



Pontificia Universidad  
**JAVERIANA**  
Cali

**IMPLEMENTACIÓN DE MODELOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE LA DEMANDA EN UNA EMPRESA MANUFACTURERA DE PRODUCTOS DE ASEO Y DESINFECCIÓN**

*Luisa Angélica Isaza Sanabria*

*Juan Andrés Murillo Cadena*

*Carlos Fabián Villa Infante*

*Proyecto Aplicado para optar al título de  
Magister en Ciencia de Datos*

Directora Isabel Cristina García Arboleda

FACULTAD DE INGENIERÍA Y CIENCIAS  
MAESTRÍA EN CIENCIA DE DATOS  
SANTIAGO DE CALI, OCTUBRE DE 2025

## FICHA RESUMEN

### TÍTULO DEL PROYECTO:

1. ÁREA DE TRABAJO: Series de tiempo
2. TIPO DE PROYECTO: Aplicado
1. ESTUDIANTE(S): Luisa Angélica Isaza Sanabria, Juan Andrés Murillo Cadena, Carlos Fabián Villa Infante
3. CORREO ELECTRÓNICO: [luisaisaza@javerianacali.edu.co](mailto:luisaisaza@javerianacali.edu.co) , [juanmurillo@javerianacali.edu.co](mailto:juanmurillo@javerianacali.edu.co) , [carlosvilla@javerianacali.edu.co](mailto:carlosvilla@javerianacali.edu.co)
4. DIRECCIÓN Y TELEFONO: Cl 37 a 18 a 21, Calarcá, Q. 320 711 6746.- Cra 1 a # 56- 94, Cali, V. 318 346 6576 – Cra 21 a 159-28 Bogotá, C. 300 530 3289
5. DIRECTOR: Dra. Isabel Cristina García Arboleda
6. VINCULACIÓN DEL DIRECTOR: Profesora de planta
7. CORREO ELECTRÓNICO DEL DIRECTOR: [Isabel.garcia@javerianacali.edu.co](mailto:Isabel.garcia@javerianacali.edu.co)
8. CO-DIRECTOR (Si aplica): N/A
9. GRUPO O EMPRESA QUE LO AVALA: Berhlan de Colombia SAS
10. OTROS GRUPOS O EMPRESAS: N/A
11. PALABRAS CLAVE (al menos 5): Pronóstico, Demanda, Machine Learning, Clúster, Ciencia de datos
12. FECHA DE INICIO: 2 febrero 2025
13. DURACIÓN ESTIMADA (En meses): 10 meses
14. RESUMEN: *Este proyecto de investigación implementó modelos de Machine Learning con el objetivo de optimizar la predicción de la demanda en una empresa manufacturera de productos de aseo y desinfección, atendiendo problemáticas como el exceso de inventarios y ventas pérdidas. A partir de un análisis exploratorio de datos se identificaron variables relevantes y patrones de comportamiento de la demanda que permitieron incrementar la precisión de los pronósticos mediante técnicas avanzadas de ciencia de datos. Los resultados obtenidos en los modelos se sistematizaron y se presentaron en un dashboard interactivo en Power BI, lo que facilitó la interpretación de la información y respaldó la toma de decisiones tanto operativas como estratégicas. En términos de impacto, la implementación de esta propuesta contribuyó a optimizar la gestión de inventarios, disminuir el riesgo de desabastecimiento y reducir los costos asociados a productos obsoletos, evidenciando la pertinencia de estas metodologías en entornos industriales caracterizados por una alta variabilidad en la demanda.*

|   |           |
|---|-----------|
| <b>1. DEFINICIÓN DEL PROBLEMA .....</b>                           | <b>6</b>  |
| 1.1. PLANTEAMIENTO DEL PROBLEMA.....                              | 6         |
| 1.2. FORMULACIÓN DEL PROBLEMA.....                                | 6         |
| <b>2. OBJETIVOS .....</b>   | <b>7</b>  |
| 2.1. OBJETIVO GENERAL.....  | 7         |
| 2.2. OBJETIVOS ESPECÍFICOS .....                                  | 7         |
| <b>3. MARCO TEÓRICO Y ANTECEDENTES .....</b>                      | <b>8</b>  |
| 3.1. MARCO TEÓRICO.....   | 8         |
| 3.1.1. CONTEXTO ORGANIZACIONAL.....                               | 8         |
| 3.1.2. PRONÓSTICO .....   | 8         |
| 3.1.3. PLANIFICACIÓN DE LA DEMANDA .....                          | 9         |
| 3.1.4. PROCESO DE S&OP .....                                      | 9         |
| 3.1.5. METODOLOGÍA CRISP- DM.....                                 | 9         |
| 3.1.6. MODELOS CLÁSICOS PARA PREDECIR LA DEMANDA .....            | 10        |
| 3.1.7. MODELOS DE MACHINE LEARNING PARA PREDECIR LA DEMANDA ..... | 12        |
| 3.1.8. MÉTODOS DE EVALUACIÓN DE PREDICCIÓN .....                  | 14        |
| 3.1.9. MODELOS DE SEGMENTACIÓN DE SERIES DE TIEMPO .....          | 15        |
| 3.1.10. COEFICIENTES DE CORRELACIÓN Y COINTEGRACIÓN .....         | 15        |
| 3.2. ANTECEDENTES .....   | 16        |
| <b>4. PREPARACIÓN Y ANÁLISIS EXPLORATORIO DE DATOS .....</b>      | <b>17</b> |
| 4.1. RECOLECCIÓN Y LIMPIEZA DE DATOS .....                        | 17        |
| 4.2. ANÁLISIS EXPLORATORIO DE DATOS .....                         | 20        |
| 4.3. ANÁLISIS DE VARIABLES EXTERNAS.....                          | 27        |
| 4.4. DEFINICIÓN DE VARIABLES DEL MODELO DE PRONÓSTICO .....       | 30        |
| 4.5. CLÚSTER DE SERIES DE TIEMPO.....                             | 31        |
| 4.6. DESCOMPOSICIÓN DE LA SERIE DE TIEMPO.....                    | 34        |
| 4.7. DEFINICIÓN DE HORIZONTES DE TIEMPO .....                     | 37        |
| <b>5. APLICACIÓN DE MODELOS DE PRONÓSTICO.....</b>                | <b>38</b> |
| <b>6. EVALUACIÓN DEL MODELO.....</b>                              | <b>40</b> |
| <b>7. VISUALIZACIÓN DEL MODELO.....</b>                           | <b>42</b> |
| <b>8. CONCLUSIONES .....</b>                                      | <b>45</b> |
| <b>9. TRABAJOS FUTUROS.....</b>                                   | <b>46</b> |
| <b>10. REFERENCIAS BIBLIOGRÁFICAS.....</b>                        | <b>48</b> |

## LISTA DE FIGURAS

|  |    |
|--|----|
| Fig. 1 Metodología CRISP DM .....                                | 9  |
| Fig. 2 Ejemplo serie SARIMA.....                                 | 12 |
| Fig. 3 Ejemplo Boosting .....                                    | 13 |
| Fig. 4 KG pedidos por año .....                                  | 21 |
| Fig. 5 Participación por categoría .....                         | 22 |
| Fig. 6 Participación en KG por centro de operación .....         | 22 |
| Fig. 7 Participación por unidad de negocio .....                 | 23 |
| Fig. 8 Cantidad de clientes por canal de atención.....           | 24 |
| Fig. 9 Participación de venta en KG por zona .....               | 25 |
| Fig. 10 Tendencia cantidad pedida en KG vs Valor \$ .....        | 26 |
| Fig. 11 Relación de ventas vs variables externas.....            | 28 |
| Fig. 12 Resultados del Test de Cointegración .....               | 29 |
| Fig. 13 Resultados regresión variables externas .....            | 30 |
| Fig. 14 Cantidad pedida en promedio mensual por clúster .....    | 32 |
| Fig. 15 Promedio Mensual pedido en Kg para cada Subclúster ..... | 33 |
| Fig. 16 Resumen por subclúster .....                             | 33 |
| Fig. 17 Descomposición STL - Subclúster A .....                  | 35 |
| Fig. 18 Descomposición STL - Subclúster B.....                   | 36 |
| Fig. 19 Descomposición STL - Subclúster C.....                   | 37 |
| Fig. 20 Dashboard Power BI.....                                  | 44 |
| Fig. 21 Flujo para pronóstico de demanda .....                   | 45 |

## LISTA DE TABLAS

|  |    |
|--|----|
| TABLA I MODELOS CLÁSICOS DE PREDICCIÓN                     | 11 |
| TABLA II MODELOS DE MACHINE LEARNING                       | 13 |
| TABLA III MÉTRICAS DE CALIDAD DE PREDICCIÓN                | 14 |
| TABLA IV VARIABLES INTERNAS PRINCIPALES                    | 18 |
| TABLA V MODELOS PREDOMINANTES POR CLÚSTER SEGÚN MENOR RMSE | 42 |

## LISTA DE ANEXOS

|   |    |
|---|----|
| Anexo A. Diccionario de Variables.....              | 52 |
| Anexo B. Resultados evaluación mejores modelos..... | 55 |

## INTRODUCCIÓN

La predicción precisa de la demanda es un desafío crítico en la gestión de cadenas de suministro, particularmente en industrias manufactureras caracterizadas por patrones de consumo volátiles y estacionales [1]. Los métodos tradicionales de pronóstico, como los modelos estadísticos, a menudo fallan en capturar patrones complejos y no lineales, así como en adaptarse a cambios rápidos del mercado, lo que lleva a ineficiencias como el exceso de stock o desabastecimientos [1].

Esta investigación aborda estos desafíos en el contexto de una empresa manufacturera de productos de limpieza y desinfección, donde los pronósticos inexactos han resultado en ineficiencias operativas. Este estudio propone la implementación de un modelo predictivo utilizando técnicas avanzadas de Machine Learning, para mejorar la precisión de los pronósticos de demanda. Al integrar el análisis exploratorio de datos, el modelado predictivo y herramientas de visualización, el proyecto busca mejorar la toma de decisiones estratégicas en la gestión de inventarios y la planificación de ventas. La investigación contribuye al campo de la ciencia de datos al evaluar la efectividad de estas técnicas en un contexto industrial altamente estacional, abordando brechas en la aplicación de Machine Learning a la predicción de demanda en tales entornos [1].

Con el fin de estructurar el proceso de investigación y desarrollo, se adoptó la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) [2], ampliamente utilizada en proyectos de analítica de datos. Esta metodología permitió organizar el trabajo en fases interrelacionadas: entendimiento del negocio, comprensión y preparación de los datos, modelado, evaluación y despliegue. Su aplicación aseguró un enfoque sistemático que facilitó la conexión entre los objetivos estratégicos de la organización y las técnicas de análisis implementadas.

Gracias a la aplicación de la metodología CRISP-DM, se logró identificar variables internas importantes que influyen en la demanda mediante un análisis detallado de datos. Con este marco, se evaluaron modelos de Machine Learning adaptados a la realidad operativa de la empresa y se implementó un dashboard interactivo en Power BI para optimizar la visualización e interpretación de los resultados de pronóstico. Estos avances generaron una mayor precisión en los pronósticos y un soporte sólido para la toma de decisiones estratégicas. En consecuencia, esta investigación aportó beneficios prácticos para la organización como conocimiento académico sobre la aplicación de Machine Learning en contextos industriales altamente estacionales.

## **1. DEFINICIÓN DEL PROBLEMA**

### **1.1. PLANTEAMIENTO DEL PROBLEMA**

En medio de disrupciones constantes en las cadenas de suministro globales y volatilidad del mercado, la previsión de la demanda se ha vuelto fundamental para las empresas que buscan consolidarse y mantenerse competitivas [1]. Los métodos tradicionales de pronóstico han mostrado limitaciones para capturar relaciones no lineales y adaptarse a cambios rápidos del mercado, lo que pone en riesgo la eficiencia operativa de las organizaciones [3].

La predicción precisa de la demanda es esencial para la gestión de la cadena de suministro, ya que permite a las empresas anticipar las necesidades del mercado y alinear sus niveles de inventario, precios y promociones con las expectativas de los consumidores [4]. Estas proyecciones son críticas para evitar el exceso de inventario y el desabastecimiento. Sin embargo, los modelos tradicionales utilizados para estas estimaciones suelen arrojar resultados inconsistentes y no logran adaptarse a las dinámicas de los datos temporales [5].

Este desafío no es ajeno al contexto colombiano, un caso crítico de ineficiencia en la previsión de la demanda se evidenció en una empresa del sector manufacturero del Eje Cafetero, que enfrentaba dificultades de acumulación de inventarios y pérdida de ventas. Por un lado, los pronósticos imprecisos generaban un excedente del 16,42%, equivalente a \$4.048 millones en productos obsoletos, lo que incrementaba los costos de almacenamiento y exponía a la compañía a pérdidas adicionales por caducidad. Por otro lado, la subestimación de la demanda en 2024 ocasionó pérdidas por \$15.574 millones, correspondientes al 5,93% de las oportunidades anuales, afectando directamente la rentabilidad y sostenibilidad del negocio.

Los indicadores de eficiencia del pronóstico, medidos mediante el RMSE (Root Mean Square Error), mostraban un desempeño deficiente: 46% en la planta principal, 52% en la Costa Atlántica y 41% en la región central del país. Esta falta de precisión no solo impactaba negativamente la rentabilidad, sino que también limitaba la capacidad de la empresa para adaptarse a los cambios del entorno. En consecuencia, se evidenciaba la necesidad de reemplazar los modelos tradicionales por enfoques más robustos de ciencia de datos, capaces de mejorar la exactitud de las previsiones, optimizar la gestión de inventarios y reducir las pérdidas tanto por exceso como por insuficiencia de stock.

### **1.2. FORMULACIÓN DEL PROBLEMA**

Pregunta general: ¿Puede un modelo de Machine Learning mejorar los pronósticos de demanda de una empresa manufacturera de productos de aseo y desinfección?

Preguntas de sistematización:

1. ¿Cuáles son las variables clave que afectan la demanda de productos de aseo y desinfección en el contexto de la empresa?
2. ¿Cómo se implementa un modelo de predicción de demanda que se ajuste a la realidad operativa de la empresa?
3. ¿Cómo se puede evaluar la precisión del modelo de pronóstico de demanda en comparación con el modelo actual para determinar el más efectivo?
4. ¿Cómo puede la empresa visualizar e interpretar fácilmente los resultados del modelo de pronóstico de demanda?

## **2. OBJETIVOS**

### **2.1. OBJETIVO GENERAL**

Implementar un modelo de Machine Learning en una empresa manufacturera de productos de aseo y desinfección que permita elaborar pronósticos de demanda a nivel SKU y cliente por semana a partir de variables internas y externas de la compañía.

### **2.2. OBJETIVOS ESPECÍFICOS**

- Integrar y realizar análisis exploratorio de datos de fuentes de información interna y externa relacionados con el negocio de manera que se pueda determinar las variables clave que afectan la demanda (Cierres de mes, temporadas, etc.) para la elaboración de los modelos de pronóstico.
- Ejecutar y entrenar modelos de predicción de demanda ajustados a la realidad de la empresa, considerando sus procesos operativos y datos disponibles.
- Aplicar métricas y métodos de evaluación para comparar la precisión de los modelos de pronóstico de demanda revisados en el negocio para asegurar la calidad de las predicciones.
- Crear una visualización en Power BI que represente el resultado del modelo predictivo de demanda integrado con los procesos de la compañía.

### **3. MARCO TEÓRICO Y ANTECEDENTES**

Se presenta una revisión de literatura de los principales conceptos, modelos de estadística clásica y de Machine Learning utilizados para predecir la demanda. Adicionalmente, se presentarán los métodos de evaluación más utilizados para determinar su calidad.

#### **3.1. MARCO TEÓRICO**

##### **3.1.1. CONTEXTO ORGANIZACIONAL**

La organización objeto de estudio es una compañía manufacturera colombiana dedicada a la producción y comercialización de productos de aseo y desinfección, con más de 25 años de experiencia en el desarrollo de soluciones orientadas al cuidado del hogar, ropa, higiene personal y limpieza institucional.

Su portafolio está dirigido a un amplio espectro de clientes, atendidos a través de múltiples canales de distribución que incluyen desde formatos retail de bajo costo (Hard Discount) hasta grandes superficies comerciales y segmentos institucionales o empresariales.

Adicionalmente, la empresa cuenta con una destacada participación en el desarrollo de marcas privadas, ofreciendo soluciones a más de 260 clientes nacionales e internacionales. Sus productos se caracterizan por mantener un equilibrio entre calidad, innovación, eficiencia logística y costo integral, bajo un enfoque de cooperación y co-creación con el cliente.

Desde el punto de vista operativo y logístico, la compañía dispone de varias plantas de producción y centros de distribución estratégicamente ubicados en distintas regiones del país. Entre sus instalaciones más representativas se destacan aquellas situadas en el Eje Cafetero y en la región del Atlántico, lo que le permite contar con una capacidad operativa robusta y una cobertura geográfica eficiente, garantizando una oferta oportuna y tiempos de entrega competitivos.

##### **3.1.2. PRONÓSTICO**

Es un método para predecir un evento o condición futura mediante el análisis de patrones y el descubrimiento de tendencias en datos anteriores y actuales, se emplean enfoques matemáticos y se aplican modelos estadísticos para poder generarlos. En el ámbito empresarial, tiene como objetivo estimar la demanda que los clientes realizan de los productos y así facilitar la asignación de presupuestos, capital y recursos humanos [6].

### 3.1.3. PLANIFICACIÓN DE LA DEMANDA

Es un proceso de gestión de la cadena de suministro que permite a la empresa proyectar la demanda futura y ajustar de forma exitosa la producción de la empresa. La planificación de la demanda debe asegurar que se tenga exactamente la cantidad correcta de inventario en el lugar adecuado, y así evitar que se agoten las existencias y anticipar los requerimientos del mercado [7].

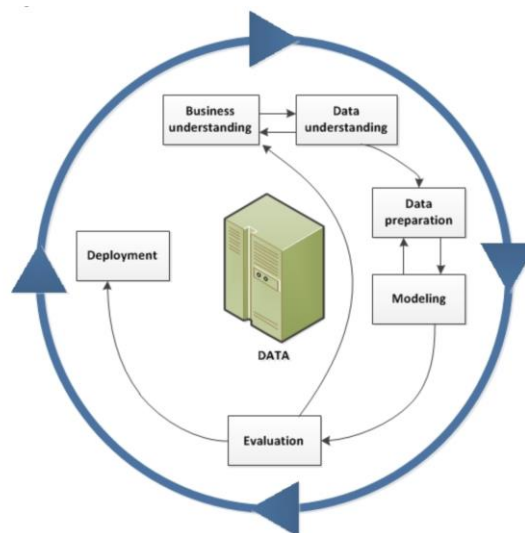
### 3.1.4. PROCESO DE S&OP

La planificación de ventas y operaciones (S&OP) es un proceso de planificación integrado que alinea la demanda, el suministro y la planificación financiera, y se gestiona como parte de la planificación maestra de una compañía. El S&OP está diseñado para apoyar la toma de decisiones ejecutivas relacionadas con la aprobación de un plan material y financiero factible y rentable.

El plan de ventas y operaciones utiliza la demanda agregada global como punto de partida y compara la demanda esperada con la oferta disponible en términos de recursos, como máquinas, mano de obra, y la capacidad material [8].

### 3.1.5. METODOLOGÍA CRISP- DM

La metodología CRISP-DM es un modelo de proceso estructurado y flexible para proyectos de minería de datos y ciencia de datos [2].



*Fig. 1 Metodología CRISP DM*  
*Fuente: <https://wiki.labnuevoleon.mx>*

La metodología CRISP-DM consta de seis fases principales, detalladas en la Fig. 1. En primer lugar, la comprensión del negocio busca definir claramente los objetivos desde la perspectiva organizacional. Posteriormente, la comprensión de los datos permite familiarizarse con la información disponible y detectar posibles problemas de calidad. Con base en ello, la preparación de los datos se centra en seleccionar, limpiar y transformar la información para que pueda ser utilizada en los modelos.

La fase de modelado consiste en aplicar técnicas estadísticas o de Machine Learning para generar modelos predictivos, los cuales son analizados en la fase de evaluación, donde se determina si cumplen los objetivos planteados y si resultan adecuados para el negocio. Finalmente, la fase de despliegue busca integrar los resultados en los procesos operativos, de modo que aporten valor en la toma de decisiones.

Estas fases no se ejecutan de manera estrictamente lineal, sino que forman un ciclo flexible que facilita ajustes y mejoras continuas en función de los hallazgos obtenidos durante el proyecto.

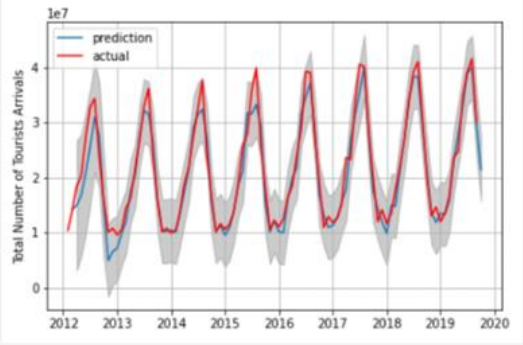
### **3.1.6. MODELOS CLÁSICOS PARA PREDECIR LA DEMANDA**

Las empresas suelen utilizar modelos estadísticos clásicos que son efectivos para pronosticar la demanda, son útiles para situaciones donde las relaciones son lineales y se dispone de pocos datos. Sin embargo, tienen limitaciones para captar patrones complejos de los datos y relaciones no lineales [9].

En la Tabla I, se relacionan los modelos clásicos de pronósticos:

TABLA I  
MODELOS CLÁSICOS DE PREDICCIÓN

| Modelo        | Definición  |
|---------------|---|
| <b>ARIMA</b>  | <p>Son ampliamente utilizados para el análisis y predicción de series temporales debido a su capacidad de capturar patrones de autocorrelación de datos univariados [9] . Los componentes básicos de los modelos ARIMA son:</p> <p><b>AR:</b> indica que la variable evolutiva de interés hace una regresión sobre sus propios valores retardados.</p> <p><b>I:</b> indica que los valores de los datos se han sustituido por la diferencia entre sus valores y los valores anteriores</p> <p><b>MA:</b> indica que el error de regresión es en realidad una combinación lineal de términos de error cuyos valores se produjeron contemporáneamente y en distintos momentos del pasado.</p> |
| <b>SARIMA</b> | <p>Una extensión de ARIMA que permite modelar directamente el componente estacional de la serie. Se trata de una serie temporal con un ciclo que se repite. SARIMA añade tres nuevos hiperparámetros para especificar la auto regresión (<b>AR</b>), la diferenciación (<b>I</b>) y la media móvil (<b>MA</b>) para el componente estacional de la serie. La Fig. 2 es un ejemplo de una serie temporal con un ciclo que se repite [10].</p>  |

|                            |  |
|----------------------------|--|
|                            |  <p><i>Fig. 2 Ejemplo serie SARIMA</i><br/>Fuente: <a href="https://medium.com/data-science/how-to-model-a-time-series-through-a-sarima-model-e7587f85929c">https://medium.com/data-science/how-to-model-a-time-series-through-a-sarima-model-e7587f85929c</a></p>  |
| <p><b>Holt Winters</b></p> | <p>Es una técnica de suavizado exponencial para series temporales que modela tres componentes: nivel, tendencia y estacionalidad. Existen dos variantes: la aditiva, adecuada para datos con estacionalidad constante, y la multiplicativa, ideal para datos con estacionalidad cuya amplitud varía con el nivel de la serie. Este modelo es útil para pronósticos a corto y mediano plazo en series con patrones estacionales y tendencias, pero requiere una estimación cuidadosa de los parámetros de suavizado para evitar sobreajuste [11].</p> |
| <p><b>Prophet</b></p>      | <p>Es un modelo de pronóstico de series temporales desarrollado por Facebook, diseñado para manejar datos con estacionalidades múltiples, tendencias no lineales y efectos de días festivos. Utiliza un enfoque basado en modelos aditivos generalizados, descomponiendo la serie en componentes de tendencia, estacionalidad y efectos exógenos [12].</p>   |

Fuente: Elaboración Propia

### 3.1.7. MODELOS DE MACHINE LEARNING PARA PREDECIR LA DEMANDA

Si bien los modelos revisados en la tabla I son efectivos para realizar pronósticos, existen herramientas más robustas para la precisión de los pronósticos en entornos cambiantes con múltiples variables no lineales. El Machine Learning, busca mejorar el análisis de datos en pro de una predicción futura, mediante el uso de algoritmos basados en información antigua o reciente.

Algunas ventajas de usar modelos de Machine Learning para series de tiempo son:

- Capacidad de capturar relaciones no lineales.

- No requieren estacionariedad.
- Pueden manejar grandes volúmenes de datos e identificar patrones [13].

En la tabla II, se relacionan algunos de los modelos Machine Learning que pueden ser utilizados en los pronósticos de demanda:

TABLA II  
MODELOS DE MACHINE LEARNING

| Modelo         | Definición  |
|----------------|---|
| <b>XGBoost</b> | <p>Es un algoritmo basado en árboles de decisión que utiliza un enfoque de boosting para mejorar su rendimiento. Dado que los árboles de decisión tienden al sobreajuste (Aprender demasiado los detalles y el ruido de los datos de entrenamiento, pierde la capacidad de generar nuevos datos [16] .) Se pueden usar métodos como el boosting para crear modelos más robustos. El boosting, combina varios árboles débiles individuales para crear una fuente de aprendizaje. Cada árbol débil, se entrena secuencialmente para corregir los errores cometidos por los modelos anteriores, luego de cientos de iteraciones, los árboles débiles se vuelven más fuertes [17].</p> <div data-bbox="456 1163 1065 1507" data-label="Diagram"> </div> <p><i>Fig. 3 Ejemplo Boosting</i><br/>Fuente: <a href="https://www.ibm.com/es-es/think/topics/xgboost">https://www.ibm.com/es-es/think/topics/xgboost</a></p> |

|                         |   |
|-------------------------|---|
| <b>Redes Neuronales</b> | Una red neuronal toma decisiones de forma similar al cerebro humano, utiliza procesos que imitan la forma en que trabajan las neuronas. Toda red neuronal consta de capas de nodos o también conocidas como neuronas artificiales, una capa de entrada, una o varias ocultas y una capa de salida. Si la salida de cualquier nodo individual está por encima del valor umbral, ese nodo se activa y envía datos a la siguiente capa de la red, de lo contrario no se pasa ningún dato a la siguiente capa [18].   |
| <b>RNN - LSTM</b>       | Una <b>RNN</b> (Red neuronal recurrente) es una red neuronal profunda entrenada con datos secuenciales o de series de tiempo para crear un modelo de Machine Learning que pueda hacer predicciones secuenciales. Sin embargo, las RNN no pueden predecir un estado actual si el estado anterior no se encuentra en el pasado reciente. Para solucionar esto, las redes neuronales <b>LSTM</b> (Long Short-Term Memory) tienen celdas en las capas ocultas de la red neuronal que tiene tres puertas: Entrada, salida y una puerta de olvido lo cual le permite capturar dependencias de largo plazo [18]. |

Fuente: Elaboración Propia

### 3.1.8. MÉTODOS DE EVALUACIÓN DE PREDICCIÓN

Un pronóstico adecuado de la demanda debe complementarse con la utilización de técnicas que evalúen la exactitud y el desempeño del modelo mediante métricas de validación. En la tabla III, se resumen las métricas más comunes [19].

TABLA III  
MÉTRICAS DE CALIDAD DE PREDICCIÓN

| <b>Métrica</b> | <b>¿Qué es?</b>  | <b>¿Por qué?</b>              | <b>¿Cuándo se usa?</b>                       |
|----------------|--|-------------------------------|--|
| <b>MSE</b>     | Mide la diferencia cuadrada promedio entre los valores estimados y reales. | Enfatiza los grandes errores. | Cuando los errores grandes son más críticos. |

|             |   |  |  |
|-------------|---|--|--|
| <b>RMSE</b> | Raíz cuadrada del MSE, en las mismas unidades que la variable de respuesta. | Interpretación más fácil de la escala de errores | Cuando la escala de error debe coincidir con la escala del objetivo. |
| <b>MAE</b>  | Diferencia absoluta promedio entre los valores estimados y reales.          | Menos sensible a los valores atípicos.           | Con muchos valores atípicos o residuos no normales.                  |

Fuente: Elaboración Propia

### 3.1.9. MODELOS DE SEGMENTACIÓN DE SERIES DE TIEMPO

La segmentación es una técnica que permite dividir un conjunto de datos en grupos homogéneos según sus características. En el contexto de datos organizados secuencialmente, se deben tener en cuenta la dimensión temporal y la evolución de los datos a través del tiempo [20].

En la segmentación de series de tiempo existen dos enfoques:

1. Segmentación temporal: Busca dividir una misma serie en tramos o subseries que presenten comportamientos distintos, ya sea por un cambio estructural, cambios de tendencia o aparición de nuevos patrones [21].
2. Segmentación por similitud: Tiene como objetivo agrupar diferentes series que presentan trayectorias similares en el tiempo. Es útil cuando se dispone de múltiples series de tiempo que se desea analizar de forma conjunta, como es el caso de ventas de productos, regiones geográficas o perfiles de consumo [22].

En cuanto a la segmentación por similitud, uno de los algoritmos más utilizados es K-Means debido a su simplicidad y eficiencia [23]. Para ser usado en series de tiempo, se requiere que las series tengan la misma longitud y que los datos estén debidamente normalizados o escalados [24].

### 3.1.10. COEFICIENTES DE CORRELACIÓN Y COINTEGRACIÓN

En el análisis de series temporales, los coeficientes de correlación permiten cuantificar la fuerza y dirección de la relación entre dos variables numéricas o entre dos series temporales [25]. El coeficiente más utilizado es el de Pearson, que mide la relación lineal entre dos variables continuas. Su valor oscila entre -1 y 1: valores cercanos a 1 indican una relación lineal positiva fuerte, valores cercanos a -1 una relación lineal negativa fuerte y valores próximos a 0 sugieren una ausencia de relación lineal significativa [26].

En contextos donde las relaciones no son estrictamente lineales o los datos no cumplen los supuestos de normalidad, también pueden emplearse otros coeficientes como el de Spearman o Kendall, que miden relaciones monótonas y son más robustos frente a valores atípicos [27].

En series temporales, la correlación puede analizarse de forma contemporánea (entre valores en el mismo instante) o con desfase (correlaciones cruzadas), para identificar relaciones rezagadas en el tiempo. No obstante, es importante tener en cuenta que la presencia de tendencias comunes o no estacionariedad puede generar correlaciones espurias, este término, atribuido a Karl Pearson, se refiere a la situación de correlación entre las variables, aunque las variables originales no estén correlacionadas o sean aleatorias [26].

Para abordar este problema se recurre al concepto de cointegración [28]. Dos o más series de tiempo no estacionarias están cointegradas si existe una combinación lineal de ellas que sea estacionaria. Esto implica que, aunque las series por separado presenten tendencias, mantienen una relación de equilibrio de largo plazo. La metodología de Engle y Granger [28] es una de las más utilizadas para verificar la cointegración entre dos series y se desarrolla en dos etapas:

1. Estimación de la relación de largo plazo: Se ajusta un modelo de regresión lineal entre las series y se obtienen los residuos de esta estimación.
2. Prueba de estacionariedad de los residuos: Si los residuos resultan estacionarios (mediante pruebas como ADF), se concluye que las series están cointegradas.

### **3.2. ANTECEDENTES**

El pronóstico de la demanda es un elemento fundamental en el proceso de toma de decisiones de cualquier organización, debido a que permite ajustar las estrategias para mejorar las ventas y la productividad [29]. En este contexto, diversos estudios han explorado el uso de modelos avanzados para aumentar la precisión de los pronósticos y la toma de decisiones. Un ejemplo de ello es el desarrollo de una herramienta de pronóstico de demanda en una cadena de retail en Turquía, donde se emplearon diversos métodos de predicción, como media móvil, ARIMA, Holt-Winters, modelos de regresión y técnicas de deep learning. En este estudio, los modelos fueron evaluados y comparados con los valores reales históricos [30]. Se asignaron pesos a cada método según su exactitud, y los mejores modelos fueron seleccionados y ponderados para obtener un pronóstico final.

De manera similar, varias investigaciones sobre el uso de redes neuronales para la predicción de ventas han mostrado resultados igualmente prometedores. En [31], se implementó un modelo de red neuronal de retropropagación (BPNN) para prever la demanda de productos de cuidado

bucal, obteniendo una alta precisión. Este enfoque destacó la capacidad de las redes neuronales para modelar patrones de demanda complejos y no lineales. Asimismo, [32] demostró la eficacia de algoritmos de Machine Learning, como el Árbol de Decisiones y el Gradient Boosted Tree (GBT), en la predicción de ventas para una tienda electrónica. El modelo GBT superó significativamente a los métodos tradicionales, alcanzando una precisión del 98%.

En 2021, se llevó a cabo una investigación utilizando el algoritmo XGBoost para el pronóstico de ventas en el sector retail, específicamente con datos de ventas de Walmart [33]. Este algoritmo destacó por su eficiencia, precisión y capacidad para manejar grandes volúmenes de datos, superando a modelos como la regresión lineal y la regresión de cresta.

En el sector alimentario colombiano, un proyecto reciente implementa técnicas avanzadas de Machine Learning, como LSTM (Long Short-Term Memory), Random Forest, Prophet y SARIMA, para mejorar el pronóstico de ventas a corto plazo [34]. Este estudio mostró una reducción significativa en el error porcentual absoluto medio (MAPE).

Finalmente, las redes neuronales profundas han sido validadas en la predicción de series de tiempo tanto lineales como no lineales. En [35], se evidenció que las redes neuronales pueden ofrecer una alta precisión en la predicción de series temporales, obteniendo errores absolutos medios inferiores al 5%. Además, en [36] se combinaron modelos ARIMA, CNN (redes neuronales convolucionales) y LSTM, obteniendo mejoras significativas en la predicción de precios del carbón.

Es así, como todas las investigaciones anteriormente mencionadas, subrayan la importancia de utilizar técnicas avanzadas de ciencia de datos para la mejora de pronósticos de manera que se permita incrementar la eficiencia operativa de gestión de inventarios y producción. El uso de técnicas de Machine Learning demuestran un impacto positivo en la mejora de la precisión de los pronósticos, especialmente en las industrias con variabilidad en la demanda.

## **4. PREPARACIÓN Y ANÁLISIS EXPLORATORIO DE DATOS**

### **4.1. RECOLECCIÓN Y LIMPIEZA DE DATOS**

Inicialmente, se integraron datos internos de la empresa provenientes de una base de datos en SQL Server, que contenía información de pedidos, clientes y productos correspondiente al periodo comprendido entre enero de 2020 y abril de 2025. Como resultado del proceso de recolección, se consolidó una tabla con 55 atributos y un total de 989,413 registros, la cual sirvió como base para el análisis.

Posteriormente, se llevó a cabo un proceso de limpieza y depuración con el objetivo de garantizar la calidad de la información. Dicho proceso incluyó la eliminación de duplicados, el tratamiento de valores nulos, la verificación de formatos y la homogeneización de categorías.

A continuación, se efectuó una revisión exhaustiva para identificar las variables más relevantes para el pronóstico de demanda a nivel de SKU y cliente, bajo una frecuencia semanal. La selección se fundamentó en el conocimiento del negocio y en la pertinencia analítica de cada variable frente al objetivo del estudio.

Como resultado, se definieron 26 variables como insumo principal para el modelado predictivo. En la Tabla IV se presenta el detalle de estas variables empleadas en el análisis exploratorio, mientras que el diccionario completo, con la descripción detallada de cada una, se encuentra disponible en el Anexo A.

TABLA IV  
 VARIABLES INTERNAS PRINCIPALES

| <b>Nombre Atributo</b>         | <b>Tipo Dato</b>   | <b>Descripción</b>  |
|--------------------------------|--------------------|---|
| <b>C.O.</b>                    | Catagórico nominal | Indicador para el centro de operación.<br>1= Planta eje Cafetero<br>5= Planta costa Atlántica<br>10= Planta Centro<br>8= Planta Valle del cauca<br>3= CEDI Antioquia<br>11= CEDI Bogotá |
| <b>Fecha</b>                   | Fecha              | Fecha creación del pedido   |
| <b>Fecha entrega</b>           | Fecha              | Fecha de entrega del pedido   |
| <b>U.M.</b>                    | Catagórico nominal | Indicador para diferenciar la unidad del pedido. Puede ser en unidades o cajas.   |
| <b>Cant. pedida</b>            | Numérico continuo  | Cantidad pedida por el cliente  |
| <b>Cant. factura</b>           | Numérico continuo  | Cantidad facturada  |
| <b>Nro. documento</b>          | Catagórico nominal | Número documento interno  |
| <b>Valor subtotal local</b>    | Numérico continuo  | Valor pedido total  |
| <b>Tipo de maquila cliente</b> | Catagórico nominal | Unidades de negocio del cliente (Marca Privada, Gran Consumo, Soluciones de Limpieza y Desinfección, Internacionalización)  |

|                                     |                    |  |
|-------------------------------------|--------------------|--|
|                                     |                    | (USA)), es la principal segmentación de la organización para clasificar el tipo de cliente.  |
| <b>id sucursal</b>                  | Catagórico nominal | Índice punto de envío cliente  |
| <b>Ciudad</b>                       | Catagórico nominal | Ciudad del punto envío del cliente   |
| <b>Zona</b>                         | Catagórico nominal | Zona geográfica del punto de envío del cliente   |
| <b>Razón social cliente factura</b> | Catagórico nominal | Razón social del cliente   |
| <b>Canal</b>                        | Catagórico nominal | Clasificación de los clientes (Grandes cadenas, Mayoristas, supermercados, distribuidor aliado, etc.) Contiene 25 tipos de clasificaciones diferentes. |
| <b>Bodega</b>                       | Catagórico nominal | Bodega de almacenamiento de la cual se descontará el inventario  |
| <b>Ítem</b>                         | Catagórico nominal | Referencia/Sku del producto  |
| <b>Pedida en KG</b>                 | Numérico continuo  | Peso en kilogramos de la cantidad pedida   |
| <b>Pedida en LT</b>                 | Numérico continuo  | Volumen en litros de la cantidad pedida  |
| <b>Factor U.M. emp.</b>             | Numérico discreto  | Cantidad de unidades que conforman una caja  |
| <b>Categoría</b>                    | Catagórico nominal | Categorización del producto de acuerdo con su tipo (Aseo, Blanqueador, Cuidado Personal, Lavandería, Crema Lavalozza, Complementarios, Insumos, NA)    |
| <b>Familia</b>                      | Catagórico nominal | Detalle de la categoría por tipo producto (Suavizante, Jabón líquido, Jabón de manos, Limpia pisos, Lavalozza Liquido, etc.). Hay 52 diferentes tipos  |
| <b>Maquila</b>                      | Catagórico nominal | Marca del producto   |
| <b>Venta Perdida</b>                | Catagórico nominal | Descripción del motivo por el cual no fue posible atender el pedido en caso de que suceda.   |
| <b>Estado docto.</b>                | Catagórico ordinal | Estado del pedido:<br>En elaboración= se encuentra en proceso de creación  |

|                        |                    |  |
|------------------------|--------------------|--|
|                        |                    | Aprobado=El pedido está pendiente por atenderse completa o parcialmente<br>Retenido= El pedido presenta una novedad<br>Cumplido=El pedido ya fue entregado/cerrado |
| <b>tipo cliente</b>    | Categorico nominal | Nacional o internacional   |
| <b>Maquila cliente</b> | Categorico nominal | Retail, Hard Discount, Linea Institucional   |

Fuente: Elaboración Propia

Durante la depuración, se revisó la presencia de valores nulos en las variables seleccionadas en donde se identificó que las variables “Cantidad pedida en kg”, “Cantidad pedida en litros”, “Categoría”, “Familia” y “Maquila” contenían registros faltantes.

La variable con mayor cantidad de datos nulos fue Familia, con 573 registros lo que representa solo el 0,05 % del total de observaciones. Dado su bajo impacto, se optó por eliminar estos registros, sin comprometer la integridad del análisis.

Adicionalmente, se aplicaron los siguientes ajustes:

- Eliminación de clientes nuevos: Se excluyeron los registros de clientes cuya primera compra se realizó hace menos de 12 meses, ya que no contaban con un historial suficiente para un pronóstico confiable. En estos casos, el pronóstico se define a partir de acuerdos comerciales previamente establecidos entre el cliente y la organización.
- Eliminación de la categoría “Insumos”: Con el fin de enfocar el análisis en los productos principales (aseo y desinfección), se excluyeron los pedidos asociados a esta categoría, que corresponden a solicitudes excepcionales de materiales como etiquetas, corrugado o fragancias, los cuales no forman parte del objetivo del estudio.
- Segmentación por unidades de negocio: Se filtraron los registros para conservar únicamente aquellos pertenecientes a las unidades de negocio Marca Privada, Gran Consumo y Soluciones de Limpieza y Desinfección, por ser las más representativas para la compañía. Los pedidos correspondientes a otras unidades no se consideraron relevantes para este análisis.

## 4.2. ANÁLISIS EXPLORATORIO DE DATOS

A partir de los datos anteriormente revisados se realiza el análisis exploratorio para tener un mayor entendimiento del negocio.

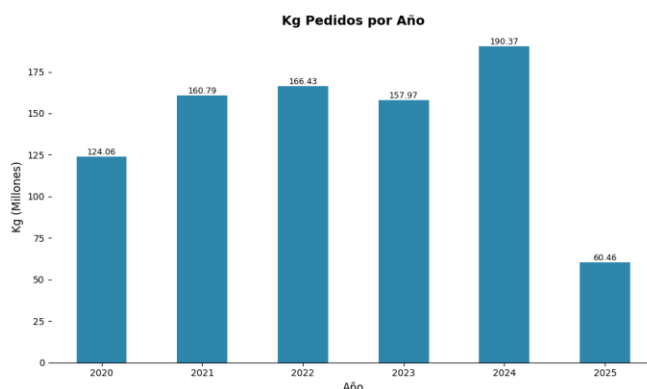
#### 4.2.1 ANÁLISIS DESCRIPTIVO Y ENTENDIMIENTO DEL NEGOCIO

La empresa, dedicada a la manufactura de productos de aseo y desinfección, cuenta con más de 25 años de trayectoria en el mercado nacional e internacional, atendiendo múltiples canales y clientes institucionales.

La Fig. 4, muestra la evolución del volumen total de pedidos medidos en kilogramos durante el periodo 2020–2025. Se observa una tendencia general de crecimiento en la demanda, alcanzando su punto máximo en 2024, con aproximadamente 190 millones de kilogramos solicitados, lo que representa el mayor volumen histórico registrado por la organización.

Este comportamiento sugiere una consolidación del mercado y una expansión sostenida en los últimos años, principalmente asociada al fortalecimiento de las marcas propias, la diversificación del portafolio y el incremento en la cobertura de clientes.

Para el año 2025, el volumen acumulado parcial asciende a 60 millones de kilogramos, lo que indica que el comportamiento anual aún está en desarrollo.



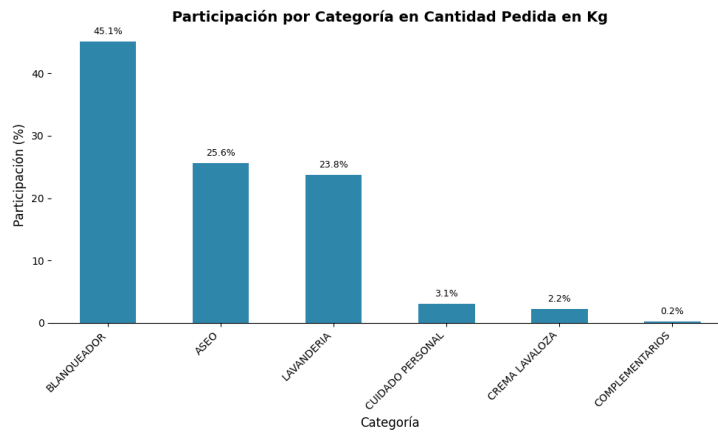
*Fig. 4 KG pedidos por año*

*Fuente: Elaboración Propia*

La Fig. 5 presenta la distribución porcentual del volumen total de pedidos medidos en kilogramos, clasificados por categoría de producto. Se evidencia una alta concentración en tres categorías principales: Blanqueador (45,1%), Aseo (25,6%) y Lavandería (23,8%) las cuales representan conjuntamente más del 94% del total solicitado.

Esto refleja que el portafolio de mayor rotación y demanda corresponde a líneas orientadas al cuidado y limpieza del hogar, consolidándose como el núcleo del negocio. Por su parte, las categorías Cuidado Personal (3,1%), Crema Lava Loza (2,2%) y Complementarios (0,2%) presentan una participación marginal en términos de volumen, lo que sugiere

oportunidades de diversificación o fortalecimiento estratégico si se busca equilibrar la mezcla de productos y aumentar la participación de segmentos con mayor valor agregado.

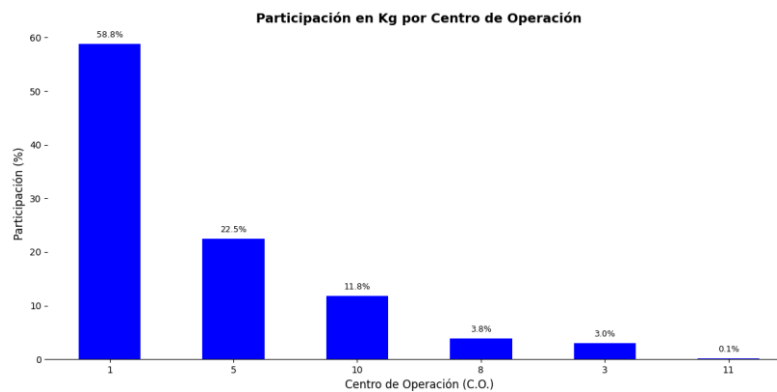


*Fig. 5 Participación por categoría*

*Fuente: Elaboración Propia*

La Fig. 6 muestra la distribución porcentual del volumen total de pedidos en kilogramos según el Centro de Operación (C.O.) responsable de su atención. Se observa una alta concentración productiva en el C.O. 1 (Planta Eje Cafetero), que participa con un 58,8% del total, consolidándose como el principal centro operativo de la organización. En segundo lugar, se encuentra el C.O. 5 (Planta Costa Atlántica) con una participación del 22,5%, seguido del C.O. 10 (Planta Centro) con 11,8%. Estos tres centros concentran conjuntamente más del 93% del volumen total, lo que indica una estructura operacional centralizada en las plantas de manufactura principales.

Los centros de distribución (CEDI) —Antioquia (C.O. 3), Bogotá (C.O. 11) y Valle del Cauca (C.O. 8)— registran participaciones marginales, reflejando que su rol operativo se orienta más a la logística y distribución, y no a la generación directa de volumen productivo.



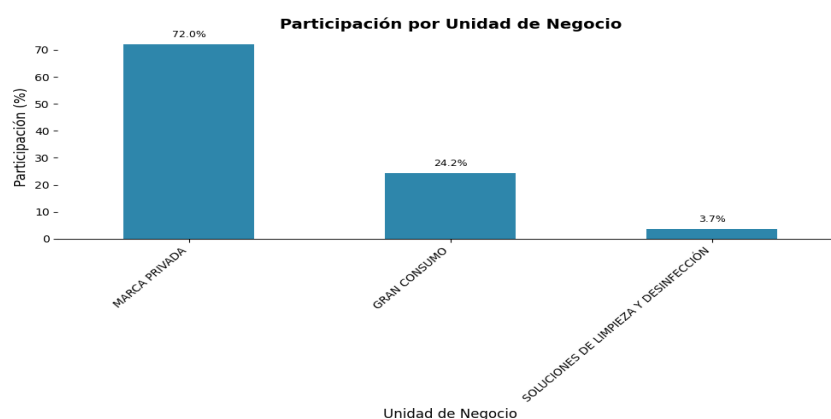
*Fig. 6 Participación en KG por centro de operación*

*Fuente: Elaboración Propia*

Desde una perspectiva comercial, la empresa organiza su operación en tres unidades de negocio que agrupan a los clientes según su perfil, nivel de compras y tipo de relación contractual. Esta segmentación permite orientar las estrategias de producción, ventas y servicio hacia las necesidades específicas de cada mercado.

La Fig. 7 muestra que el segmento de Marca Privada representa la principal fuente de demanda, con una participación del 72,0% sobre el total de kilogramos pedidos. En este grupo se concentran los clientes de mayor volumen, entre ellos los formatos de bajo costo (Hard Discount) y las grandes cadenas nacionales e internacionales, para quienes la compañía desarrolla productos exclusivos bajo sus propias marcas. Su relevancia estratégica se refleja en un promedio mensual cercano a 12 millones de kilogramos solicitados, impactando directamente la planificación productiva y logística.

En segundo lugar, la unidad de Gran Consumo aporta el 24,2% del volumen total, incluyendo supermercados, grandes superficies, cadenas regionales y distribuidores. Este segmento contribuye de manera significativa a la diversificación del portafolio y al posicionamiento en canales de retail, combinando volumen y margen competitivo. Finalmente, la unidad de Soluciones de Limpieza y Desinfección participa con el 3,7%, enfocándose en clientes institucionales como hoteles, hospitales, centros comerciales y distribuidores especializados, los cuales demandan productos concentrados y de alto valor agregado.



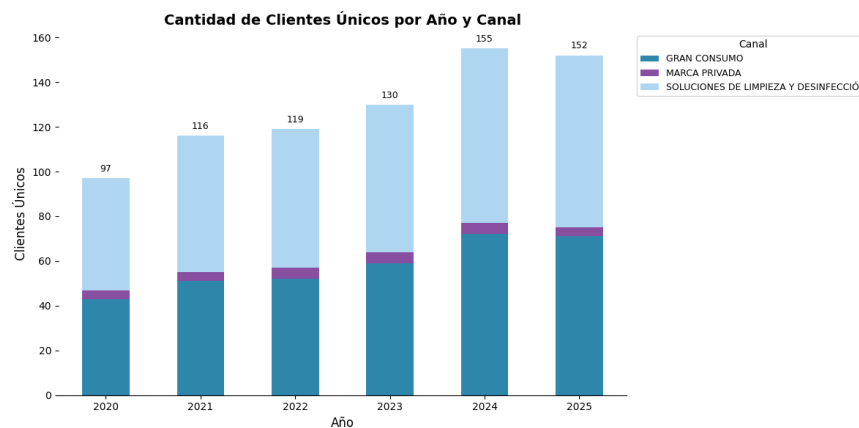
**Fig. 7 Participación por unidad de negocio**

*Fuente: Elaboración Propia*

A lo largo del periodo 2020–2025, la compañía ha mostrado un crecimiento sostenido en su base de clientes únicos, reflejo del fortalecimiento de su presencia comercial y la diversificación de canales. El total de clientes pasó de 97 en 2020 a 155 en 2024, con una ligera variación en 2025 (152 clientes), lo que evidencia una tendencia positiva en la expansión del mercado atendido.

La Fig. 8 muestra que el canal Gran Consumo concentra el mayor número de clientes, presentando un crecimiento constante a lo largo de los años. Este comportamiento confirma su rol como principal motor de ampliación de cobertura y penetración comercial, especialmente en supermercados, cadenas regionales y distribuidores.

El canal de Soluciones de Limpieza y Desinfección ha venido consolidándose progresivamente, representando un segmento especializado orientado a clientes institucionales, con potencial de desarrollo en mercados de nicho como hospitales, hoteles y centros comerciales. En contraste, el canal de Marca Privada, aunque cuenta con una base reducida de clientes, agrupa a los principales compradores en términos de volumen.

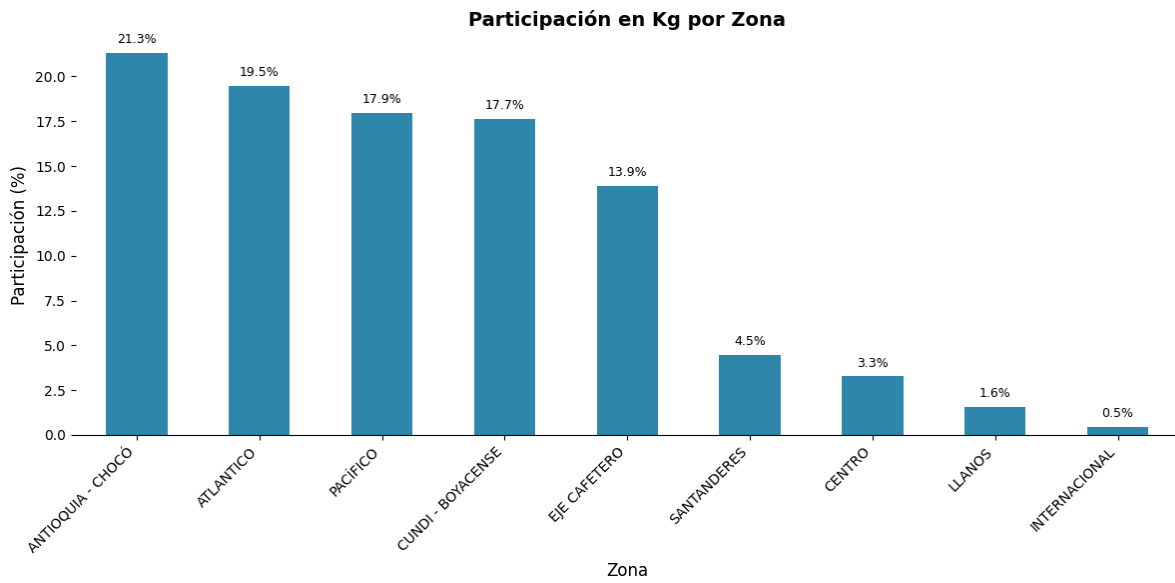


*Fig. 8 Cantidad de clientes por canal de atención*

*Fuente: Elaboración Propia*

El análisis por zona permite identificar la distribución territorial de la demanda, evidenciando el peso relativo que cada región aporta al volumen total de pedidos. De acuerdo con la Fig. 9, se observa una concentración significativa en las zonas Antioquia–Chocó (21,3%), Atlántico (19,5%), Pacífico (17,9%) y Cundinamarca–Boyacá (17,7%), las cuales en conjunto representan más del 76% del total de kilogramos solicitados. Estas regiones constituyen los mercados más relevantes para la organización, ya sea por su densidad poblacional, dinamismo comercial o presencia de grandes clientes institucionales y cadenas retail.

En una segunda línea se ubican las zonas Eje Cafetero (13,9%), Santanderes (4,5%) y Centro (3,3%), que aportan una participación complementaria, consolidando la cobertura nacional de la compañía. Por su parte, las regiones de Llanos (1,6%) e Internacional (0,5%) presentan una participación marginal, las cuales se encuentran en crecimiento.

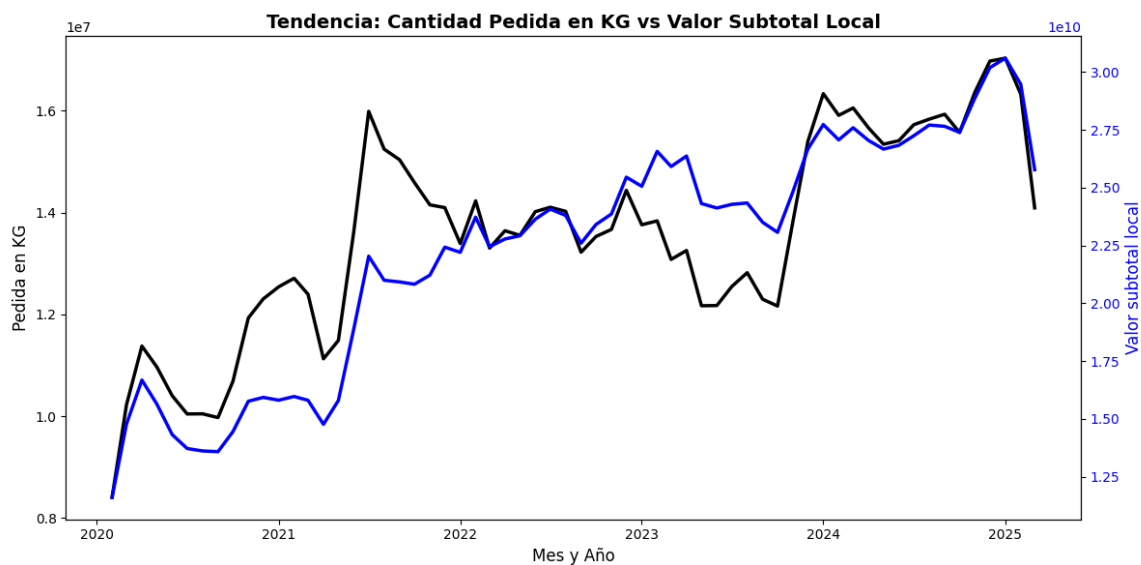


*Fig. 9 Participación de venta en KG por zona*  
*Fuente: Elaboración Propia*

#### 4.2.2 Visualización de la serie de tiempo

El comportamiento conjunto de las variables cantidad pedida (en kilogramos) y valor subtotal local (valor del pedido en dinero) entre enero de 2020 y 2025 evidencia una tendencia general creciente a lo largo del periodo analizado, lo que refleja un incremento sostenido en la demanda y en el valor de las ventas.

Sin embargo, en la Fig. 10 se observa que entre finales de 2022 e inicios de 2024 se amplía la brecha entre el volumen pedido y el valor total de los pedidos, lo que indica un mayor crecimiento relativo del valor respecto al volumen. Esta diferencia se atribuyó principalmente a ajustes en la estructura de precios, especialmente en los clientes del segmento de Marca Privada, que concentran aproximadamente el 72% del total de kilogramos solicitados. Dicho comportamiento muestra que la variación del valor total pedido no solo está determinada por el incremento en la cantidad demandada, sino también por estrategias de reajuste de precios y posibles cambios en la mezcla de productos.



**Fig. 10 Tendencia cantidad pedida en KG vs Valor \$**

Fuente: Elaboración Propia

### 4.3. ANÁLISIS DE VARIABLES EXTERNAS

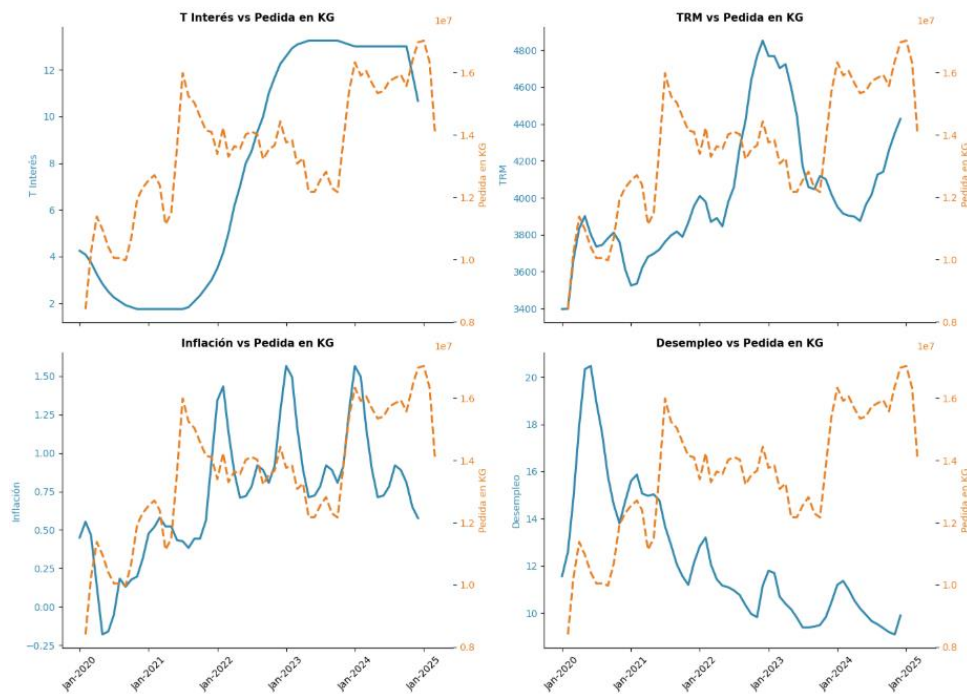
En la gestión de la demanda de productos de consumo, pueden intervenir factores externos que afectan tanto el comportamiento de la demanda como la precisión de los pronósticos. Estos elementos, como condiciones macroeconómicas o cambios en el entorno, generan variaciones que los modelos predictivos deben considerar para mejorar su exactitud [4].

En este estudio, la variable dependiente seleccionada fue la cantidad pedida (en kilogramos), al representar directamente el volumen de la demanda. Es importante tener en cuenta que los precios de los productos pueden verse afectados no solo por cambios en el volumen de los pedidos, sino también por el comportamiento del Índice de Precios al Consumidor (IPC), que refleja el efecto de la inflación sobre el costo de los bienes.

Con este enfoque, se evaluó la relación entre la cantidad de kilos solicitados y una serie de indicadores externos, agrupando los datos a nivel mensual para mantener coherencia con la frecuencia de publicación de las variables macroeconómicas. Las variables externas consideradas fueron:

- IPC (Índice de Precios al Consumidor)
- TRM (Tasa Representativa del Mercado)
- Tasa de desempleo
- Tasa de interés

Al comparar mensualmente la cantidad de kilos pedidos frente a cada uno de estos indicadores (ver Fig. 11), se observa que la tasa de desempleo presenta una relación inversa con la demanda: cuando el desempleo aumenta, los kilos solicitados tienden a disminuir, lo cual sugiere una sensibilidad del consumo a las condiciones del mercado laboral. En contraste, las demás variables presentan patrones menos consistentes, aunque se evidencian ciertos efectos de la inflación y de la TRM sobre el comportamiento general de la demanda.



**Fig. 11** Relación de ventas vs variables externas

Fuente: Elaboración Propia

Con el propósito de evaluar la existencia de una relación de equilibrio a largo plazo entre la demanda (expresada como la cantidad pedida en kilogramos) y las variables macroeconómicas externas, se aplicó la prueba de cointegración de Engle–Granger [28].

En la Fig. 12 se presentan los resultados obtenidos. En todos los casos, el valor-p resultó mayor a 0.05, por lo que no se rechaza la hipótesis nula de no cointegración. Esto indica que no existe evidencia estadística de una relación de equilibrio a largo plazo entre las variables externas consideradas (tasa de interés, TRM, inflación y desempleo) y la demanda en kilos.

Estos resultados sugieren que las fluctuaciones de la demanda no dependen de manera estable de las condiciones macroeconómicas analizadas, al menos en el horizonte temporal del estudio. Por lo tanto, las variables externas no presentan cointegración con la variable dependiente, lo que implica que sus efectos sobre la demanda podrían ser de corto plazo o no lineales, en lugar de estructurales o permanentes.

RESULTADOS DEL TEST DE COINTEGRACIÓN (Engle-Granger)

T Interés:  
Estadístico de prueba: -2.2033  
Valor-p: 0.4226  
→ No hay evidencia de cointegración

TRM:  
Estadístico de prueba: -2.0523  
Valor-p: 0.5013  
→ No hay evidencia de cointegración

Inflación:  
Estadístico de prueba: -0.9762  
Valor-p: 0.9073  
→ No hay evidencia de cointegración

Desempleo:  
Estadístico de prueba: -1.8054  
Valor-p: 0.6271  
→ No hay evidencia de cointegración

*Fig. 12 Resultados del Test de Cointegración*

*Fuente: Elaboración Propia*

Posteriormente, se realizó una regresión lineal múltiple para evaluar el impacto de las variables macroeconómicas externas sobre la cantidad pedida en kilogramos, con el fin de determinar su viabilidad para ser incorporadas en el modelo de pronóstico. (Ver Fig. 13)

El modelo presenta un  $R^2$  de 0.252, lo que indica que únicamente el 25.2% de la variabilidad en la demanda puede ser explicada por las variables independientes consideradas (tasa de interés, TRM, inflación y desempleo). Este bajo nivel de ajuste sugiere que las variables externas analizadas tienen un poder explicativo limitado sobre la demanda.

A pesar de ello, el modelo resulta estadísticamente significativo en conjunto ( $p = 0.002$ ), lo cual indica que, colectivamente, las variables incluidas tienen alguna capacidad para explicar la variación en la demanda. No obstante, al analizar cada variable de forma individual, ninguna resulta significativa al nivel del 5%, siendo la tasa de desempleo la que más se aproxima a la significancia ( $p = 0.064$ ), lo que sugiere una posible relación negativa con la cantidad pedida (coeficiente = -268,900).

Estos resultados refuerzan la conclusión previa del análisis de cointegración, indicando que las variables externas no mantienen una relación estable ni estadísticamente significativa con la demanda en el horizonte temporal evaluado, por lo que su inclusión en modelos de pronóstico podría tener un impacto limitado.

OLS Regression Results

```

=====
Dep. Variable:          Pedida en KG      R-squared:              0.252
Model:                  OLS              Adj. R-squared:         0.198
Method:                 Least Squares    F-statistic:            4.708
Date:                   Sun, 28 Sep 2025    Prob (F-statistic):     0.00240
Time:                   22:48:11         Log-Likelihood:         -975.79
No. Observations:      61              AIC:                    1962.
Df Residuals:           56              BIC:                    1972.
Df Model:                4
Covariance Type:       nonrobust
=====

```

|           | coef       | std err  | t      | P> t  | [0.025    | 0.975]   |
|-----------|------------|----------|--------|-------|-----------|----------|
| const     | 1.269e+07  | 4.26e+06 | 2.977  | 0.004 | 4.15e+06  | 2.12e+07 |
| T Interés | -1.775e+04 | 1.05e+05 | -0.169 | 0.866 | -2.28e+05 | 1.93e+05 |
| TRM       | 828.0884   | 1022.845 | 0.810  | 0.422 | -1220.917 | 2877.094 |
| Inflación | 1.051e+06  | 7.8e+05  | 1.347  | 0.183 | -5.11e+05 | 2.61e+06 |
| Desempleo | -2.689e+05 | 1.42e+05 | -1.889 | 0.064 | -5.54e+05 | 1.63e+04 |

```

=====
Omnibus:                1.529    Durbin-Watson:          0.908
Prob(Omnibus):          0.466    Jarque-Bera (JB):      1.025
Skew:                   -0.310   Prob(JB):               0.599
Kurtosis:               3.137    Cond. No.               6.02e+04
=====

```

*Fig. 13 Resultados regresión variables externas*

*Fuente: Elaboración Propia*

Finalmente, los resultados de las pruebas de cointegración y regresión lineal evidencian que las variables macroeconómicas externas consideradas no son suficientes ni estadísticamente relevantes para explicar la variación observada en la cantidad pedida en kilogramos.

#### 4.4. DEFINICIÓN DE VARIABLES DEL MODELO DE PRONÓSTICO

Con base en el análisis del negocio y la evaluación estadística de las variables disponibles, se identificaron las variables que resultan significativas para el desarrollo del modelo de pronóstico:

- **Fecha:** Permite estructurar la información en formato de serie de tiempo, siendo esencial para la aplicación de métodos de pronóstico temporal y la identificación de patrones estacionales y tendencias.
- **Razón social cliente factura:** Identifica de manera única al cliente asociado a cada transacción, posibilitando el análisis segmentado por cliente y la generación de pronósticos desagregados.
- **Ítem:** Representa el producto comercializado, necesario para realizar pronósticos específicos por referencia o SKU.

- **Cantidad Pedida en KG:** Corresponde a la variable objetivo del modelo, es decir, la que se desea pronosticar. Se seleccionó esta métrica porque refleja el volumen real de compra, a diferencia del valor monetario, que puede estar distorsionado por variaciones en los precios o estrategias comerciales.

No se incluyeron variables externas, dado que, como se evidenció en los análisis de cointegración y regresión lineal, los factores macroeconómicos disponibles no aportan información significativa ni explican adecuadamente la variabilidad en la cantidad pedida en kilogramos. Por tanto, el modelo se centrará en las variables internas y en los patrones históricos de demanda, los cuales representan de forma más precisa el comportamiento real del negocio.

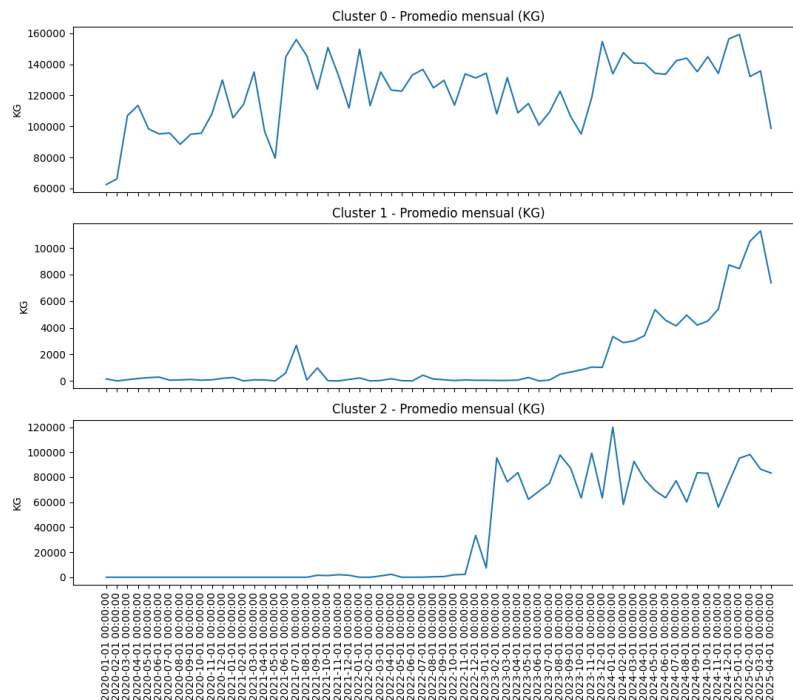
#### 4.5. CLÚSTER DE SERIES DE TIEMPO

Basándose en el estudio de Subramanian, Kayalvizhi y Thangarasu [37], que resaltó la importancia de la segmentación y el análisis de componentes temporales para mejorar la precisión de los pronósticos de demanda, se procedió a identificar grupos de clientes con patrones de comportamiento similares a partir de sus historiales de compra.

Para ello, se construyeron series temporales mensuales de la variable Pedida en KG por cliente, las cuales reflejaron el volumen de producto solicitado a lo largo del tiempo. Posteriormente, se aplicaron técnicas de reducción de dimensionalidad (PCA) para capturar las características más relevantes de la forma de cada serie, seguidas del algoritmo de agrupamiento K-Means, con el fin de identificar patrones homogéneos de comportamiento. Se evaluaron diferentes configuraciones en cuanto al número de clústeres, seleccionando aquella que presentó mayor coherencia interpretativa y estabilidad temporal.

Del análisis inicial surgieron tres grupos principales, resumidos en la Fig. 14.

- **Clúster 0 – Clientes constantes a lo largo del tiempo:** presentaron un comportamiento de compra estable, con volúmenes moderados y fluctuaciones controladas.
- **Clúster 1 – Clientes con crecimiento progresivo:** iniciaron con volúmenes bajos o nulos y mostraron una tendencia ascendente sostenida en los últimos períodos.
- **Clúster 2 – Clientes con picos recientes:** se caracterizaron por incrementos abruptos en la demanda en fechas específicas, seguidos de descensos.



**Fig. 14 Cantidad pedida en promedio mensual por clúster**

*Fuente: Elaboración Propia*

Debido a su relevancia operativa y al interés por patrones estables y representativos, se decidió profundizar en el Clúster 0, realizando una segunda segmentación interna para identificar subgrupos más homogéneos. Nuevamente se aplicaron PCA y K-Means, fijando tres subclústeres (Fig. 15).

En el caso de los clústeres 1 y 2 no se realizó una descomposición estacional ni un análisis detallado de sus componentes, ya que agrupan clientes con comportamientos de demanda aún inestables o en etapas tempranas de desarrollo comercial. Desde la perspectiva del modelo de Sales & Operations Planning (S&OP), estos grupos se gestionan principalmente a través de un enfoque colaborativo con el equipo de ventas, donde las proyecciones se elaboran considerando información cualitativa, conocimiento del mercado y planes comerciales específicos, más que mediante patrones estadísticos. Por esta razón, su pronóstico se apoya en la planeación conjunta con el área comercial, complementando el enfoque cuantitativo aplicado al Clúster 0, que concentra los clientes con mayor estabilidad y representatividad histórica.



**Fig. 15 Promedio Mensual pedido en Kg para cada Subclúster**  
Fuente: Elaboración Propia

En la Fig. 16 se resumen los subgrupos obtenidos:

```
=== Resumen por subcluster (C0, KG originales) ===
```

| subcluster | n_clientes | prom_mensual  | minimo    | maximo      |
|------------|------------|---------------|-----------|-------------|
| A          | 6          | 67076.989583  | 1780534.0 | 6951278.0   |
| B          | 87         | 142880.301545 | 31772.0   | 308185910.0 |
| C          | 14         | 17133.572545  | 60234.0   | 8062298.0   |

| subcluster | meses_compra_prom | cv_prom  |
|------------|-------------------|----------|
| A          | 62.333333         | 0.429727 |
| B          | 56.287356         | 0.741954 |
| C          | 49.571429         | 0.748126 |

**Fig. 16 Resumen por subclúster**  
Fuente: Elaboración Propia

- Subclúster A – Clientes de consumo medio y comportamiento estable: Incluyó 6 clientes con un promedio mensual cercano a 67.000 kg, un coeficiente de variación de 0,43 y compras distribuidas de forma constante en el tiempo. Mostraron un patrón estable con ligeras fluctuaciones y tendencia moderada.
- Subclúster B – Clientes de alto volumen y mayor variabilidad: Conformó el grupo más representativo (87 clientes), con un promedio mensual de 142.880 kg y  $cv = 0,74$ . Aunque

presentaron oscilaciones mensuales, mantuvieron una tendencia de crecimiento y aportaron la mayor proporción del volumen total.

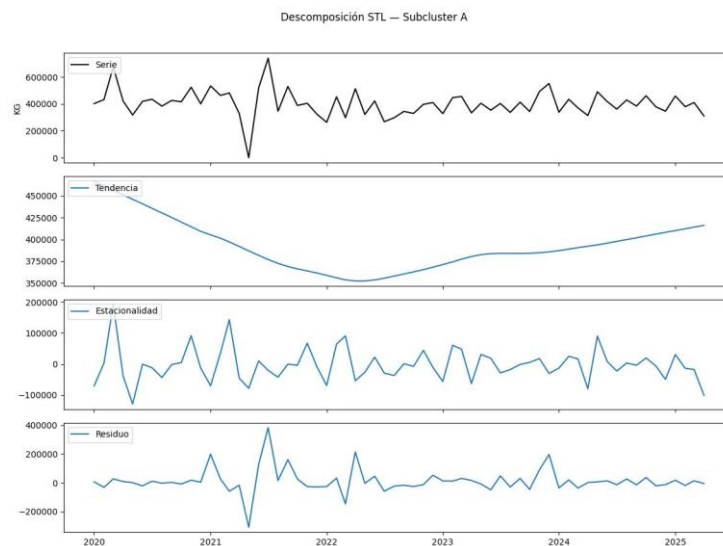
- Subclúster C – Clientes de bajo volumen y comportamiento irregular: Se compuso de 14 clientes con un promedio mensual de 17.133 kg y  $cv = 0,75$ . Evidenciaron una demanda mucho más volátil y concentraciones de consumo en meses específicos, posiblemente asociadas a compras estacionales o puntuales.

#### **4.6. DESCOMPOSICIÓN DE LA SERIE DE TIEMPO**

Con el propósito de profundizar en la comprensión de los patrones temporales presentes en cada subgrupo identificado dentro del Clúster 0, se aplicó una descomposición STL (Seasonal-Trend decomposition using Loess) a las series temporales mensuales de la variable Pedida en KG. Esta técnica permitió separar cada serie en tres componentes fundamentales: tendencia, estacionalidad y residuo, facilitando el análisis individual de los efectos a largo plazo, los patrones periódicos y las fluctuaciones aleatorias.

### Subclúster A – Comportamiento estable con leve recuperación reciente

En la Fig. 17, correspondiente al Subclúster A, se observa una tendencia decreciente entre 2020 y 2023, seguida de una recuperación moderada hacia finales del período analizado. Este comportamiento sugiere una contracción inicial en los volúmenes solicitados, asociada a variaciones de precios, y una posterior estabilización. La componente estacional mostró fluctuaciones recurrentes, aunque de magnitud moderada, indicando la presencia de ciclos temporales, pero con un peso relativo menor frente a la tendencia, los residuos evidenciaron picos aislados, atribuibles a variaciones puntuales de la demanda sin patrón sistemático.

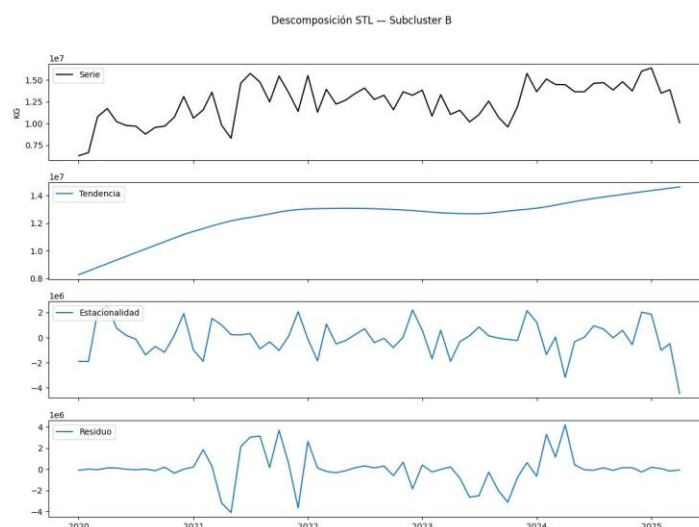


*Fig. 17 Descomposición STL - Subclúster A*

*Fuente: Elaboración Propia*

## Subclúster B – Alta demanda y crecimiento sostenido

En la Fig. 18, correspondiente al Subclúster B, se identificó una tendencia ascendente pronunciada desde 2020 hasta mediados de 2023, con una leve estabilización posterior. Este grupo, conformado por los clientes de mayor volumen, mostró una clara trayectoria de expansión de la demanda, lo cual puede reflejar la consolidación de relaciones comerciales o una mayor penetración del portafolio de productos. El componente estacional presentó variaciones marcadas (de hasta  $\pm 2$  millones de kilogramos), lo que indica una estacionalidad significativa, posiblemente asociada a ciclos comerciales o estacionales específicos (por ejemplo, promociones o campañas institucionales), los residuos mantuvieron un comportamiento relativamente estable, sin evidencias de choques estructurales relevantes.

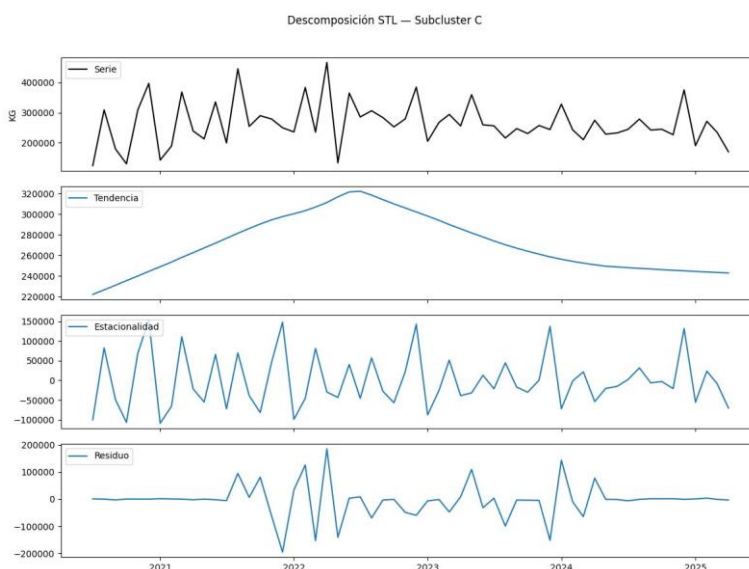


**Fig. 18** Descomposición STL - Subclúster B

Fuente: Elaboración Propia

### Subclúster C – Picos iniciales y tendencia descendente

La Fig. 19 muestra la descomposición del Subclúster C, caracterizado por clientes de bajo volumen y comportamiento irregular. Se evidenció una tendencia creciente hasta 2022, seguida de una caída sostenida hacia 2025. Este patrón sugiere que la demanda de este grupo fue transitoria, posiblemente asociada a compras esporádicas o contratos de corta duración. La estacionalidad presentó oscilaciones frecuentes, aunque de menor amplitud respecto al Subclúster B, lo cual denota un patrón menos estructurado, los residuos mantuvieron una dispersión considerable, confirmando la volatilidad de este segmento.



*Fig. 19 Descomposición STL - Subclúster C*

*Fuente: Elaboración Propia*

#### 4.7. DEFINICIÓN DE HORIZONTES DE TIEMPO

Con el propósito de evaluar y ejecutar los modelos de pronóstico desarrollados, se establecieron tres horizontes temporales: 4, 8 y 12 semanas. La selección de estos periodos respondió tanto a la estructura del ciclo de planeación de la compañía como a los requerimientos del proceso de Sales and Operations Planning (S&OP), el cual demanda diferentes niveles de anticipación para respaldar las decisiones tácticas, operativas y estratégicas [8].

El horizonte de corto plazo (4 semanas) se orientó a decisiones tácticas inmediatas, tales como el ajuste de planes de producción, la gestión de inventarios y la respuesta ágil ante variaciones en la demanda. En este nivel, los pronósticos permiten reaccionar oportunamente ante cambios de mercado o desviaciones respecto al plan.

El horizonte de mediano plazo (8 semanas) permitió una planificación más sólida de los recursos productivos y logísticos, facilitando la coordinación interfuncional entre áreas como compras, producción, distribución y ventas. Este nivel de anticipación garantiza una mayor eficiencia en la asignación de recursos y en la disponibilidad de capacidades operativas.

Por su parte, el horizonte de largo plazo (12 semanas) se alineó con actividades de carácter estratégico, como la formulación de presupuestos, la definición de estrategias comerciales y la evaluación de inversiones estructurales, entre ellas la adquisición de maquinaria o la expansión de capacidad instalada.

Esta segmentación de horizontes permitió entrenar, aplicar y evaluar los modelos de pronóstico en distintos contextos de decisión, garantizando su relevancia práctica y su ajuste a las necesidades de planeación de la organización.

## **5. APLICACIÓN DE MODELOS DE PRONÓSTICO**

Una vez definidos los horizontes de pronóstico y establecidas las segmentaciones por subclúster, se procedió con la implementación de los modelos con el objetivo de estimar la demanda semanal en kilogramos a nivel de cliente y categoría. La implementación del modelo se realizó en Python 3.10, integrando librerías especializadas como statsmodels, xgboost, prophet y scikit-learn.

En una primera fase, se evaluó la posibilidad de construir los pronósticos con el mayor nivel de detalle posible, a nivel SKU–cliente. No obstante, las pruebas iniciales mostraron que este enfoque implicaría tiempos de cómputo excesivos. Por ejemplo, el entrenamiento de 405 modelos tomó cerca de nueve horas, y la proyección hacia un esquema completo de 300 SKU y 180 clientes habría requerido más de un mes de procesamiento continuo. Ante esta limitación, se decidió trabajar a un nivel de agregación superior (categoría dentro de cada subclúster), lo que permitió mantener la viabilidad técnica y la estabilidad del proceso sin comprometer la capacidad analítica del modelo.

A partir de la información histórica comprendida entre enero de 2020 y abril de 2025, se construyeron series temporales semanales con un conjunto de variables que facilitaron la identificación de patrones y relaciones relevantes. Entre estas variables se incluyeron rezagos de una a doce semanas, medias móviles, indicadores de calendario como semana, mes y trimestre, y variables categóricas asociadas a los clientes y las categorías de producto. Con el fin de preservar la estructura temporal, las series se dividieron en un conjunto de entrenamiento equivalente al 80% de los datos y otro de validación con el 20% restante, manteniendo siempre la secuencia cronológica.

Posteriormente, en la fase de modelado se adoptó una estrategia comparativa que combinó enfoques clásicos y modernos para capturar tanto las tendencias y estacionalidades como los efectos no lineales de las series. Entre los modelos clásicos se emplearon los modelos SARIMAX y Holt-Winters. El modelo SARIMAX, se calibró mediante una búsqueda sistemática de hiperparámetros, seleccionando las combinaciones que minimizaron los criterios de información y los errores en el conjunto de validación. Mientras el modelo Holt-Winters permitió descomponer las series en sus componentes de nivel, tendencia y estacionalidad, ajustando sus parámetros de suavizamiento mediante una búsqueda manual en rejilla orientada a minimizar el error cuadrático medio.

Asimismo, se incorporó el modelo Prophet, que permitió capturar puntos de cambio estructurales y múltiples estacionalidades (semanal, mensual y anual), regulando el hiperparámetro `changepoint_prior_scale` para equilibrar la flexibilidad del modelo frente al riesgo de sobreajuste. Además, se exploraron redes neuronales recurrentes de tipo Elman y Jordan, diseñadas para capturar dependencias temporales más complejas. La arquitectura Elman se implementó mediante una capa SimpleRNN con activación tanh y dropout, mientras que la red Jordan se construyó a través de una celda personalizada que incorpora retroalimentación desde la salida anterior (feedback), lo que le permite modelar relaciones de largo plazo y dinámicas no lineales. Ambas redes fueron entrenadas con el optimizador Adam y pérdida MSE, utilizando datos escalados y transformados en ventanas temporales deslizantes, y su selección final se realizó con base en el desempeño obtenido en el conjunto de validación temporal.

Finalmente, con el fin de capturar interacciones no lineales entre las variables derivadas, se implementó el algoritmo XGBoost Regressor, el cual fue entrenado sobre un conjunto ampliado de características y calibrado mediante la función `ParameterGrid` de scikit-learn, explorando diversas combinaciones de hiperparámetros como la profundidad de los árboles, la tasa de aprendizaje, el número de estimadores y los parámetros de muestreo, seleccionando aquellas configuraciones con el menor error en validación.

Para garantizar la comparabilidad, cada modelo fue evaluado en tres horizontes temporales: corto plazo (4 semanas), mediano plazo (8 semanas) y largo plazo (12 semanas), alineados con los distintos niveles de planificación táctica, operativa y estratégica de la organización. Los resultados evidenciaron que no existía un único modelo superior en todos los casos. Por ejemplo, Las series con estacionalidad marcada fueron mejor representadas por Holt-Winters o SARIMAX, mientras que aquellas con mayor variabilidad o presencia de cambios estructurales se beneficiaron de la flexibilidad de Prophet, las RNN de tipo Elman y Jordan, y especialmente del XGBoost Regressor. Esta combinación permitió obtener pronósticos robustos y consistentes, que sirvieron como base para los procesos posteriores de planeación de la demanda y desagregación hacia el nivel cliente.

A partir de la selección del mejor modelo por subclúster y categoría, se generó una base de datos consolidada en formato CSV, la cual integra en un mismo esquema el histórico semanal y los

pronósticos multihorizonte (4, 8 y 12 semanas). Esta estructura, diseñada para su uso en Power BI, facilita el análisis jerárquico “de lo general a lo particular”: primero a nivel Cluster×Categoría para validar tendencias y estacionalidades globales, y luego a nivel Cliente×Categoría para analizar la desagregación operativa del plan. El proceso comenzó asegurando la consistencia temporal de los datos mediante la normalización de fechas, la eliminación de duplicidades y la restricción a un rango de años válido, finalmente, se reconstruyeron las series semanales a una frecuencia anclada a lunes (W-MON), garantizando la alineación entre los índices de entrenamiento y proyección.

Durante la fase de proyección, se reutilizaron los mejores modelos y configuraciones seleccionadas, evitando una nueva búsqueda de hiperparámetros. Además, se incorporó un paso de robustecimiento que recortó valores atípicos mediante un filtro MAD y forzó la no negatividad en las predicciones, garantizando coherencia operativa. En el caso del XGBoost, se aplicó un enfoque recursivo: cada semana proyectada se agregó al histórico para recalcular rezagos y características, produciendo trayectorias más realistas y dinámicas. En paralelo, los modelos clásicos, Prophet y UCM realizaron proyecciones directas sobre el histórico completo ajustado.

Finalmente, los pronósticos agregados por subclúster y categoría fueron desagregados al nivel cliente mediante un esquema proporcional dinámico, basado en las participaciones relativas observadas en las últimas cuatro semanas. La distribución se realizó con una lógica en cascada de 4,8 y 12 semanas y una consulta al histórico completo cuando fue necesario. Este enfoque mantuvo la coherencia entre los totales y las asignaciones individuales, privilegiando la información más actual. Con los datos históricos y pronosticados, se consolidó una base de datos en formato CSV a nivel semanal, incorporando campos de control como el tipo de dato (histórico o pronóstico), el horizonte temporal, el modelo seleccionado y la participación aplicada. Esta estructura permitió analizar, en Power BI, la evolución histórica y los escenarios futuros en una sola visualización, conservando trazabilidad y soporte analítico para la toma de decisiones.

## **6. EVALUACIÓN DEL MODELO**

La evaluación del modelo constituye una etapa fundamental dentro del proceso de modelado, ya que permite validar el desempeño de las técnicas seleccionadas y establecer con evidencia cuantitativa cuál de ellas ofrece la mayor precisión y estabilidad al pronosticar la demanda. En este proyecto se seleccionó RMSE como métrica principal de evaluación de los modelos porque permite medir el error del pronóstico en las mismas unidades del negocio (kilogramos), lo que facilita la interpretación de los resultados para los equipos de planeación y producción.

Además, RMSE penaliza de manera más fuerte los errores grandes, característica especialmente relevante en contextos de demanda con alta variabilidad o con picos abruptos donde un error elevado puede traducirse en desabastecimientos, excesos de inventario o impactos financieros significativos. Esta propiedad permitió diferenciar con mayor precisión cuál modelo respondía

mejor ante fluctuaciones relevantes del negocio. Asimismo, RMSE resultó adecuado para comparar modelos con distinta naturaleza (SARIMAX, Holt-Winters, Prophet, RNN, XGBoost), pues mantiene consistencia en la escala y permite evaluar desempeño a través de los diferentes horizontes (4, 8 y 12 semanas). Su sensibilidad a grandes desviaciones facilitó la selección de modelos que no solo ajustaran bien la tendencia global, sino que también capturaran adecuadamente las variaciones críticas de la demanda.

Los resultados obtenidos, presentados en el Anexo B, mostraron que no existe un modelo único que funcione mejor en todos los casos. No obstante, se identificaron patrones claros según el tipo de subclúster y la complejidad de las series de demanda.

En términos generales, XGBoost fue el modelo que ofreció un desempeño más estable y preciso, logrando los menores errores en la mayoría de los escenarios. Su comportamiento fue especialmente destacado en los subclústeres A y C, donde la demanda presenta una alta variabilidad y picos de consumo repentinos. En el subclúster A, por ejemplo, dentro de la categoría Aseo, XGBoost logró reducir de forma notable el error respecto a los métodos tradicionales, adaptándose mejor a los cambios de nivel y a los ciclos estacionales. Algo similar ocurrió en el subclúster C, donde en la categoría Cuidado Personal superó de manera consistente a modelos como Prophet y SARIMAX, mostrando una mayor capacidad para responder ante comportamientos irregulares.

Por otro lado, el modelo SARIMAX obtuvo mejores resultados en el subclúster B, caracterizado por una demanda más estable y predecible. En categorías como Aseo, Blanqueador y Lavandería, este modelo logró representar adecuadamente las tendencias y estacionalidades, manteniendo un error bajo y constante en los distintos horizontes de proyección.

El modelo Prophet, en cambio, mostró un mejor desempeño en categorías con estacionalidades múltiples o con cambios estructurales más marcados, como Crema Lavalozza y Cuidado Personal. En estos casos, su flexibilidad permitió capturar los ciclos de mediano y largo plazo, especialmente en horizontes de 8 y 12 semanas, donde otros modelos tendían a sobre-reaccionar ante los picos locales.

Finalmente, las redes neuronales Elman y Jordan ofrecieron resultados competitivos en horizontes medianos y largos, mostrando una buena capacidad de ajuste. Sin embargo, al extender el horizonte de predicción, su error aumentó progresivamente. Aunque fueron eficaces en ciertos subclústeres, en promedio presentaron una menor estabilidad en comparación con XGBoost.

En conjunto, estos hallazgos permitieron consolidar una visión más clara del comportamiento de cada modelo según el tipo de serie y el contexto del subclúster. Los resultados globales se resumen en la Tabla V.

TABLA V  
 MODELOS PREDOMINANTES POR CLÚSTER SEGÚN MENOR RMSE

| Subclúster | Modelo predominante (según menor RMSE) | Comportamiento general   |
|------------|--|--|
| A          | XGBoost                                | Mejor desempeño en series con fluctuaciones moderadas y patrones mixtos. |
| B          | SARIMAX                                | Estabilidad en contextos de demanda regular y alta estacionalidad.       |
| C          | XGBoost                                | Alta precisión en series con picos, ruido y patrones no lineales.        |

Fuente: Elaboración Propia

En síntesis, los resultados demuestran que la elección del modelo óptimo debe ser contextual, ya que la estructura temporal, el nivel de ruido y el grado de variabilidad de la demanda determinan cuál algoritmo ofrece el mejor rendimiento en términos de error absoluto medido mediante RMSE. Esta evidencia valida la pertinencia de la estrategia adoptada, basada en la selección diferenciada del modelo por subclúster, categoría y horizonte de predicción, como mecanismo para capturar la complejidad inherente a la demanda.

En consecuencia, la combinación de métodos clásicos y modelos de Machine Learning permitió equilibrar precisión y estabilidad: mientras SARIMAX resultó más adecuado en contextos de demanda estructurada y predecible, XGBoost ofreció una mayor capacidad para ajustarse a comportamientos no lineales, picos y cambios abruptos. A partir de esta evaluación, se procedió a integrar los modelos seleccionados, incorporando tanto el histórico semanal como las proyecciones multihorizonte (4, 8 y 12 semanas), con el fin de facilitar su análisis y visualización en Power BI, asegurando trazabilidad, coherencia y soporte analítico para la toma de decisiones estratégicas en el proceso de planeación de la demanda.

## 7. VISUALIZACIÓN DEL MODELO

Con el objetivo de integrar los modelos predictivos desarrollados en Python dentro de los procesos operativos de la empresa y facilitar su uso en la planeación de la demanda, se diseñó un dashboard interactivo en Power BI, como se muestra en Fig. 20. Esta herramienta se eligió por su facilidad para conectarse a distintas fuentes de datos, su compatibilidad con el entorno Microsoft utilizado por la compañía y la posibilidad de compartir informes de manera sencilla entre los equipos de trabajo.

Dicho tablero reúne en un mismo espacio tanto la información histórica como las proyecciones generadas por los modelos seleccionados, permitiendo una lectura clara y un análisis comparativo entre lo observado y lo estimado. Su elemento principal es un gráfico de líneas que muestra

la evolución semanal de la demanda en kilogramos (KG) junto con las proyecciones en tres horizontes distintos: 4, 8 y 12 semanas para visualizar el comportamiento real de la demanda y contrastarlo con los resultados esperados, identificando patrones, tendencias o posibles desviaciones.

Para facilitar la exploración de los datos, el dashboard incorpora filtros interactivos que permiten segmentar la información por categoría de producto, subclúster y horizonte de pronóstico, brindando al usuario la posibilidad de ajustar la vista según sus necesidades de análisis. Además, se añadió una tabla detallada que permite revisar los valores proyectados a nivel de cliente, lo cual resulta útil para realizar análisis más específicos o comparar el comportamiento entre diferentes segmentos.

Un aspecto clave del diseño fue la organización jerárquica de la información, que permite navegar de manera intuitiva desde niveles generales hasta detalles particulares. Esto significa que el análisis puede comenzar, por ejemplo, en un nivel de subclúster o categoría de producto, para luego profundizar hacia niveles más detallados, como cliente. Gracias a esta estructura, el tablero facilita tanto la identificación de tendencias globales como el análisis puntual de situaciones específicas

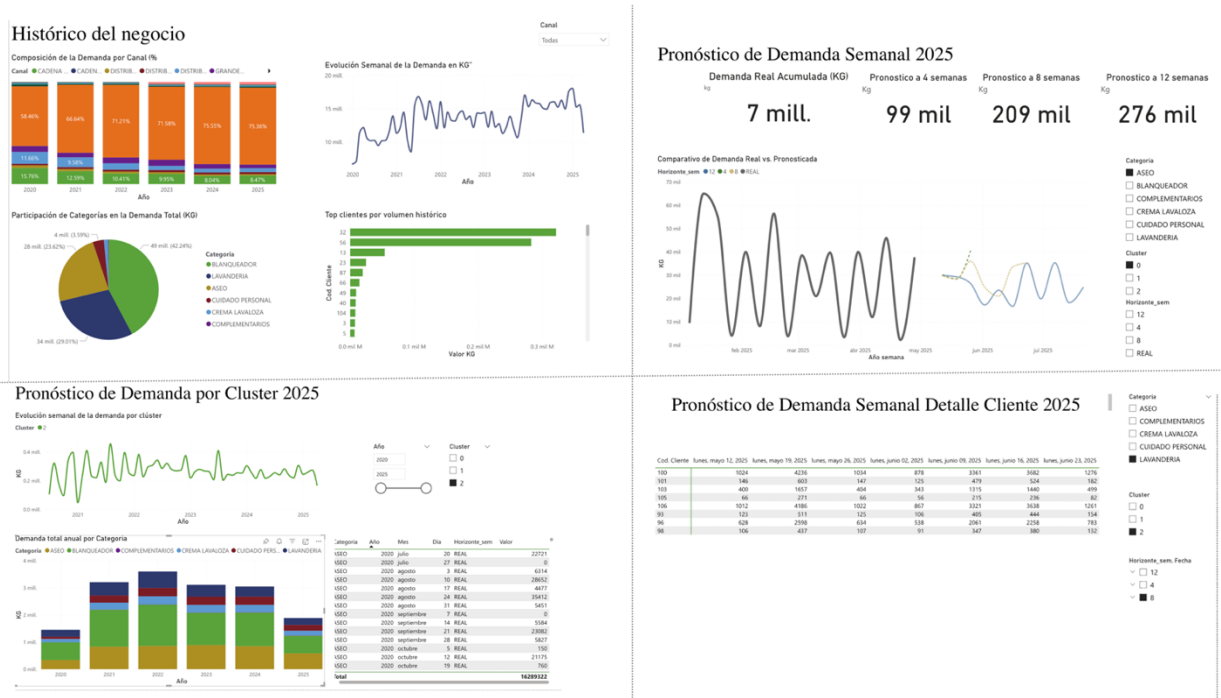


Fig. 20 Dashboard Power BI  
Fuente: Elaboración Propia

Más allá de su función visual, el dashboard también cumple un rol analítico y de validación, ya que permite comparar las proyecciones con el comportamiento real de la demanda y monitorear el desempeño de los modelos a lo largo del tiempo. En consecuencia, esta herramienta se convierte en un apoyo directo para la toma de decisiones en los procesos de planeación, producción, inventarios y gestión comercial, al ofrecer una visión consolidada, actualizada y de fácil interpretación.

En conjunto, el dashboard desarrollado en Power BI no solo organiza la información de forma clara y accesible, sino que además la transforma en un recurso estratégico que ayuda a la empresa a anticiparse a cambios en la demanda, validar sus pronósticos y mejorar la precisión de sus decisiones operativas y tácticas.

Finalmente, para ilustrar la integración de todas las etapas del proyecto, la Fig. 21 presenta el flujo general del proceso de pronóstico, el cual sintetiza los pasos descritos en los capítulos anteriores y evidencia cómo estos se articulan en la visualización del modelo mediante el dashboard. Esta fase constituye el cierre del ciclo analítico, donde los resultados se ponen a disposición de los equipos de planeación y ventas para su análisis y retroalimentación, permitiendo el ajuste continuo de los modelos y el fortalecimiento del sistema de pronóstico.



*Fig. 21 Flujo para pronóstico de demanda*

*Fuente: Elaboración Propia*

## 8. CONCLUSIONES

El proyecto dejó en claro que el problema de la predicción de la demanda no era simplemente un ejercicio de análisis de datos, sino un elemento central para mejorar la planeación y reducir costos dentro de la organización. Lo interesante es que los resultados muestran que la estrategia de segmentar por subclúster y categoría, sumada a la elección del mejor modelo según el horizonte y a la optimización de sus hiperparámetros, permitió capturar la diversidad de comportamientos en el consumo y ajustar el pronóstico a la realidad de cada segmento.

Un hallazgo importante fue la solidez que mostró XGBoost a lo largo del análisis. En la mayoría de los casos se mantuvo como el modelo más confiable, especialmente en los subclústeres A y C, donde logró capturar con precisión la dinámica de categorías como Aseo, Complementarios, Cuidado Personal y Lavandería. En el subclúster A, por ejemplo, no solo se destacó en Aseo, Cuidado Personal y Lavandería, sino que en la categoría de Complementarios incluso superó a modelos más complejos como las redes neuronales tipo Elman. Algo similar se observó en el subclúster C, donde volvió a ofrecer resultados consistentes en diferentes categorías y horizontes, manteniendo un buen control sobre la variabilidad gracias a un ajuste adecuado de sus parámetros.

El caso del subclúster B fue diferente: aquí SARIMAX se comportó mejor en varias categorías (Aseo, Blanqueador, Complementarios y Lavandería), lo que indica que en este grupo la dinámica

temporal resulta más estable y se adapta mejor a modelos clásicos excepto en Crema Lavalozza y Cuidado Personal, modelos más modernos como Prophet, Elman o el propio XGBoost mostraron ventajas en ciertos horizontes, confirmando que cada segmento tiene particularidades que requieren enfoques distintos.

Estos resultados reafirman que no existe un modelo único que sirva para todo el universo de clientes y productos. El rendimiento depende directamente del segmento, la categoría y el horizonte de predicción. No obstante, se identifica un patrón claro: XGBoost, apoyado en la optimización de hiperparámetros, se consolidó como la alternativa más robusta en escenarios de alta variabilidad y ruido, mientras que SARIMAX sigue siendo una opción confiable en series con patrones más regulares y predecibles.

En conclusión, la estrategia de segmentar, optimizar hiperparámetros, evaluar múltiples modelos y quedarse con el más adecuado en cada situación representó un avance significativo en la precisión del pronóstico.

## **9. TRABAJOS FUTUROS**

Actualmente el proyecto ya cuenta con un código automatizado que selecciona el mejor modelo y ajusta sus hiperparámetros para cada combinación de subclúster, categoría y horizonte. El reto ahora es fortalecer este esquema, incorporando algoritmos de última generación y probando técnicas de ensamblado que permitan aprovechar lo mejor de varios modelos al mismo tiempo. Con ello, el proceso de pronóstico ganará en robustez y flexibilidad para adaptarse a cambios en el comportamiento de la demanda.

Un siguiente paso importante es avanzar hacia pronósticos probabilísticos. Pasar de entregar un valor único a trabajar con intervalos de confianza daría a las áreas de negocio un marco más realista para tomar decisiones bajo riesgo, considerando el costo real que implica un quiebre de inventario o un exceso de stock.

También se abre la oportunidad de enriquecer los modelos con variables externas más cercanas al negocio, como promociones, cambios de precios, campañas de mercadeo, datos de sell-out de los clientes e incluso estacionalidades locales. Integrar estas señales de forma estructurada permitiría anticipar mejor las variaciones bruscas que hoy en día resultan difíciles de capturar.

Otra línea de trabajo clave será fortalecer la automatización y el monitoreo del sistema. La adopción de prácticas de automatización de procesos de analítica permitirá orquestar procesos como la limpieza de datos, el reentrenamiento periódico de modelos, el seguimiento de métricas y la detección de desvíos en tiempo real para garantizar que el sistema no solo funcione bien en pruebas, sino que se mantenga confiable y vigente en el largo plazo.

Un aspecto adicional que no debe dejarse de lado es la explicabilidad. Dotar al modelo de herramientas que permitan entender qué factores están impulsando la demanda en cada segmento ayudará a generar confianza en las áreas comerciales y de operaciones, para que sea más que un número, el pronóstico podrá convertirse en un insumo de diálogo y toma de decisiones.

Finalmente, será fundamental medir el impacto del sistema en métricas de negocio, no únicamente en indicadores estadísticos. Evaluar mejoras en la rotación de inventarios, reducción de quiebres o márgenes optimizados permitirá demostrar el verdadero valor de esta solución para la compañía.

En conjunto, estos pasos marcarán la transición de un modelo experimental a un sistema integral de gestión de la demanda, con la capacidad de adaptarse a diferentes clientes, productos y horizontes, y de conectar de manera directa con la rentabilidad y el nivel de servicio al cliente.

## 10. REFERENCIAS BIBLIOGRÁFICAS

- [1] D. Ivanov, «Viable supply chain model: integrating agility, resilience and sustainability perspectives—lessons from and thinking beyond the COVID-19 pandemic,» [En línea]. Available: <https://doi.org/10.1007/s10479-020-03640-6>. [Último acceso: 29 Octubre 2024].
- [2] IBM, «IBM,» IBM, 17 08 2021. [En línea]. Available: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>. [Último acceso: 9 7 2025].
- [3] D. A. V. A. M. Z. B. D. K. B. S. B. T. C. B. R. J. B. J. B. J. E. B. e. a. F. Petropoulos, «Forecasting: theory and practice,» 20 01 2022. [En línea]. Available: <https://doi.org/10.1016/j.ijforecast.2021.11.001>. [Último acceso: 29 10 2024].
- [4] R. O. O. B. a. C. M. K. Douaioui, «Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review,» 26 09 2024. [En línea]. Available: <https://doi.org/10.3390/asi7050093>. [Último acceso: 29 10 2024].
- [5] A. F. A.-G. B. I. S. Z. E. Kamal, «Machine Learning Fusion and Data Analytics Models for Demand Forecasting in the Automotive Industry: A Comparative Study,» 03 06 2023. [En línea]. Available: <https://doi.org/10.54216/FPA.120102>. [Último acceso: 29 10 2024].
- [6] R. Caballar and C. Stryker, «¿Qué es la previsión?,» IBM, 22 07 2024. [En línea]. Available: <https://www.ibm.com/mx-es/think/topics/forecasting#:~:text=IBM&text=Caballar%2C%20Cole%20Stryker,%20es%20la%20previsi%C3%B3n%3Fmodelos%20estad%C3%ADsticos%20para%20generar%20predicciones>. [Último acceso: 21 11 2024].
- [7] IBM, «¿Qué es la planificación de la demanda?,» [En línea]. Available: <https://www.ibm.com/mx-es/topics/demand-planning>. [Último acceso: 21 11 2024].
- [8] Oracle, «¿Qué es la planificación de ventas y operaciones (S&OP)?,» [En línea]. Available: <https://www.oracle.com/co/scm/supply-chain-planning/sales-operations-planning/s-and-op/>. [Último acceso: 20 11 2024].
- [9] D. Osorio Aristizábal, «Comparación de modelos de series temporales ARIMA, SARIMAX y LSTM para la predicción del índice COLCAP,» 08 10 2024. [En línea]. Available: <https://hdl.handle.net/10784/34670>. [Último acceso: 01 11 2024].
- [10] M. Ali, «Tutorial de previsión de series temporales,» datacamp, 11 09 2024. [En línea]. Available: <https://www.datacamp.com/es/tutorial/tutorial-time-series-forecasting>. [Último acceso: 21 11 2024].

- [11] D. C. Montgomery, C. L. Jennings, and M. Kulahci, «Exponential Smoothing for Seasonal Data,» de *Introduction to Time Series Analysis and Forecasting, 2nd ed.*, New Jersey , WILEY, 2015, p. 277.
- [12] B. S. J. Taylor, «Forecasting at Scale,» *The American Statistician*, vol. 72, nº 1, pp. 37-45, 2018.
- [13] D. Hinestroza Ramírez, «El Machine Learning a través de los tiempos, y los aportes a la humanidad,» 2018. [En línea]. Available: <https://hdl.handle.net/10901/17289>. [Último acceso: 17 11 2024].
- [14] B. Bowerman, R. O'Connell, and A. Koehler, *Regression Analysis and Forecasting*, 1st ed., California: Duxbury Press, 2005.
- [15] I. H Witten , E. Frank and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques 2nd edition.*, San Francisco: Morgan Kaufmann Publishers, 2005.
- [16] IBM, «¿Qué es el sobreajuste?,» [En línea]. Available: <https://www.ibm.com/es-es/topics/overfitting>. [Último acceso: 21 11 2024].
- [17] IBM, «¿Qué es XGBoost?,» [En línea]. Available: <https://www.ibm.com/es-es/topics/xgboost>. [Último acceso: 21 11 2024].
- [18] IBM, «¿Qué son las redes neuronales?,» [En línea]. Available: <https://www.ibm.com/es-es/topics/neural-networks>. [Último acceso: 21 11 2024].
- [19] A. Kumar, «MSE vs RMSE vs MAE vs MAPE vs R-Squared: When to Use?,» 2024 08 2024. [En línea]. Available: <https://vitalflux.com/mse-vs-rmse-vs-mae-vs-mape-vs-r-squared-when-to-use/>. [Último acceso: 2024 11 18].
- [20] E. C. S. H. D. & P. M. Keogh, «Segmenting time series: A survey and novel approach,» de *Data Mining in Time Series Databases*, World Scientific, 2004, pp. 1-21.
- [21] R. F. P. & E. I. A. Killick, «Optimal detection of changepoints with a linear computational cost,» *Journal of the American Statistical Association*, pp. 1590-598, 2012.
- [22] S. S. A. S. & W. T. Y. Aghabozorgi, «Time-series clustering – A decade review. Information Systems,» nº 53, pp. 16-38, 2015.
- [23] N. Watanabe, «A k-means method for trends of time series: An application to time series of COVID-19 cases in Japan,» *Japanese Journal of Statistics and Data Science*, nº 5, pp. 303-319, 2022.
- [24] F. F. G. W. G. I. N. A. E. C. Y. & K. E. Petitjean, «Dynamic Time Warping averaging of time series allows faster and more accurate classification,» *ata Mining and Knowledge Discovery*, vol. 30, nº 2, pp. 476-509, 2016.

- [25] D. N. Gujarati y D. C. Porter, «Naturaleza del análisis de regresión,» de *Econometría*, México, McGraw-Hil, 2010, p. 20.
- [26] J. D. S., «Correlación,» *Revista Chilena de Anestesia*, vol. 43, nº 2, pp. 116-121, 2014.
- [27] C. C. a. C. Dehon, «Influence functions of the Spearman and Kendall correlation measures,» *Statistical Methods & Applications*, vol. 19, p. 497–515, 2010.
- [28] R. F. E. a. C. W. J. Granger, «Co-integration and Error Correction: Representation, Estimation, and Testing,» *Econometrica*, vol. 55, nº 2, pp. 251-276, 1987.
- [29] M. Gurnani, Y. Korke, P. Shah, S. Udmale, V. Sambhe, and S. Bhirud, «Forecasting of sales by using fusion of Machine Learning techniques,» 24 02 2017. [En línea]. Available: 10.1109/ICDMAI.2017.8073492. [Último acceso: 11 11 2024].
- [30] Z. H. Kilimci, A. O. Akyuz, M. Uysal, S. Akyokus, M. O. Uysal, B. Atak Bulbul y M. A. Ekmis, «An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain,» 26 05 2019. [En línea]. Available: <https://doi.org/10.1155/2019/9067367>. [Último acceso: 30 10 2024].
- [31] S. Vhatkar and J. Dias, «Oral-care goods sales forecasting using artificial neural network model,» 2016. [En línea]. Available: <https://doi.org/10.1016/j.procs.2016.03.031>. [Último acceso: 25 10 2024].
- [32] S. Cheriyan, S. Ibrahim, S. Mohanan, and S. Treesa, «Intelligent sales prediction using Machine Learning techniques,» 16 08 2018. [En línea]. Available: 10.1109/ICCECOME.2018.8659115. [Último acceso: 06 11 2024].
- [33] X. Dairu and Z. Shilong, «Machine Learning model for sales forecasting by using XGBoost,» 15 01 2021. [En línea]. Available: 10.1109/ICCECE51280.2021.9342304. [Último acceso: 10 11 2024].
- [34] U. C. Morales, «Pronóstico de ventas a corto plazo en el sector de consumo masivo a través de modelos de Machine Learning,» 11 01 2024. [En línea]. Available: <https://www.palermo.edu/ingenieria/Pdf2010/CyT9/02.pdf>. [Último acceso: 18 11 2024].
- [35] L. Escobar R., J. Valdes H., and S. Zapata C., «Redes neuronales artificiales en predicción de series de tiempo. Una aplicación a la industria,» 2009. [En línea]. Available: <https://www.palermo.edu/ingenieria/Pdf2010/CyT9/02.pdf>. [Último acceso: 18 11 2024].
- [36] L. Ji, Y. Zou, K. He, and B. Zhu, «Carbon futures price forecasting based with ARIMA-CNN-LSTM model,» 01 2019. [En línea]. Available: <https://doi.org/10.1016/j.procs.2019.11.254>. [Último acceso: 20 10 2024].

- [37] K. & T. G. Subramanian, « Developing a Forecasting Model for Retailers Based on Customer Segmentation using Data Mining Techniques,» *International Journal of Trend in Scientific Research and Development*, pp. 151-155, 2018.
- [38] N. Dasari, «<https://towardsdatascience.com/>,» 9 7 2025. [En línea]. Available: <https://towardsdatascience.com/time-series-forecasting-made-simple-part-3-1-stl-decomposition-understanding-initial-trend-and-seasonality-prior-to-loess-smoothing>. [Último acceso: 4 8 2025].
- [39] C. & M. M. V. Pires, «Enhancing Water Management: A Comparative Analysis of Time Series Prediction Models for Distributed Water Flow in Supply Networks,» *Water*, vol. 16, nº 13, p. 1827, 2024.
- [40] K. M. y N. Moroke, «Comparative Study of Holt-Winters Triple Exponential Smoothing and Seasonal Arima: Forecasting Short-Term Seasonal Car Sales in South Africa,» *Risk Governance & Control: Financial Markets & Institutions*, vol. 6, nº 1, 2016.
- [41] K. K. K. H. D. D. Emir Zunic, «Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data,» *International Journal of Computer Science & Information Technology*, vol. 12, nº 2, 2020.

## ANEXOS

### Anexo A. Diccionario de Variables

| <b>Nombre atributo</b>          | <b>Tipo Dato</b>   | <b>Descripción</b>   |
|---------------------------------|--------------------|--|
| <b>C.o.</b>                     | Catagórico nominal | Indicador para el centro de operación.<br>1= Planta eje Cafetero<br>5= Planta costa Atlántica<br>10= Planta Centro<br>3= CEDI Antioquia<br>11= CEDI Bogotá |
| <b>Fecha</b>                    | Fecha              | Fecha creación del pedido  |
| <b>Fecha entrega</b>            | Fecha              | Fecha de entrega del pedido  |
| <b>Orden de compra</b>          | Catagórico nominal | Numero orden de compra del cliente   |
| <b>Desc. Condición de pago</b>  | Catagórico ordinal | Días máximo de préstamo  |
| <b>Fecha actualización</b>      | Fecha              | Fecha de última actualización del pedido   |
| <b>Notas docto</b>              | Catagórico nominal | Notas del pedido   |
| <b>U.m.</b>                     | Catagórico nominal | Indicador para diferenciar la unidad del pedido. Puede ser en unidades o cajas.  |
| <b>Cant. Pedida</b>             | Numérico continuo  | Cantidad pedida por el cliente   |
| <b>Cant. Factura</b>            | Numérico continuo  | Cantidad facturada   |
| <b>Nro. Documento</b>           | Catagórico nominal | Numero documento interno   |
| <b>Valor subtotal local</b>     | Numérico continuo  | Valor pedido total   |
| <b>Valor pendiente subtotal</b> | Numérico continuo  | Valor cantidades pendientes por entregar   |
| <b>Cant. Pendiente</b>          | Numérico continuo  | Cantidad pendiente por entregar  |
| <b>Desc. Sucursal despacho</b>  | Catagórico nominal | Nombre punto envío cliente   |
| <b>Id sucursal</b>              | Catagórico nominal | Índice punto de envío cliente  |
| <b>Tipo de maquila cliente</b>  | Catagórico nominal | Unidades de negocio del cliente (Marca Privada, Gran Consumo, Soluciones de Limpieza y Desinfección, Internacionalización)                                 |

|                                     |                    |   |
|-------------------------------------|--------------------|---|
|                                     |                    | (USA)), es la principal segmentación de la organización para clasificar el tipo de cliente.   |
| <b>Lead time</b>                    | Numérico discreto  | Es la cantidad de días pactado con el cliente, en el cual la organización se compromete a atender el pedido desde la recepción del pedido hasta la entrega. |
| <b>Ciudad</b>                       | Categorico nominal | Ciudad del punto envío del cliente  |
| <b>DESC depto.</b>                  | Categorico nominal | Departamento del punto de envío del cliente   |
| <b>Zona</b>                         | Categorico nominal | Zona geográfica del punto de envío del cliente  |
| <b>Pais</b>                         | Categorico nominal | País del punto de envío del cliente   |
| <b>Cliente despacho</b>             | Categorico nominal | Nit del cliente   |
| <b>Razón social cliente factura</b> | Categorico nominal | Razón social del cliente  |
| <b>Nombre vendedor</b>              | Categorico nominal | Nombre del vendedor   |
| <b>Cant. Disponible</b>             | Numérico continuo  | Cantidad inventario existente de la referencia solicitada de acuerdo con la bodega  |
| <b>Canal</b>                        | Categorico nominal | Clasificación de los clientes (Grandes cadenas, Mayoristas, supermercados, distribuidor aliado, etc.) Contiene 25 tipos de clasificaciones diferentes.      |
| <b>Cant. Comprom.</b>               | Numérico continuo  | Cantidad inventario reservado para el pedido  |
| <b>Bodega</b>                       | Categorico nominal | Bodega de almacenamiento de la cual se descontará el inventario   |
| <b>Ítem</b>                         | Categorico nominal | Referencia/Sku del producto   |
| <b>Desc. Ítem</b>                   | Categorico nominal | Nombre del producto   |
| <b>Notas ítem</b>                   | Categorico nominal | Nombre del producto detallado   |
| <b>Pedida en KG</b>                 | Numérico continuo  | Peso en kilogramos de la cantidad pedida  |
| <b>Pendiente en KG</b>              | Numérico continuo  | Peso en kilogramos de la cantidad pendiente   |

|                                   |                    |   |
|-----------------------------------|--------------------|---|
| <b>Factura en KG</b>              | Numérico continuo  | Peso en kilogramos de la cantidad facturada   |
| <b>Pedida en LT</b>               | Numérico continuo  | Volumen en litros de la cantidad pedida   |
| <b>Pendiente en LT</b>            | Numérico continuo  | Volumen en litros de la cantidad pendiente  |
| <b>Factor U.M. emp.</b>           | Numérico discreto  | Cantidad de unidades que conforman una caja   |
| <b>Categoría</b>                  | Categórico nominal | Categorización del producto de acuerdo con su tipo (Aseo, Blanqueador, Cuidado Personal, Lavandería, Crema Lavalozza, Complementarios, Insumos, NA)   |
| <b>Familia</b>                    | Categórico nominal | Detalle de la categoría por tipo producto (Suavizante, Jabón líquido, Jabón de manos, Limpia pisos, Lavalozza Liquido, etc.). Hay 52 diferentes tipos   |
| <b>Tamaño</b>                     | Numérico continuo  | Volumen de la presentación de producto en mililitros.   |
| <b>Maquila</b>                    | Categórico nominal | Marca del producto  |
| <b>Código barra principal</b>     | Categórico nominal | Número del código de barras   |
| <b>Venta perdida</b>              | Categórico nominal | Descripción del motivo por el cual no fue posible atender el pedido en caso de que suceda.  |
| <b>Estado docto.</b>              | Categórico ordinal | Estado del pedido:<br>En elaboración= se encuentra en proceso de creación<br>Aprobado=El pedido esta pendiente por atenderse completa o parcialmente<br>Retenido= El pedido presenta una novedad<br>Cumplido=El pedido ya fue entregado/cerrado |
| <b>Fecha factura desde pedido</b> | Fecha              | Fecha ultima factura asociado al pedido   |
| <b>Factura desde pedido</b>       | Categórico nominal | Numero de la última factura asociada al pedido  |
| <b>Cumplimiento</b>               | Numérico           | % cum cantidades facturadas vs pedidas  |
| <b>F_desc_motivos_otros</b>       | Categórico nominal | Motivo de cierre de pedidos masivos por novedades en el sistema   |

|                                  |                    |  |
|----------------------------------|--------------------|--|
| <b>Fecha retenido manual</b>     | Fecha              | Fecha de la última vez que se retiene un pedido  |
| <b>Fecha aprobación retenido</b> | Fecha              | Fecha del momento en el que pasa de retenido a aprobado  |
| <b>Segmento</b>                  | Categorico nominal | Detalle para la unidad de negocio “gran consumo” (G-30, Tradicional, G-30 internacional)   |
| <b>Tipo cliente</b>              | Categorico nominal | Nacional o internacional   |
| <b>Tipo_maquila_Cliente</b>      | Categorico nominal | Retail, Hard Discount, Linea Institucional   |
| <b>Reprogramado</b>              | Categorico binario | (Si, No) para identificar si el pedido está asociado a un pedido anterior que tuvo que ser cerrado para ser entregado en otra fecha. |

Fuente: Elaboración Propia

#### *Anexo B. Resultados evaluación mejores modelos*

| <b>Subclúster</b> | <b>Categoría</b> | <b>Horizonte_sem</b> | <b>Mejor_Modelo</b> | <b>RMSE</b> |
|-------------------|------------------|----------------------|---------------------|-------------|
| 0                 | ASEO             | 4                    | XGBOOST             | 4902.95     |
| 0                 | ASEO             | 8                    | XGBOOST             | 6295.65     |
| 0                 | ASEO             | 12                   | XGBOOST             | 8800.33     |
| 0                 | BLANQUEADOR      | 4                    | XGBOOST             | 21982.35    |
| 0                 | BLANQUEADOR      | 8                    | SARIMAX             | 11543.6     |
| 0                 | BLANQUEADOR      | 12                   | XGBOOST             | 18564.27    |
| 0                 | COMPLEMENTARIOS  | 4                    | XGBOOST             | 3683.9      |
| 0                 | COMPLEMENTARIOS  | 8                    | XGBOOST             | 2738.27     |
| 0                 | COMPLEMENTARIOS  | 12                   | ELMAN               | 1875.92     |
| 0                 | CREMA LAVALOZA   | 4                    | XGBOOST             | 3944.3      |
| 0                 | CREMA LAVALOZA   | 8                    | XGBOOST             | 4465.64     |

|   |                  |    |         |           |
|---|------------------|----|---------|-----------|
| 0 | CREMA LAVALOZA   | 12 | XGBOOST | 6198.49   |
| 0 | CUIDADO PERSONAL | 4  | XGBOOST | 2195.49   |
| 0 | CUIDADO PERSONAL | 8  | XGBOOST | 3600.25   |
| 0 | CUIDADO PERSONAL | 12 | XGBOOST | 4302.92   |
| 0 | LAVANDERIA       | 4  | XGBOOST | 8792.6    |
| 0 | LAVANDERIA       | 8  | XGBOOST | 10853.79  |
| 0 | LAVANDERIA       | 12 | XGBOOST | 11131.69  |
| 1 | ASEO             | 4  | SARIMAX | 130900.66 |
| 1 | ASEO             | 8  | SARIMAX | 126621.5  |
| 1 | ASEO             | 12 | XGBOOST | 195916.92 |
| 1 | BLANQUEADOR      | 4  | SARIMAX | 293732.28 |
| 1 | BLANQUEADOR      | 8  | SARIMAX | 219761.21 |
| 1 | BLANQUEADOR      | 12 | SARIMAX | 202412.9  |
| 1 | COMPLEMENTARIOS  | 4  | SARIMAX | 712.09    |
| 1 | COMPLEMENTARIOS  | 8  | SARIMAX | 941.98    |
| 1 | COMPLEMENTARIOS  | 12 | XGBOOST | 2355.92   |
| 1 | CREMA LAVALOZA   | 4  | PROPHET | 11660.84  |
| 1 | CREMA LAVALOZA   | 8  | SARIMAX | 17748.2   |
| 1 | CREMA LAVALOZA   | 12 | XGBOOST | 21349.52  |
| 1 | CUIDADO PERSONAL | 4  | SARIMAX | 8325.79   |
| 1 | CUIDADO PERSONAL | 8  | ELMAN   | 22695.76  |
| 1 | CUIDADO PERSONAL | 12 | XGBOOST | 49280.03  |

|   |                  |    |         |           |
|---|------------------|----|---------|-----------|
| 1 | LAVANDERIA       | 4  | SARIMAX | 143000.66 |
| 1 | LAVANDERIA       | 8  | SARIMAX | 140876.38 |
| 1 | LAVANDERIA       | 12 | SARIMAX | 153198    |
| 2 | ASEO             | 4  | XGBOOST | 6424.45   |
| 2 | ASEO             | 8  | XGBOOST | 9205.26   |
| 2 | ASEO             | 12 | XGBOOST | 9769.29   |
| 2 | BLANQUEADOR      | 4  | SARIMAX | 8523.38   |
| 2 | BLANQUEADOR      | 8  | XGBOOST | 12253.8   |
| 2 | BLANQUEADOR      | 12 | XGBOOST | 16975.15  |
| 2 | COMPLEMENTARIOS  | 4  | XGBOOST | 102.87    |
| 2 | COMPLEMENTARIOS  | 8  | XGBOOST | 148.78    |
| 2 | COMPLEMENTARIOS  | 12 | HW      | 108.15    |
| 2 | CREMA LAVALOZA   | 4  | SARIMAX | 1849.85   |
| 2 | CREMA LAVALOZA   | 8  | XGBOOST | 3570.77   |
| 2 | CREMA LAVALOZA   | 12 | XGBOOST | 4061.15   |
| 2 | CUIDADO PERSONAL | 4  | XGBOOST | 2211.93   |
| 2 | CUIDADO PERSONAL | 8  | XGBOOST | 4441.45   |
| 2 | CUIDADO PERSONAL | 12 | XGBOOST | 4609.05   |
| 2 | LAVANDERIA       | 4  | XGBOOST | 2554.16   |
| 2 | LAVANDERIA       | 8  | XGBOOST | 5767.14   |
| 2 | LAVANDERIA       | 12 | XGBOOST | 5939.79   |

Fuente: Elaboración Propia