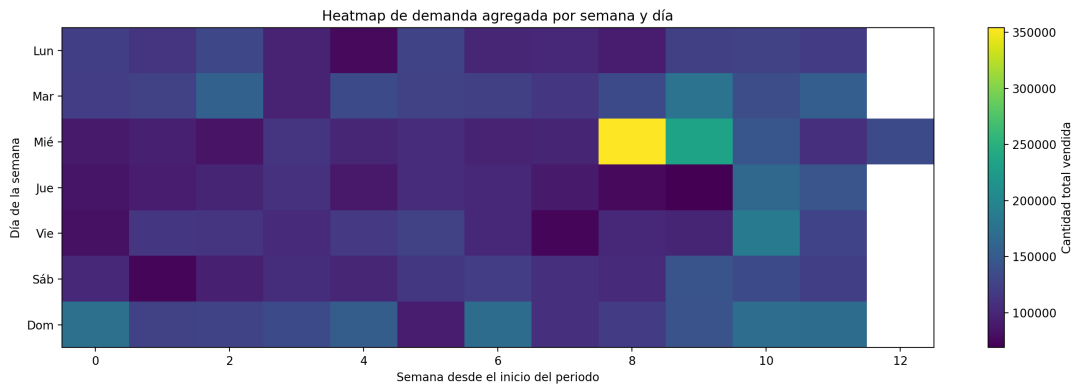


Figura 5.8: Villa Café: heatmap de demanda agregada por semana y día de la semana.

La Figura 5.7 y la Figura 5.8 complementan esta lectura al reorganizar la serie por semana y día de la semana. Esta visualización permite identificar con mayor claridad si los picos aparecen de forma aislada o si tienden a concentrarse en determinados tramos del calendario. En ambos negocios se observa una estructura intra-semanal visible, aunque no idéntica, lo que refuerza la pertinencia de incorporar variables de calendario en la fase de modelado. Asimismo, la recurrencia de episodios de mayor intensidad confirma que la serie agregada presenta cambios de nivel que no se explican únicamente por ruido aleatorio.

En conjunto, este diagnóstico temporal agregado fortalece el EDA por dos razones. Primero, muestra de forma explícita que la demanda completa del negocio sí posee estructura temporal interpretable antes del entrenamiento. Segundo, conecta de manera más directa la exploración inicial con la lógica del pipeline posterior: la

utilidad de los rezagos y promedios móviles, el valor del calendario y, en el caso de Villa Café, la conveniencia de considerar señales contextuales adicionales cuando la serie exhibe mayor variabilidad y cambios de régimen más notorios.

5.1.3. Criterio de representatividad: control por familia

Tras el ranking inicial, se incorporó una restricción de representatividad con el fin de evitar que el Top-10 quedara dominado por un único tipo de producto. Se definió una variable `familia` a partir del texto de `producto` mediante normalización y agrupación semántica básica, y se construyó el Top-10 recorriendo el ranking en orden, aceptando candidatos mientras cada familia no excediera **dos** apariciones.

Esta decisión se justifica por tres razones: (i) **reducción de redundancia**, ya que presentaciones muy similares tienden a compartir patrones y podrían inducir un conjunto Top-10 poco diverso; (ii) **cobertura operativa**, al incorporar referencias críticas con dinámicas distintas (p. ej., plaza, huevos, bebidas, cigarrillos); y (iii) **validez del diagnóstico**, dado que un EDA basado en una sola familia puede sesgar la lectura de variables hacia un comportamiento particular. En consecuencia, la restricción por familia actúa como un criterio complementario de representatividad y no como sustituto del ordenamiento por `score_final`.

5.1.4. Top-10 final por negocio y validación del panel producto—día

Aplicando las reglas anteriores (ranking por `score_final` + restricción por familia) se obtuvieron los Top-10 finales por negocio:

Cuadro 5.3: Top-10 final de productos en Mercaya

Rank	Productoid	Producto
1	378	Papa Negra x gr (plaza)
2	410	Pollo Pechuga x gr (plaza)
3	223	Huevo AAA Unidad (huevos)
4	2116	Arepas Unidad (arepas)
5	2009	Arepas Paquete Pequeño (arepas)
6	224	Huevo AAA Cubeta x30 (huevos)
7	2204	Cigarrillo Unidad (cigarrillo)
8	594	Éxito Leche Entera x900ml (exito)
9	1345	Costeña Lata x310cm3 (costeña)
10	1519	Coca Cola Plástico x1.5l (coca)

Cuadro 5.4: Top-10 final de productos en Villa Café

Rank	Productoid	Producto
1	410	Pollo Pechuga x gr (plaza)
2	378	Papa Negra x gr (plaza)
3	223	Huevo AAA Unidad (huevos)
4	2116	Arepas Unidad (arepas)
5	224	Huevo AAA Cubeta x30 (huevos)
6	2009	Arepas Paquete Pequeño (arepas)
7	2204	Cigarrillo Unidad (cigarrillo)
8	594	Éxito Leche Entera x900ml (exito)
9	20	Mustang Rothmans x10 (mustang)
10	1345	Costeña Lata x310cm3 (costeña)

Posteriormente, el dataset completo se filtró a estos identificadores. Un resultado clave del EDA fue la **validación de integridad del panel**: el subconjunto Top-10 quedó con **850 filas por negocio**, consistente con 85 días $\times$ 10 productos. Esta verificación es crítica para pronóstico diario porque confirma que existe **una observación por producto y día** (incluyendo ceros), condición necesaria para construir lags coherentes, ventanas móviles y un protocolo de evaluación temporal sin fuga [9].

5.1.5. Diagnóstico de variables explicativas: correlaciones como lectura preliminar

Una vez consolidado el panel Top-10, se realizó un **diagnóstico exploratorio** para entender qué tan alineadas están las variables explicativas con la demanda diaria (*cantidad*). Este análisis no se usó como un mecanismo de selección definitiva, sino como una **primera lectura** para: (i) verificar que la historia de demanda (*lags* y ventanas *rolling*) estuviera capturando señal temporal plausible, y (ii) identificar posibles redundancias entre predictores antes de avanzar a la evaluación formal dentro del pipeline [12].

En coherencia con el Marco Teórico, se evitó depender de una única medida de asociación. Por ello, se calcularon tres correlaciones complementarias, que tienden a comportarse distinto frente a no linealidades, asimetrías y valores extremos:

- **Pearson**, como referencia de asociación lineal.
- **Spearman**, basada en rangos, útil cuando la relación es monotónica pero no necesariamente lineal.

- **Kendall**, también por rangos y generalmente más conservadora en presencia de ruido [12].

El patrón general fue consistente: las asociaciones más altas con **cantidad** se concentraron en variables que resumen la **historia reciente de ventas**, particularmente los *lags* y los agregados *rolling* (incluyendo promedios móviles). En términos operativos, esto sugiere que, para los productos seleccionados, una fracción importante de la variación diaria puede anticiparse a partir del comportamiento inmediato previo (memoria de corto plazo), lo cual es esperable en series minoristas cuando se trabaja a nivel *producto-día*.

En contraste, las variables exógenas de **calendario** (p. ej., `quincena` y `dia_habil`) y de **clima** (p. ej., temperatura y descripciones de clima) mostraron asociaciones cercanas a cero bajo estas métricas bivariadas. Sin embargo, este resultado no se interpreta como evidencia para descartarlas. En este tipo de datos, una correlación baja puede coexistir con aportes relevantes cuando: (i) el efecto es **no lineal** (por ejemplo, solo se activa en ciertos rangos), (ii) el impacto es **producto-dependiente** (cambia según la referencia), o (iii) la contribución aparece principalmente mediante **interacciones** con la historia de demanda u otras variables [12]. Por esta razón, las correlaciones se asumieron como un **insumo de orientación** y no como un criterio de exclusión automática; la utilidad real de cada bloque de variables se confirmó posteriormente mediante evaluación temporal consistente dentro del pipeline.

5.1.6. MI: dependencia no lineal y relevancia de productoid

Como complemento al diagnóstico basado en correlaciones, se calculó la **información mutua (MI)** entre cada variable explicativa y la variable objetivo **cantidad**, con el propósito de identificar **dependencias potenciales que no necesariamente son lineales** [12]. En este contexto, MI se interpreta como una medida de *relevancia informativa*: valores más altos sugieren que la variable aporta señal útil para explicar la demanda, incluso cuando la relación no puede describirse adecuadamente mediante una asociación lineal.

Los resultados mostraron que los predictores que resumen el comportamiento reciente tienden a concentrar mayor MI. En particular, los agregados construidos con ventanas móviles (por ejemplo, `rolling_mean_7`) destacaron entre las variables más informativas, lo cual es coherente con su función de capturar el **nivel reciente** y cambios de corto plazo en la demanda [12]. Este hallazgo, más que reemplazar la evaluación predictiva, sirvió como criterio práctico para priorizar bloques de variables en la exploración posterior.

La MI elevada de **productoid** sugiere que el problema es **multiserie**: cada producto tiene patrones propios y el identificador aporta señal útil cuando se codifica adecuadamente (p. ej., *one-hot*) [10, 12]. Por ello, este hallazgo respalda el uso de un **modelo global multiserie**, manteniendo validación temporal estricta y

preprocesamiento ajustado solo con entrenamiento para evitar fuga temporal [9, 10].

5.1.7. Evaluación temporal del aporte de variables exógenas mediante *lags* ($h=1$) sin fuga temporal

Dado que el horizonte operativo principal es $h = 1$ (Sección 3.1.4), el EDA incluyó una verificación específica del aporte potencial de variables exógenas bajo la restricción *ex-ante*. Para ello, se construyó la variable objetivo futura $y_{p,t+1}$ mediante un desplazamiento por producto y se evaluó su asociación con cada exógena x observada en instantes no posteriores al momento de decisión, es decir, $x_{p,t-\ell}$, con $\ell \in \{0, 1, 2, 3, 7, 14\}$. Este conjunto de *lags* permite contrastar un efecto inmediato ($\ell = 0$) frente a efectos diferidos plausibles (por ejemplo, respuestas con retraso o acumulación), preservando el orden temporal y evitando incorporar información del futuro [9, 12].

La evaluación se ejecutó con un esquema de *backtesting* con ventana expansiva, usando: (i) un histórico mínimo inicial de 28 días, (ii) un bloque de evaluación de 14 días y (iii) *folds* secuenciales. En cada *fold* se calculó una medida de asociación entre $y_{p,t+1}$ y $x_{p,t-\ell}$, y luego se promedió el resultado a lo largo de los cortes temporales. Para cada exógena se definió el *best lag* como el valor de ℓ que maximiza dicha asociación promedio, con el objetivo de identificar, de forma trazable, el rezago más consistente dentro del periodo analizado.

En términos empíricos, las asociaciones observadas fueron reducidas en comparación con la señal capturada por la historia del *target*. Este patrón sugiere que el aporte de exógenas —si existe— podría ser débil, no lineal o dependiente del producto, por lo que estos resultados se asumen como un filtro exploratorio y no como evidencia concluyente. En consecuencia, la contribución de exógenas se valida de manera definitiva en la etapa de entrenamiento comparativo, contrastando modelos **con** y **sin** exógenas bajo el mismo protocolo temporal y el mismo pipeline [12].

5.1.8. Implicaciones para el pipeline: hipótesis verificables y conjuntos candidatos de variables

Los diagnósticos del EDA se asumen aquí como una **fuentes de decisiones para el modelado**: permiten proponer supuestos plausibles sobre dónde está la señal predictiva y traducirlos en comparaciones controladas. En coherencia con el enfoque de **pipeline** y evaluación temporal establecido en el Marco Teórico, estos supuestos no se toman como conclusiones, sino como **hipótesis verificables** que se validan (o se descartan) bajo el mismo protocolo de entrenamiento, preprocesamiento y validación [12].

- **Hipótesis H1 (señal endógena dominante a corto plazo).** Para el

horizonte $h = 1$, la mayor parte de la capacidad predictiva proviene de la **historia reciente de demanda**, representada mediante *lags* y agregaciones *rolling*.

- **Hipótesis H2 (aporte exógeno marginal y dependiente del producto).** Las variables exógenas (calendario y, cuando aplica, clima) podrían aportar mejoras **marginales** o **condicionadas** a ciertos productos, principalmente al interactuar con la dinámica reciente o al capturar cambios recurrentes del contexto operativo (p. ej., efectos de quincena o patrones estacionales).
- **Hipótesis H3 (conveniencia del enfoque global multiserie).** La inclusión de **productoid** permite modelar diferencias estructurales entre productos y, al mismo tiempo, aprovechar patrones compartidos; por tanto, es razonable evaluar un **modelo global multiserie** como alternativa sostenible frente a modelos independientes [10].

A partir de estas hipótesis, se definieron **conjuntos candidatos de variables** para ser comparados de forma justa dentro del mismo pipeline (mismo preprocesamiento y mismo esquema temporal), priorizando el criterio de *ex-ante* (solo información disponible hasta t) y controlando el riesgo de *data leakage* [12]. En particular, se consideraron tres familias de entrada:

- **C1: Endógenas.** Variables derivadas de la demanda (bloques de *lags* y *rolling*, y señales de recencia cuando aplique), junto con **productoid** en el enfoque global.
- **C2: Endógenas + calendario.** C1 más variables calendario interpretables (p. ej., *dia*, *dia_habil*, *quincena* y componentes calendario derivados de *fecha* cuando se incluyan en el diseño).
- **C3: Endógenas + calendario + clima.** C2 más variables meteorológicas disponibles y alineadas temporalmente con el día t (o sus versiones rezagadas si corresponde), para evaluar si aportan señal adicional sin comprometer la operatividad del pronóstico.

Estos conjuntos se operacionalizan más adelante como *feature sets* (según la convención **L_ R_ M_ CAL_ WX**) y constituyen la base de la comparación experimental previa al ajuste fino de hiperparámetros.

5.1.9. Salida del EDA: insumos listos para entrenamiento

El EDA se planteó con un propósito operativo: cerrar la fase exploratoria dejando **insumos directamente utilizables** en el entrenamiento comparativo, sin perder trazabilidad sobre *qué se seleccionó y por qué*. Como resultado, se consolidaron cinco productos principales que definen el punto de partida del pipeline:

- **Top-10 por negocio listos para modelado.** Se definió, para cada negocio, el subconjunto de productos a entrenar bajo criterios reproducibles de **rotación** e **intermitencia**, complementados con una restricción de **representatividad por familia**. Esta selección reduce el ruido de productos con comportamiento esporádico extremo y concentra el análisis en ítems relevantes para la operación.
- **Panel diario completo para el subconjunto analizado.** Para los Top-10 se construyó un panel *producto-día* sin vacíos, con **850 observaciones por negocio**. Esta estructura es condición necesaria para calcular *lags* y agregaciones *rolling* de forma consistente, evitando sesgos por días faltantes y garantizando comparabilidad entre modelos.
- **Prioridad empírica de la historia de demanda.** El análisis cuantitativo evidenció que la señal predictiva más estable proviene de la memoria reciente, reflejada en `lag_`, `rolling_` y `rolling_mean_`. Este hallazgo orienta el diseño de *feature sets* y explica por qué el entrenamiento parte de bloques endógenos como componente central.
- **Sustento del enfoque multiserie.** Un diagnóstico complementario mediante **información mutua (MI)** mostró dependencia no lineal relevante asociada al identificador del producto, lo que respalda la inclusión de `productoid` en un esquema **global multiserie** para aprender patrones compartidos sin perder diferenciación entre productos.
- **Evaluación temporal sin fuga para variables exógenas.** La revisión temporal se realizó bajo el horizonte principal $h = 1$, cuidando que cualquier señal externa se incorpore de manera *ex-ante*. En los casos donde no se dispone de predictores futuros, se estableció el uso de versiones **rezagadas** para evitar *data leakage* y mantener una evaluación realista.

En conjunto, estos insumos permiten que la fase de entrenamiento inicie sobre bases **justificadas, auditables** y coherentes con el marco de **pipelines** y **validación temporal** presentado en el Capítulo 3.1 [9, 12].

CAPÍTULO 5. ANÁLISIS EXPLORATORIO Y SELECCIÓN DE CARACTERÍSTICAS Y PRODUCTOS

productoid	cantidad	día	día_habil	quincena	temp.c	clima_desc	producto	lag_1	lag_2	...	rolling_7	rolling_mean_2	rolling_mean_3	rolling_mean_4	rolling_mean_5	rolling_mean_6	rolling_mean_7	ano	mes	día-num
1	20	5.0	3	1.0	0.0	21.81	Mudang Rothmans x10	0.0	2.0	...	8.0	1.0	1.0	1.0	1.2	1.0	1.1	2025	10	29
185	223	40.0	3	1.0	0.0	21.81	Huevo AAA Unidad	61.0	155.0	...	452.0	108.0	98.3	91.5	84.2	70.2	64.6	2025	10	29
186	224	9.0	3	1.0	0.0	21.81	Huevo AAA Cubeta x30	20.0	17.0	...	65.0	18.5	15.7	14.8	12.6	10.5	9.3	2025	10	29
299	378	8550.0	3	1.0	0.0	21.81	Papa Negra x gr	5895.0	7041.0	...	42782.0	6468.0	7935.3	7999.2	7208.4	6007.0	6111.7	2025	10	29
326	410	9265.0	3	1.0	0.0	21.81	Pollo Pechuga x gr	13405.0	6675.0	...	79766.0	10040.0	14460.0	15360.2	14748.2	12290.2	11395.1	2025	10	29
492	594	2.0	3	1.0	0.0	21.81	Exito Leche Entera x900ml	4.0	5.0	...	16.0	4.5	3.3	3.0	3.0	2.5	2.3	2025	10	29
1166	1345	1.0	3	1.0	0.0	21.81	Costeña Lata x310cm3	2.0	2.0	...	5.0	2.0	1.7	1.2	1.0	0.8	0.7	2025	10	29
1720	2009	9.0	3	1.0	0.0	21.81	Arepas Paquete Pequeño	6.0	10.0	...	44.0	8.0	9.0	9.0	8.4	7.0	6.3	2025	10	29
1817	2116	64.0	3	1.0	0.0	21.81	Arepas Unidad	83.0	54.0	...	286.0	68.5	63.7	59.0	53.0	44.2	40.9	2025	10	29
1897	2204	11.0	3	1.0	0.0	21.81	Cigarrillo Unidad	4.0	6.0	...	59.0	5.0	11.7	12.5	11.8	9.8	8.4	2025	10	29

10 rows × 30 columns

Figura 5.9: Villa Café: muestra del conjunto de datos producto–día con variables de calendario y clima, e ingeniería de características mediante rezagos (*lags*) y agregaciones móviles (*rolling*).

productoid	cantidad	día	día_habil	quincena	temp.c	clima_desc	producto	lag_1	lag_2	...	rolling_7	rolling_mean_2	rolling_mean_3	rolling_mean_4	rolling_mean_5	rolling_mean_6	rolling_mean_7	ano	mes	día-num
179	223	98.0	3	1.0	0.0	21.8	Huevo AAA Unidad	42.0	90.0	...	410.0	66.0	67.7	73.8	69.6	58.0	58.6	2025	10	29
180	224	12.0	3	1.0	0.0	21.8	Huevo AAA Cubeta x30	9.0	19.0	...	78.0	14.0	15.3	15.8	14.8	12.3	11.1	2025	10	29
292	378	4435.0	3	1.0	0.0	21.8	Papa Negra x gr	8595.0	5961.0	...	42083.0	7278.0	8675.3	7048.2	7212.6	6010.5	6011.9	2025	10	29
320	410	6735.0	3	1.0	0.0	21.8	Pollo Pechuga x gr	12845.0	9880.0	...	53101.0	11362.5	12025.0	11138.8	10085.0	8404.2	7585.9	2025	10	29
487	594	1.0	3	1.0	0.0	21.8	Exito Leche Entera x900ml	3.0	3.0	...	14.0	3.0	3.0	2.5	2.8	2.3	2.0	2025	10	29
1163	1345	2.0	3	1.0	0.0	21.8	Costeña Lata x310cm3	0.0	1.0	...	9.0	0.5	3.0	2.2	1.8	1.5	1.3	2025	10	29
1296	1519	4.0	3	1.0	0.0	21.8	Coca Cola Plástico x1.5l	1.0	0.0	...	12.0	0.5	1.0	1.2	2.0	1.7	1.7	2025	10	29
1715	2009	7.0	3	1.0	0.0	21.8	Arepas Paquete Pequeño	10.0	14.0	...	62.0	12.0	14.0	13.2	11.0	9.2	8.9	2025	10	29
1810	2116	47.0	3	1.0	0.0	21.8	Arepas Unidad	55.0	70.0	...	387.0	62.5	72.7	80.2	69.0	57.5	55.3	2025	10	29
1891	2204	11.0	3	1.0	0.0	21.8	Cigarrillo Unidad	5.0	13.0	...	54.0	9.0	11.3	11.8	10.8	9.0	7.7	2025	10	29

10 rows × 30 columns

Figura 5.10: Mercaya: muestra del conjunto de datos producto–día con variables de calendario y clima, e ingeniería de características mediante rezagos (*lags*) y agregaciones móviles (*rolling*).

Capítulo 6

Entrenamiento y Evaluación Inicial de los Modelos Predictivos de Demanda

6.1. Entrenamiento inicial de modelos con todos los predictores (línea base)

Esta fase establece la **línea base reproducible** del experimento: un punto de referencia común para cuantificar, de manera controlada, el efecto de (i) la selección sistemática de variables y (ii) la optimización posterior del modelado. En consecuencia, el objetivo aquí no es maximizar el desempeño a toda costa, sino fijar una referencia sólida sobre el panel Top-10 de cada negocio, utilizando el conjunto completo de predictores definido como `ORIGINAL_FULL`.

Bajo esta lógica, el entrenamiento inicial se ejecuta respetando estrictamente el protocolo temporal ya adoptado en el documento y encapsulando el preprocesamiento dentro de un **pipeline**, de forma que cada transformación se ajuste únicamente con el conjunto de entrenamiento y se aplique de manera consistente sobre validación y prueba. Con ello, las comparaciones entre alternativas quedan sustentadas en cambios reales del modelado, y no en variaciones accidentales del procedimiento.

Para garantizar comparabilidad, se mantuvieron constantes tres elementos durante esta fase:

- (a) **El mismo *holdout* temporal** (mismo bloque final reservado para evaluación fuera de muestra).
- (b) **El mismo pipeline de preprocesamiento**, incluyendo el tratamiento de variables numéricas y categóricas (p. ej., imputación y codificación consistente).
- (c) **El mismo conjunto de entrada `ORIGINAL_FULL`**, sin reducción ni selección adicional de variables.

Con estos controles, el resultado de esta sección se interpreta como la **referencia mínima exigente** a partir de la cual se juzgan las mejoras posteriores: si una estrategia de selección de variables o un ajuste adicional de hiperparámetros produce ganancias, dichas ganancias pueden atribuirse con mayor confianza a esa decisión específica, manteniendo estable el resto del pipeline. Además, esta línea base permite identificar desde el inicio posibles **trade-offs** entre precisión, tiempo de entrenamiento y complejidad efectiva, que son relevantes para la viabilidad práctica del enfoque en PYMEs comercializadoras.

6.1.1. Protocolo experimental implementado en el pipeline

El entrenamiento y la evaluación se ejecutaron bajo un **protocolo único** para **Mercaya** y **Villa Café**, con el propósito de asegurar que las comparaciones entre modelos y *feature sets* fueran consistentes y, sobre todo, cronológicamente válidas en el panel *producto-día*.

En primer lugar, se aplicó un **holdout temporal real por fechas únicas** con `TEST_SIZE=0.20`. A diferencia de una partición aleatoria por filas, el corte se realizó sobre el bloque final de días, evitando que observaciones del *mismo día* quedaran repartidas entre entrenamiento y prueba. Esta decisión reduce el riesgo de fuga temporal asociada a compartir contexto (calendario y clima) y preserva una lectura operativa del problema: entrenar con información del pasado y evaluar en días posteriores. En ambos negocios, el corte temporal fue:

- **TRAIN:** 2025-10-29 → 2026-01-04 (n=680)
- **TEST:** 2026-01-05 → 2026-01-21 (n=170)

donde n corresponde al número de observaciones del panel *producto-día*; para el subconjunto Top-10, se tienen 10 filas por día.

En segundo lugar, se implementó un **pipeline** de preprocesamiento para garantizar que *todas* las transformaciones se ajustaran **exclusivamente con TRAIN** y se aplicaran **sin cambios** sobre TEST. En particular, se empleó imputación por **mediana** para variables numéricas e imputación por **moda** para variables categóricas, seguida de **One-Hot Encoding** para `clima_desc` y, cuando el diseño lo incluye, `productoid`. Esta elección responde a que ambas variables se interpretan como **nominales** (no poseen orden natural): tratarlas como números impondría relaciones artificiales; en cambio, la codificación one-hot conserva el significado de cada categoría y permite capturar tanto efectos por tipo de clima como diferencias sistemáticas entre productos en el enfoque multiserie.

En contraste, las variables calendario `día` (día de la semana) y `día-num` (día del mes) se mantuvieron como **enteros discretos** (1–7 y 1–31, respectivamente), dado que se interpretan como posiciones dentro de la semana/mes y los ensambles

basados en árboles pueden explotarlas mediante particiones por rangos, sin aumentar innecesariamente la dimensionalidad del espacio de entrada.

Adicionalmente, la búsqueda de subconjuntos de variables y la optimización de hiperparámetros se realizaron **solo dentro de TRAIN** mediante validación cruzada temporal. Esta restricción no es únicamente metodológica: busca aproximar el proceso a un **entorno real**, en el cual las decisiones de modelado (qué variables usar y cómo parametrizar el algoritmo) deben tomarse únicamente con la información **disponible hasta ese momento**, sin acceso a datos futuros. De este modo, el desempeño estimado refleja un escenario más cercano a la operación, donde el modelo se entrena con historial y se enfrenta a días posteriores no observados.

Una vez completada esta búsqueda interna, cada alternativa candidata se entrenó con TRAIN completo y se evaluó *out-of-sample* en TEST para cuantificar su desempeño en el tramo más reciente.

Como consecuencia de la codificación categórica, el conjunto base ORIGINAL_FULL (28 variables crudas) se expandió a 41 *features* efectivas (raw→eff: 28→41). Reportar esta expansión resulta pertinente porque refleja la complejidad real con la que entrenan los modelos tabulares y el costo asociado tanto al ajuste como a la optimización dentro del pipeline.

6.1.2. Predictores incluidos en ORIGINAL_FULL y justificación

El *feature set* ORIGINAL_FULL se utiliza como **línea base** porque conserva, sin recortes iniciales, los tres bloques de señal definidos para el pronóstico: **memoria de demanda**, **calendario** y **contexto ambiental**. Por tanto, el conjunto reúne las variables ya construidas en el alistamiento del panel *producto-día*: calendario (*dia*, *dia_habil*, *quincena*, *ano*, *mes*, *dia-num*), clima (*temp_c*, *clima_desc*) y memoria (*lags lag_1..lag_7* y agregaciones *rolling_2..rolling_7*, *rolling_mean_2..rolling_mean_7*).

Adicionalmente, se incluye **productoid** para reflejar el carácter **multiserie**: permite que un **modelo global** aprenda diferencias sistemáticas entre productos (niveles y patrones) sin entrenar un modelo independiente por cada uno. Aunque **productoid** es un identificador numérico, se interpreta como una etiqueta **nominal** (sin magnitud) y su representación se gestiona dentro del **pipeline** de preprocesamiento, manteniendo consistencia entre entrenamiento y prueba bajo el mismo protocolo temporal.

6.1.3. Modelos comparados y criterio de línea base

Se compararon cuatro aproximaciones, todas entrenadas con el *feature set* ORIGINAL_FULL y evaluadas bajo el mismo **pipeline** y el mismo **holdout temporal**:

- **Linear**: regresión lineal como referencia, por su interpretabilidad y costo computacional mínimo.
- **RF_TUNED**: Random Forest entrenado dentro del pipeline y optimizado mediante búsqueda aleatoria en dos fases.
- **XGB_TUNED**: XGBoost optimizado en dos fases, con objetivo `reg:squarederror`.
- **LGBM_TUNED**: LightGBM optimizado en dos fases, con objetivo `regression`.

En coherencia con la definición adoptada en el documento, esta “línea base” (*baseline*) mantiene fijos **ORIGINAL_FULL** y el protocolo temporal del pipeline; por tanto, las diferencias observadas en esta etapa se interpretan principalmente como efecto del algoritmo y su ajuste, y sirven como referencia para cuantificar ganancias posteriores por selección de variables (reducción de dimensionalidad) y ajustes adicionales del modelado.

6.1.4. Métricas globales y tratamiento de MAPE/WAPE

El desempeño en el conjunto de prueba (*test*) se reportó con **MAE**, **MSE**, R^2 , R^2_{adj} y **WAPE (%)**. En este trabajo se **descarta el uso de MAPE** como criterio de comparación debido a su inestabilidad numérica e interpretativa cuando existen demandas bajas o cercanas a cero, una condición frecuente en el panel *producto-día*. En adelante, la lectura principal se apoya en **MAE** y **WAPE**, y se complementa con R^2 , **tiempo de entrenamiento** e inspección visual de predicciones para identificar patrones de error relevantes.

6.1.5. Resultados globales en test con ORIGINAL_FULL

En esta subsección se resume el desempeño en test del *feature set* completo **ORIGINAL_FULL** (raw→eff: 28→41), siguiendo el criterio de reporte establecido en la Sección 6.1.4. En consecuencia, el resumen prioriza **MAE**, **WAPE**, R^2 y **tiempo de entrenamiento**.¹

¹Los reportes completos del pipeline incluyen el conjunto total de métricas (incluyendo **MSE** y R^2_{adj}); aquí se presenta una síntesis orientada a la comparación global.

Cuadro 6.1: Villa Café: resumen global en test con ORIGINAL_FULL (28→41).

Modelo	MAE	MSE	R^2	WAPE (%)	Tiempo (s)
RF_TUNED	1355.22	$1,4288 \times 10^7$	0.6989	43.41	282.89
LGBM_TUNED	1420.73	$1,7199 \times 10^7$	0.6376	45.51	120.73
XGB_TUNED	1514.64	$1,8075 \times 10^7$	0.6191	48.52	14.79
Linear	1697.14	$1,5073 \times 10^7$	0.6824	54.37	0.01

Villa Café (test, ORIGINAL_FULL). **Lectura.** En Villa Café, **RF_TUNED** obtiene la mejor precisión global en el conjunto completo (**menor MAE** y **menor WAPE**, junto con el **mayor R^2**), a costa de registrar el mayor tiempo de entrenamiento. **XGB_TUNED** reduce de manera marcada el costo computacional, pero en esta configuración completa presenta un error agregado superior. **Linear**, por su parte, es prácticamente instantáneo y conserva un R^2 competitivo, aunque con MAE/WAPE más altos, lo que sugiere limitaciones para capturar no linealidades e interacciones relevantes bajo el conjunto completo de predictores.

Cuadro 6.2: Mercaya: resumen global en test con ORIGINAL_FULL (28→41).

Modelo	MAE	MSE	R^2	WAPE (%)	Tiempo (s)
XGB_TUNED	1046.58	$8,7141 \times 10^6$	0.7923	34.88	15.94
RF_TUNED	1075.68	$8,1555 \times 10^6$	0.8056	35.85	267.28
LGBM_TUNED	1277.66	$1,1235 \times 10^7$	0.7322	42.58	140.04
Linear	1425.98	$8,9300 \times 10^6$	0.7871	47.53	0.02

Mercaya (test, ORIGINAL_FULL). **Lectura.** En Mercaya, **XGB_TUNED** presenta el mejor balance precisión–costo dentro del conjunto completo, con **menor MAE** y **menor WAPE** y un tiempo de entrenamiento bajo. **RF_TUNED** alcanza el mejor ajuste global en términos de R^2 , pero con un costo computacional sustancialmente mayor. **Linear** mantiene su ventaja de rapidez, aunque con error agregado superior; **LGBM_TUNED**, bajo esta configuración completa, resulta menos competitivo en MAE/WAPE.

6.1.6. Inspección visual de predicciones de Villa Café

Además de las métricas, la evaluación de la **línea base** en **Villa Café** con el set completo **ORIGINAL_FULL** (raw→eff: 28→41) incorporó visualizaciones orientadas al diagnóstico aplicado: (i) serie agregada por fecha (suma de productos) de demanda real vs. predicha, (ii) dispersión global predicha vs. real con referencia $y = x$, y (iii) series real vs. predicha por producto para evidenciar heterogeneidad y casos difíciles. Estas gráficas complementan el resumen numérico al permitir ubicar temporalmente

CAPÍTULO 6. ENTRENAMIENTO Y EVALUACIÓN INICIAL DE LOS MODELOS PREDICTIVOS DE DEMANDA

episodios de mayor desviación y reconocer productos con dinámica particularmente compleja.



Figura 6.1: Villa Café: comparación global (agregado por fecha) de demanda real vs. predicha para modelos entrenados con el set completo ORIGINAL_FULLL.

En la Figura 6.1, se observa que los modelos siguen la **tendencia general** de la demanda agregada, pero tienden a **suavizar la amplitud** ante cambios abruptos. En particular, frente a **picos pronunciados** la predicción suele subestimar la magnitud máxima, y en descensos rápidos aparece una sobreestimación transitoria, lo cual es consistente con un efecto de *inercia* asociado a la memoria temporal (*lags/rolling*) cuando la variación real responde a factores operativos no plenamente representados en los predictores disponibles.

La comparación entre familias refuerza esta lectura. El modelo **Linear** produce trayectorias más regulares y evidencia limitaciones ante **no linealidades** y **cambios de régimen**. En contraste, los ensambles basados en árboles (en especial **Random Forest**) se ajustan mejor a variaciones, aunque los mayores desvíos se concentran en días donde la demanda se aparta bruscamente del patrón reciente.

Finalmente, la desagregación por producto permite identificar cuáles referencias concentran mayor dificultad de pronóstico y en qué tramos del periodo de prueba se presentan las discrepancias más relevantes. En Villa Café, esta lectura se documenta explícitamente a continuación mediante las series Real vs. Pred por producto para

CAPÍTULO 6. ENTRENAMIENTO Y EVALUACIÓN INICIAL DE LOS MODELOS PREDICTIVOS DE DEMANDA

cada modelo (Figuras 6.2–6.5), aportando insumos concretos para la discusión posterior sobre estrategias de mejora del desempeño.

Análisis producto a producto de Villa Café: demanda real vs. predicha por modelo

La inspección **producto a producto** es especialmente relevante en un escenario multiserie, porque la decisión operativa de reposición se ejecuta por referencia y el error puede distribuirse de manera desigual entre productos (heterogeneidad de escala, intermitencia y sensibilidad a picos). Por ello, las Figuras 6.2–6.5 se usan como diagnóstico aplicado del comportamiento de cada familia de modelos sobre el Top-10 de Villa Café durante el periodo de prueba.

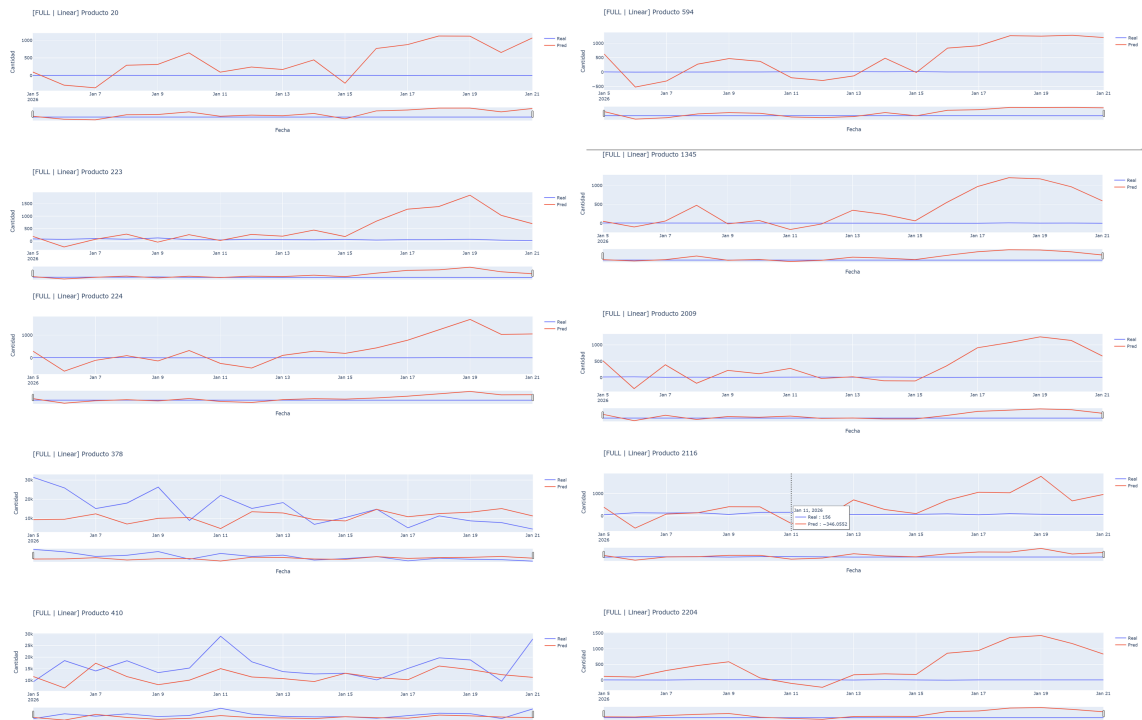


Figura 6.2: Villa Café (FULL, ORIGINAL_FULL). Predicción vs. demanda real por producto (Top-10) para **Linear**.

CAPÍTULO 6. ENTRENAMIENTO Y EVALUACIÓN INICIAL DE LOS MODELOS PREDICTIVOS DE DEMANDA

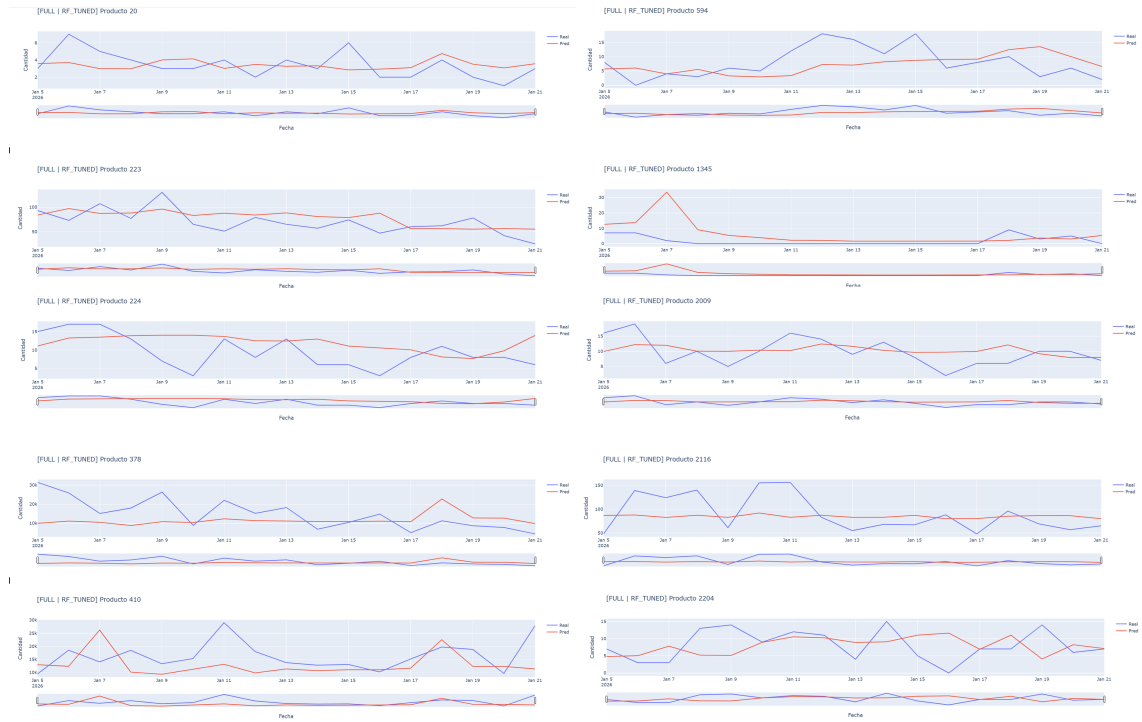


Figura 6.3: **Villa Café (FULL, ORIGINAL_FULL)**. Predicción vs. demanda real por producto (Top-10) para **RF_TUNED**.

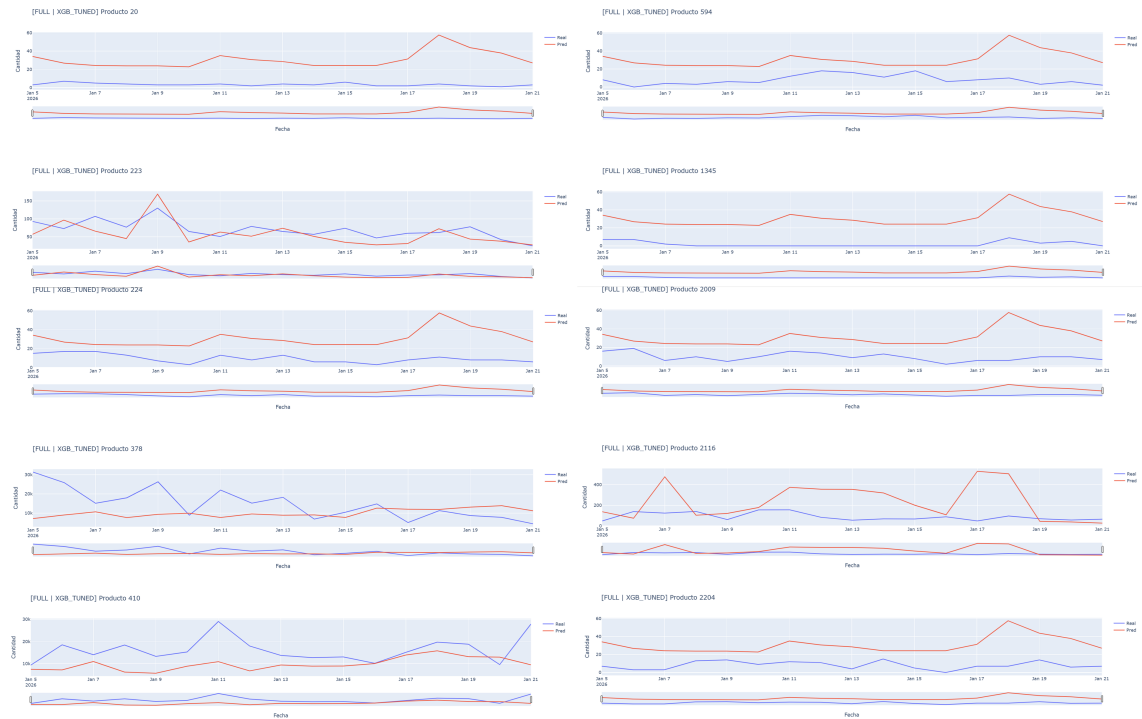


Figura 6.4: **Villa Café (FULL, ORIGINAL_FULL)**. Predicción vs. demanda real por producto (Top-10) para **XGB_TUNED**.

CAPÍTULO 6. ENTRENAMIENTO Y EVALUACIÓN INICIAL DE LOS MODELOS PREDICTIVOS DE DEMANDA

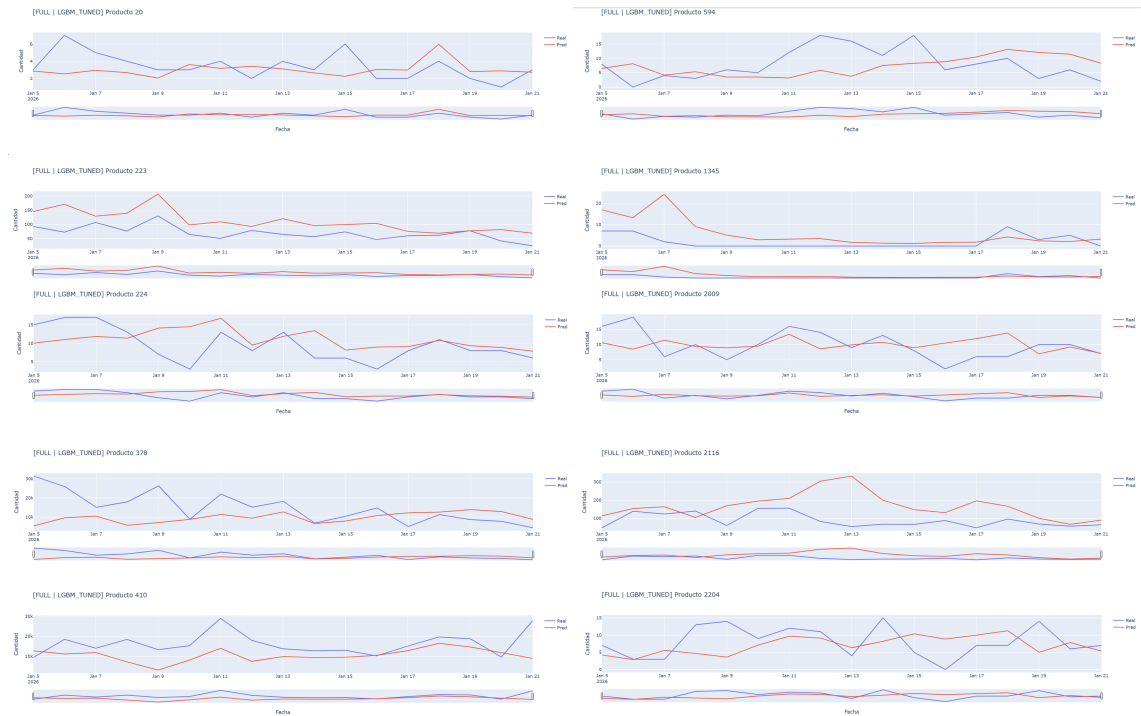


Figura 6.5: **Villa Café (FULL, ORIGINAL_FULL)**. Predicción vs. demanda real por producto (Top-10) para **LGBM_TUNED**.

Criterios de lectura (diagnóstico aplicado). Para cada referencia se priorizan cinco señales:

- Nivel (sesgo):** predicción sistemáticamente por encima/por debajo de la serie real (sobrestock vs. rupturas).
- Amplitud:** captura de picos y caídas vs. *suavización* (subestimar máximos y/o sobreestimar mínimos).
- Sincronía:** reacción ante cambios de régimen sin retraso (inercia por *lags/rolling*).
- Intermitencia:** comportamiento ante demandas bajas, ceros o ventas esporádicas.
- Coherencia operativa:** plausibilidad del pronóstico (no negatividad y rangos consistentes con el histórico inmediato).

Linear (Figura 6.2). En la inspección por producto se evidencian **limitaciones de calibración** bajo el set completo: para varias referencias, la trayectoria predicha se separa del nivel real y puede generar valores fuera de rango operativo. En términos prácticos, esto reduce su confiabilidad como señal directa de reposición sin **post-procesamiento** (p. ej., imponer $\hat{y} \leftarrow \max(0, \hat{y})$) y/o sin regularización. En esta línea

base, su utilidad principal es funcionar como **referencia de costo computacional mínimo**.

RF_TUNED (Figura 6.3). Random Forest suele entregar predicciones **más plausibles** y cercanas al rango observado. No obstante, aparece el patrón típico de ensambles promediadores: (i) **suavización** frente a variaciones abruptas y (ii) tendencia a **subestimar picos** (o a sostener niveles moderados tras una caída rápida). Operativamente, esto implica que, ante productos con alta variabilidad, la reposición basada solo en $\hat{y}$ puede requerir **stock de seguridad** para mitigar riesgo de ruptura en días pico.

XGB_TUNED (Figura 6.4). XGBoost muestra capacidad para seguir tendencias, pero en varias referencias se aprecia un **sesgo positivo** (predicciones por encima del nivel real). En productos con dinámica intermitente, también puede mantener un nivel predicho distinto de cero incluso cuando la demanda observada es baja. Desde inventario, este patrón eleva el riesgo de **sobrestock** si se aplica una política de reposición proporcional al pronóstico diario. Este hallazgo respalda que, en etapas posteriores, se revisen decisiones de regularización y control de complejidad (profundidad, mínimos por hoja, *learning rate*) y se contraste si la selección de variables reduce dicho sesgo.

LGBM_TUNED (Figura 6.5). LightGBM tiende a producir curvas **estables** y con buen compromiso entre seguimiento de tendencia y suavización. Sin embargo, pueden aparecer **sobreestimaciones localizadas** (picos pronosticados que no se materializan), lo cual sugiere sensibilidad a patrones del histórico reciente que no se repiten en el bloque de prueba. Operativamente, un pico “falso” puede inducir compras innecesarias; por ello, conviene complementar el pronóstico con reglas de negocio (límites de incremento diario) y, en etapas posteriores, evaluar si la reducción del set de variables mejora la robustez.

Hallazgos transversales (Villa Café). La lectura producto a producto permite establecer cuatro conclusiones prácticas:

- **Heterogeneidad de escala:** coexisten referencias de distinta magnitud; por ello, el desempeño global puede ocultar errores relevantes en productos pequeños o quedar dominado por unos pocos de alta demanda.
- **Intermitencia y ceros:** algunas referencias presentan ventas esporádicas; los modelos globales tienden a “rellenar” con niveles promedio, lo que explica parte del error y sugiere estrategias futuras específicas (p. ej., enfoque en dos etapas: probabilidad de venta + cantidad).
- **Picos y cambios bruscos:** los ensambles basados en árboles suelen suavizar amplitud (subestimar máximos y sobreestimar mínimos), lo que obliga a

interpretar $\hat{y}$ como señal central y complementar con **stock de seguridad** cuando el costo de ruptura es alto.

- **Coherencia operativa:** modelos que producen valores fuera de rango requieren correcciones explícitas para uso productivo; en esta línea base, los ensambles (RF/XGB/LGBM) tienden a entregar pronósticos más plausibles que el modelo lineal.

En síntesis, este análisis por producto (Villa Café) no reemplaza el resumen global, pero determina qué tan **operacionalmente confiable** resulta cada alternativa en el Top-10: identifica referencias donde el pronóstico es utilizable y aquellas donde persisten sesgos (sobre/infraestimación) que deben abordarse en las fases siguientes mediante selección de variables, ajustes adicionales del modelado y/o reglas operativas de reposición.

Conclusión breve por producto (Villa Café): por qué las curvas se comportan así

Aunque las diferencias entre modelos afectan la precisión, la **forma de las curvas** en las Figuras 6.2–6.5 está dominada, principalmente, por la **naturaleza operativa de cada referencia**: (i) productos **a granel** y perecederos tienden a mostrar variabilidad y picos; (ii) productos **de consumo básico** presentan niveles más estables con estacionalidad semanal; y (iii) referencias **discretas/intermitentes** exhiben ceros o caídas bruscas (frecuentemente compatibles con baja frecuencia de compra o rupturas de inventario). En este contexto, la **suavización** y la **inercia** observadas en varios modelos son consistentes con el uso dominante de rezagos y ventanas móviles (*lags/rolling*), que tienden a “promediar” cambios abruptos.

- **(410) Pollo Pechuga x gr (plaza).** Al ser un producto **a granel** y de compra recurrente, su demanda suele presentar **variación diaria** y **picos puntuales**; por ello, es común observar **subestimación de máximos** cuando el modelo suaviza la amplitud del cambio.
- **(378) Papa Negra x gr (plaza).** Similar al pollo, la venta **por gramo** induce oscilaciones por ajustes operativos (compra por necesidad inmediata), lo que explica tramos con cambios rápidos que los modelos tienden a seguir con **retraso** o con menor amplitud.
- **(223) Huevo AAA Unidad (huevos).** Producto de **consumo estable** con estacionalidad semanal moderada; por ello, las curvas suelen verse **más regulares** y los modelos tienden a mantener buen seguimiento del nivel, salvo en picos anómalos.
- **(2116) Arepas Unidad (arepas).** Suele comportarse como un consumo **frecuente** pero con amplitud limitada; por eso, los gráficos tienden a mos-

trar **mesetas** y variaciones suaves, donde la principal dificultad es capturar **incrementos cortos** sin sobreestimar.

- **(224) Huevo AAA Cubeta x30 (huevos)**. A diferencia del huevo por unidad, la compra por cubeta suele ser **menos frecuente** y en **bloques**; esto genera curvas con **picos discretos** y más ceros relativos, donde el modelo puede “rellenar” con valores pequeños por efecto de promedios históricos.
- **(2009) Arepas Paquete Pequeño (arepas)**. Similar a la cubeta, suele ser una compra **por ocasión** (paquetes) y no necesariamente diaria; esto explica mayor **intermitencia** y el riesgo de **sobreestimación** cuando el modelo sostiene un nivel positivo ante días sin compra.
- **(2204) Cigarrillo Unidad (cigarrillo)**. Es una venta **discreta** con alta rotación, pero sensible a disponibilidad; caídas bruscas o ceros son compatibles con **rupturas** o cambios de flujo, y los modelos tienden a recuperarse con **inercia** tras un quiebre.
- **(594) Éxito Leche Entera x900ml (exito)**. Producto empacado de consumo relativamente **constante**; por ello, las curvas suelen ser estables y la desviación típica se asocia a **pequeños picos** que pueden quedar suavizados.
- **(20) Mustang Rothmans x10 (mustang)**. Al ser un empaque específico, su compra suele ser **menos diaria** que el cigarrillo unidad; esto genera **intermitencia** y explica por qué algunos modelos mantienen un nivel distinto de cero aun cuando se observan días sin demanda.
- **(1345) Costeña Lata x310cm3 (costeña)**. La demanda de bebidas en lata suele concentrar incrementos en **días específicos**; por ello, se observan picos que los modelos tienden a **subestimar** si el histórico reciente no anticipa el cambio con claridad.

En conjunto, estas diferencias por producto justifican que la comparación de modelos no se interprete únicamente con métricas agregadas: la **forma de la serie** (intermitencia, picos discretos, estabilidad) condiciona el tipo de error esperado (sobrestock vs. ruptura) y, por tanto, la lectura operativa del pronóstico.

6.1.7. Conclusión de la línea base de Villa Café

Bajo el **protocolo temporal** y el **pipeline** ya descritos, la línea base en Villa Café con el conjunto completo ORIGINAL_FULLL deja una referencia clara para interpretar las mejoras posteriores. La inspección global sugiere que, aun con este set, persisten **picos de demanda** cuya magnitud tiende a quedar subestimada y, en cambios bruscos de nivel, aparece un ajuste más gradual. Esta evidencia es coherente con un comportamiento sensible al contexto y con factores operativos que no quedan completamente explicitados en los predictores disponibles.

En los resultados agregados, **RF_TUNED** se posiciona como la alternativa con mejor desempeño global dentro de esta referencia, aunque con un **costo de entrenamiento** mayor frente a opciones más livianas. En adelante, esta configuración con **ORIGINAL_FULL** se utiliza como punto de comparación directo para cuantificar la ganancia atribuible a la **selección sistemática de variables** y a la **optimización conjunta** planteadas en la fase siguiente.

6.1.8. Inspección visual de predicciones de Mercaya

En **Mercaya**, la evaluación de la **línea base** con el set completo **ORIGINAL_FULL** (raw→eff: 28→41) se apoyó en las mismas visualizaciones descritas en la subsección anterior; por ello, aquí se enfatiza la lectura específica del negocio a partir de la serie agregada por fecha.



Figura 6.6: Mercaya: comparación global (agregado por fecha) de demanda real vs. predicha para modelos entrenados con el set completo **ORIGINAL_FULL**.

En la Figura 6.6 se observa una trayectoria agregada **más regular**, lo que facilita que las predicciones sigan con mayor estabilidad la serie real. Las discrepancias se concentran, principalmente, en (i) **subestimaciones puntuales** cuando aparecen picos de demanda y (ii) **sobreestimaciones transitorias** después de descensos

rápidos, un patrón coherente con la dependencia en **rezagos** (*lags*) y **ventanas móviles** (*rolling*) cuando la demanda cambia más rápido que su historial inmediato.

Por familia de modelos, **Linear** tiende a reproducir la tendencia promedio, pero pierde precisión cuando la relación entre predictores y demanda exige capturar **no linealidades** e **interacciones**. En contraste, los ensambles basados en árboles (en particular **XGB_TUNED** y **RF_TUNED**) se ajustan mejor a variaciones de corto plazo en el agregado, en línea con el resumen global reportado para **ORIGINAL_FULL**. La desagregación por producto permite identificar referencias donde el pronóstico resulta más exigente y, sobre todo, ubicar patrones de error con implicación directa en reposición (sub/sobreestimación por referencia).

Análisis producto a producto (Mercaya): demanda real vs. predicha por modelo

La inspección **producto a producto** es crítica en Mercaya porque el Top-10 combina referencias con **escalas distintas** (productos de alta rotación frente a productos de menor volumen), y esta heterogeneidad puede ocultarse en métricas agregadas. Por ello, las Figuras 6.7–6.10 se utilizan como diagnóstico aplicado del desempeño por referencia en el periodo de prueba.



Figura 6.7: Mercaya (**FULL**, **ORIGINAL_FULL**). Predicción vs. demanda real por producto (Top-10) para **Linear**.

CAPÍTULO 6. ENTRENAMIENTO Y EVALUACIÓN INICIAL DE LOS MODELOS PREDICTIVOS DE DEMANDA

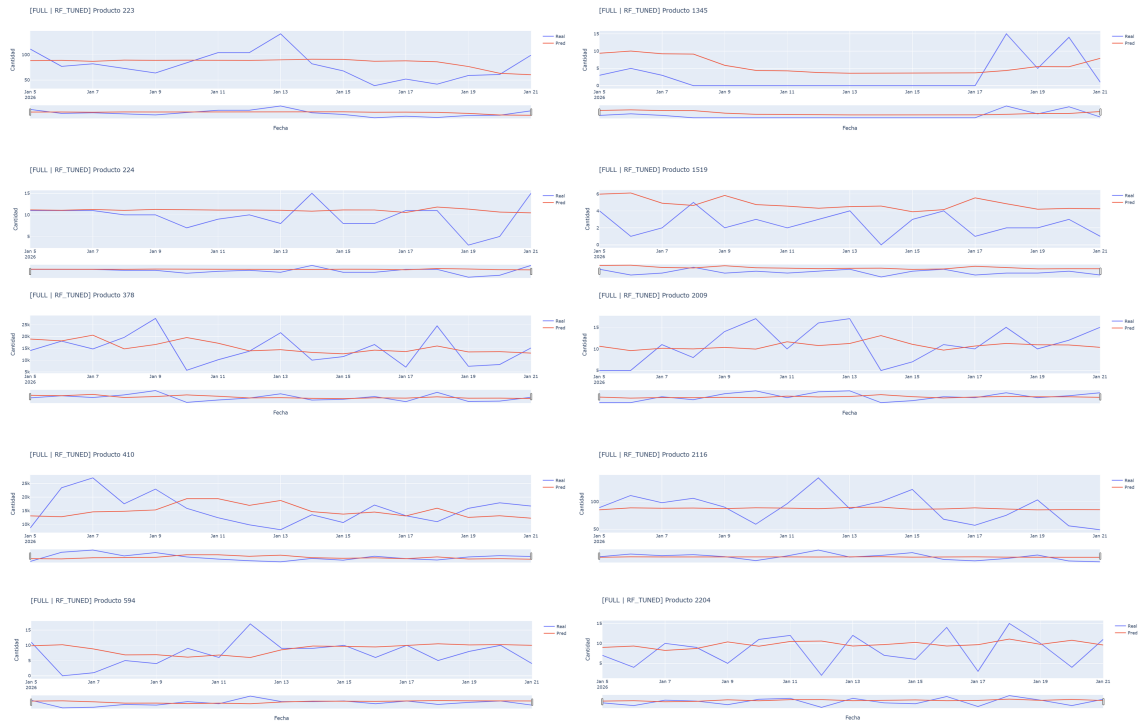


Figura 6.8: Mercaya (**FULL**, **ORIGINAL_FULLL**). Predicción vs. demanda real por producto (Top-10) para **RF_TUNED**.



Figura 6.9: Mercaya (**FULL**, **ORIGINAL_FULLL**). Predicción vs. demanda real por producto (Top-10) para **XGB_TUNED**.

CAPÍTULO 6. ENTRENAMIENTO Y EVALUACIÓN INICIAL DE LOS MODELOS PREDICTIVOS DE DEMANDA



Figura 6.10: Mercaya (**FULL**, **ORIGINAL_FULL**). Predicción vs. demanda real por producto (Top-10) para **LGBM_TUNED**.

Criterios de lectura (diagnóstico aplicado). Para cada referencia se priorizan cinco señales:

- Nivel (sesgo):** predicción sistemáticamente por encima/por debajo del nivel real (sobrestock vs. rupturas).
- Amplitud:** captura de picos y caídas vs. *suavización* (subestimar máximos y/o sobreestimar mínimos).
- Sincronía:** reacción ante cambios de régimen sin retraso (inercia por *lags/rolling*).
- Heterogeneidad por escala:** consistencia tanto en productos de alto volumen como en referencias pequeñas.
- Coherencia operativa:** pronósticos plausibles (no negatividad, ausencia de explosiones puntuales no justificadas, rangos compatibles con el histórico inmediato).

Linear (Figura 6.7). En Mercaya, el modelo lineal global evidencia **limitaciones de calibración por escala:** en varias referencias la trayectoria predicha se separa del nivel real y puede generar valores poco plausibles para operación. Este patrón es

consistente con la incapacidad del modelo lineal para capturar **no linealidades** e **interacciones** presentes en la dinámica de demanda. En consecuencia, su utilidad en esta etapa se mantiene como **referencia de mínima complejidad**, más que como candidato operativo directo sin correcciones (p. ej., no negatividad y límites de variación).

RF_TUNED (Figura 6.8). Random Forest tiende a entregar predicciones **más estables y plausibles** por producto, manteniéndose en rangos coherentes con el histórico inmediato. Sin embargo, aparece el patrón típico de ensambles promediadores: **suaviza la amplitud** y puede **subestimar picos** o sostener niveles moderados tras descensos rápidos. Operativamente, esto sugiere interpretar $\hat{y}$ como señal central y complementar con **stock de seguridad** cuando los picos sean críticos.

XGB_TUNED (Figura 6.9). La desagregación por producto permite identificar un hallazgo relevante: se observan **picos pronosticados de gran magnitud y sincronizados en el tiempo** en varias referencias, que no necesariamente se materializan con la misma intensidad en la demanda real. Este patrón es especialmente riesgoso para reposición porque podría inducir **sobrestock simultáneo** en múltiples productos. Por ello, en etapas posteriores resulta pertinente reforzar controles de complejidad y robustez (profundidad, mínimos por hoja, regularización) y evaluar si la **selección de variables** reduce la aparición de picos espurios.

LGBM_TUNED (Figura 6.10). LightGBM exhibe un compromiso intermedio: suele producir trayectorias **más regulares** y menos extremas que XGBoost, manteniendo plausibilidad operativa en una parte importante de las referencias. No obstante, persisten dos patrones: (i) **subestimación de máximos** en productos con picos pronunciados y (ii) **sesgos localizados** en referencias de menor volumen (predicciones que se sostienen por encima del nivel real en tramos de baja demanda). En términos de inventario, esto se traduce en riesgo de ruptura en días pico (si no hay stock de seguridad) y riesgo de sobrestock en productos pequeños (si la reposición es proporcional al pronóstico).

Hallazgos transversales (Mercaya). El análisis por producto permite sintetizar cuatro conclusiones operativas:

- **La escala importa:** el desempeño no es uniforme; productos de alto volumen pueden dominar la lectura global, mientras que referencias pequeñas concentran sesgos relativos.
- **Los ensambles suavizan:** RF y LGBM suelen ser más plausibles, pero subestiman picos; se recomienda complementar con **stock de seguridad** cuando el costo de ruptura sea alto.

- **Picos sincronizados como señal de alerta:** en XGB_TUNED aparecen picos simultáneos en varias referencias; es un riesgo operativo y debe controlarse con regularización/robustez y con selección de variables.
- **Valor del diagnóstico:** más allá de “qué modelo gana en promedio”, se identifica **dónde falla** y cómo se traduce a decisiones por producto (ruptura vs. sobrestock).

En síntesis, este apartado complementa las métricas globales: permite evaluar la **confiabilidad operativa** por referencia en Mercaya y deja evidencia concreta para justificar las decisiones de la fase siguiente (selección de variables y optimización conjunta del modelado).

Conclusión breve por producto (Mercaya): por qué las curvas se comportan así

En las Figuras 6.7–6.10, la **morfología** de cada curva se explica, principalmente, por la **naturaleza del producto** y su patrón de compra: referencias **a granel** (plaza) tienden a variar con mayor amplitud; básicos como huevo/arepa muestran mayor estabilidad; y productos empacados/bebidas pueden presentar **picos puntuales** (ocasiones específicas) que, bajo un pipeline dominado por rezagos, tienden a ser **suavizados**. En ese marco, discrepancias abruptas o días con demanda baja pueden ser compatibles con **intermitencia** o **restricciones operativas** (p. ej., disponibilidad).

- **(378) Papa Negra x gr (plaza).** La venta **por gramo** tiende a mostrar variabilidad diaria; los picos y cambios bruscos se reflejan como tramos donde el modelo suele **seguir con menor amplitud** por efecto de suavización.
- **(410) Pollo Pechuga x gr (plaza).** Producto a granel y perecedero, con compras ajustadas a necesidad inmediata; esto explica oscilaciones y picos donde es frecuente observar **subestimación de máximos** y ajustes graduales posteriores.
- **(223) Huevo AAA Unidad (huevos).** Presenta demanda **más continua** y patrón semanal; por ello las curvas suelen ser regulares, y los errores aparecen cuando hay picos puntuales no sostenidos por el histórico inmediato.
- **(2116) Arepas Unidad (arepas).** Referencia con consumo frecuente; típicamente mantiene **nivel estable** y variación moderada, por lo que la forma del gráfico suele estar dominada por pequeñas ondulaciones semanales más que por saltos extremos.
- **(2009) Arepas Paquete Pequeño (arepas).** Compra por paquete suele ser **más episódica**; por eso aparecen segmentos con intermitencia, donde el modelo puede **sobreestimar** si “hereda” un nivel promedio desde rezagos recientes.

- **(224) Huevo AAA Cubeta x30 (huevos).** Compra menos frecuente y en bloques; esto genera **picos discretos** y más ceros relativos, lo que favorece predicciones que se ven **suavizadas** o con “piso” positivo en modelos globales.
- **(2204) Cigarrillo Unidad (cigarrillo).** Venta discreta de alta rotación, pero sensible a disponibilidad; caídas pronunciadas o ceros son compatibles con **rupturas** o cambios de flujo, y los modelos suelen recuperarse con **inercia** tras el quiebre.
- **(594) Éxito Leche Entera x900ml (exito).** Producto empacado de consumo relativamente estable; por ello, los gráficos tienden a ser suaves y los errores suelen concentrarse en **picos pequeños** o cambios cortos.
- **(1345) Costeña Lata x310cm3 (costeña).** Producto asociado a ocasiones específicas; es común observar picos que los modelos **subestiman** si no hay señal fuerte en rezagos (o los *lags* capturan un patrón distinto al del bloque de prueba).
- **(1519) Coca Cola Plástico x1.5l (coca).** Bebida con demanda potencialmente sensible a picos puntuales; esto se traduce en curvas con incrementos localizados donde el modelo suele **suavizar** el máximo o desplazarlo levemente por dependencia en memoria temporal.

En síntesis, cada gráfica refleja una combinación de (i) **patrón de compra por unidad de venta** (granel vs. unidad vs. paquete), (ii) **frecuencia/intermitencia** y (iii) **respuesta del modelo a cambios abruptos** (suavización e inercia por *lags/rolling*). Este cierre por referencia justifica por qué, aun con un *feature set* completo, algunos productos son intrínsecamente más difíciles de pronosticar y requieren lectura operativa (stock de seguridad, límites de variación diaria) en fases posteriores.

6.1.9. Conclusión de la línea base de Mercaya

En **Mercaya**, la línea base con `ORIGINAL_FULL` fija una referencia cuantitativa y visual bajo el *holdout* temporal del pipeline, coherente con el uso real (predecir futuro con información pasada). A nivel agregado, la demanda es relativamente regular; así, su variación se explica principalmente por **memoria reciente** y **calendario**, y los desajustes se concentran en cambios abruptos puntuales (picos o saltos de nivel).

En desempeño y costo, el *boosting* —en particular **XGB_TUNED**— ofrece el mejor compromiso **precisión–tiempo**, mientras que **RF_TUNED** logra un ajuste competitivo con mayor costo de entrenamiento. Esta línea base se usa para cuantificar la ganancia de la **selección de conjuntos compactos** y la **optimización conjunta** (precisión, eficiencia y mantenibilidad) en la fase siguiente.

Capítulo 7

Optimización y Evaluación conjunta: features, hiperparámetros y eficiencia

7.1. Pipeline integral de entrenamiento, selección de variables y optimización

Como se discutió en el Marco Teórico, el **pronóstico de demanda** en **PYMEs comercializadoras** se interpreta como una decisión operativa (desabastecimiento vs. sobrestock) y, por tanto, debe evaluarse con separación temporal estricta para evitar **data leakage**. Además, el desempeño observado depende del **pipeline completo**—ingeniería de variables, validación temporal, selección de variables y optimización de hiperparámetros— y no únicamente del algoritmo. Con base en estas premisas, en esta sección se describe **cómo se operacionalizan** dichos criterios en un procedimiento reproducible que integra: comparación de modelos supervisados, búsqueda controlada de subconjuntos de variables y optimización de hiperparámetros bajo restricciones de costo computacional.

El pipeline se aplica en dos escenarios de negocio (**Mercaya** y **Villa Café**) bajo el mismo protocolo metodológico, de modo que las diferencias en los resultados se interpreten como rasgos del fenómeno (señales de calendario, sensibilidad contextual y estructura de la demanda) y no como artefactos del proceso de evaluación.

7.1.1. Propósito y criterios de diseño

Con el panel **producto–día** ya construido y el esquema de evaluación temporal definido, este capítulo responde dos preguntas prácticas: (i) si es posible trabajar con *feature sets* más compactos sin sacrificar (e incluso mejorando) el desempeño, y (ii) qué tan conveniente es optimizar los modelos sobre esos subconjuntos cuando se dispone de un presupuesto computacional acotado. Para ello, se plantea un **pipeline**

integral que permite: (a) comparar modelos bajo el mismo marco temporal, (b) hacer explícito el intercambio precisión-complejidad (dimensionalidad efectiva y tiempo de cómputo), y (c) derivar una recomendación operativa reproducible para el pronóstico diario ($h = 1$) en **Mercaya** y **Villa Café**.

El diseño del pipeline se guía por tres criterios:

- **Generalización:** privilegiar evidencia fuera de muestra y minimizar el sobreajuste mediante validación temporal.
- **Trazabilidad y reproducibilidad:** asegurar que corridas, modelos y *feature sets* sean comparables bajo el mismo protocolo.
- **Viabilidad operativa:** considerar de forma explícita el costo (tiempo) y la complejidad (dimensionalidad efectiva) en la toma de decisiones.

Bajo estos criterios, el pipeline se organiza en tres etapas consecutivas, cada una con un objetivo experimental distinto:

1. **EtapA A (referencia):** entrenamiento con `ORIGINAL_FULL` para fijar un punto de comparación cuantitativo.
2. **EtapA B (Feature Search, 2 fases):** exploración controlada de *feature sets* para identificar alternativas compactas competitivas.
3. **EtapA C (Tuning, 2 fases):** optimización sistemática de hiperparámetros sobre los *feature sets* priorizados por desempeño y por simplicidad.

7.1.2. Validaciones operativas para la búsqueda y la optimización

Antes de iniciar la exploración de variables y la optimización, el pipeline incorpora verificaciones sencillas para asegurar que los resultados sean atribuibles a los cambios experimentales (conjuntos de variables e hiperparámetros) y no a inconsistencias del insumo o de la ejecución.

Control de esquema y copia de trabajo

En primer lugar, se comprueba que el *dataframe* contiene todas las variables definidas como disponibles en `AVAILABLE_FEATURES`. A partir de este insumo validado, se crea una copia de trabajo que aísla la ejecución del pipeline frente a modificaciones accidentales durante la preparación y el entrenamiento. Con este control se evita comparar *feature sets* sobre bases no equivalentes y se garantiza que las diferencias observadas provengan del subconjunto de variables evaluado o del ajuste del modelo.

Ancla temporal para particiones y validación cruzada

Para garantizar particiones coherentes en el *holdout* y en la validación cruzada temporal, el pipeline requiere una columna de fecha válida que actúe como ancla cronológica, manteniendo la unidad de análisis diaria (**producto–día**).

7.1.3. Validación cruzada temporal por fecha como unidad indivisible

En la evaluación de *feature sets* y en el ajuste de hiperparámetros, el pipeline utiliza **validación cruzada temporal** construida **por fechas**. En concreto, cada *split* agrupa todas las observaciones de un mismo día del panel *producto–día*, de modo que esa fecha queda completa en entrenamiento o completa en validación (sin fraccionarse). Con ello se mantiene la coherencia temporal del experimento y se evita **data leakage** por compartir el mismo contexto diario (p. ej., variables de calendario y clima).

Para equilibrar la estabilidad de la comparación y el costo computacional, se definieron tres configuraciones de validación temporal con niveles crecientes de exigencia:

- **Búsqueda de variables (Fase 1):** se utiliza un número reducido de particiones temporales para realizar un **cribado inicial rápido**, cuyo objetivo es descartar alternativas claramente débiles sin incurrir en un costo elevado.
- **Búsqueda de variables (Fase 2):** se incrementa el número de particiones temporales para obtener un **ranking más estable** de los subconjuntos de variables que superaron el filtro inicial, reduciendo la dependencia de un periodo específico.
- **Optimización de hiperparámetros:** se mantiene una cantidad de particiones comparable a la fase anterior, buscando un **compromiso** entre consistencia en la evidencia y viabilidad del tiempo de ejecución durante el ajuste fino del modelo.

7.1.4. Complejidad efectiva: dimensionalidad posterior al preprocesamiento

En este trabajo, la complejidad de un *feature set* no se interpreta únicamente por el número de variables *crudas*, sino por la **dimensionalidad efectiva** que resulta tras el preprocesamiento dentro del **pipeline** (raw→eff). Por ello, en cada experimento se reporta el número de *features* efectivas después de las transformaciones aplicadas (p. ej., imputación y codificación) y se utiliza como un indicador operativo de complejidad.

Esta medida complementa las métricas de desempeño, ya que permite contrastar alternativas considerando también el costo práctico asociado a entrenar y mantener modelos con mayor dimensionalidad.

7.1.5. Etapa B: selección de variables en dos fases (exploración controlada)

La Etapa B responde a una necesidad práctica del experimento: antes de optimizar modelos, es indispensable **reducir el espacio de búsqueda** a un grupo de conjuntos de variables que, con evidencia preliminar, parezcan realmente prometedores. Por eso, en esta etapa no se busca todavía el “mejor resultado final”, sino construir una base sólida para decidir **qué alternativas vale la pena optimizar** en la Etapa C.

En coherencia con el Marco Teórico, la selección se organiza alrededor de los **bloques de señal** ya definidos (memoria de demanda, calendario y contexto ambiental) y de la convención de **etiquetas** que permite reconstruir qué información contiene cada alternativa. En esta sección, por tanto, no se repiten esas definiciones; el foco está en explicar **cómo se operacionaliza** esa estructura para generar candidatos, evaluarlos con validación temporal y elegir los que avanzan.

Búsqueda guiada por bloques: una exploración interpretable

En lugar de probar variables de forma aislada, la exploración se diseña para comparar **combinaciones coherentes** de información: más o menos memoria reciente, inclusión o exclusión de calendario y aporte del contexto ambiental. Esta decisión vuelve la búsqueda **auditable** (se entiende qué cambió entre alternativas) y facilita que los resultados se interpreten como evidencia sobre el aporte relativo de cada bloque, y no como hallazgos accidentales.

Generación de candidatos y controles para asegurar comparabilidad

A partir de la estructura anterior se generan múltiples conjuntos candidatos. Para que la comparación sea justa y viable, el pipeline aplica tres controles operativos:

- **Conjuntos equivalentes:** si dos alternativas terminan siendo iguales (mismas variables), se evalúa una sola vez, evitando repetir cómputo sin ganar información.
- **Consistencia del enfoque multiserie:** se mantiene constante la formulación global del problema entre candidatos, de manera que cualquier diferencia observada provenga del *feature set* y no de un cambio de planteamiento.

- **Presupuesto computacional:** si el número de alternativas excede el presupuesto definido, se trabaja con una muestra amplia pero acotada, seleccionada de forma reproducible, y se conserva siempre el conjunto completo como referencia interna.

Con esto se obtiene un conjunto de candidatos suficientemente diverso para explorar estructuras, pero lo bastante acotado como para ejecutar la etapa sin comprometer la viabilidad computacional.

Cómo se ordenan los candidatos: precisión, estabilidad y costo

El objetivo central de la Etapa B es **ordenar** conjuntos de variables usando únicamente información del periodo de entrenamiento, mediante un esquema de validación temporal. Para cada candidato se reportan tres señales complementarias:

- **Precisión (criterio principal):** error absoluto medio promedio en las particiones temporales, por su lectura directa en unidades y su estabilidad para comparar alternativas.
- **Estabilidad:** variación del error entre particiones, útil para distinguir conjuntos que funcionan “bien en promedio” de aquellos que mantienen un comportamiento más consistente ante cambios de régimen.
- **Costo:** tiempo de evaluación bajo el mismo protocolo, para identificar alternativas cercanas en desempeño pero más eficientes.

Así, el ranking no se limita a “quién tiene el menor error”, sino que también incorpora **robustez** y **viabilidad** como criterios relevantes para una solución aplicable.

Fase 1: filtrado amplio con evaluación ligera

En la Fase 1 se evalúa un conjunto amplio de candidatos con una validación temporal más ligera y un modelo fijo (sin optimización). La intención es que el ordenamiento dependa principalmente de **las variables** y no de decisiones finas de ajuste. Como salida se obtiene un listado preliminar con precisión, estabilidad y tiempo, a partir del cual se define un grupo reducido de alternativas para revisión.

Fase 2: re-evaluación del grupo preseleccionado

En la Fase 2 se toma únicamente el grupo preseleccionado y se re-evalúa con una validación temporal más exigente. Este paso reduce la probabilidad de que

una alternativa haya quedado bien posicionada por un resultado favorable puntual y produce un ranking **más confiable** para decidir qué conjuntos avanzan a la optimización.

Criterio de paso a la Etapa C: precisión y simplicidad

Con el ranking robusto se define el grupo final de conjuntos que pasan a la Etapa C combinando dos rutas:

1. **Ruta por precisión:** se eligen los candidatos con mejor desempeño promedio en la fase robusta.
2. **Ruta por simplicidad:** se incluyen alternativas más compactas y/o más eficientes cuando resultan cercanas en desempeño, porque pueden ser preferibles en operación por facilidad de mantenimiento y menor costo.

Este criterio de selección formaliza el propósito del capítulo: avanzar hacia una solución que no solo sea precisa, sino también **implementable** bajo restricciones reales.

7.1.6. Etapa C: optimización de modelos en dos fases por conjunto de variables

Una vez definida, en la Etapa B, la lista de *feature sets* que vale la pena llevar a la práctica (por **precisión** y por **simplicidad**), el siguiente paso consiste en **ajustar los modelos** sobre esos conjuntos bajo el mismo protocolo temporal. La intención de esta etapa no es “probar combinaciones sin criterio”, sino obtener, para cada familia de modelos, una configuración **estable** y **reproducible** que mantenga controlado el costo computacional.

Para lograrlo, la optimización se organiza en **dos fases** que se complementan. Primero se realiza una búsqueda amplia que permite ubicar rápidamente zonas del espacio de hiperparámetros donde el modelo funciona bien; luego, se refina alrededor de la mejor zona encontrada para reducir sensibilidad a una única corrida y consolidar un resultado más consistente.

Qué se optimiza y cómo se evalúa

En términos prácticos, el ajuste se concentra en hiperparámetros que determinan: (i) la **capacidad** del modelo para capturar patrones no lineales, (ii) los mecanismos de **regularización** que evitan sobreajuste, y (iii) componentes de **muestreo** que aportan robustez cuando la señal es ruidosa o parcialmente redundante.

La comparación entre configuraciones se hace siempre con **validación cruzada temporal** y usando como criterio el **error absoluto medio promedio** en validación. Esta elección mantiene una lectura directa en unidades de demanda y evita problemas de interpretación asociados a métricas porcentuales en contextos con ventas bajas o intermitentes.

Fase 1: búsqueda amplia bajo presupuesto

En la primera fase se evalúa un conjunto amplio de configuraciones dentro de rangos deliberadamente extensos. La idea es cubrir alternativas conservadoras y agresivas, de modo que el procedimiento no quede sesgado hacia un único nivel de complejidad. El resultado de esta fase no es un “modelo final”, sino una **región candidata** (o un pequeño grupo de configuraciones cercanas) donde el desempeño en validación temporal es consistentemente favorable.

Random Forest. En esta familia se ajustan, principalmente, parámetros que controlan: (a) el tamaño del ensamble (número de árboles), (b) la complejidad de cada árbol (profundidad y condiciones mínimas para dividir), y (c) la diversidad interna del ensamble (fracción de variables consideradas en las particiones y esquemas de muestreo). En demanda diaria con variabilidad alta, estos elementos suelen ser determinantes para estabilizar predicciones sin “memorizar” ruido.

Modelos de gradiente (XGBoost / LightGBM). En estos modelos la optimización se concentra en el balance entre aprendizaje y complejidad: número de iteraciones y tasa de aprendizaje (que deben analizarse de manera conjunta), parámetros de estructura del árbol (profundidad / hojas) y umbrales mínimos que actúan como regularización. Adicionalmente, se consideran fracciones de submuestreo de filas y columnas, que tienden a mejorar la generalización cuando hay señales correlacionadas (por ejemplo, memoria temporal junto con calendario y variables ambientales).

Fase 2: refinamiento local para consolidar el resultado

La segunda fase parte de la mejor región identificada en la búsqueda amplia y realiza un ajuste más fino en su vecindario. En la práctica, esto implica variar de forma controlada los hiperparámetros más influyentes (por ejemplo, complejidad del árbol y ritmo de aprendizaje) alrededor de los valores que ya mostraron buen desempeño. Esta estrategia aprovecha mejor el presupuesto computacional: en lugar de dispersar esfuerzo en configuraciones poco plausibles, lo concentra en mejorar marginalmente una solución que ya es competitiva y en reducir la variabilidad del resultado final.

Criterios de selección y trazabilidad

La selección de hiperparámetros y la estructura en dos fases responden a tres criterios operativos:

- **Evidencia fuera de muestra:** toda decisión se apoya en validación temporal, manteniendo la coherencia con el uso real (entrenar con pasado y validar en periodos posteriores).
- **Control del costo:** el tiempo de ajuste se trata como un recurso limitado; por ello, la búsqueda se diseña para ser amplia al inicio y más focalizada al final.
- **Reproducibilidad:** se mantienen constantes las particiones temporales, el criterio de puntuación y el protocolo de preprocesamiento, de forma que las mejoras observadas puedan atribuirse al ajuste del modelo y al *feature set*, y no a cambios del procedimiento.

Salida de la Etapa C

La salida de esta etapa es un conjunto de resultados por combinación (*feature set*, *modelo*) que resume: el desempeño obtenido bajo validación temporal (promedio y variabilidad entre particiones), el costo computacional asociado al ajuste y el desempeño final en el conjunto de prueba una vez fijada la mejor configuración. Esta matriz permite comparar alternativas entre familias de modelos manteniendo trazabilidad sobre **qué variables** se usaron y **cómo** se ajustó cada modelo.

7.1.7. Criterios de selección final: precisión vs. eficiencia

La selección final se aborda como una decisión **multi-criterio**. En coherencia con los criterios de **viabilidad operativa** definidos previamente (costo de cómputo y complejidad del modelo), la **precisión** se mantiene como objetivo primario, pero se explicita que, en un despliegue real, también importan el **tiempo de entrenamiento** y la **mantenibilidad** (cantidad de señales y esfuerzo de actualización). Por esta razón, los resultados se reportan priorizando **MAE** como criterio principal y usando **WAPE (%)** como lectura operativa complementaria.

Para cada negocio, el proceso entrega tres alternativas, de modo que la elección final dependa de la restricción dominante:

- **Mejor por familia de modelo:** se selecciona, para cada familia, la configuración con **menor MAE** en el conjunto de prueba, verificando que el **WAPE (%)** sea competitivo.

- **Más rápido dentro del 5 % del mejor:** se identifica una alternativa con desempeño *cercano* al óptimo, pero con menor costo de entrenamiento. La cercanía se define con un margen del 5 % respecto al mejor resultado, usando **WAPE (%)** como referencia operativa.
- **Menos variables dentro del 5 % del mejor:** se selecciona una alternativa que reduzca la complejidad (menor número de variables) sin una pérdida significativa de desempeño, manteniéndose dentro del mismo margen del 5 % respecto al mejor resultado (según **WAPE (%)**).

Adicionalmente, se determina un **mejor global** (considerando todas las familias y conjuntos de variables) y se reportan alternativas cercanas bajo los dos criterios operativos anteriores. En consecuencia, la salida no se limita a un único “ganador”: se presenta un **portafolio de decisiones justificadas** que permite escoger entre *precisión máxima, rapidez de entrenamiento o simplicidad/mantenibilidad*.

7.2. Resultados: Mercaya

Bajo el mismo horizonte de pronóstico ($h = 1$) y el protocolo de evaluación temporal definido en la Sección 7.1, los resultados muestran que, en **Mercaya**, la mayor parte de la capacidad predictiva se concentra en dos fuentes de señal: (i) la **memoria de corto plazo** de la demanda y (ii) un **calendario compacto**. En particular, las variables climáticas no resultan necesarias para alcanzar el mejor desempeño global, lo que sugiere una dinámica predominantemente **endógena**: el comportamiento reciente de ventas, modulado por regularidades del ciclo mensual, explica gran parte de la demanda del día siguiente.

7.2.1. Conjunto ganador: variables finales e interpretación

En Mercaya, los ganadores por familia de modelo convergen en el mismo conjunto `L4\_R3\_M0\_CAL3\_WX0`. Esta convergencia es valiosa por dos razones: primero, entrega una alternativa operativa clara (un conjunto reutilizable); y segundo, respalda que la señal dominante identificada es **estable** y no depende de una arquitectura específica. Dado que la nomenclatura y el significado de cada bloque (**L**, **R**, **M**, **CAL**, **WX**) ya fueron establecidos en el Marco Teórico, aquí se enfatiza la lectura específica de Mercaya a partir de las variables efectivamente seleccionadas.

Cuadro 7.1: Mercaya: conjunto ganador L4_R3_M0_CAL3_WX0 y variables incluidas.

Bloque	Variables
Memoria (rezagos)	lag_1, lag_2, lag_3, lag_4
Memoria (ventanas)	rolling_2, rolling_3, rolling_4
Calendario	dia, quincena, mes
Identidad	productoid

Lectura del conjunto (qué aporta para predecir y_{t+1}). En términos prácticos, este conjunto responde a tres necesidades del problema en Mercaya: (i) conservar memoria suficiente para representar el nivel y la dirección reciente de la demanda, (ii) capturar regularidades del ciclo mensual con pocas variables calendario, y (iii) sostener el enfoque de **modelo global multiserie** mediante el identificador del producto.

- **Rezagos inmediatos (lag_1–lag_4).** Indican que la información más útil está en el comportamiento de los últimos días. En Mercaya, incluir cuatro rezagos consecutivos permite que el modelo reaccione a cambios recientes sin depender de una memoria más larga que, para $h = 1$, no aporta una ganancia clara.
- **Ventanas cortas (rolling_2–rolling_4).** Complementan a los rezagos al resumir el volumen reciente y amortiguar ruido diario. En la práctica, estas ventanas ayudan a estabilizar la predicción cuando hay fluctuaciones transitorias sin cambiar el nivel subyacente.
- **Calendario compacto (mes, quincena, dia).** Sugiere que la variación relevante está asociada a regularidades **mensuales y quincenales** (p. ej., ciclo de ingresos/rotación), y que no se requiere un calendario más fino para explicar el día siguiente bajo este horizonte.
- **Identificador del producto (productoid).** Mantiene la capacidad del modelo global para diferenciar patrones y escalas entre productos, evitando entrenar un modelo independiente por cada referencia.

Ausencias informativas (qué sugiere lo que no quedó seleccionado). Más que una lista de variables descartadas, estas ausencias ayudan a delimitar qué señal *no* fue determinante para generalizar en el *holdout*:

- **Clima ausente en el óptimo (WX0).** En Mercaya, el posible efecto climático parece ser débil, indirecto o ya absorbido por la memoria y el calendario. Operativamente, esto es una ventaja: reduce dependencia de fuentes externas y simplifica el despliegue.

- **Memoria más larga ausente** (`lag_5..lag_7` y `rolling_5..rolling_7`). Extender la memoria no mejora lo suficiente para $h = 1$ o introduce un suavizamiento que resta sensibilidad ante cambios rápidos.
- **Calendario adicional ausente** (p. ej., `ano`, `dia-num`). Indica que, para este negocio y horizonte, el patrón relevante ya queda capturado por `mes/quincena` y una variable de día, y que añadir más granularidad no aporta una mejora consistente fuera de muestra.

7.2.2. Ganadores por modelo

Cuadro 7.2: Mercaya: mejor configuración por familia en prueba

Modelo	Feature set	MAE	WAPE %	R^2	R^2_{adj}	MSE	Tiempo (s)	Vars crudas
RF_TUNED	L4\R3\M0_CAL3\WX0	933.04	31.10	0.8321	0.8096	$7,04 \times 10^6$	228.98	11
XGB_TUNED	L4\R3\M0_CAL3\WX0	984.10	32.80	0.7967	0.7694	$8,53 \times 10^6$	9.59	11
LGBM_TUNED	L4\R3\M0_CAL3\WX0	1111.63	37.05	0.7473	0.7134	$1,06 \times 10^7$	51.79	11
Linear	L4\R3\M0_CAL3\WX0	1117.42	37.24	0.8212	0.7972	$7,50 \times 10^6$	0.01	11

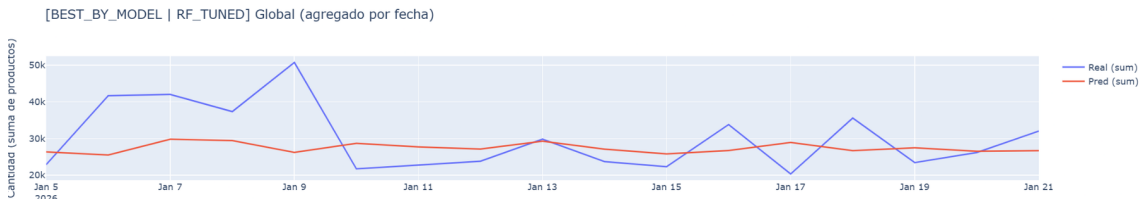


Figura 7.1: Mercaya: mejor configuración del modelo Random Forest (RF_TUNED). Comparación global entre demanda real y predicha agregada por fecha.

Interpretación. En Mercaya, el mejor desempeño global corresponde a **RF_TUNED**, al presentar el menor **MAE** y **WAPE %** y, simultáneamente, el menor **MSE** entre alternativas, lo cual implica mejor control de errores grandes (episodios de sobre/subestimación de alta magnitud). En inventarios, este balance es especialmente deseable porque reduce el error típico y mitiga el riesgo de decisiones extremas de compra.

La Figura 7.1 complementa la lectura numérica y permite diagnosticar *cómo* se materializa el error. Se observa que la predicción sigue la trayectoria general, pero con un patrón consistente de **suavizamiento**: los máximos pronunciados tienden a subestimarse y las caídas abruptas tienden a amortiguarse. Este comportamiento es coherente con un modelo que combina lags y ventanas para priorizar estabilidad frente a variaciones transitorias. Operativamente, esto evita reaccionar de forma

exagerada ante picos aislados; sin embargo, también sugiere que los residuales más grandes se concentrarán en días con cambios rápidos, donde conviene complementar el pronóstico con reglas de negocio (p. ej., stock de seguridad o umbrales de reposición).

En comparación, **XGB_TUNED** ofrece un costo de entrenamiento mucho menor, pero cede precisión (MAE/WAPE %) y aumenta el MSE, evidenciando menor robustez ante errores grandes. Por su parte, **Linear** mantiene un R^2 relativamente alto, pero exhibe MAE/WAPE % elevados, consistente con limitaciones para capturar no linealidades e interacciones (producto–calendario–memoria).

7.2.3. Alternativas globales cercanas: rapidez y menos variables

Además del mejor global, se seleccionaron dos alternativas que permanecen dentro de un margen del 5 % respecto al óptimo (en WAPE), privilegiando criterios operativos: (i) **menor tiempo de entrenamiento** y (ii) **menor número de variables crudas**. Esta comparación es especialmente útil cuando el modelo debe recalibrarse con frecuencia o ejecutarse en entornos con restricciones de cómputo y de mantenimiento.

Cuadro 7.3: Mercaya: mejor global, alternativas dentro del 5 % y mención de una opción de entrenamiento muy rápido

Tipo	Modelo	Feature set	MAE	WAPE %	Tiempo (s)	Vars crudas
Mejor global	RF_TUNED	L4\R3\M0\CAL3\WX0	933.04	31.10	228.98	11
Más rápido (5 %)	RF_TUNED	L2\R2\M3\CAL4\WX0	972.55	32.41	210.11	12
Menos vars (5 %)	RF_TUNED	L2\R3\M0\CAL2\WX1	979.67	32.65	225.92	9
Mención (muy rápido)	XGB_TUNED	L4\R3\M0\CAL3\WX0	984.10	32.80	9.59	11

Lectura del compromiso (trade-off). Dentro del umbral del 5 %, la alternativa **más rápida** reduce el tiempo de entrenamiento (228.98 s $\rightarrow$ 210.11 s) con un aumento leve del error (WAPE 31.10 % $\rightarrow$ 32.41 %). En términos prácticos, esta opción es conveniente cuando la prioridad es *iterar más rápido* (reentrenamientos, pruebas de sensibilidad o despliegues frecuentes) sin alejarse del desempeño del óptimo.

Por su parte, la alternativa de **menos variables** reduce la dimensionalidad cruda (11 $\rightarrow$ 9) con una penalización marginal (WAPE 31.10 % $\rightarrow$ 32.65 %), lo cual favorece mantenibilidad: menos señales que monitorear, menor esfuerzo de actualización y menor exposición a fallas por variables externas.

Mención: una alternativa muy rápida con XGBoost y el mismo conjunto del óptimo. Al revisar el *top* de configuraciones también aparece **XGB_TUNED**

usando exactamente el mismo conjunto del mejor global, L4_R3_M0_CAL3_WX0, con un desempeño cercano (WAPE 32.80 %) pero con un tiempo de entrenamiento **drásticamente menor** (9.59 s). Aunque queda ligeramente por encima del umbral del 5 % (por décimas), es una opción relevante cuando la restricción dominante es el **tiempo**. Este resultado es coherente con la naturaleza de XGBoost: suele entrenar con gran eficiencia gracias a implementaciones altamente optimizadas y a que el *boosting* puede lograr buena precisión con configuraciones de árboles relativamente compactas, mientras que Random Forest tiende a requerir muchos árboles para estabilizar el ensamble, elevando el costo computacional.

Cuadro 7.4: Mercaya: variables en alternativas cercanas e interpretación

Feature set	Variables incluidas	Lectura principal
L2_R2_M3_CAL4_WX0	lag_1, lag_2, rolling_2, rolling_3, rolling_mean_2..rolling_mean_4, dia, dia_habil, quincena, mes, productoid	Reemplaza parte de la memoria fina por (lag_3, lag_4) por resúmenes más estables del nivel reciente (rolling_mean). La inclusión de dia_habil sugiere que, al compactar memoria, un calendario operativo (hábil/no hábil) ayuda a recuperar parte de la variación explicable.
L2_R3_M0_CAL2_WX1	lag_1, lag_2, rolling_2, rolling_3, rolling_4, dia, ano, temp_c, productoid	La parsimonia se logra reduciendo el bloque de calendario y agregando una señal ambiental numérica (temp_c). Interpretativamente, esto sugiere que el contexto puede aportar cuando se simplifica el conjunto, aunque no sea imprescindible para el óptimo.

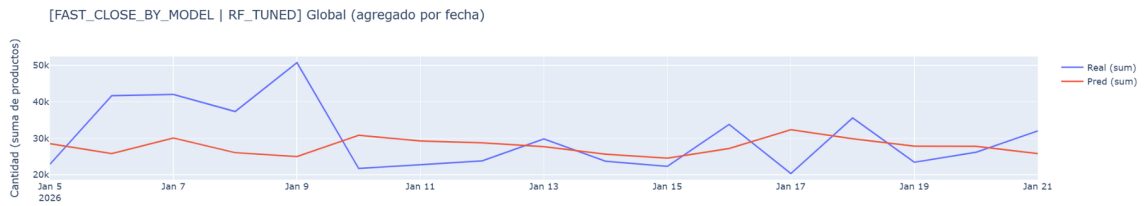


Figura 7.2: Mercaya: alternativa global *más rápida* dentro del 5 % (RF_TUNED). Comparación entre demanda real y predicha agregada por fecha.

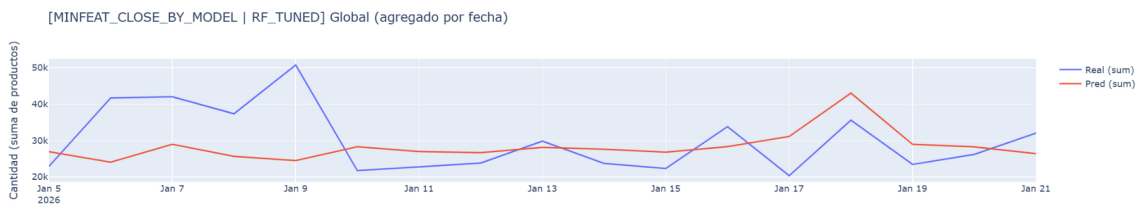


Figura 7.3: Mercaya: alternativa global con *menos variables* dentro del 5 % (RF_TUNED). Comparación entre demanda real y predicha agregada por fecha.

Qué significa que cambien las variables en las alternativas. Lectura cualitativa a nivel global. En ambas alternativas (Figuras 7.2 y 7.3) se mantiene el patrón ya observado: el modelo sigue el nivel general de la serie agregada, pero tiende a **suavizar** extremos (picos y caídas abruptas). La diferencia es de *grado*: al priorizar rapidez o parsimonia, parte del detalle puntual se sustituye por resúmenes del comportamiento reciente o por señales más gruesas, lo cual concentra las mayores discrepancias en días con cambios rápidos de nivel.

En la alternativa rápida L2\R2\M3\CAL4\WX0, el uso de promedios móviles favorece estabilidad: cuando la demanda cambia de forma relativamente persistente, estos resúmenes capturan bien el nivel y reducen variabilidad día a día; el costo es perder sensibilidad ante cambios bruscos. En la alternativa de menos variables L2\R3\M0\CAL2\WX1, la reducción del calendario y la incorporación de `temp_c` tienden a producir un agregado más conservador: útil para operación simple, pero con menor capacidad de respuesta frente a episodios atípicos.

Implicación operativa. Estas alternativas funcionan bien cuando se busca una planificación estable (por ejemplo, como base para reposición rutinaria), porque evitan reacciones exageradas ante ruido diario. Sin embargo, como el error tiende a concentrarse en extremos, conviene acompañarlas con salvaguardas operativas: **stock de seguridad** para picos (mitigar subabastecimiento) y reglas/umbrales para descensos rápidos (evitar sobreabastecimiento). En síntesis, la eficiencia se paga principalmente con menor sensibilidad a la volatilidad, más que con pérdida del nivel medio.

Diagnóstico desagregado por producto y modelo (Mercaya)

Además de la comparación agregada por fecha (suma de productos), se incluye un diagnóstico **desagregado por producto** para inspeccionar el comportamiento del modelo en series con **escalas** y **patrones** distintos (alta rotación estable, rotación intermitente y picos puntuales). Este contraste es relevante porque un buen ajuste global puede ocultar errores sistemáticos en subconjuntos de productos; por ejemplo, subestimación de picos en referencias altamente volátiles, o **sobreestimaciones puntuales** en series de baja demanda.



Figura 7.4: Mercaya: diagnóstico desagregado por producto para el **mejor global** (RF_TUNED). Comparación de demanda real vs. predicha (series individuales).

CAPÍTULO 7. OPTIMIZACIÓN Y EVALUACIÓN CONJUNTA: FEATURES, HIPERPARÁMETROS Y EFICIENCIA



Figura 7.5: Mercaya: diagnóstico desagregado por producto para la alternativa **más rápida** dentro del 5 % (RF_TUNED). Comparación de demanda real vs. predicha (series individuales).

CAPÍTULO 7. OPTIMIZACIÓN Y EVALUACIÓN CONJUNTA: FEATURES, HIPERPARÁMETROS Y EFICIENCIA



Figura 7.6: Mercaya: diagnóstico desagregado por producto para la alternativa con **menos variables** dentro del 5 % (RF_TUNED). Comparación de demanda real vs. predicha (series individuales).

Hallazgos. Al inspeccionar las series individuales (Figuras 7.4–7.6) se observan patrones consistentes:

- **Amortiguación del ruido (suavizamiento).** En productos de mayor rotación y escala, la predicción tiende a ser más estable que la serie real: **reduce la variabilidad** y suele **subestimar picos** y **sobreestimar valles**. Este comportamiento es coherente con un conjunto de variables dominado por señales de memoria (rezagos y ventanas), que introduce inercia para ganar robustez fuera de muestra.
- **Heterogeneidad por escala.** Coexisten productos con magnitudes muy altas y productos con demanda baja. En consecuencia, discrepancias relativas pequeñas pueden traducirse en errores absolutos grandes en series de alta escala, mientras que en series de baja demanda el error se concentra en **episodios puntuales** (días atípicos) más que en el nivel promedio.
- **Mayor dificultad ante cambios rápidos de nivel.** Cuando la demanda presenta saltos o quiebres, el ajuste puede reaccionar con retraso o con menor amplitud, lo que sugiere que parte del error se asocia a eventos operativos no observados explícitamente (p. ej., promociones, ruptura de inventario, reposición tardía o cambios de exhibición).

- **Picos espurios en la alternativa con menos variables.** En la configuración con menos variables (Figura 7.6) se observa, en varios productos de baja demanda, un **pico predicho pronunciado y sincronizado** (alrededor del 18 de enero) que no está presente en la serie real. Este patrón sugiere sensibilidad a una señal común (p. ej., calendario) o a combinaciones de variables que, al reducir el *feature set*, generan **sobreestimaciones puntuales** (falsos positivos).

Comparación (mejor global vs. rápida vs. menos variables). El contraste entre escenarios permite identificar cómo se materializa el intercambio precisión–eficiencia:

- **Mejor global.** Presenta el equilibrio más consistente entre seguimiento del nivel y control de variabilidad: reproduce adecuadamente la trayectoria general de cada producto y concentra los residuales más grandes en episodios aislados (picos o quiebres), manteniendo estabilidad en el resto del periodo.
- **Alternativa más rápida.** En la muestra visual presentada, mantiene una forma **muy similar** a la del mejor global (y en varios productos resulta prácticamente indistinguible), lo que sugiere que, en este escenario, la ganancia de eficiencia se logra con un impacto marginal sobre el ajuste por producto.
- **Alternativa con menos variables.** Es la que exhibe diferencias más notorias: tiende a un ajuste **más conservador** (aproximación al nivel medio) y, adicionalmente, aparecen **sobreestimaciones puntuales** (picos espurios) en productos de baja demanda. En consecuencia, el error se desplaza con mayor probabilidad hacia episodios extremos (picos/caídas) aun cuando se conserve un desempeño global competitivo.

Implicación operativa. En Mercaya, este diagnóstico refuerza que el modelo funciona bien como **línea base estable** para reposición diaria, pues evita reacciones exageradas ante ruido de corto plazo. No obstante, dado que los residuales más grandes se concentran en **extremos y cambios rápidos de nivel**, se recomienda acompañar el pronóstico con salvaguardas: (i) **stock de seguridad** para mitigar subestimación en picos (reducir riesgo de desabastecimiento), (ii) reglas de revisión ante variaciones abruptas (umbrales sobre cambios recientes o alertas cuando la demanda observada se aparte del patrón esperado) y (iii) en configuraciones con reducción de variables, validaciones adicionales para evitar **picos espurios** que podrían inducir **sobrestock** si la reposición se basa exclusivamente en el pronóstico.

Lectura complementaria. En general, el modelo reproduce adecuadamente el **nivel** y la **dirección** de la demanda en productos de comportamiento relativamente estable, mientras que en productos con **cambios abruptos** tiende a presentar un efecto de **suavizamiento** (subestimación de picos y amortiguación de caídas). Adicionalmente, al reducir el conjunto de variables pueden emerger **sobreestimaciones puntuales** en series de baja demanda. En términos operativos, esta evidencia es consistente

con la lectura global: alternativas cercanas conservan desempeño promedio, pero las discrepancias más relevantes se concentran en episodios de alta variabilidad o en eventos atípicos; por ello, se recomienda acompañar el pronóstico con salvaguardas (p. ej., stock de seguridad y reglas de excepción) para reducir el riesgo de decisiones subóptimas de reposición.

7.3. Resultados: Villa Café

Bajo el mismo horizonte ($h = 1$) y el mismo protocolo de evaluación temporal, en **Villa Café** se confirma que la mayor parte de la capacidad predictiva proviene de la **memoria temporal** de la demanda. Sin embargo, a diferencia de **Mercaya**, el mejor desempeño global se obtiene cuando el modelo incorpora una señal exógena *categorica* asociada al estado del clima (`clima_desc`). En términos prácticos, esto sugiere una dinámica menos puramente endógena: además del historial reciente y de la rutina del calendario, existen *regímenes* ambientales (o su efecto indirecto sobre afluencia y preferencias) que ayudan a explicar variaciones del consumo de un día a otro.

7.3.1. Conjunto ganador: variables finales e interpretación

En Villa Café, el mejor global corresponde al conjunto `L4\_R2\_M0\_CAL3\_WX1`, el cual combina una memoria que mezcla corto plazo con una referencia semanal y, adicionalmente, activa el bloque de clima mediante `clima_desc`. La Tabla 7.5 resume las variables incluidas.

Cuadro 7.5: Villa Café: conjunto `L4\_R2\_M0\_CAL3\_WX1` y variables incluidas

Bloque	Variables
Memoria (lags)	<code>lag_1</code> , <code>lag_3</code> , <code>lag_5</code> , <code>lag_7</code>
Memoria (ventanas)	<code>rolling_2</code> , <code>rolling_3</code>
Calendario	<code>dia</code> , <code>dia_habil</code> , <code>quincena</code>
Contexto	<code>clima_desc</code>
Identidad	<code>productoid</code>

Qué significa que estas variables estén (y otras no). Como se discutió en el marco metodológico (bloques de señal y etiquetas `\cdot`), aquí interesa resaltar qué información resultó realmente necesaria para explicar y_{t+1} en este negocio:

- **Lags espaciados (1,3,5,7).** La selección sugiere que no basta con una “continuidad diaria” corta; el modelo se beneficia de anclas que recogen persistencias

y regularidades cercanas a una semana, además de cambios de nivel que no siempre se manifiestan de forma suave día a día.

- **Ventanas cortas (2–3).** Estas ventanas aportan un resumen del nivel muy reciente sin “diluir” demasiado la señal; es decir, estabilizan el pronóstico, pero conservan sensibilidad ante cambios relativamente rápidos.
- **Calendario operativo (dia_habil y quincena).** La combinación apunta a que la rutina (hábil/no hábil) y el ciclo de ingresos son relevantes para el consumo cotidiano; en comparación con un calendario más extenso, aquí se prioriza lo que suele tener lectura directa en operación.
- **clima_desc como marcador de régimen.** Su presencia indica que el contexto ambiental introduce variaciones sistemáticas (directas o indirectas) que no quedan completamente absorbidas por la memoria y el calendario, por lo que aporta capacidad explicativa adicional.
- **productoid.** Se mantiene el enfoque multiserie: el modelo requiere la identidad del producto para aprender diferencias estructurales entre ítems (escala y patrón) dentro de un entrenamiento global.

Ausencias informativas. La ausencia de ventanas largas es coherente con la necesidad de responder a cambios de nivel más rápidos, evitando un suavizamiento excesivo.

7.3.2. Ganadores por modelo: métricas resumidas

Cuadro 7.6: Villa Café: mejor configuración por familia en prueba

Modelo	Feature set	MAE	WAPE %	R^2	R^2_{adj}	MSE	Tiempo (s)	Vars crudas
RF_TUNED	L4_R2_M0_ CAL3_WX1	1099.53	35.22	0.7840	0.7483	$1,02 \times 10^7$	200.74	11
XGB_TUNED	L3_R3_M2_ CAL3_WX1	1284.24	41.14	0.7221	0.6715	$1,32 \times 10^7$	10.01	13
LGBM_TUNED	L3_R2_M2_ CAL4_WX1	1180.99	37.83	0.7327	0.6841	$1,27 \times 10^7$	55.31	13
Linear	L3_R3_M0_ CAL2_WX0	1163.84	37.28	0.7588	0.7300	$1,14 \times 10^7$	0.01	9

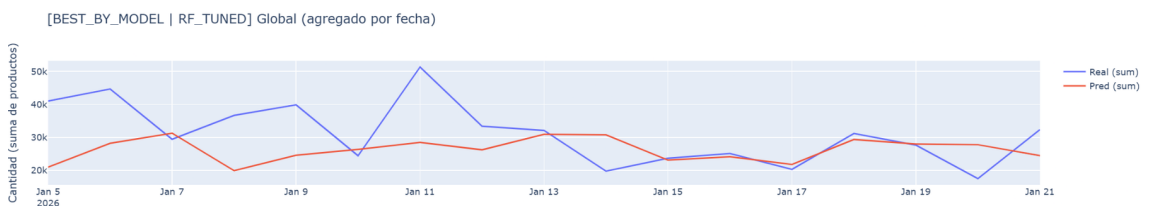


Figura 7.7: Villa Café: desempeño global del mejor modelo en prueba (RF_TUNED). Comparación entre demanda real y predicha agregada por fecha (suma de productos).

Interpretación. En Villa Café, **RF_TUNED** ofrece el mejor balance: logra el menor MAE y el menor WAPE %, con un ajuste global sólido ($R^2 \approx 0,78$). Frente a **Mercaya**, el WAPE % más alto y el R^2 menor sugieren una demanda menos “endógena” y más influida por el contexto; en esa línea, la presencia de **clima_desc** en el conjunto ganador respalda la existencia de *regímenes* ambientales que desplazan el consumo.

Operativamente, el modelo constituye una base adecuada para reposición diaria, pero conviene acompañarlo con **stock de seguridad** y **umbrales de revisión** en días de mayor variabilidad, donde el pronóstico tiende a suavizar picos y ajustes abruptos.

La Figura 7.7 complementa la lectura numérica: al igual que en Mercaya, se observa un seguimiento adecuado del nivel general, pero con un suavizamiento que hace más visibles los residuales en días de cambios rápidos. Operativamente, esto sugiere acompañar el pronóstico con salvaguardas (por ejemplo, **stock de seguridad** y **umbrales/reglas** para episodios de alta volatilidad), de forma que el modelo se use como base de decisión sin subestimar escenarios extremos.

7.3.3. Alternativas globales cercanas: rapidez y menos variables

Además del mejor global, se identificaron alternativas que permanecen dentro de un margen del 5 % respecto al óptimo (en **WAPE %**), con el propósito de privilegiar criterios operativos: (i) **menor tiempo de entrenamiento** (útil cuando se reentrena con frecuencia) y (ii) **menor número de variables crudas** (útil cuando se busca simplicidad de mantenimiento).

Cuadro 7.7: Villa Café: mejor global y alternativas cercanas (5 %)

Tipo	Modelo	Conjunto de variables	MAE	WAPE %	Tiempo (s)	Vars crudas
Mejor global	RF_TUNED	L4\R2\M0\CAL3\WX1	1099.53	35.22	200.74	11
Más rápido (5 %)	RF_TUNED	L4\R0\M3\CAL4\WX0	1115.81	35.75	198.58	12
Menos vars (5 %)	RF_TUNED	L4\R2\M0\CAL3\WX1	1099.53	35.22	200.74	11

Lectura del compromiso (trade-off). La alternativa **más rápida** reduce ligeramente el tiempo de entrenamiento (200.74 s $\rightarrow$ 198.58 s) manteniéndose dentro del umbral del 5 %, con un aumento pequeño del error relativo agregado (WAPE % 35.22 $\rightarrow$ 35.75). En la práctica, esta opción resulta atractiva cuando la prioridad es *iterar con mayor agilidad* (reentrenamientos, pruebas de sensibilidad o comparaciones recurrentes) sin alejarse del desempeño del óptimo.

Nota interpretativa (parsimonia). Bajo el criterio de **menor número de variables dentro del 5 % del mejor global**, no aparece un conjunto más parsimonioso

distinto al ganador; por ello, la alternativa de menos variables coincide con el **mejor global**. Esto sugiere que el conjunto óptimo ya es *compacto* y que retirar variables adicionales degrada el desempeño más allá del margen permitido.

Cuadro 7.8: Villa Café: variables en alternativas cercanas e interpretación

Conjunto de variables	Variables incluidas	Lectura principal
L4_R0_M3_CAL4_WX0 (<i>rápida</i>)	lag_1, lag_2, lag_4, lag_7, rolling_mean_2..rolling_mean_4, dia, dia_habil, ano, mes, productoid	Sustituye clima_desc por estacionalidad (ano, mes) y un resumen más estable del nivel reciente (rolling_mean). Mantiene un WAPE % cercano con tiempo marginalmente menor, lo que sugiere que una parte del contexto puede aproximarse con calendario y suavización; sin embargo, el óptimo conserva clima_desc, indicando que el <i>régimen</i> climático aporta información adicional cuando se busca el mejor desempeño.
L4_R2_M0_CAL3_WX1 (<i>menos variables</i>)	lag_1, lag_3, lag_5, lag_7, rolling_2, rolling_3, dia, dia_habil, quincena, clima_desc, productoid	Coincide con el mejor global dentro del umbral del 5 %. En términos prácticos, sugiere que los lags espaciados y clima_desc son componentes <i>difíciles de remover</i> si se desea conservar el desempeño en Villa Café.

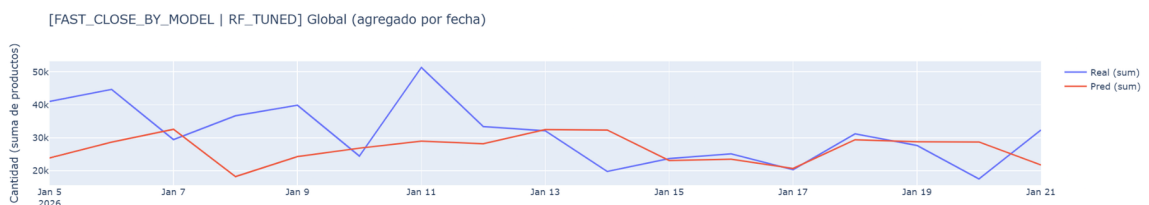


Figura 7.8: Villa Café: alternativa global *más rápida* (RF_TUNED). Comparación global entre demanda real y predicha agregada por fecha.

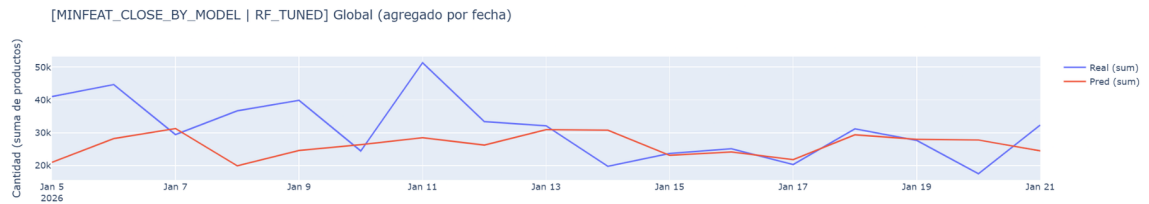


Figura 7.9: Villa Café: alternativa global con *menos variables* (RF_TUNED). Comparación global entre demanda real y predicha agregada por fecha.

Qué cambia en las variables de la alternativa rápida. Lectura cualitativa.

La lectura cualitativa sigue el mismo patrón descrito previamente en Mercaya: a nivel agregado, las alternativas tienden a **suavizar extremos** (picos y caídas abruptas), concentrando los residuales más grandes en días de **cambio rápido de nivel**. La diferencia, en Villa Café, es que este comportamiento se vuelve más sensible a *cambios de régimen*: la alternativa rápida prescinde de `clima_desc` y compensa con calendario (`ano/mes`) y promedios móviles, lo cual funciona cuando el contexto actúa de forma relativamente persistente; pero el hecho de que el óptimo mantenga `clima_desc` sugiere que existen días donde, aun con historial similar, el contexto climático cambia el consumo esperado.

Implicación operativa. Si el objetivo es robustez ante regímenes, el conjunto óptimo con `clima_desc` es la opción preferible. Si la prioridad es *agilidad* (reentrenar y comparar con frecuencia) o reducir dependencia de esa señal exógena, la alternativa rápida ofrece un desempeño cercano dentro del umbral, con una formulación que descansa más en calendario y suavización. En ambos casos, conviene mantener las salvaguardas operativas ya discutidas (p. ej., stock de seguridad y reglas/umbrales) para mitigar errores en extremos.

Análisis desagregado por producto: mejor global, alternativa rápida y alternativa con menos variables de Villa Café

Además de la comparación agregada por fecha, se realiza un diagnóstico **por producto** en el conjunto de prueba con el fin de verificar si el desempeño global se mantiene de manera consistente a nivel de ítem o si existen referencias donde el error se concentra. Este análisis es relevante porque la agregación puede **ocultar** comportamientos heterogéneos: productos de alta rotación (mayor escala), productos de baja demanda e **intermitencia**, y series con **picos aislados** o cambios bruscos.

En particular, la Figura 7.10 presenta el escenario *mejor global*, la Figura 7.11 presenta la alternativa *más rápida* dentro del 5%, y la Figura 7.12 presenta la alternativa con *menos variables* dentro del 5%.

CAPÍTULO 7. OPTIMIZACIÓN Y EVALUACIÓN CONJUNTA: FEATURES, HIPERPARÁMETROS Y EFICIENCIA

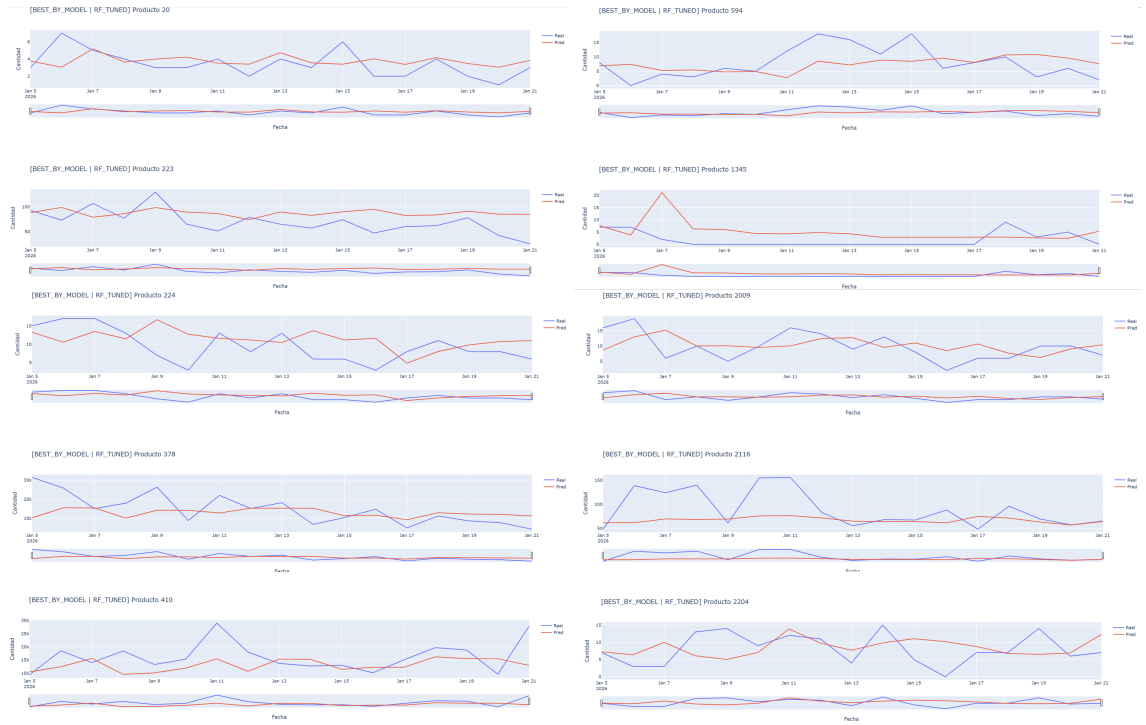


Figura 7.10: Villa Café: comparación desagregada por producto entre demanda real y predicha en prueba. Escenario *mejor global* (RF_TUNED).



Figura 7.11: Villa Café: comparación desagregada por producto entre demanda real y predicha en prueba. Escenario *más rápido dentro del 5%* (RF_TUNED).

CAPÍTULO 7. OPTIMIZACIÓN Y EVALUACIÓN CONJUNTA: FEATURES, HIPERPARÁMETROS Y EFICIENCIA

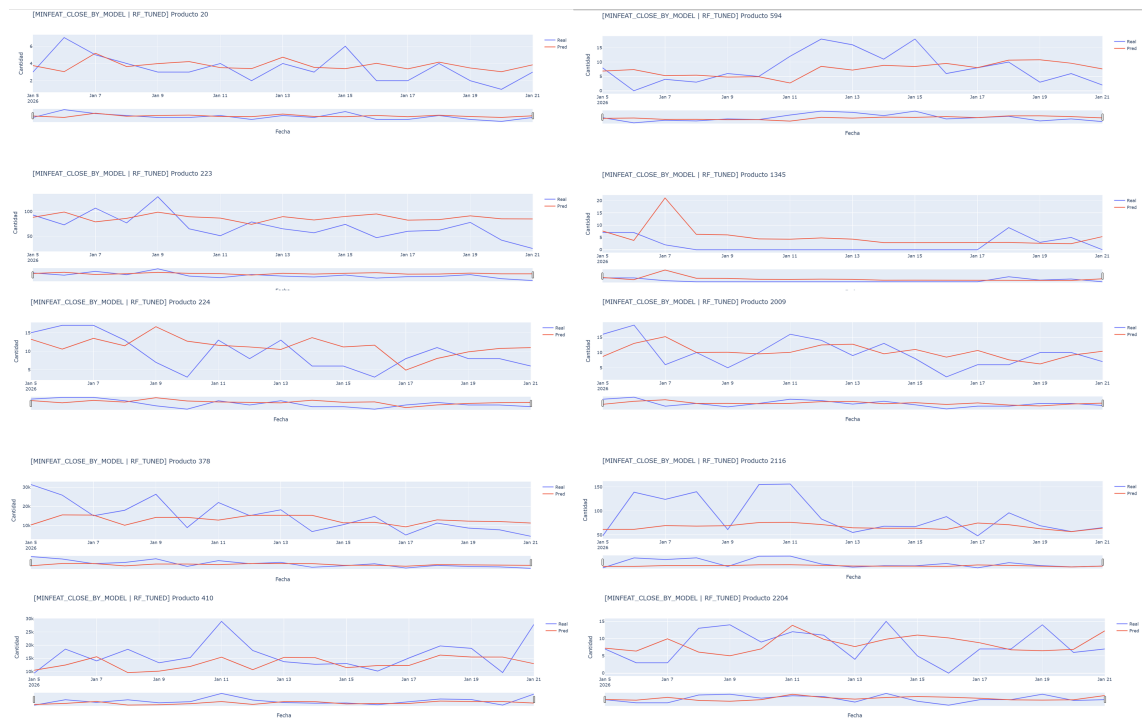


Figura 7.12: Villa Café: comparación desagregada por producto entre demanda real y predicha en prueba. Escenario *menos variables dentro del 5 %* (RF_TUNED).

Hallazgos principales (lectura por producto). La inspección desagregada (Figuras 7.10–7.12) confirma patrones operativamente relevantes:

- **Seguimiento del nivel con suavizamiento.** En la mayoría de productos, las predicciones preservan el **nivel promedio** y, en términos generales, la **dirección** de la serie; sin embargo, tienden a **suavizar extremos**. Esto se manifiesta como **subestimación de picos** (p. ej., en series de mayor escala) y **amortiguación** de variaciones abruptas, concentrando los residuales en días de cambio rápido.
- **Heterogeneidad por escala.** Se observan productos con escalas muy distintas (desde unidades hasta magnitudes significativamente mayores). En consecuencia, en productos de alta escala, discrepancias relativas moderadas pueden traducirse en **errores absolutos** grandes; mientras que, en productos de baja escala, el error se hace más visible en **eventos puntuales** (picos aislados) que en el nivel medio.
- **Demanda baja/intermitente y sesgo hacia valores medios.** En productos con demanda baja o con periodos cercanos a cero, el modelo tiende a generar predicciones *conservadoras* (cercanas al promedio reciente), lo que puede inducir (i) **sobreestimación** en días con demanda nula y (ii) **subestimación** cuando ocurre un salto repentino. En estos casos también pueden aparecer **picos predichos no observados** en la serie real, lo que sugiere que el ajuste

privilegia regularidades globales por encima de eventos raros específicos de un ítem.

Comparación entre escenarios (mejor vs. rápido vs. menos variables). A nivel de producto, las tres alternativas exhiben un comportamiento **muy similar** en la muestra visual considerada, lo cual sugiere que el intercambio precisión–eficiencia es **marginal** en este escenario:

- **Mejor global.** Proporciona el ajuste más consistente en términos de equilibrio entre seguimiento del nivel y control de la variabilidad, concentrando los mayores residuales en episodios de picos o quiebres.
- **Más rápido.** Mantiene una forma global prácticamente equivalente a la del mejor global en los productos ilustrados, por lo que resulta adecuado cuando la restricción dominante es el **tiempo de iteración** (reentrenamientos frecuentes o ejecución operativa).
- **Menos variables.** Conserva el comportamiento general observado en las otras alternativas; no obstante, al reducir señales, puede reforzar un ajuste **más parsimonioso** (mayor aproximación al promedio) en productos con alta variabilidad o intermitencia, donde el error tiende a concentrarse en picos y transiciones.

Implicación operativa. La evidencia por producto respalda el uso del *mejor global* como opción más estable cuando se prioriza robustez, y respalda la alternativa *más rápida* cuando se requiere **reentrenar o iterar** con mayor frecuencia. En los tres escenarios, los productos con picos pronunciados o intermitencia justifican salvaguardas: (i) **stock de seguridad** para mitigar riesgo de subabastecimiento por subestimación de picos, y (ii) **reglas de excepción** (alertas/umbrales) para revisión manual ante variaciones bruscas o predicciones atípicas, reduciendo el riesgo de decisiones subóptimas de reposición (desabastecimiento vs. sobrestock).

7.4. Comparación integral entre Mercaya y Villa Café

7.4.1. Comparación global de desempeño

La Tabla 7.9 resume, para cada negocio, la configuración global ganadora en prueba bajo el mismo horizonte ($h = 1$) y el mismo protocolo temporal. Esta comparación es relevante porque permite separar qué parte del desempeño proviene

del *procedimiento* (fijo e idéntico) y qué parte se explica por la *dinámica propia* de cada negocio.

Cuadro 7.9: Mejor global por negocio (prueba)

Negocio	Modelo	Conjunto de variables	MAE	WAPE %	R^2	MSE	Tiempo (s)
Mercaya	RF_TUNED	L4_R3_M0_CAL3_WX0	933.04	31.10	0.8321	$7,04 \times 10^6$	228.98
Villa Café	RF_TUNED	L4_R2_M0_CAL3_WX1	1099.53	35.22	0.7840	$1,02 \times 10^7$	200.74

Lectura comparativa. Mercaya exhibe menor MAE y menor WAPE %, junto con un R^2 más alto. En términos prácticos, esto sugiere que la demanda del día siguiente es, en promedio, *más predecible* a partir de la información reciente del producto y un calendario compacto; es decir, el comportamiento está más cerca de una *dinámica endógena* (dominada por su propio historial). En Villa Café, el error típico y el error relativo agregado aumentan, y el R^2 disminuye, lo que es consistente con una demanda donde influyen componentes adicionales que no siempre se reflejan de forma estable en el historial inmediato.

Justificación desde el conjunto seleccionado. La diferencia entre negocios no se observa solo en las métricas, sino también en la estructura del conjunto ganador: Mercaya alcanza el óptimo sin clima (WX0), mientras que Villa Café incorpora `clima_desc` (WX1). Esta evidencia sugiere que, en Villa Café, el contexto actúa como un *marcador de régimen* que desplaza la demanda esperada incluso cuando el historial reciente es similar, lo cual tiene implicaciones directas para reposición (mayor riesgo de desviaciones operativas si se ignora el contexto).

7.4.2. Diferencias en señal dominante (a partir de las variables seleccionadas)

La comparación por bloques permite interpretar *qué tipo de señal* aporta más valor predictivo en cada negocio:

- **Contexto climático.** Mercaya alcanza el mejor desempeño sin clima (WX0); Villa Café retiene `clima_desc` (WX1), lo que indica que el contexto contribuye a explicar variaciones que no quedan capturadas completamente por memoria y calendario.
- **Calendario.** Mercaya incluye `mes`, lo que es consistente con un componente mensual explícito; Villa Café refuerza `dia_habil` y `quincena`, variables de lectura operativa que suelen reflejar cambios en rutina y ciclo de ingresos.
- **Memoria temporal.** Mercaya prioriza continuidad inmediata (`lag_1..lag_4`), mientras que Villa Café utiliza lags espaciados e incorpora `lag_7`, señal coherente con patrones cercanos a semana.

Implicación. Estas diferencias no se interpretan como “un negocio mejor que otro”, sino como evidencia de que la estructura de la demanda cambia entre contextos. Por ello, mantener el mismo marco experimental en ambos escenarios fortalece la conclusión: el contraste observado proviene del fenómeno y no de cambios metodológicos.

7.5. Efecto de seleccionar conjuntos compactos frente a entrenar con el conjunto completo

En el capítulo de línea base se entrenó con el conjunto completo `ORIGINAL_FULL` (28 variables crudas; 41 variables efectivas tras codificación). Sobre esa referencia, esta sección cuantifica el efecto de usar conjuntos compactos seleccionados y optimizados: aquí el objetivo no es repetir la motivación teórica, sino mostrar **qué cambia en la evidencia en prueba** cuando se reduce complejidad de manera controlada.

7.5.1. Mejor global vs. conjunto completo

Para aislar el efecto de la selección de variables, la comparación se hace manteniendo constante la familia de modelo (`RF_TUNED`). Así, cualquier ganancia observada se atribuye principalmente a (i) eliminar señales redundantes o inestables y (ii) concentrar la señal dominante bajo el mismo protocolo temporal.

Cuadro 7.10: Mejora del mejor global (`RF_TUNED`) frente al conjunto completo `ORIGINAL_FULL`

Métrica	Mercaya <code>ORIGINAL_FULL</code>	Mercaya Mejor	Villa Café <code>ORIGINAL_FULL</code>	Villa Café Mejor
MAE	1075.68	933.04	1355.22	1099.53
WAPE %	35.85	31.10	43.41	35.22
MSE	$8,16 \times 10^6$	$7,04 \times 10^6$	$1,43 \times 10^7$	$1,02 \times 10^7$
R^2	0.8056	0.8321	0.6989	0.7840
Tiempo (s)	262.74	228.98	276.47	200.74
Vars crudas	28	11	28	11
Vars efectivas	41	20	41	24

Lectura. En ambos negocios, el conjunto compacto mejora simultáneamente precisión y eficiencia: baja MAE y WAPE %, reduce MSE, sube R^2 y disminuye el tiempo de entrenamiento. Esto es consistente con la idea de que el conjunto completo no solo “trae más señal”, sino también más *complejidad* (dimensión efectiva) y mayor exposición a variables que no generalizan igual al *holdout*. En términos de decisión operativa, el resultado es relevante porque permite lograr un pronóstico más confiable con menor dependencia de un gran número de insumos.

Cuadro 7.11: Mejora relativa del mejor global (RF_TUNED) frente a ORIGINAL_FULL

Indicador	Mercaya	Villa Café
Mejora MAE (%)	$\approx 13,26 \%$	$\approx 18,87 \%$
Mejora WAPE % (%)	$\approx 13,26 \%$	$\approx 18,86 \%$
Reducción MSE (%)	$\approx 13,64 \%$	$\approx 28,3 \%$
ΔR^2 (abs.)	$\approx +0,0265$	$\approx +0,0851$
Reducción tiempo (%)	$\approx 12,85 \%$	$\approx 27,4 \%$
Reducción vars crudas (%)	$\approx 60,7 \%$	$\approx 60,7 \%$
Reducción vars efectivas (%)	$\approx 51,2 \%$	$\approx 41,5 \%$

Justificación adicional. La mejora es especialmente marcada en Villa Café (mayor aumento de R^2 y mayor reducción de MSE). Esto respalda que, en un contexto más sensible al régimen, un conjunto compacto bien seleccionado no solo simplifica, sino que ayuda a evitar que el modelo “persiga” señales menos transferibles a días futuros.

7.5.2. Alternativas cercanas vs. conjunto completo base line

Además del mejor global, es útil contrastar alternativas cercanas que priorizan rapidez o mantenibilidad. La Tabla 7.12 muestra que incluso estas opciones —con decisiones operativas distintas— superan con claridad al conjunto completo ORIGINAL_FULL.

Cuadro 7.12: Alternativas cercanas (RF_TUNED) comparadas contra el conjunto completo ORIGINAL_FULL

Negocio	Escenario	Conjunto de variables	MAE	WAPE %	Tiempo (s)	Vars crudas
Mercaya	Completo (ORIGINAL_FULL)	—	1075.68	35.85	262.74	28
Mercaya	Más rápido (5 %)	L2_R2_M3_CAL4_WX0	972.55	32.41	210.11	12
Mercaya	Menos variables (5 %)	L2_R3_M0_CAL2_WX1	979.67	32.65	225.92	9
Villa Café	Completo (ORIGINAL_FULL)	—	1355.22	43.41	276.47	28
Villa Café	Más rápido (5 %)	L4_R0_M3_CAL4_WX0	1115.81	35.75	198.58	12
Villa Café	Menos variables (5 %)	L4_R2_M0_CAL3_WX1	1099.53	35.22	200.74	11

Conclusión de la comparación. En ambos negocios, las alternativas compactas mejoran la evidencia en prueba frente a ORIGINAL_FULL. En Mercaya, el margen permite escoger entre rapidez o parsimonia sin salir del umbral de cercanía; en Villa Café, el mejor global ya es compacto y, por tanto, la ruta operativa más clara es decidir si se preserva el régimen climático (máxima estabilidad) o si se sustituye por estacionalidad y suavización cuando la restricción dominante es el cómputo.

7.6. Conclusión integral de resultados

En síntesis, la evidencia en prueba confirma que la predicción diaria a un día ($h = 1$) se sostiene, en ambos negocios, sobre **memoria temporal reciente** y **calendario**. La diferencia clave es el rol del contexto: en Mercaya el óptimo prescinde de clima (WX0), mientras que en Villa Café la inclusión de `clima_desc` (WX1) aporta estabilidad adicional al capturar regímenes que desplazan la demanda.

Desde una perspectiva operativa, esta estructura explica por qué **RF_TUNED** se posiciona como mejor global: su capacidad de modelar interacciones y heterogeneidad por producto se traduce en menor error típico y menor error relativo agregado bajo separación temporal estricta, sin requerir el conjunto completo **ORIGINAL_FULL**. Además, el experimento aporta valor práctico al no entregar una única “respuesta”, sino un conjunto de opciones justificadas: máxima precisión (mejor global) y alternativas cercanas cuando la restricción dominante sea reentrenamiento frecuente, cómputo disponible o mantenibilidad del sistema.

Capítulo 8

Conclusiones

Con base en los resultados obtenidos en Mercaya y Villa Café, se concluye que el proceso desarrollado **cumple el propósito central del trabajo: desarrollar y evaluar** un enfoque de Inteligencia Artificial (IA) para predecir la demanda diaria ($h = 1$) de productos de alta rotación en dos PYMEs minoristas de Ibagué. Más que “presentar un modelo ganador”, el aporte se concreta en un **procedimiento reproducible y justificable** que transforma registros operativos en una serie producto–día, compara alternativas bajo evaluación temporal realista y reporta resultados con una lectura operativa del error (precisión, tiempo de entrenamiento y complejidad/mantenibilidad).

Alcance y lectura de la evidencia. Las conclusiones se sustentan en una evaluación *out-of-sample* mediante **holdout temporal** (bloque final de fechas). La selección de variables y el ajuste de hiperparámetros se realizaron **solo con TRAIN**, usando validación cruzada temporal para evitar fuga de información, y el desempeño final se observó en **TEST**. Por tanto, los resultados deben entenderse como **evidencia offline** para el periodo analizado; su robustez puede fortalecerse en trabajo futuro con backtesting sobre múltiples cortes temporales o con un esquema TRAIN/VALIDACIÓN/TEST más estricto.

La preparación de datos fue determinante para modelar con coherencia. Un hallazgo transversal fue que la calidad del dato no es un detalle accesorio: **construir una serie producto–día consistente** (tipos, integridad temporal y faltantes) fue una condición para que las métricas reflejaran patrones reales de demanda y no irregularidades del registro. En contexto minorista, duplicados o discontinuidades pueden introducir cambios artificiales que el modelo interpreta como señal; por ello, el alistamiento de datos aportó estabilidad y trazabilidad al proceso.

La señal útil se concentró en un núcleo compacto de variables. La búsqueda de subconjuntos mostró que, para este problema y este histórico, los mejores resultados se obtuvieron con conjuntos relativamente compactos (aprox. 9–12 variables crudas antes de codificación). La mayor parte de la información predictiva se concentró en: (i) **memoria temporal** (`lag_k`, `rolling/rolling_mean`), (ii) **calendario** (`dia`, `quincena`, `dia_habil`, `mes`), y (iii) **identidad del producto** (`productoid`) para sostener el enfoque multiproducto. Consistente con esto, al comparar el conjunto completo (28 crudos; 41 efectivos) frente a conjuntos compactos, se observó una mejora en ambos negocios: en Mercaya el MAE se redujo de 1075.68 a 933.04 y el WAPE % de 35.85 a 31.10 (con R^2 de 0.8056 a 0.8321); en Villa Café el MAE pasó de 1355.22 a 1099.53 y el WAPE % de 43.41 a 35.22 (con R^2 de 0.6989 a 0.7840). Esto sugiere que simplificar ayudó a reducir ruido y redundancias, mejorando la generalización en el tramo de prueba.

Desempeño comparativo y lectura metodológica. Bajo el protocolo y periodo evaluados, **Random Forest (RF_TUNED)** **mostró el mejor equilibrio global observado** en ambos negocios (MAE y WAPE % en prueba, y desempeño competitivo en MSE). Este resultado es coherente con la presencia de no linealidades, interacciones entre memoria y calendario, y heterogeneidad entre productos; no obstante, se resalta que su consolidación definitiva requiere ampliar el histórico y reevaluar en periodos adicionales. Asimismo, la comparación entre negocios sugiere que el aporte de variables exógenas es **contextual**: en Mercaya el mejor caso no requirió clima (`L4\R3\M0\CAL3\WX0`), mientras que en Villa Café sí apareció `clima_desc` en el mejor conjunto (`L4\R2\M0\CAL3\WX1`); para un despliegue real, esta señal debería provenir de pronóstico meteorológico o emplearse rezagada.

Implicaciones operativas: apoyo a la decisión, con foco en alta rotación. La evaluación temporal permitió evitar conclusiones optimistas y mantener una lectura operativa: el desempeño observado debe entenderse como una **reducción medible de incertidumbre** útil para reposición, sin prometer exactitud total (especialmente ante picos o eventos no observados). Además, se observó que el error varía por producto, y que el enfoque en **alta rotación** es metodológicamente coherente y operativamente relevante: son series más estables y de mayor volumen, donde el impacto potencial del error evitado es mayor. Finalmente, reportar alternativas cercanas al óptimo pero **más simples** (menos variables o menor tiempo de entrenamiento, dentro del umbral del 5 % sobre WAPE %) fortalece la aplicabilidad en PYMEs, donde sostenibilidad y mantenibilidad suelen ser tan importantes como la precisión.

Síntesis. Se construyó y evaluó un enfoque predictivo de demanda diaria multiproducto ($h = 1$) en dos PYMEs, identificando un núcleo compacto de variables relevantes y mejoras de generalización mediante selección/optimización. En el periodo analizado, Random Forest mostró buen equilibrio; se requiere validación prospectiva y ampliar el histórico para consolidar su desempeño en operación.

Capítulo 9

Trabajo Futuro

A partir de los resultados y de las limitaciones identificadas durante las etapas de alistamiento, limpieza, selección de variables, entrenamiento y evaluación temporal en Mercaya y Villa Café, se proponen líneas de trabajo futuro orientadas a: **(i) fortalecer la aplicabilidad operativa**, **(ii) ampliar la robustez ante ruido y cambios de régimen** y **(iii) maximizar el impacto en la gestión de inventarios** de PYMEs. Estas extensiones se formulan en coherencia con el enfoque del trabajo: decisiones **justificables, reproducibles y sostenibles**, compatibles con restricciones reales de infraestructura, calidad del dato y capacidades técnicas del comercio minorista local.

9.1. Extensión del período de observación y construcción de un histórico de mayor cobertura

Una línea prioritaria consiste en ampliar el horizonte de datos. En este trabajo, el modelado se desarrolló con una ventana equivalente a un **trimestre**, determinada por la disponibilidad y continuidad de la captura. Como trabajo futuro, se recomienda consolidar un histórico **significativamente mayor** (por ejemplo, 6–12 meses o más), por tres razones: (i) capturar **estacionalidad anual** y ciclos de calendario de mayor escala; (ii) aumentar la exposición a **eventos raros pero operativamente relevantes** (p. ej., cambios de proveedor, variaciones de surtido, ajustes de precio, interrupciones de operación); y (iii) disminuir la sensibilidad del modelo a particularidades del período observado, mejorando la **validez externa** y la estabilidad de generalización.

Un histórico más amplio también habilita evaluaciones más exigentes: validación con **múltiples cortes temporales** (*rolling forecasting origin* o ventana deslizante), análisis de degradación por *drift* y comparación de políticas de reentrenamiento (por ejemplo, mensual vs. por ventana móvil). En consecuencia, ampliar el histórico no es solo “tener más datos”, sino **mejorar la capacidad del sistema para aprender**

patrones estructurales y sostener desempeño ante variaciones típicas de operación en PYMEs.

9.2. Validación prospectiva y evaluación en operación

Se recomienda evolucionar la evaluación hacia un esquema **prospectivo**, en el que el modelo se ejecute de forma continua durante varias semanas o meses, registrando diariamente: (i) la predicción generada, (ii) la demanda observada y (iii) la decisión de compra efectivamente tomada. Este paso es clave porque traduce métricas técnicas (MAE/WAPE %) en evidencia de utilidad bajo condiciones reales: promociones, sustituciones por disponibilidad, cambios de proveedor, variaciones de surtido, eventos locales y ruido operacional.

Adicionalmente, la validación prospectiva permite medir indicadores cercanos a la operación, tales como: **frecuencia de rupturas de stock, exceso de inventario, merma y rotación** por categoría, así como el grado de adopción (uso efectivo de la recomendación) y su relación con el resultado final. Con ello, el análisis deja de centrarse únicamente en “qué tan bien predice” y pasa a responder “**qué tan útil es para decidir**”.

9.2.1. Incorporación explícita de eventos, intervenciones y cambios estructurales

Un frente de mejora consiste en representar explícitamente intervenciones operativas y cambios estructurales: indicadores de promoción (tipo, duración, alcance), variaciones de precio, descuentos, cambios de proveedor, quiebres previos de inventario, sustituciones por falta de producto y cambios de horarios/operación. Esta información es relevante porque una fracción importante del error tiende a concentrarse cuando la demanda deja de seguir patrones regulares; por tanto, incorporar causas operativas explícitas puede reducir residuales y mejorar estabilidad, especialmente en productos sensibles a precio, disponibilidad o campañas.

En particular, una mejora crítica es distinguir entre **baja venta por baja demanda** y **baja venta por desabastecimiento**. Cuando no se identifica el quiebre de stock, la variable objetivo se “trunca” y el modelo aprende una dinámica sesgada. Incorporar señales de disponibilidad (o reglas de detección de quiebres) permitiría aproximar mejor el consumo potencial y, con ello, mejorar la interpretabilidad y la utilidad del pronóstico.

9.2.2. Pronóstico multi-horizonte alineado con ciclos reales de abastecimiento

Aunque el horizonte de un día ($h = 1$) es útil para decisiones diarias, se recomienda extender el alcance hacia predicción **multi-horizonte** (p. ej., $t+3$, $t+7$ y $t+14$). Esta extensión se justifica porque la reposición en PYMEs suele estar condicionada por *lead time* del proveedor, compras semanales, descuentos por volumen y restricciones logísticas. Un esquema multi-horizonte permitiría planificar pedidos con anticipación, negociar cantidades con mejor previsión y ajustar inventario bajo ventanas de decisión más realistas.

Además, habilita estrategias diferenciadas: reabastecimiento frecuente para productos de alta rotación y planificación semanal/quincenal para referencias con menor rotación o proveedores con mayor tiempo de entrega, reduciendo decisiones reactivas y mejorando coherencia entre predicción y calendario operativo de compras.

9.2.3. De la predicción a la decisión: recomendación de compra integrada con políticas de inventario

Un trabajo futuro clave es transformar la salida del modelo en recomendaciones accionables de compra, integrando inventario actual, stock mínimo, vida útil (perecederos), *lead time* del proveedor, capacidad de almacenamiento y reglas de unidades de empaque. Esta evolución es necesaria porque el objetivo operativo no es únicamente minimizar error, sino reducir costos reales asociados a **rupturas de stock**, **sobreabastecimiento** y **merma**.

En consecuencia, se sugiere integrar el pronóstico con políticas de control (p. ej., puntos de reorden, reglas (s, S) o umbrales dinámicos) calibradas por producto según variabilidad, margen, criticidad y costo de ruptura. El resultado esperado es pasar de “predecir demanda” a **soportar decisiones de reposición** con criterios transparentes y auditables (pronóstico $\rightarrow$ nivel objetivo $\rightarrow$ cantidad sugerida).

9.2.4. Incertidumbre, cuantiles y toma de decisiones basada en riesgo

Se recomienda incorporar estimación explícita de incertidumbre mediante intervalos de predicción o cuantiles (p. ej., P_{50} , P_{80} , P_{90}), usando enfoques como *quantile regression* (en boosting) o calibración mediante *conformal prediction*. Esta extensión es especialmente pertinente porque, en operación, el costo de subestimar (venta perdida y pérdida de fidelización) rara vez es equivalente al de sobreestimar (capital inmovilizado y merma).

Con cuantiles calibrados, sería posible definir **stocks de seguridad** por producto

de manera explícita, priorizar decisiones robustas en referencias volátiles y ajustar políticas según tolerancia al riesgo y criticidad (por ejemplo, usar cuantiles altos en productos con alta penalización por ruptura).

9.2.5. Segmentación por comportamiento y esquemas híbridos de modelado

Dado que el desempeño no es homogéneo entre productos, se propone avanzar hacia segmentación por comportamiento de demanda: productos estables, intermitentes y altamente volátiles. Un trabajo futuro consiste en diseñar un esquema híbrido que mantenga un **modelo global multiproducto** (para capturar patrones comunes y reducir costo de mantenimiento), complementado por modelos especializados o reglas específicas por grupo.

Esta estrategia reduce el riesgo de que la señal mayoritaria del modelo global opaque dinámicas particulares y permite decisiones diferenciadas: reglas conservadoras para series intermitentes, umbrales y seguridad reforzada para referencias volátiles y, en alta rotación, mayor dependencia del pronóstico como guía principal.

9.2.6. Monitoreo, *drift* y mantenimiento del modelo en escenarios reales

En escenarios reales, la demanda cambia por inflación, competencia, cambios de surtido, estacionalidad y hábitos de consumo. Por ello, se recomienda incorporar monitoreo de *drift* (cambios en la distribución de variables y/o en la relación variable-objetivo), alertas de degradación de métricas y políticas de reentrenamiento (mensual, por ventana móvil o por evento).

Además, conviene formalizar un protocolo mínimo de mantenimiento que especifique: frecuencia de actualización, datos mínimos requeridos, validaciones de integridad y criterios de despliegue (por ejemplo, reemplazar el modelo solo si supera umbrales de mejora y estabilidad). Sin monitoreo, un modelo puede deteriorarse de manera silenciosa y perder utilidad operativa aun si su desempeño histórico fue adecuado.

9.2.7. Trazabilidad y gobernanza del dato para reducir fallas operativas

Una mejora transversal es fortalecer la trazabilidad del dato operacional: control de duplicados, detección de quiebres de stock, tratamiento sistemático de devoluciones/anulaciones y consistencia entre ventas, inventario y compras. En PYMEs, el desempeño del modelo depende directamente del flujo de captura; inconsistencias

pequeñas pueden introducir degradaciones fuertes y difíciles de diagnosticar.

En consecuencia, se recomienda formalizar reglas de validación automáticas y reportes de calidad (p. ej., alertas por ventas cero anómalas, saltos abruptos de inventario, precios atípicos, productos sin historial suficiente), con el objetivo de hacer el sistema sostenible sin requerir intervención manual constante.

9.2.8. Medición ambiental y sostenibilidad con indicadores verificables

Aunque es razonable inferir que una mejor planeación reduce desperdicio, se propone medir explícitamente el componente ambiental con indicadores verificables: merma (unidades descartadas), vencimientos, volumen de residuos y frecuencia de reposiciones urgentes antes y después de adoptar el sistema. Esto conectaría el aporte técnico con impacto tangible y medible, y permitiría priorizar acciones por categoría (perecederos vs. no perecederos) donde el costo del error tiene consecuencias directas.

9.2.9. Escalabilidad y generalización a más PYMEs en el contexto local

Finalmente, se recomienda ampliar la validación hacia más PYMEs de Ibagué u otros contextos cercanos para evaluar transferibilidad: qué variables se mantienen relevantes, qué familias de modelos generalizan mejor y qué tan sensible es el rendimiento a diferencias en portafolio, ubicación y hábitos de compra. Esta extensión fortalecería la validez externa del enfoque y permitiría convertir el proceso en un marco reutilizable para adopción a mayor escala, manteniendo el principio central del trabajo: construir soluciones de IA aplicables y sostenibles en negocios reales.

Síntesis de trabajos futuros. En síntesis, estas líneas buscan evolucionar el aporte desde un sistema predictivo técnicamente consistente hacia un **sistema de decisión de inventario robusto, medible y sostenido en el tiempo**: (i) histórico más amplio para capturar estacionalidad y eventos raros, (ii) evaluación prospectiva en operación, (iii) incorporación de eventos e intervenciones, (iv) manejo explícito de incertidumbre, y (v) monitoreo de *drift* y gobernanza del dato. De esta manera, el enfoque podría maximizar su utilidad práctica en PYMEs: **anticipar demanda, mejorar reposición y reducir simultáneamente rupturas de stock y sobreabastecimiento** mediante un proceso transparente y reproducible.

Bibliografía

- [1] Bancóldex. ¿qué es una pyme? Disponible en: <https://www.bancoldex.com/es/que-es-una-pyme-1338>, 2024. Último acceso: 30 de enero de 2025.
- [2] Byron Patricio Andrade Clavijo and María Julieta Guerrero Cepeda. Evaluación de la gestión de inventarios y facturación en el almacén peregrine falcón. *Ciencia Latina Revista Científica Multidisciplinar*, Volumen 7, Número 2 (Marzo–Abril 2023), 03 2023.
- [3] Christoph Szedlak, Patrick Pötters, and Bert Leyendecker. Application of artificial intelligence in small and medium-sized enterprises. In: *Proceedings of the 5th North American International Conference on Industrial Engineering and Operations Management (IEOM)*, 08 2020.
- [4] Angelica Fidencia Zambrano-Ponce and David Alejandro Zaldumbide-Peralvo. Disponibilidad de inventarios frente a la demanda en productos de tiendas tuti. *593 Digital Publisher CEIT*, 9(2), 228–244, 03 2024.
- [5] Cristian Caraballo Monterrosa and Katherin Pérez Tovar. El impacto de la inteligencia artificial y la automatización en la eficiencia operativa de las empresas manufactureras del caribe colombiano (2020–2024). Monografía de especialización (Especialización en Alta Gerencia y Desarrollo Organizacional), 2025. Asesor: Francisco José Ariza Guerra.
- [6] Sebastián Pinto Castro. Integración de la inteligencia artificial en la gestión de inventarios para pequeñas empresas: Mejora de la eficiencia y reducción de costos. Trabajo de grado, Programa de Ingeniería Industrial, 2024.
- [7] Armando Guío Español, Elena Tamayo Uribe, Pablo Gómez Ayerbe, and María Paula Mujica. Marco ético para la inteligencia artificial en colombia. Departamento Administrativo de la Presidencia de la República (Presidencia de la República de Colombia). Documento elaborado con apoyo de CAF y BID, 10 2021. Versión final. Octubre de 2021.
- [8] Jimmy Ernesto Fernández Cabrera, Aramis Alfonso Llanes, and Rafael A. Ramos Gómez. Procedimiento para el pronóstico de la demanda en una mipyme cubana. *EASI: Engineering and Applied Sciences in Industry*, 1(2):14–22, 12 2022. doi: 10.53591/easi.v1i2.1783.

- [9] Rob J. Hyndman and George Athanasopoulos. Forecasting: Principles and practice. Online textbook, Monash University, Australia, 04 2018. Second edition. Preface dated April 2018; online version last updated 5 September 2018; print version last updated 8 May 2018.
- [10] Tim Januschowski, Jan Gasthaus, Yuyang Wang, David Salinas, Valentin Flunkert, Michael Bohlke-Schneider, and Laurent Callot. Criteria for classifying forecasting methods. *International Journal of Forecasting*, 36:167–177, 2020. doi: 10.1016/j.ijforecast.2019.05.008.
- [11] Tim Januschowski, Yuyang Wang, Kari Torkkola, Timo Erkkilä, Hilaf Hasson, and Jan Gasthaus. Forecasting with trees. *International Journal of Forecasting*, 38:1473–1481, 2022. doi: 10.1016/j.ijforecast.2021.10.004. Open access (CC BY-NC-ND 4.0).
- [12] Max Kuhn and Kjell Johnson. *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press, Boca Raton, FL, 2020. ISBN 978-1-138-07922-9. Chapman & Hall/CRC Data Science Series. Taylor & Francis Group.
- [13] Matthias Feurer and Frank Hutter. Hyperparameter optimization. In Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren, editors, *Automated Machine Learning: Methods, Systems, Challenges*, pages 3–33. Springer, 2019. doi: 10.1007/978-3-030-05318-5_1.
- [14] F. Javier Medrano Avedillo. Comparación de algoritmos para la predicción de demanda del número de vehículos de una empresa logística. Trabajo Fin de Máster (Ingeniería Industrial), Escuela Técnica Superior de Ingeniería, Universidad de Sevilla, 12 2022. Tutor: Daniel Gutiérrez Reina.
- [15] Jhon Wilder Lezcano Echeverri. Modelo de predicción de venta en una compañía textil con técnicas de machine learning, 2025. Proyecto de Grado, Maestría en Ciencias de los Datos y la Analítica.
- [16] Uriel Mauricio Castro Morales. Pronóstico de ventas a corto plazo en el sector de consumo masivo a través de modelos de machine learning, 2024. Trabajo de grado, Maestría en Inteligencia Analítica para la Toma de Decisiones (MIIA).
- [17] Jenny Paola Martínez-Pérez. Predicción de tendencias de consumo en moda sostenible en colombia mediante inteligencia artificial. Documento académico (prototipo en Streamlit) en el marco del Diplomado en Tendencia de Marketing e Inteligencia Artificial, s.f. El documento no muestra un año de publicación explícito en la primera página.
- [18] Nicolas A. Acero O. Drinkmaster pos: Optimización de ventas y gestión de inventario. Seminario de profundización, Programa de Ingeniería de Sistemas, Universidad Cooperativa de Colombia, Sede Villavicencio - Meta, 09 2024. Asesor técnico: Ing. Giovanny J. Parrado C.; Asesora metodológica: Ing. Francy Y. Patiño M.

- [19] Andrés Santiago Krovina Rueda. Optimización de inventarios y pronóstico de demanda en el sector retail: una aplicación con modelos predictivos sobre datos simulados. Trabajo de investigación, Escuela de Negocios, Maestría en Inteligencia de Negocios y Ciencia de Datos, 2025. Documento en PDF. Profesor: Mario Salvador González. Último acceso: 6 de febrero de 2026.
- [20] Md Sabbirul Haque, Md Shahedul Amin, and Jonayet Miah. Retail demand forecasting: A comparative study for multivariate time series. Manuscrito en PDF, 2023. Incluye autores y afiliaciones (IEEE; University of Tennessee Chattanooga; University of South Dakota). © Copyright by authors, 2023. All rights reserved.