



Pontificia Universidad
JAVERIANA
Cali

**Modelo espacio temporal para la predicción de la demanda de emergencias médicas en
Bogotá**

Milena Andrea Riaño Sepúlveda – Código 8986485
Wendy Dayanna Mendoza Bautista – Código 8985604
Andrea Estefanía Amaya Garzón – Código 8986238

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

Director
Juan Camilo Paz Roa, PhD.

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, DICIEMBRE 5 DE 2024

TABLA DE CONTENIDO

1	DEFINICIÓN DEL PROBLEMA	2
1.1	PLANTEAMIENTO DEL PROBLEMA.....	2
1.2	FORMULACIÓN DEL PROBLEMA.....	3
2	OBJETIVOS DEL PROYECTO	4
2.1	OBJETIVO GENERAL	4
2.2	OBJETIVOS ESPECÍFICOS	4
3	MARCO TEÓRICO Y ANTECEDENTES.....	5
3.1	MARCO TEÓRICO	5
3.1.1	Contexto de los Sistemas de Emergencias Médicas.	6
3.1.2	Demanda de Servicios de Emergencias Médicas en Bogotá.	7
3.1.3	Modelos de pronóstico de demanda espaciotemporal.	7
3.1.4	Modelos de aprendizaje automático.	7
3.2	ANTECEDENTES.....	9
4	IDENTIFICACIÓN Y SELECCIÓN DE VARIABLES CLAVE	14
4.1	FUENTES DE DATOS.....	14
4.2	SELECCIÓN Y PROCESAMIENTO DATOS	14
4.2.1	Procesamiento de los Datos.....	15
4.2.2	Fuente Principal	18
4.2.3	Datos Secundarios y Complementarios	20
4.3	ANÁLISIS EXPLORATORIO DE LA BASE DE DATOS.....	22
5	EVALUACIÓN DE MODELOS PREDICTIVOS.....	33
5.1	SELECCIÓN DE LA UNIDAD ESPACIAL Y TEMPORAL	34
5.2	DIVISIÓN DE DATOS PARA ENTRENAMIENTO Y PRUEBA	34
5.3	METODOLOGÍA DE EVALUACIÓN	34
5.4	SELECCIÓN DE MODELOS.....	36
5.4.1	Modelos Tradicionales	36
5.4.2	Modelos De Aprendizaje Automático	37
5.5	COMPARACIÓN DE RESULTADOS	38
5.6	EXPLORACIÓN DE OTROS MODELOS	39
5.7	CAMBIO DE UNIDAD ESPACIAL: DE UPZ A UPL.....	40
5.7.1	Evaluación de Modelos Tradiciones	42
5.7.2	Evaluación de Modelos de Aprendizaje Automático	43
5.7.3	Comparación Final de Resultados	43
6	IMPLEMENTACIÓN DEL MODELO PREDICTIVO ESPACIOTEMPORAL.....	45
6.1	CRITERIOS DE SELECCIÓN	45
6.2	VARIABLES FINALES.....	46
6.3	IMPLEMENTACION DEL MODELO XGBOOST REGRESSOR	46
6.4	DESPLIEGUE DEL MODELO.....	49
7	DESARROLLO DEL TABLERO DE CONTROL INTERACTIVO	54
7.1	ESTRUCTURA DEL TABLERO	54
7.1.1	Hoja No. 1: Visualización de Incidentes por UPL y Parámetros Temporales	54

7.1.2	Hoja No. 2: Comparación de Datos históricos y las predicciones del modelo.....	57
7.2	FUNCIONALIDADES DINÁMICAS DEL TABLERO	58
8	CONCLUSIONES Y TRABAJOS FUTUROS.....	61
8.1	CONCLUSIONES.....	61
8.2	TRABAJOS FUTUROS	62
9	REFERENCIAS BIBLIOGRÁFICAS	63

LISTA DE FIGURAS

Figura 1.	Clasificación de los modelos de pronóstico por tipo.....	6
Figura 2.	Flujo del modelo de aprendizaje automático.....	8
Figura 3.	Mapa de Bogotá por UPZ y UPL.....	21
Figura 4.	Proporción de incidentes con despacho de ambulancia.....	14
Figura 5.	Flujo de trabajo para ejecución del proyecto.....	16
Figura 6.	Porcentaje de Incidentes por Localidad.	22
Figura 7.	Mapa de Cantidad de Incidentes por UPZ.....	23
Figura 8.	Distribución por Tipo de Incidente.....	24
Figura 9.	Distribución por Prioridad Inicial y Final.	24
Figura 10.	Tendencia del Promedio Diario de Incidentes por Mes y Año.	25
Figura 11.	Tendencia de Incidentes por Mes (2022-2024).....	26
Figura 12.	Distribución de Servicios de Emergencias por Hora.....	27
Figura 13.	Distribución de incidentes por hora y día de la semana.	28
Figura 14.	Total de Incidentes por Día de la Semana y Período del Día.....	29
Figura 15.	Mapa de distribución Demográfica por Grupos Etarios en UPZ.....	30
Figura 16.	Mapa de correlación entre las variables.	31
Figura 17.	Flujo metodológico aplicado	33
Figura 18.	Cantidad de Incidentes por UPL.	42
Figura 19.	Comparación de las métricas de desempeño de los modelos.	44
Figura 20.	Datos reales Vs pronostico del modelo.....	47
Figura 21.	Importancia de variables.....	48
Figura 22.	Mapa de incidentes – Tablero de control.....	55
Figura 23.	Incidentes por fecha – Tablero de control.....	55
Figura 24.	Incidentes por Hora del Día – Tablero de Control.	56
Figura 25.	Incidentes por UPL – Tablero de Control.....	56
Figura 26.	Comparativo Incidentes - Tablero de control.	57
Figura 27.	Incidentes Reales Vs Predicción - Tablero de Control.....	57
Figura 28.	Mapa predicción - Tablero de Control.....	58
Figura 29.	Vista descriptiva – Tablero de control.	59
Figura 30.	Vista predictiva – Tablero de control.....	60

LISTA DE TABLAS

Tabla 1. Matriz de Variables Clave para Modelos de Predicción.....	17
Tabla 2. Matriz de variables del CRUE.	19
Tabla 3. Resultados ANOVA.	32
Tabla 4. Comparación de resultados	38
Tabla 5. Desempeño de los Modelos	43
Tabla 6. Variables del modelo final.....	46

INTRODUCCIÓN

Bogotá, una ciudad que alberga a más de siete millones de habitantes, enfrenta importantes desafíos en la eficiencia de su sistema de servicios de emergencia. Factores como la congestión vehicular crónica y la creciente demanda de servicios médicos de urgencia provocan tiempos de respuesta que superan los estándares internacionales recomendados.

En este contexto, es crucial identificar soluciones que permitan anticipar la demanda de emergencias y distribuir los recursos de manera más eficiente. Predecir la cantidad de servicios de ambulancias que se necesitan en un intervalo de tiempo y una zona específica de la ciudad, ayuda con la gestión eficaz de los automotores, es por eso que se desarrolló un modelo predictivo espaciotemporal que permite prever la demanda servicios de emergencias médicas en un intervalo de 12 horas y a nivel de Unidad de Planeamiento Zonal (UPZ), una división geográfica de Bogotá utilizada para planificar el desarrollo urbano y la gestión de recursos en la ciudad como la estratificación, el comercio, el uso del suelo, las viviendas y las condiciones de vida de sus habitantes [1]. Sin embargo, durante el desarrollo del proyecto, se identificó que los modelos con datos ajustados a la Unidad de Planeamiento Local (UPL) —una unidad espacial más amplia que la UPZ— ofrecían un mejor desempeño predictivo. Por ello, se adoptó la UPL como la unidad espacial de análisis principal, optimizando la precisión de la demanda.

La metodología del proyecto combina la ciencia de datos con un enfoque riguroso en el aprendizaje automático y la estadística. Partiendo del entendimiento del problema, la consolidación de datos primarios y secundarios, y la evaluación de algoritmos avanzados como Redes Neuronales, Gradient Boosting Machine y Random Forest, junto a modelos estadísticos tradicionales como la Distribución Binomial Negativa. Estas técnicas, ampliamente utilizadas para modelar datos espaciotemporales, destacan por su capacidad para capturar la variabilidad tanto en el tiempo como en el espacio.

La relevancia de este trabajo radica en la posibilidad de mejorar la capacidad del sistema de emergencias para prever los patrones de demanda y, por tanto, optimizar el uso de los recursos disponibles. Al poder anticipar de manera más precisa la demanda en distintos momentos y zonas de la ciudad, se busca reducir los tiempos de respuesta y mejorar la eficiencia del sistema, permitiendo así una atención más rápida y eficaz para la población. Este proyecto, al integrar algoritmos de aprendizaje automático con métodos estadísticos tradicionales, contribuye a una visión moderna y completa para enfrentar el reto de predecir la demanda de emergencias en una ciudad tan grande y compleja como Bogotá. Además, busca ofrecer una base sólida para la toma de decisiones informadas por parte de las autoridades locales, con el fin de mejorar tanto la planificación como la distribución de los recursos disponibles en el sistema de emergencias.

1 DEFINICIÓN DEL PROBLEMA

1.1 PLANTEAMIENTO DEL PROBLEMA

La capital colombiana es una de las ciudades más pobladas del país, con una población fija de aproximadamente 7.9 millones de habitantes. Además, Bogotá experimenta un notable incremento en su actividad diaria debido a que se registran más de 13 millones de viajes diarios, muchos de ellos generados por población flotante debido a actividades económicas, educativas y de transporte [2]. Esta dinámica genera una alta demanda de servicios médicos de emergencia, los cuales son atendidos por un parque automotor de emergencia compuesto por 873 ambulancias entre públicas y privadas. De estas, 645 son de soporte vital básico y 228 son unidades de soporte vital avanzado. La asignación y movilización de estos recursos críticos está a cargo del Centro Regulador de Urgencias y Emergencias (CRUE), que gestiona la línea de emergencia 123 para coordinar las respuestas ante incidentes médicos urgentes. Sin embargo, pese a los esfuerzos por optimizar estos servicios, el sistema enfrenta retos significativos.

Uno de los principales retos es la congestión vehicular, que según estudios de INRIX (2022) coloca a Bogotá como una de las ciudades con peor tráfico del mundo, con un promedio de 122 horas al año perdidas en embotellamientos por habitante [3]. Además, la alta demanda de servicios médicos de emergencia a menudo pone a prueba la capacidad de respuesta del sistema, lo que puede llevar a tiempos de espera prolongados y a la saturación de los servicios de urgencias en hospitales y clínicas, resultando en retrasos que pueden ser determinantes en el pronóstico del paciente.

Este factor impacta directamente en los tiempos de respuesta de las ambulancias, que superan en muchos casos el estándar internacional de 8 minutos recomendado por la Organización Mundial de la Salud (OMS). En la Ciudad el tiempo promedio de respuesta oscila entre 15 y 25 minutos de acuerdo con el CRUE [4], y este tiempo puede ser mayor si se toman en cuenta factores como horas pico, debido al aumento en el tráfico, o zonas de la ciudad de difícil acceso. En una denuncia radicada recientemente por la concejal Diana Diago [5], se menciona que, para servicios con prioridad crítica y alta, el tiempo de respuesta promedio es de 31 minutos, mientras que para incidentes clasificados con prioridad media la espera promedio es de 1 hora y 6 minutos. Estos tiempos prolongados no solo comprometen la atención inmediata, sino que también repercuten en la saturación de los servicios hospitalarios, ya que los retrasos en el transporte de pacientes impactan en la capacidad de los hospitales para liberar recursos, generando un efecto acumulativo.

Para mejorar la eficacia del sistema, se han implementado tecnologías de información que permiten el rastreo y la asignación dinámica de recursos. Sin embargo, sigue habiendo un amplio margen de mejora en términos de infraestructura, integración de datos y protocolos de actuación para responder de manera más eficiente y oportuna a las emergencias médicas en la capital colombiana.

Este panorama presenta un escenario donde la implementación de modelos de pronóstico de demanda espaciotemporal, basados en técnicas avanzadas de análisis de datos, se vuelve esencial para mejorar la eficiencia del sistema. Dicha implementación podría significar un cambio paradigmático en la gestión de los servicios de emergencia, apuntando a una asignación diaria de recursos más predictiva y estratégica para cada zona de la ciudad, que no solo atienda la emergencia actual, sino que prepare la respuesta para las próximas, mejorando así la calidad del servicio y la oportuna atención a los pacientes.

1.2 FORMULACIÓN DEL PROBLEMA

¿Cómo se puede pronosticar la demanda diaria de servicios de emergencias médicas en la ciudad por zonas, utilizando modelos espacio temporales?

- ¿Cómo identificar y seleccionar los datos necesarios para construir una base de datos robusta, que refleje adecuadamente el impacto de factores como condiciones meteorológicas, eventos especiales y la densidad del tráfico en los patrones de demanda temporal y espacial de los servicios de emergencias médicas en Bogotá?
- ¿Cuáles son los modelos predictivos que pueden ser empleados para que aporten adecuadamente a la solución del problema?
- ¿Cuál de los modelos probados muestra un mejor desempeño en cuanto a la métrica de comparación?
- ¿De qué forma se pueden representar efectivamente los resultados para garantizar una comprensión clara y una lectura intuitiva de los pronósticos del modelo?

2 OBJETIVOS DEL PROYECTO

2.1 OBJETIVO GENERAL

Diseñar un modelo predictivo espaciotemporal, orientado a prever la demanda de servicios de emergencias médicas diarias en Bogotá a nivel de UPZ; con esto apoyar la toma de decisiones de la Secretaría de Salud de Bogotá en cuanto a estrategias de prevención e intervención y así mejorar la eficiencia operativa del sistema de emergencias médicas.

2.2 OBJETIVOS ESPECÍFICOS

- I. Identificar las variables que impactan la demanda de emergencias médicas de Bogotá mediante la consulta de fuentes primarias y secundarias que permita la construcción de la base de datos y su análisis exploratorio.
- II. Evaluar un conjunto de modelos seleccionados de aprendizaje automático más relevantes en la literatura para identificar aquel que brinda mejor desempeño en la predicción de emergencias médicas en Bogotá.
- III. Implementar un modelo predictivo basado en aprendizaje automático, para determinar su aplicabilidad y rendimiento en la proyección de la demanda de servicios de emergencias médicas en las UPZ de Bogotá.
- IV. Elaborar un tablero de control interactivo que muestre los resultados sobre los pronósticos del modelo, facilitando la interpretación de estos para una toma de decisiones rápida.

3 MARCO TEÓRICO Y ANTECEDENTES

3.1 MARCO TEÓRICO

En esta sección, se establecen las bases contextuales y analíticas sobre las cuales se sustenta este proyecto. Inicialmente, se examinó el contexto del sistema de emergencias médicas en la ciudad de Bogotá, abordando a su vez, el comportamiento de la demanda de estos servicios a nivel nacional. De igual forma, se exploraron las teorías y aplicaciones de modelos de pronóstico de demanda de servicios de emergencias médicas con un enfoque espaciotemporal para zonas urbanas.

En primer lugar, es importante definir dos conceptos que se desarrollan a lo largo de la investigación y que permiten tener un mejor entendimiento de esta, tales como demanda, que en términos de servicios de salud está definida como “La población que, de acuerdo con una necesidad de atención, requiere o potencialmente revela un consumo de servicios de salud” [6]. Ahora bien, si se pone en contexto de emergencia médica, se refiere a la cantidad de solicitudes o necesidades de atención médica inmediata en situaciones críticas o urgentes como accidentes, enfermedades graves, lesiones repentinas o cualquier situación que requiera atención médica de forma rápida para preservar la vida del paciente.

Existen varios factores que pueden influir en la demanda de servicios de emergencias médicas:

- Tráfico: tiene diferentes influencias como accidentes de tránsito, puesto que, un aumento de tránsito está asociado a un mayor riesgo de accidentes, asimismo puede generar retrasos en los tiempos de respuesta ya que pueden ser más largos de lo esperado, representando un impacto negativo pues la capacidad oportuna de respuesta es crucial en una emergencia médica.
- Clima: Las condiciones climáticas juegan un papel importante en el despacho de ambulancias, ya que pueden influir tanto en la frecuencia de incidentes como en los tiempos de respuesta [7]. Durante lluvias intensas, es más probable que ocurran accidentes de tráfico o incidentes de salud. Además, el mal tiempo puede dificultar el desplazamiento de las ambulancias, incrementando los tiempos de llegada al lugar del incidente debido a calles congestionadas, inundadas o condiciones de poca visibilidad.
- Eventos naturales (catástrofes): puede tener un impacto significativo de acuerdo con la magnitud del evento, en el mismo se puede generar desplazamientos, lesiones físicas, heridas traumáticas, entre otros, colapsando el sistema de emergencias de la ciudad.
- Eventos especiales: implican la congregación de grandes grupos de personas en un solo

lugar, pueden abarcar desde conciertos y eventos deportivos, hasta manifestaciones y festivales, siendo un generador de emergencias o de retraso en la movilidad.

- Densidad poblacional: las áreas urbanas densamente pobladas pueden experimentar una mayor demanda de servicios de emergencias médicas debido a la alta probabilidad de accidentes, lesiones y eventos médicos agudos en la población.
- Factores socioeconómicos: las poblaciones con limitado acceso a servicios de salud preventiva tienden experimentar mayores tasas de emergencias médicas.

Por otra parte, el concepto de pronóstico hace referencia a “Una estimación cuantitativa o cualitativa de uno o varios factores (variables) que conforman un evento futuro, con base en información actual o del pasado” [8]. En el contexto de los servicios de emergencias médicas, el pronóstico de demanda hace referencia a la estimación anticipada de la cantidad y tipo de incidentes o emergencias médicas que es probable que ocurran en un área geográfica específica durante un período de tiempo determinado. En la Figura 1, (adaptado de capítulo XIII modelos de pronósticos [9]), se muestra la clasificación de los modelos de pronóstico por tipo, estos tipos de pronósticos son esenciales para la planificación y asignación eficiente de recursos en servicios de emergencia, como ambulancias, personal médico y equipos especializados, entre otros.

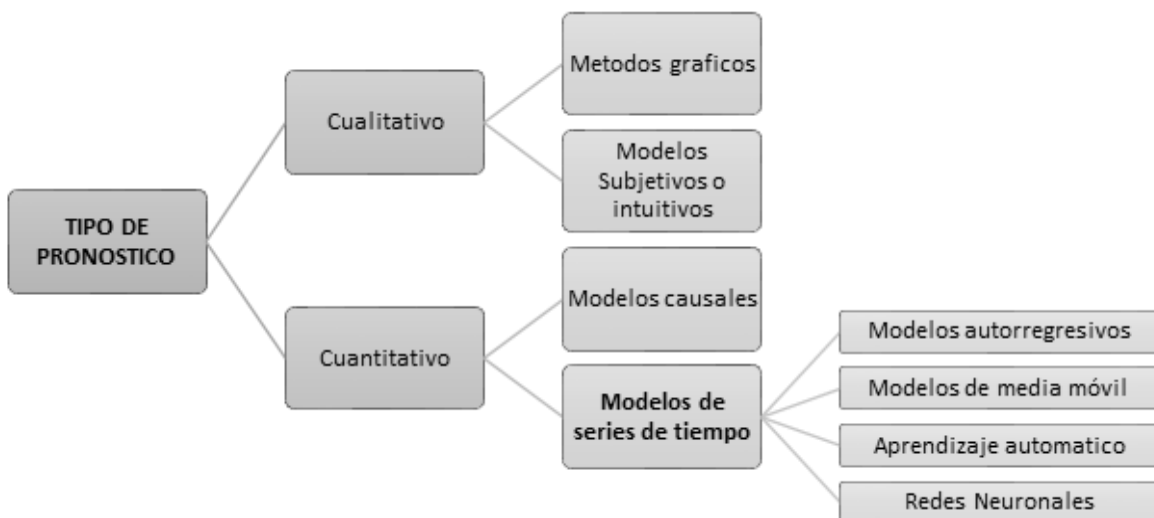


Figura 1. Clasificación de los modelos de pronóstico por tipo.

3.1.1 Contexto de los Sistemas de Emergencias Médicas.

Una emergencia médica se define como “Aquella situación que afecta o pone en peligro la salud o la vida de una o de muchas personas” [10]. Esta situación puede darse en la calle (público) o en el hogar (privado), y es ahí cuando se inicia la asistencia por parte de los sistemas asignados,

para el caso de Bogotá el NUSE (número único de seguridad y emergencias) o la línea telefónica 123, que gestiona emergencias tanto de Policía, Movilidad, Bomberos, Gestión de Riesgos, Secretaría de Salud. Esta última, representada por la Dirección de Urgencias y Emergencias médicas, es la encargada de coordinar la atención prehospitalaria, asistencia médica y el transporte de enfermos o accidentados a hospitales y clínicas; esto a través del CRUE que es la agencia encargada de recibir las solicitudes y organizar la prestación del servicio.

Teniendo en cuenta lo anterior, cuando se presenta una llamada o solicitud de atención por la comunidad donde hay una situación de emergencia, esta ingresa al Centro de Comando Control Comunicaciones y Cómputo (C4) a través la Línea de Emergencias 123; desde allí, es transferida al Centro Operativo de la Subdirección CRUE, con el fin de gestionar la respuesta que se requiera según el incidente presentado [11]. Los incidentes se clasifican en dos tipos:

- Incidentes con despacho: corresponden a aquellos incidentes que tienen asignación de un vehículo de emergencias dirigido al sitio de ocurrencia del incidente.
- Incidentes sin despacho: corresponden a incidentes cuya respuesta es gestionada a través de asesoría telefónica por los Técnicos Asistenciales en Regulación Médica con el direccionamiento técnico del médico regulado.

3.1.2 Demanda de Servicios de Emergencias Médicas en Bogotá.

Datos reportados por la Secretaría Distrital de Salud en los Boletines de Indicadores periódicos, resaltan que para el año 2021 se presentaron alrededor de 145.000 casos que involucraron una situación de emergencia médica en Bogotá; De estos, el 53% terminaron en la atención médica del paciente, representando la atención con traslado el 40% y la atención sin traslado el 13%. El 47% restante de casos representa los despachos fallidos, es decir, aquellos donde se asigna el recurso, pero luego es cancelado por el paciente o cuando al llegar al evento ya no corresponde a desplazamiento hacia algún centro de atención [11].

3.1.3 Modelos de pronóstico de demanda espaciotemporal.

Existen diversos modelos de pronóstico de demanda, por lo que la elección del modelo depende de varios factores, como la disponibilidad de datos, la naturaleza de la demanda, el horizonte temporal y otros elementos contextuales, por tanto, es fundamental realizar una revisión de los modelos que más se ajustan a la necesidad del proyecto.

3.1.4 Modelos de aprendizaje automático.

Puede definirse como “una rama de la inteligencia artificial cuyo fin es dotar a una máquina, a través de algoritmos, de la capacidad de entrenar y aprender a partir de datos” [12]. Este tipo

de modelo busca desarrollar algoritmos que le permitan a la maquina aprender a reconocer determinados patrones imitando la capacidad que tienen las personas de aprender mediante ejemplos. Estos modelos utilizan datos como entrada y ajustan sus parámetros internos para realizar tareas específicas sin estar programados explícitamente para cada una de ellas, lo que permite la solución de tareas como la predicción para nuevos casos. Además, otra bondad de estos algoritmos es la propiedad de corregir con nuevos entrenamientos los errores que el programa pueda cometer en pasos anteriores, con lo cual mejora su rendimiento. La Figura 2 muestra el flujo utilizado para el desarrollo de los modelos de aprendizaje automático.

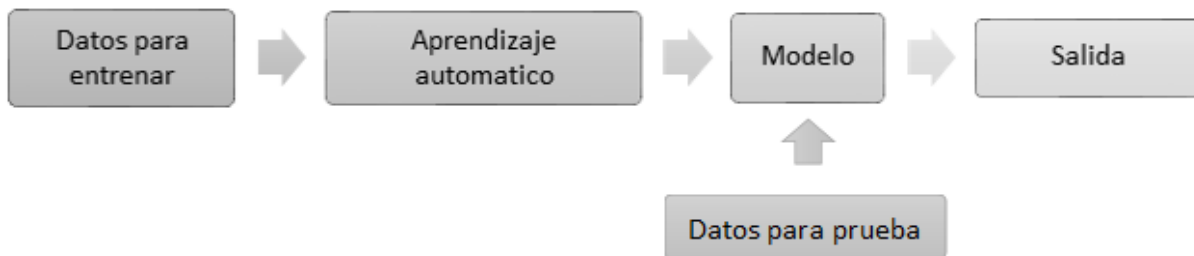


Figura 2. Flujo del modelo de aprendizaje automático.

3.1.4.1 Modelos de aprendizaje profundo

Los modelos de aprendizaje profundo (Deep Learning), son definidos por Karthik R. y Abhishek S. (2019) como “una subárea del aprendizaje automático que basa sus raíces en modelos de redes neuronales” [13]. Dentro de las arquitecturas más comunes están las redes neuronales multicapa (MLP), redes neuronales recurrentes (RNN) y las redes neuronales convolucionales (CNN), cada una con sus particularidades y casos de uso, en donde su principal característica es la identificación de patrones no lineales capaces de aprender de representaciones de alto nivel de abstracción como texto, imágenes, video o audio. Por lo general, su estructura más simple se compone de: capa de entrada, capas ocultas, y capa de salida [14].

En el contexto de pronósticos de demanda, las redes neuronales pueden ser aplicadas para prever patrones y comportamientos futuros en datos históricos de demanda. A continuación, se describen algunas arquitecturas de las redes neuronales más usadas para este tipo de tarea, teniendo en cuenta la investigación realizada por Halima B., Ilham S., Said A. [15].

- **Multilayer Perceptron (MLP):** Para esta arquitectura de redes neuronales multicapa, en primer lugar, la capa de entrada envía los datos de entrada a los nodos ocultos de la capa oculta, en esta fase se utilizan múltiples capas sucesivas que se alimentan unas a otras en dirección, desde la capa de entrada a la capa de salida. Adicionalmente, en sus capas ocultas se pueden definir diferentes funciones de activación para modelar la no

linealidad de los datos dependiendo del problema a solucionar, desde tareas de clasificación, diagnóstico, predicción y reconocimiento de imágenes. Para el caso, este tipo de redes pueden resultar bastante útiles al momento de identificar los complejos patrones de demanda de emergencias médicas.

- Redes Neuronales Recurrentes (RNN): Son un tipo de arquitectura de redes neuronales diseñadas para trabajar con datos secuenciales o temporales, tiene conexiones recurrentes que permiten recordar información anterior a través del tiempo.
 - Long Short-Term Memory (LSTM): Dentro de la arquitectura de las RNN, pero diseñadas para superar el problema de desvanecimiento del gradiente, las LSTMs son efectivas para capturar dependencias a largo plazo en datos temporales.
- Redes Neuronales Convolucionales (CNN): Son un tipo especializado de arquitectura de redes neuronales diseñada principalmente para procesar y analizar tareas cognitivas, como reconocimiento de imágenes, procesamiento de lenguaje natural, series de tiempo, etc. Aunque son más conocidas por su éxito en tareas de visión por computadora, también se han aplicado con éxito en dominios como el procesamiento de series temporales para predicción de demanda de eventos.

3.2 ANTECEDENTES

Este apartado proporciona una revisión crítica de estudios previos que han examinado la demanda y gestión de servicios de emergencias médicas, así como los modelos analíticos utilizados para abordar estos problemas, destacando las contribuciones y limitaciones encontradas. Además, se profundiza en la aplicación de técnicas estadísticas convencionales y actuales orientadas a predecir la demanda de servicios, desarrollando así un fundamento sólido para el desarrollo de un modelo predictivo adaptado a las necesidades específicas y contextuales de Bogotá.

En la investigación realizada por Sergio González y Valentina Velásquez [16], se evalúan los modelos tradicionales versus modelos de inteligencia artificial para pronósticos de demanda utilizando series de tiempo, de tendencia, cíclicas, estacionales y erráticas en los sectores de construcción, transporte, medio ambiente, economía, seguridad y entretenimiento, con diferentes tipos de periodicidades según la información recolectada, diaria, semanal o mensual, teniendo en cuenta bases de datos de contaminación, crimen, nacimientos, comportamiento del dólar, temperaturas, pasajeros, con el fin de determinar el modelo que más se ajusta al tipo de dato y el tipo de demanda que se desea pronosticar. Se utilizaron los modelos Holt Winters y ARIMA (Modelo autorregresivo integrado de media móvil), los cuales se consideran tradicionales

en contraste con los modelos: CNN, LSTM y MLP, presentando como resultado una tabla comparativa que permite apreciar el modelo que más se ajusta a la naturaleza de los datos. Si bien, este estudio ofrece una visión de los diferentes modelos que se pueden utilizar para pronósticos de demanda, solo brinda un panorama general, no profundiza en la naturaleza de los datos para nuestro caso de un modelo espacio- temporal.

El trabajo de Paulo Abreu, Daniel Santos y Ana Barbosa-Póvoa [17], aborda los desafíos en la gestión eficiente de los servicios médicos de emergencia (EMS). Destacando la importancia de la previsión de la demanda para respaldar decisiones estratégicas en la asignación de recursos escasos, como vehículos y personal. La demanda de EMS se ve afectada por factores temporales, contextuales y geográficos, y la planificación efectiva requiere comprender los patrones que estos generan sobre la demanda. El aumento de la carga de trabajo en los centros de despacho de emergencias, debido al crecimiento demográfico y al envejecimiento de la población, ha generado la necesidad de gestionar recursos de manera más eficiente. Su objetivo general era responder a la pregunta ¿Existen diferencias en los patrones de demanda espaciotemporales para diferentes niveles de prioridad de llamadas y tipos de vehículos? Para resolverlo se propuso el uso de redes neuronales (NN), explorando su aplicación en el contexto específico de Portugal. El artículo destaca la falta de exploración de la demanda de servicios de ambulancia con enfoque en NN y aborda el impacto de la demanda compartida en áreas geográficas vecinas en la predicción de la demanda de vehículos de emergencia. Utiliza dos modelos de pronóstico NN, como resultados se obtuvo que los modelos basados en redes neuronales demostraron precisión, especialmente en la predicción de volúmenes de llamadas y despachos de vehículos, y destacan la importancia de considerar la demanda compartida por zonas vecinas.

El trabajo de Chinchay, Franco y Rosas [18], propone un modelo de gestión que contribuya con el cierre de la brecha de atención prehospitalaria en Lima (Perú). En este se plantea un modelo econométrico espacial a partir de una regresión lineal e incorporando: Modelo de rezago espacial (SLM) y Modelo de error espacial (SEM). El modelo planteado tiene variables espaciales, como accidentes de tránsito y ubicación de la población por distrito, así como de caracterización, tales como enfermedades crónicas y grupo etario de los pacientes atendidos. Adicionalmente, analiza las problemáticas en la oferta de servicios prehospitalarios con el fin de plantear el modelo de gestión óptimo que permitiera identificar, clasificar y utilizar de manera eficiente los recursos del SAMU (Sistema de Atención Móvil de Urgencia) en Perú.

La investigación realizada por J. Barrera Gutiérrez [19], planteo como objetivo determinar la cantidad óptima y la distribución de ambulancias en Bogotá, buscando responder de manera oportuna y eficiente a los incidentes de salud que requirieran de atención prehospitalaria durante el año 2018. Este estudio se enfocó en el uso de modelos dinámicos y técnicas de simulación de eventos discretos, prefiriéndolos sobre los modelos de programación lineal y

modelos clásicos de transporte-trasbordo, debido a su capacidad para integrar variables críticas como el número de ambulancias, el tiempo de espera, y su ubicación geográfica como variables predictoras. Estos modelos dinámicos fueron fundamentales para proyectar el funcionamiento del sistema a partir de datos históricos y para capturar la complejidad del sistema de emergencias médicas de Bogotá.

Se establecieron criterios específicos para evaluar las soluciones, incluyendo el tiempo promedio de llegada de las ambulancias, la efectividad en la asignación de recursos, la insatisfacción en la atención y el uso de ambulancias, cada uno con una ponderación determinada [19]. Finalmente, este estudio aporta una perspectiva valiosa sobre la optimización operativa de recursos y proporciona hallazgos relevantes en cuanto a la utilización de técnicas de simulación de eventos discretos, además, abre la posibilidad para el análisis predictivo que prevea futuras necesidades en los servicios de emergencias médicas en la capital y sugiere la utilidad de aplicar técnicas analíticas más avanzadas en este campo.

De otra parte, el estudio realizado por el International Journal of Environmental [20] sobre la predicción de ambulancias para el día siguiente al actual en diferentes regiones, resalta la importancia del uso de variables externas que influyen en la dinámica del servicio de emergencias. Algunas de estas, que aplican en general a las zonas urbanas, son los sectores demarcados como distritos financieros, culturales o residenciales. A su vez, otro factor influyente es el temporal, como el día de la semana, día del mes o mes del año, ya que la demanda suele experimentar periodicidades, haciendo que el análisis sea más complejo y requiera el uso de métodos avanzados como Machine Learning para identificar patrones en los datos y lograr la predicción de la demanda de ambulancias.

Para este caso, los modelos considerados son: Promedio móvil regional, que estima la demanda del día siguiente; Regresión lineal, que encuentra el hiperplano de mejor ajuste entre las muestras de datos de múltiples características; Support vector regression (SVR) que a diferencia del método anterior, captura las relaciones no lineales entre variables; Multi-layer Perceptrón, una arquitectura neuronal estándar que se compone esencialmente de una secuencia de capas lineales; Radial Basis Function Network (RBFN) una variación de las Redes Neuronales Artificiales (ANN); y por último, Light Gradient Boosting Machine (LightGBM), uno de los métodos más eficientes de árbol de decisión. Las métricas utilizadas para evaluar la precisión de estos son: error porcentual absoluto ponderado (WAPE), error absoluto medio (MAE), y error cuadrático medio (MSE) [20].

Entre los resultados se encontró que el modelo que mostró mejor desempeño fue el LightGBM, permitiendo identificar que los predictores más importantes eran la región, el día de la semana, día del mes, mes del año, y la demanda de los siete días previos. Además, se observó que para

los años 2006 y 2016, la demanda de ambulancias tuvo un crecimiento exponencial cercano al 23%, el promedio diario regional de demanda de ambulancias fue de 6.33, la mediana de la edad de pacientes atendidos fue de 55 años, y el incidente más común fue traumatismo, seguido por problemas relacionados con el sistema nervioso. Lo anterior demuestra que el conjunto de variables demográficas, temporales y espaciales, influyen en la ocurrencia de eventos y su naturaleza, y resultan esenciales para obtener estimaciones fiables y precisas sobre la predicción de demanda de servicios de emergencias médicas.

Los autores Fekadu L. Bayisa, Markus Ådahl, Patrik Rydén, Ottmar Cronie [21], se centran en la importancia de optimizar los tiempos de respuesta de ambulancias para mejorar la atención prehospitalaria y reducir la mortalidad. Para el caso de Suecia, existen dificultades para cumplir con los objetivos de tiempo de respuesta, especialmente en el norte del país. La ubicación estratégica de las ambulancias y su programación en el tiempo se presentan como un problema de optimización para minimizar los tiempos de respuesta en toda la región. El objetivo principal del trabajo es describir la dinámica espaciotemporal de las llamadas, identificar regiones hotspot (con alta actividad de incidentes) y desarrollar un modelo de pronóstico a corto plazo. Se subraya la importancia de utilizar simulaciones realistas para asignar óptimamente los recursos prehospitalarios y diseñar estrategias de despacho. El enfoque propuesto, basado en procesos log-gaussianos de Cox (LGCP) no homogéneos, se justifica por su capacidad para abordar la agrupación espacial y temporal en los datos de llamadas de emergencia. Con un conjunto de datos único de 444,283 llamadas de ambulancia en las regiones más septentrionales de Suecia durante 2014 a 2018, se utiliza un LGCP que consta de componentes espaciales y temporales deterministas, y un campo de intensidad aleatorio log-gaussiano separable estocástico. La estimación de los parámetros se realiza mediante una estimación de contraste mínimo y suavizado kernel cuártico. Se adopta un modelo de regresión de Poisson para modelar la variación temporal, y se verifica la elección del modelo mediante una prueba de Monte Carlo. En resumen, el estudio preliminar demuestra la viabilidad de los LGCP para modelar y prever llamadas solicitando asistencia médica, con aplicaciones potenciales en la optimización de recursos prehospitalarios.

La anterior revisión permite tener una visión más amplia de cómo se ha abordado el tema de emergencias médicas en diferentes países desde distintos ángulos y sus pronósticos a través de modelos estadísticos tanto tradicionales como de aprendizaje de automático, los cuales sirven de guía para la selección de modelos a evaluar o por el contrario descartar modelos que no estén enfocados al objetivo del proyecto. Asimismo, proporcionan una revisión bibliográfica significativa en cuanto al problema tratado en esta investigación, contribuyendo a la construcción del marco de referencia.

En resumen, este proyecto se distingue de los proyectos revisados, puesto que busca incluir criterios adicionales a la información médica, robusteciendo la base de datos al incorporar fuentes secundarias tales como condiciones meteorológicas, datos demográficos y eventos especiales en la ciudad. Aunque la mayoría de los estudios se centran en describir las dinámicas de los patrones espaciotemporales sobre la demanda, este trabajo prioriza la integración de variables adicionales con el propósito de enriquecer el análisis predictivo de emergencias médicas en Bogotá por unidad espacial y temporal.

4 IDENTIFICACIÓN Y SELECCIÓN DE VARIABLES CLAVE

La precisión de un modelo predictivo depende en gran medida de la correcta identificación y selección de las variables que influyen en el fenómeno estudiado. En el contexto de la demanda de servicios de emergencias médicas en Bogotá, es esencial identificar los factores que impactan la ocurrencia de emergencias a nivel espacial y temporal. Este capítulo describe el proceso seguido para identificar, seleccionar y analizar las variables clave que se utilizan en la construcción del modelo predictivo espaciotemporal.

4.1 FUENTES DE DATOS

Para identificar las variables relevantes que afectan la demanda de servicios de emergencias en Bogotá, se consultaron diversas fuentes de datos, tanto primarias como secundarias. Estas fuentes incluyen registros históricos de llamadas, población por UPZ y UPL, información sobre eventos especiales, condiciones meteorológicas, y datos demográficos. Estas fuentes permiten captar los factores que afectan la demanda de servicios de emergencia, proporcionando un panorama detallado de las variables que influyen en los patrones de demanda, cada una de las cuales se detalla a continuación:

4.2 SELECCIÓN Y PROCESAMIENTO DATOS

La base de datos proporcionada por el CRUE abarca el período entre enero de 2022 a junio de 2024 con un total de 1.538.318 registros. De estos, únicamente 371.646 registros, que representan aproximadamente el 24% corresponden a incidentes que involucraron el desplazamiento de un vehículo, específicamente ambulancias, como se muestra en la figura 4.

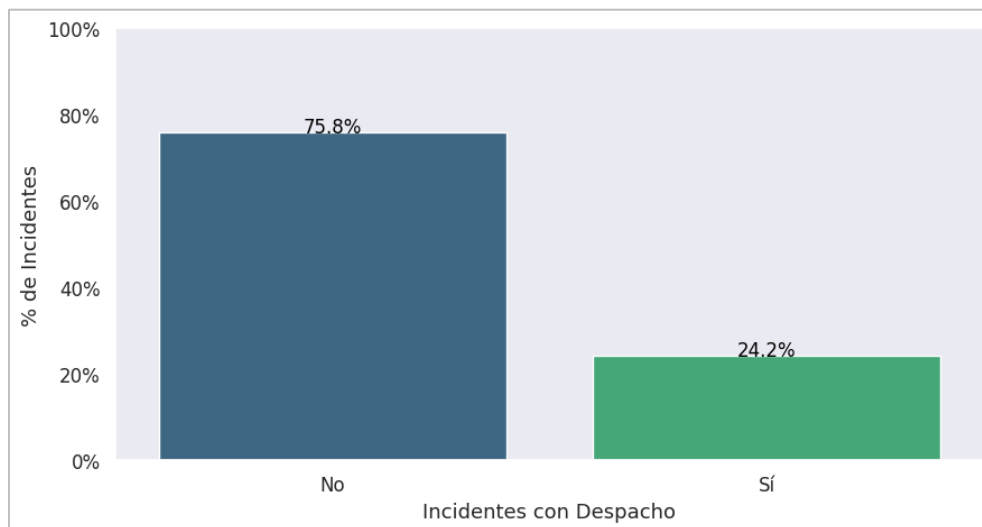


Figura 3. Proporción de incidentes con despacho de ambulancia.

Dado que el presente estudio se enfoca en un modelo espaciotemporal, se decidió utilizar como insumo solo los incidentes que implicaron un desplazamiento vehicular. Esta decisión se fundamenta en el hecho de que solo estos incidentes poseen un componente espacial, esencial para el análisis, mientras que los incidentes atendidos desde el centro de llamadas del CRUE carecen de dicho componente, al ser gestionados exclusivamente por vía telefónica.

Selección de Variables de Eventos Especiales: Para el análisis de eventos especiales, se seleccionaron las siguientes variables clave: la fecha de inicio del evento, el total de aforo, y el barrio donde se realizó el evento, incluyendo la UPZ correspondiente a cada barrio para permitir un análisis espacial más preciso. Se crearon dos variables adicionales para mejorar la comprensión de la influencia de los eventos en la demanda de servicios de emergencia: "Número de eventos", que contabiliza los eventos realizados en el mismo día para cada UPZ, y "Aforo acumulado", que suma la cantidad total de asistentes de todos los eventos realizados en un día específico dentro de la misma UPZ.

Selección de Variables de Datos Climáticos: En el análisis de los datos climáticos se consideraron las siguientes variables clave: el código de la estación meteorológica, la fecha de observación, el valor observado de las precipitaciones (dado en milímetros), y las coordenadas geográficas (latitud y longitud) de cada estación con un rango de cobertura de 11 kilómetros, con el fin de incorporar las mismas a la fuente primaria.

4.2.1 Procesamiento de los Datos

El procesamiento de los datos es una etapa vital para garantizar la calidad y la precisión del análisis posterior [28]. En este estudio, se siguió un enfoque sistemático para el manejo de los datos obtenidos de las diferentes fuentes. Inicialmente, se realizó una limpieza de los datos para eliminar registros duplicados, corregir valores erróneos y manejar los datos faltantes. En la Figura 5 se muestra el flujo de trabajo utilizado.



Figura 4. Flujo de trabajo para ejecución del proyecto.

- I. Limpieza de Datos:
 - Eliminación de Duplicados: Se identificaron y eliminaron registros duplicados en todas las bases de datos. Este paso fue fundamental para evitar sesgos en los resultados y asegurar que cada incidente, evento o registro climático fuera considerado una única vez.
 - Corrección de Valores Erróneos: Se revisaron y corrigieron valores atípicos o inconsistentes, como coordenadas geográficas fuera de rango, fechas que no correspondían con los períodos de observación, o clasificaciones incorrectas de incidentes y eventos.
 - Manejo de Datos Faltantes: Se implementaron estrategias específicas para manejar los datos faltantes. En algunos casos, los valores ausentes fueron imputados utilizando métodos estadísticos, mientras que, en otros, se optó por excluir registros incompletos cuando su inclusión podía comprometer la calidad del análisis. Para los casos donde la UPZ no pudo ser asignada mediante georreferenciación, se realizó una imputación proporcional basada en la distribución poblacional dentro de la localidad correspondiente, ajustando los valores en función del peso demográfico de cada UPZ. En cuanto a las fechas en las que no se registraron eventos especiales o precipitaciones, se les asignó un valor de 0, para reflejar la ausencia de actividad en esos días.
- II. Normalización y Transformación de Variables:
 - Formateo de Fechas y Horas: Las fechas y horas de los incidentes, eventos especiales y datos climáticos fueron convertidas a un formato unificado (YYYY-MM-DD HH:MM:SS), permitiendo una correcta comparación y análisis temporal.
 - Conversión de Coordenadas: Las coordenadas geográficas (latitud y longitud) fueron revisadas y normalizadas para asegurar su coherencia a lo largo de todos los conjuntos de datos, permitiendo así un análisis espacial preciso.
 - Unificación: Se llevó a cabo la estandarización de variables categóricas clave en la base principal, incluyendo el nombre de la localidad, el tipo de Incidente y la prioridad asignada, con el fin de unificar criterios y homogenizar los datos.
- III. Integración de Fuentes de Datos:
 - Asignación de UPZ: En la base de datos de eventos especiales, se añadió la UPZ para cada evento. Esto se logró a través de una llave que relacionaba cada barrio con su UPZ correspondiente. En los datos climáticos y en la base principal, se asignó la UPZ

correspondiente mediante la georreferenciación de las coordenadas geográficas. Los registros que no contaban con coordenadas, nombre de localidad o barrio fueron descartados ya que no aportaban el componente espacial necesario para el análisis.

- **Cálculo de Variables Derivadas:** Se crearon nuevas variables derivadas para enriquecer el análisis. Por ejemplo, se calculó el número de eventos por día y el número de eventos del día anterior por UPZ, así como la cantidad de eventos especiales y el aforo total acumulado en eventos simultáneos. Estas variables permiten capturar la intensidad y la concentración de actividades en determinadas áreas y fechas. Asimismo, a partir de la fecha se crearon variables como el número del día del mes, el día de la semana, mes y año, franja horaria de 12 horas, y días festivos en Colombia.
- **Unificación de Escalas:** Se aseguró que todas las variables, especialmente aquellas provenientes de distintas fuentes, estuvieran en la misma escala y unidades de medida, permitiendo una correcta comparación e interpretación.
- **Integración de la Cantidad de Población:** Para la base de datos de población, se realizó la inserción mediante una llave compuesta por la fecha y la UPZ, garantizando la coherencia en la combinación de las diversas fuentes de información.
- **Verificación de Coherencia Espaciotemporal:** Se validó que los datos integrados fueran coherentes tanto en el espacio como en el tiempo. Esto incluyó la revisión de la correspondencia entre los incidentes y las condiciones climáticas reportadas, así como la congruencia entre la ubicación de los eventos y su impacto en la demanda de servicios de emergencia.

Tabla 1. Matriz de Variables Clave para Modelos de Predicción.

Variable	Tipo Dato	Descripción	Fuente
Cod_UPZ	Entero	Código de la UPZ	SDS
Rango_Hora_9am-9pm	Binario	1: 9am - 9pm; 0: 9pm-9am	CRUE
Prioridad_Inicial_Moda_Baja	Entero	Cantidad de incidentes con Prioridad inicial baja	CRUE
Prioridad_Inicial_Moda_Media	Entero	Cantidad de incidentes con Prioridad inicial media	CRUE
Prioridad_Inicial_Moda_Alta	Entero	Cantidad de incidentes con Prioridad inicial alta	CRUE
Prioridad_Inicial_Moda_Crítica	Entero	Cantidad de incidentes con Prioridad inicial crítica	CRUE
Tipo_Incidente_Herido	Entero	Cantidad de incidentes – Herido	CRUE
Tipo_Incidente_TM	Entero	Cantidad de incidentes – Trastorno Mental	CRUE
Tipo_Incidente_DifRes	Entero	Cantidad de incidentes – Dificultad Respiratoria	CRUE
Tipo_Incidente_Enfermo	Entero	Cantidad de incidentes – Enfermo	CRUE
Tipo_Incidente_Convulsión	Entero	Cantidad de incidentes – Convulsión	CRUE
Tipo_Incidente_Paro	Entero	Cantidad de incidentes – Paro	CRUE

día_semana	Entero	Número del día de la semana (1-7)	CRUE
día_del_mes	Entero	Número del día del mes (1-31)	CRUE
día_del_año	Entero	Número del día del año (1-365)	CRUE
mes_del_año	Entero	Número del mes del año (1-12)	CRUE
Festivo	Binario	Variable calculada - días festivos en CO.	Propia
Infantes_0_4	Entero	Población de infantes por UPZ	DANE
Niños_5_14	Entero	Población de niños por UPZ	DANE
Adolescentes_15_19	Entero	Población de adolescentes por UPZ	DANE
Adultos_20_59	Entero	Población de adultos por UPZ	DANE
Adultos_mayores_60	Entero	Población de adultos mayores por UPZ	DANE
Total_Hombres	Entero	Población de hombres por UPZ	DANE
Total_Mujeres	Entero	Población de mujeres por UPZ	DANE
Total_población	Entero	Población total por UPZ	DANE
Densidad_poblacional	Entero	Cantidad de personas por Km2	Propia
ValorObservado_Max	Decimal	Valor de precipitación (mm)	IDEAM
Cant_eventos_esp	Entero	Cantidad de eventos especiales asociados	SUGA
Aforo_total	Entero	Capacidad total de eventos especiales por día	SUGA
Incidentes_día_anterior	Entero	Cantidad de incidentes registrados el día anterior	Propia
Total_Incidentes	Entero	Variable calculada	Propia
Broma	Binario	Circunstancia modificadora del evento - Broma	CRUE

La selección de estas variables es crucial para entender tanto la ocurrencia de incidentes como el contexto en el que se desarrollan, permitiendo una evaluación de los factores que influyen en la demanda de servicios de emergencia en Bogotá. Las variables seleccionadas abarcan aspectos detallados del incidente, como su código, prioridad, tipo de incidente y fecha de ocurrencia, además de datos geográficos y demográficos como la localidad, UPZ y distribución poblacional por edad y género. La inclusión de variables temporales, como la fecha y el valor observado de precipitaciones, permite identificar patrones estacionales y su impacto en la incidencia de emergencias, lo que es esencial para desarrollar modelos predictivos robustos. Variables relacionadas con la capacidad de respuesta, como la cantidad diaria de eventos especiales y su aforo total, también facilitan un análisis más preciso de la gestión de recursos en función de las condiciones de la ciudad.

4.2.2 Fuente Principal

La fuente primaria de datos proviene de los registros históricos de llamadas al CRUE de Bogotá. Estos datos fueron obtenidos a través de una solicitud formal presentada ante la Secretaría Distrital de Salud de Bogotá y su Comité de Ética de Investigación. La solicitud incluyó una explicación detallada del propósito del proyecto y los protocolos a implementar en cuanto a las medidas de seguridad de datos. Los registros fueron proporcionados tras la aprobación del comité de ética.

La presentación al Comité de Ética incluyó una elaboración documental donde se evidenciarán consideraciones éticas, evaluación de riesgos, el Protocolo de Seguridad para la Confidencialidad de la Información y los acuerdos sobre conflictos de interés por parte de cada integrante del equipo. También se abordó la confidencialidad de la base de datos, subrayando que su uso está restringido y no es de libre circulación. Además, se delinearón los productos esperados, asegurando a la Secretaría de Salud que la información proporcionada será utilizada de manera adecuada conforme a los más altos estándares éticos.

Los datos proporcionados por el CRUE abarcan el período de enero de 2022 a junio de 2024, dentro de los cuales se incluyen variables que describen diversos aspectos de cada incidente como el código del incidente, el tipo de incidente, la prioridad asignada inicialmente y al cierre de la atención, el tipo de recurso despachado, y las fechas y horas de inicio y cierre del incidente.

Además, se registran detalles sobre el estado final del incidente, la localidad y la zona donde ocurrió, así como las coordenadas geográficas que brindan el componente espacial necesario para el análisis. También se incluyen datos sobre circunstancias modificadoras que pueden haber afectado la atención del incidente, es decir, si el proceso de despacho fue cancelado por el paciente, si se realizó un traslado particular, si se identificó la llamada como broma, entre otras. Esta recopilación de variables permitió un análisis detallado, facilitando la comprensión de los patrones espaciales y temporales en la demanda de servicios de emergencias médicas en Bogotá (Tabla 2).

Tabla 2. Matriz de variables del CRUE.

Nombre de la variable	Descripción de la variable	Tipo dato
Código del incidente	Identificador único del incidente	Entero
Tipo de incidente	Cód. asignado por el operador del número único de seguridad y emergencia NUSE	Entero
Nombre del incidente	Nombre del tipo de incidente de acuerdo con el código asignado	Texto
Prioridad inicial	Se da de acuerdo con el tipo de incidente, la determina el operador que recibe la llamada	Texto
Prioridad final	Se da de acuerdo con el tipo de incidente, la determina el medico regulador	Texto
Hora y fecha inicial	Fecha y hora de ocurrencia del incidente	Fecha
Cod. recurso	Código del recurso	Texto
Tipo vehículo	Tipología del vehículo (básico o medicalizado)	Texto
Red	Tipo de empresa que gestiona la ambulancia (privada o pública)	Texto
Hora y fecha inicial	Fecha y hora de ocurrencia del incidente	Fecha
Hora y fecha de extincid	Fecha y hora de solicitud de la atención	Fecha
Hora y fecha 518	Fecha y hora de despacho	Fecha
Hora y fecha 523	Fecha y hora de llegada al lugar del incidente	Fecha

Hora y fecha 518b	Fecha y hora de inicio del traslado d43	Fecha
Hora y fecha 523b	Fecha y hora de llegada a la institución de salud	Fecha
CR508	Fecha y Hora recepción del paciente servicio de urgencias	Fecha
Hora y fecha terminate	Fecha y hora de solicitud de recepción del paciente	Fecha
Hora y fecha de cierre	Fecha y hora de solicitud de cierre del incidente	Fecha
Código de cierre	Código con el que se cierra el incidente	Texto
Nombre de cierre final	Nombre del tipo de cierre final de acuerdo con el código asignado	Texto
Incidente con despacho	Categorización de acuerdo con la asignación de vehículo de atención	Texto
Localidad de ocurrencia	Localidad de ocurrencia del incidente	Texto
Zona del incidente	Zona del incidente	Texto
X coordenada	Coordenada de geoposicionamiento	Entero
Y coordenada	Coordenada de geoposicionamiento	Entero
Circunstancia modificadora	Cód. circunstancia modificadora que amplía la caracterización del incidente	Texto
Nombre Circ. Mod.	Nombre de la circunstancia modificadora	Texto

4.2.3 Datos Secundarios y Complementarios

Se recopilaron datos secundarios y complementarios que enriquecen el análisis y permiten una comprensión más amplia de los factores que influyen en la demanda de servicios de emergencia en Bogotá.

- I. **Eventos Especiales:** La base de datos utilizada para el análisis proviene del Sistema Único de Gestión para el Registro, Evaluación y Autorización de Actividades de Aglomeración de Público en el Distrito Capital, abarcando el período de enero de 2015 hasta el 30 de junio de 2024. Para el análisis, se utilizaron únicamente los datos que coinciden con el período de la fuente principal. Esta base incluye información detallada sobre los eventos culturales y recreativos realizados en Bogotá como fechas de inicio, nombres y descripciones de los eventos, lugar de realización (dada por barrios de la ciudad), capacidad de aforo, y la clasificación del evento, entre otros [22].
- II. **Condiciones Meteorológicas:** Las variables climáticas, como las precipitaciones, fueron consideradas debido a su influencia potencial en la congestión vehicular que genera en la ciudad. Los datos fueron recopilados a partir de reportes históricos proporcionados por el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) [23].
- III. **Población:** Esta información fue obtenida de las proyecciones de población de Bogotá realizadas por el Departamento Administrativo Nacional de Estadística (DANE). Estas proyecciones, extraídas para el periodo en estudio, incluyen datos detallados sobre grupos etarios, género, y el total de la población por cada localidad, UPZ y UPL. Cabe resaltar que el análisis de la población de Bogotá por UPZ y UPL se incorporó como un

factor relevante, ya que áreas con mayor concentración de habitantes pueden experimentar una demanda más alta de servicios de emergencia [24].

- IV. Área Urbana: La extensión del área urbana de Bogotá fue obtenida de la base de datos proporcionada por el portal Laboratorio Urbano - Bogotá, accesible en su plataforma de datos abiertos. Esta información incluye detalles geográficos de las UPZs y UPLs, y su relación con la estructura urbana de la ciudad. De allí se extrae la extensión urbana de cada unidad de planeamiento en Km^2 , lo que permite el cálculo de la densidad poblacional para analizar su influencia en la demanda de emergencias médicas [25].
- V. Mapa de Bogotá: Para el análisis espacial, se utilizó un mapa de Bogotá que delimita el territorio geográfico por cada una de las Subredes Integradas de Servicios de Salud, segmentado por UPZ y UPL. Este mapa es obtenido de La Infraestructura de Datos Espaciales de Bogotá (IDECA) proporcionada por la Secretaría Distrital de Salud [26]. La figura 3 ilustra como está dividida la ciudad por UPL en los colores, y UPZ en las formas dentro de cada color.

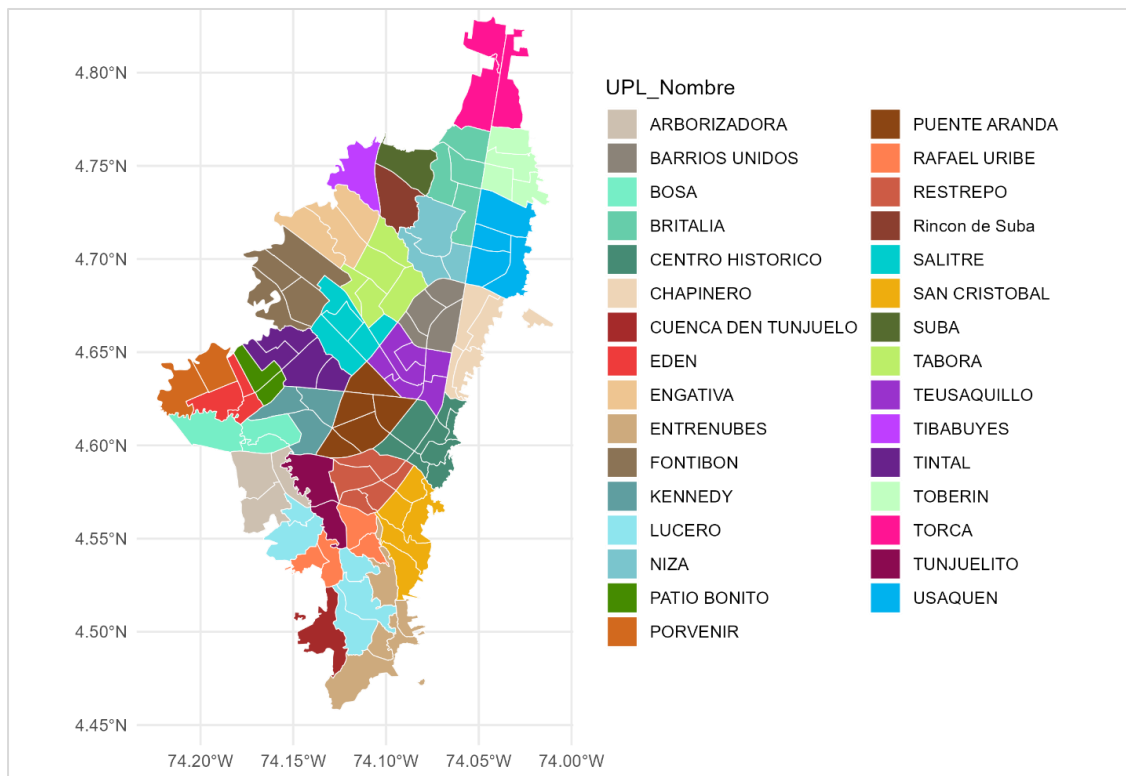


Figura 5. Mapa de Bogotá por UPZ y UPL.

- VI. Barrios por UPZ: Se utilizaron datos de barrios organizados por UPZ para establecer una correspondencia entre los eventos especiales y las UPZs, y por UPLs consecuentemente.

Esta información fue obtenida a través de la página Corporación para el Control Social (CONTRIAL) [27].

4.3 ANÁLISIS EXPLORATORIO DE LA BASE DE DATOS

A partir de la construcción de la base de datos integrando las variables seleccionadas como lo muestra la tabla 2, se realiza un análisis exploratorio para evaluar su comportamiento tanto de forma individual como en relación con otras variables. Este análisis incluye la visualización de distribuciones, la identificación de posibles datos atípicos y la exploración de relaciones cruzadas entre variables, permitiendo identificar patrones preliminares y entender mejor cómo cada variable podría influir en la demanda de servicios.

- **Localidad:** Se examinó la distribución de incidentes por localidad (figura 6), identificando que las cinco localidades que registran la mayor cantidad de emergencias médicas son Kennedy, Suba, Engativá, Bosa y Ciudad Bolívar. En contraste, localidades como La Candelaria y Antonio Nariño reportan las menores incidentes.

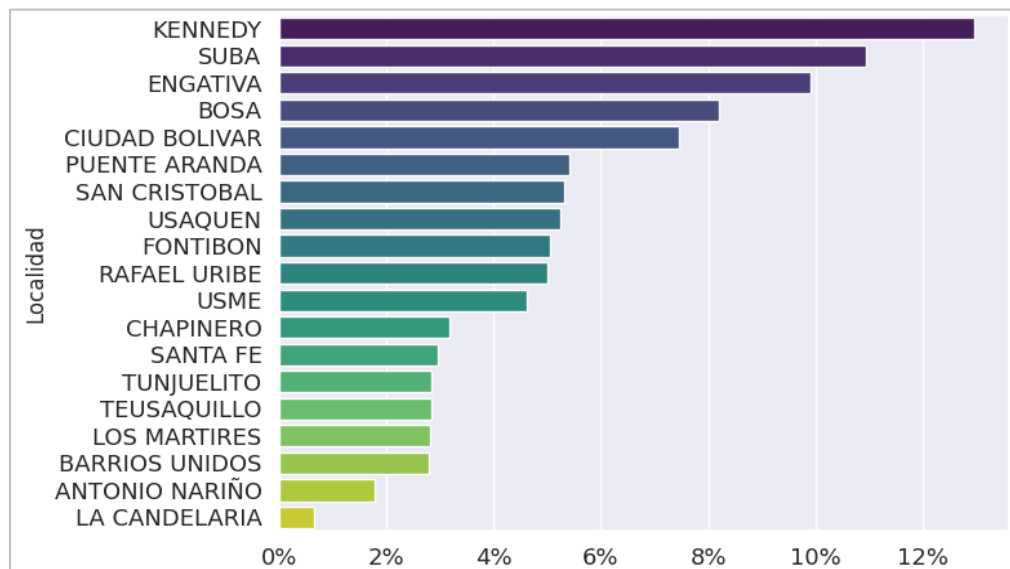


Figura 6. Porcentaje de Incidentes por Localidad.

- **UPZ:** Los datos mostrados en la figura 7, revelan que las UPZ con mayor cantidad de incidentes son El Rincón de la localidad de Suba con 11,681 registros, representando el 3.14% del total, seguida por Bosa Central con 10,439 casos (2.81%) y Bosa Occidental con 9,095 incidentes (2.45%). Tibabuyes tiene 9,046 incidentes (2.43%), mientras que Patio Bonito presenta 8,535 (2.3%). Estas UPZs concentran una parte significativa de los eventos reportados, indicando áreas con alta demanda de atención y recursos.

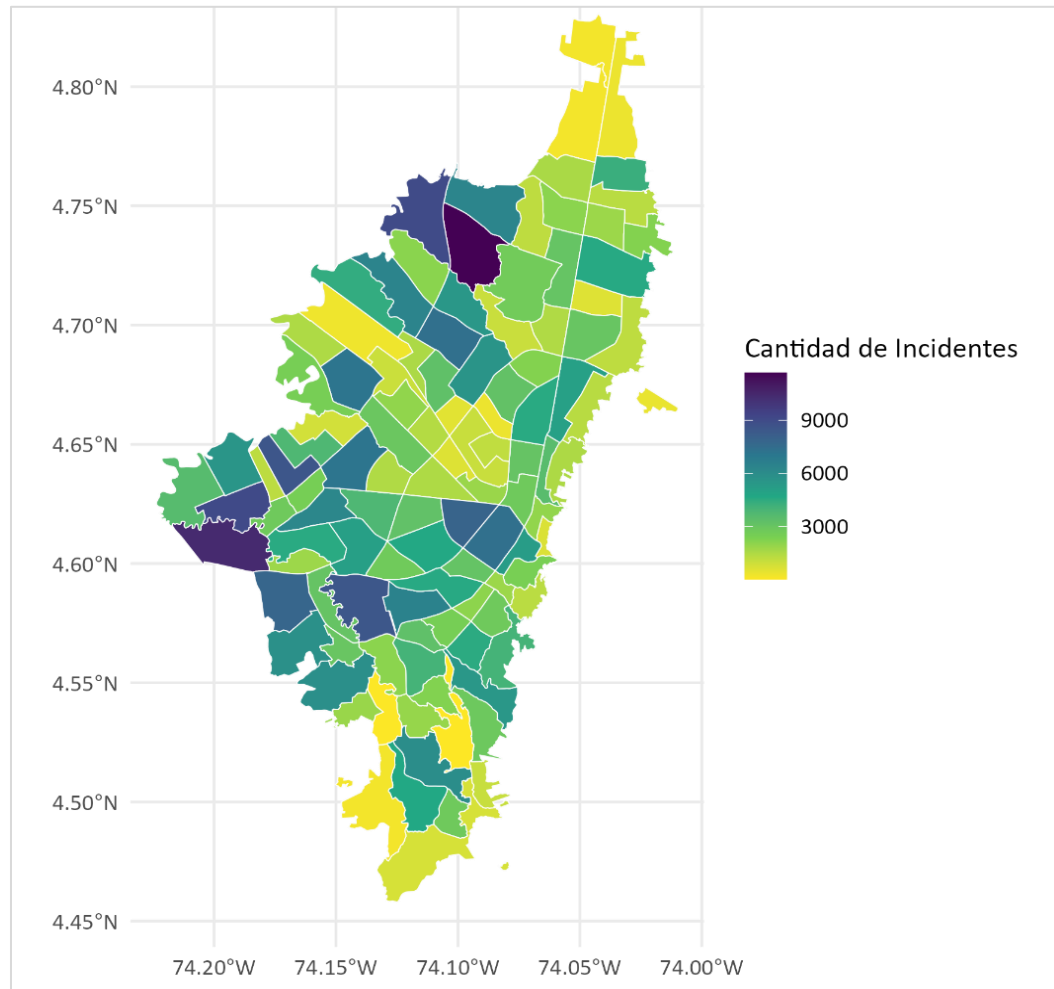


Figura 7. Mapa de Cantidad de Incidentes por UPZ.

- La Figura 8 muestra la distribución de frecuencias en el tipo de incidente registrado en el sistema. Esta diversidad de necesidades médicas urgentes resalta la complejidad y la necesidad de recursos adecuados para ofrecer atención efectiva y oportuna.

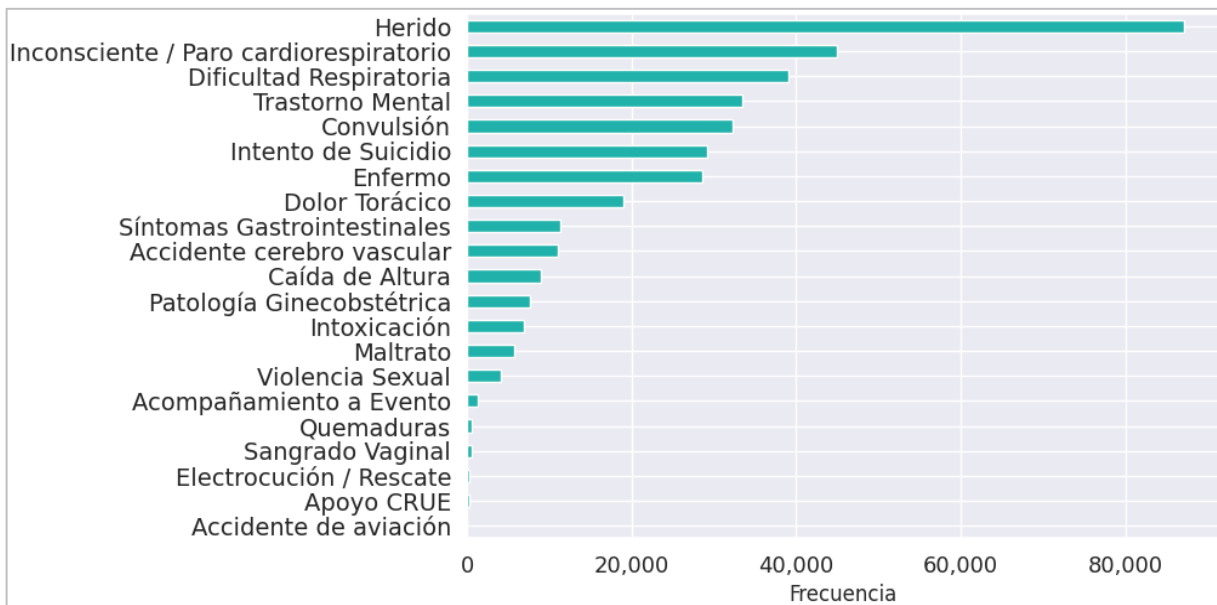


Figura 8. Distribución por Tipo de Incidente.

- Prioridad:** Es importante aclarar que esta se determina en dos momentos distintos durante la atención de un incidente. Inicialmente, la prioridad es asignada por el C4 antes de ser transferida al CRUE. Posteriormente, el CRUE asigna la prioridad final según las circunstancias específicas al cierre de cada caso. Esta categoría se asigna según la gravedad del caso y el tipo de incidente, siguiendo los lineamientos preestablecidos por el CRUE.

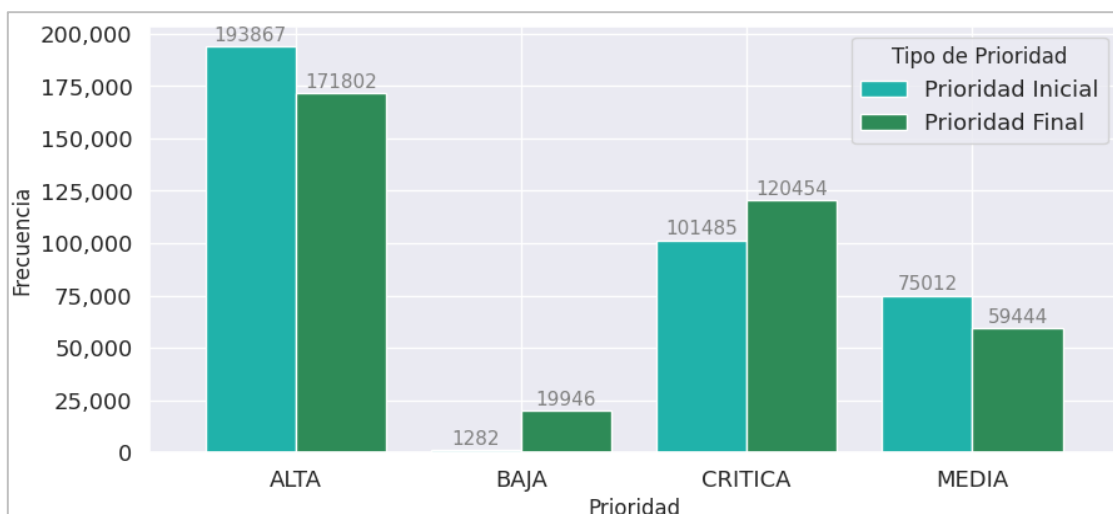


Figura 9. Distribución por Prioridad Inicial y Final.

Como se muestra en la Figura 9, la prioridad de las emergencias que requieren desplazamiento se distribuye mayormente en categorías de urgencia alta y crítica. En

Bogotá, la distribución por prioridad de los incidentes de emergencia revela una considerable carga de situaciones que demandan atención inmediata. La prioridad alta encabeza la lista con un total de 171,802 casos al cierre del incidente, es decir, la prioridad final asignada. Le sigue de cerca la categoría crítica, con 120,454 incidentes reportados. Estas cifras destacan la urgencia y la gravedad de muchas de las emergencias a las que responden los servicios médicos en la ciudad.

- **Incidentes Registrados por Mes:** Se creó una gráfica para mostrar la cantidad de incidentes registrados en cada mes durante el período de estudio. Esta gráfica permite identificar patrones estacionales o tendencias a lo largo del tiempo, proporcionando una perspectiva temporal de la frecuencia de emergencias médicas.

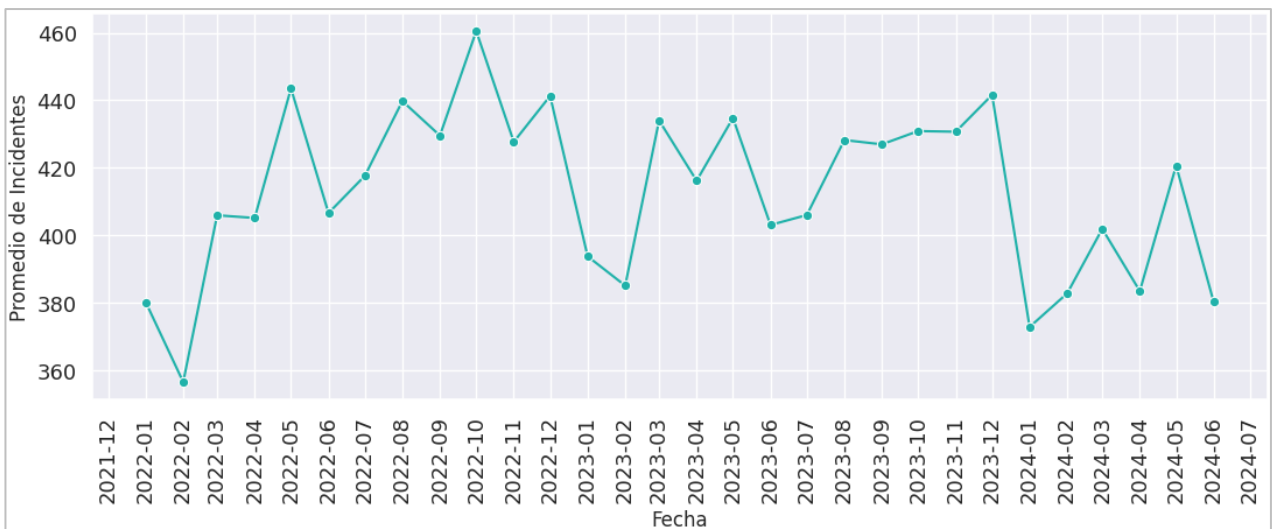


Figura 10. Tendencia del Promedio Diario de Incidentes por Mes y Año.

La Figura 10 refleja el promedio diario de incidentes desde enero de 2022 hasta junio de 2024, mostrando una tendencia con fluctuaciones a lo largo del tiempo. En general, se observa que el promedio diario de incidentes varía entre 350 y 460. Estos valores sugieren la existencia de una tendencia estacional, donde los promedios más altos se registran en los meses de mayo y octubre, mientras que los valores más bajos se observan en enero y febrero de cada año.

Asimismo, se aprecia que los meses con mayor promedio diario de emergencias médicas son mayo y octubre. En Bogotá, estos meses coinciden con temporadas de mayor lluvia, lo que podría estar relacionado con un aumento de accidentes de tránsito o enfermedades respiratorias, que son comunes durante las épocas de lluvias intensas. Las vías resbaladizas y las malas condiciones de tráfico en días lluviosos podrían incrementar la frecuencia de accidentes, lo que explica los picos en los incidentes reportados.

En contraste, los meses de enero y febrero presentan un promedio diario menor de emergencias. Este descenso podría explicarse por la disminución de la actividad en la ciudad durante las vacaciones de fin de año, cuando una parte considerable de la población sale de Bogotá. Menos tráfico y una menor densidad de personas en la ciudad contribuyen a la reducción de incidentes. Estos patrones se pueden evidenciar de mejor manera en la figura 11.

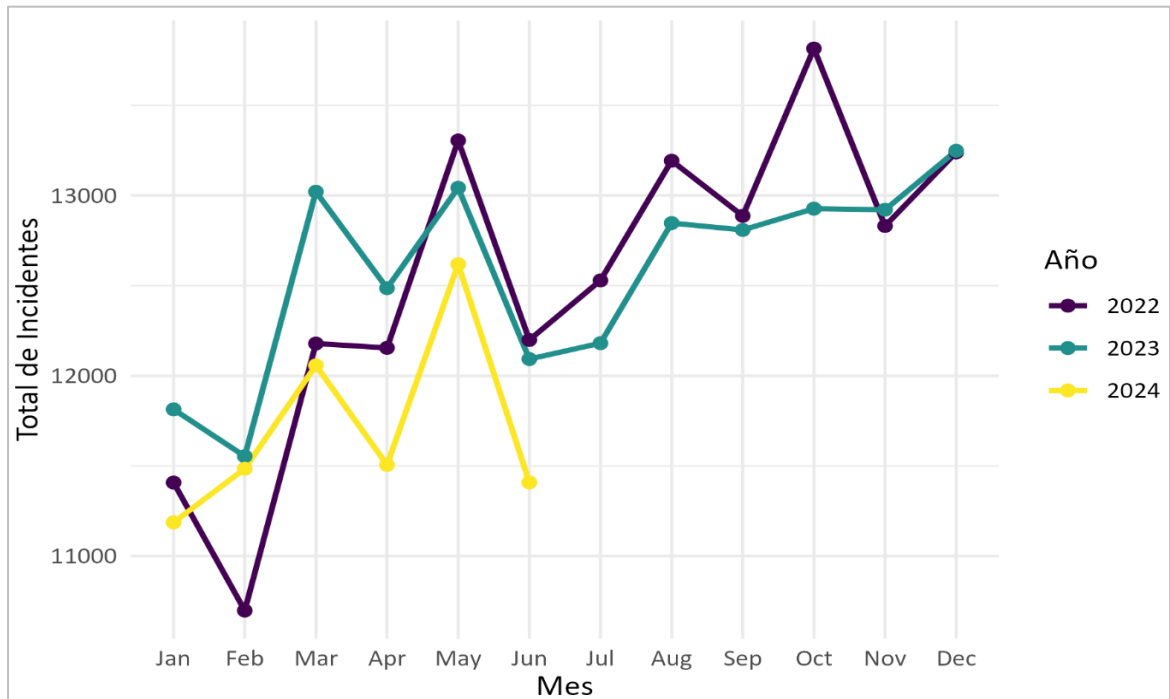


Figura 11. Tendencia de Incidentes por Mes (2022-2024).

- Distribución de incidentes por hora:** La figura 12 proporciona el número total de incidentes médicos reportados en Bogotá distribuidos a lo largo de las 24 horas del día. Cada columna muestra el número de incidentes en una hora específica, desde la medianoche (hora 0) hasta la medianoche del siguiente día (hora 23).

El análisis de la distribución horaria de incidentes médicos revela dos picos notables a lo largo del día: uno en la mañana y otro en la tarde-noche. Estos picos indican momentos de alta demanda de servicios médicos que pueden estar relacionados con diversos factores.

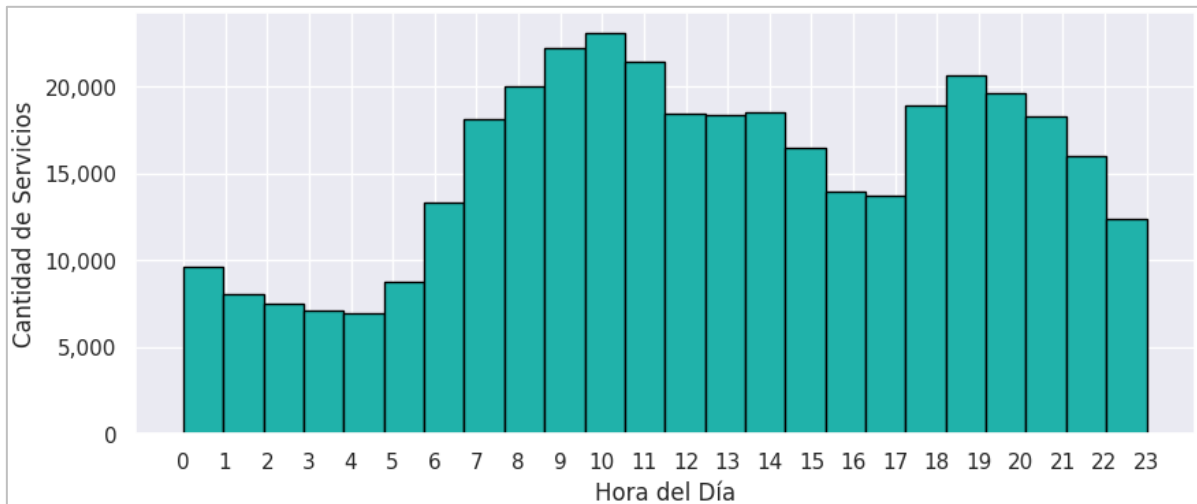


Figura 12. Distribución de Servicios de Emergencias por Hora.

El primer pico significativo en la cantidad de incidentes se observa en la mañana, especialmente entre las 8:00 a.m. y las 11:00 a.m. Las horas 8, 9 y 11 reportan 18,153, 20,036 y 21,405 incidentes, respectivamente. La hora 10:00 a.m. presenta el pico más alto a lo largo del día con una alta incidencia de 22,233 incidentes.

El segundo pico importante se registra en la tarde y noche, particularmente entre las 7:00 p.m. y las 9:00 p.m. La hora 7:00 p.m. muestra 20,610 incidentes, la hora 8:00 p.m. tiene 19,642 incidentes y la hora 9:00 p.m. presenta 18,305 incidentes.

Los picos en la mañana y en la tarde-noche sugieren que la incidencia de emergencias médicas en Bogotá está estrechamente relacionada con los patrones de actividad diaria y social. Durante la mañana, el aumento puede estar impulsado por el inicio de las rutinas diarias y el tráfico asociado, mientras que en la tarde-noche, el incremento puede reflejar el impacto de la actividad social y el transporte en la ciudad. Entender estos patrones es crucial para la planificación de recursos de emergencia y la optimización de la respuesta médica en la ciudad.

- **Distribución de incidentes por hora y día de la semana:** Al analizar la distribución de incidentes por hora y por día de la semana en Bogotá, se observa que, en términos generales, los patrones son consistentes con la distribución general de incidentes. Sin embargo, se identifican particularidades notables en la franja horaria de las primeras horas del día, desde la medianoche hasta las 5:00 a.m. en los diferentes días.

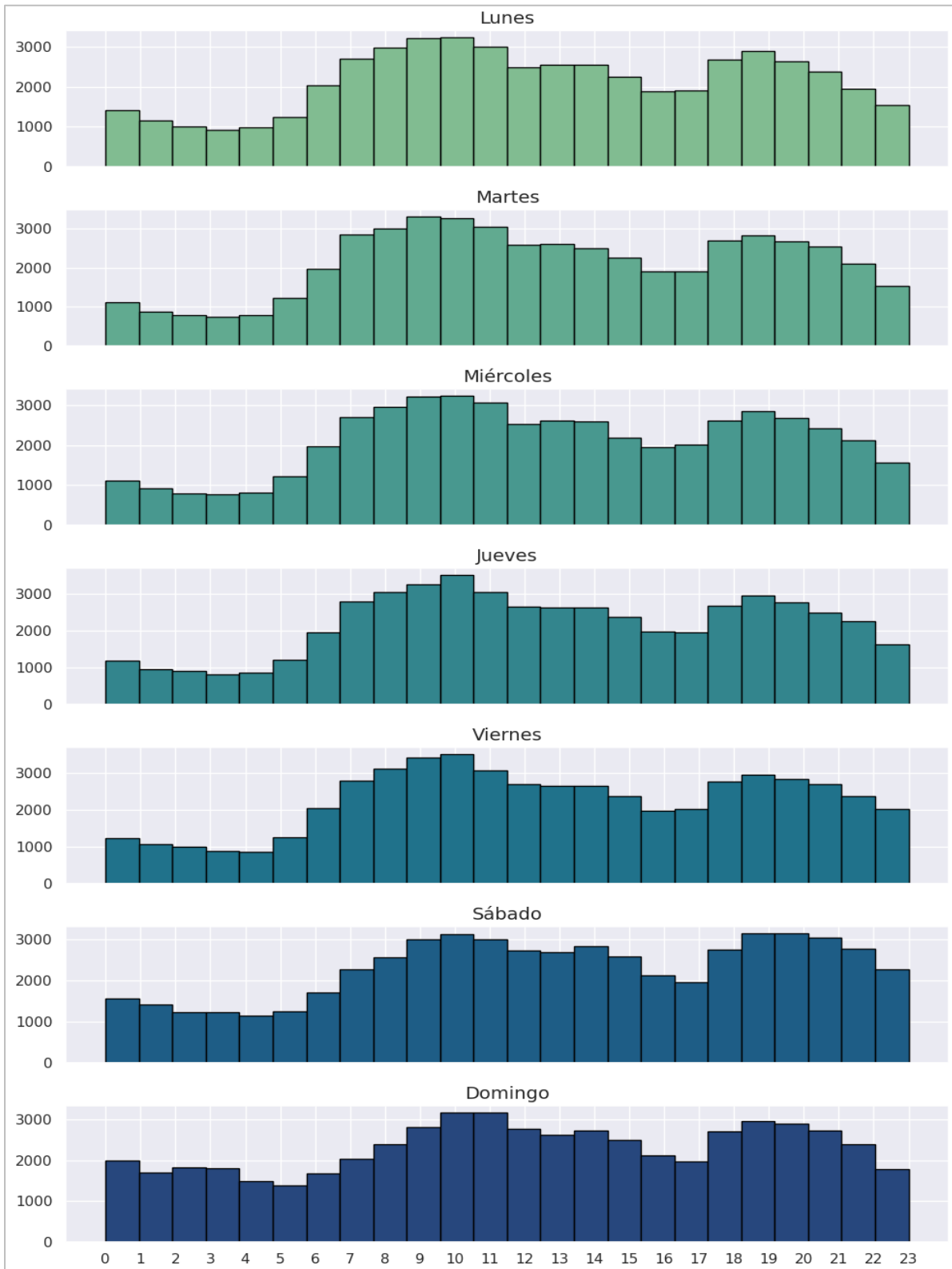


Figura 13. Distribución de incidentes por hora y día de la semana.

En la Figura 13 se observa que la distribución de incidentes por hora y día de la semana tiene un comportamiento relativamente consistente, con un patrón estable durante la mayor parte de la semana. Sin embargo, a partir del viernes, se observa un cambio en este comportamiento, pues los incidentes comienzan a aumentar notablemente durante las horas de la noche, lo que se extiende en la madrugada del sábado. Este incremento es aún más pronunciado el sábado, donde la actividad se mantiene elevada hasta las primeras horas del domingo, superando el pico de incidentes que se presentan en un día entre semana. Por otro lado, los eventos registrados el domingo disminuyen progresivamente a medida que avanza el día, volviendo a niveles más bajos y estables en comparación con el pico observado el sábado.

Este patrón de comportamiento sugiere que los días de fin de semana, especialmente las noches de viernes y sábado, representan un período crítico con un aumento significativo en la demanda de servicios de emergencia, lo que puede estar asociado con un mayor número de actividades sociales, eventos culturales, consumo de alcohol y desplazamientos que suelen ocurrir durante estos días de la semana.

- **Distribución de incidentes por día de la semana y franjas del día:** En la figura 14 se reafirman los patrones distintivos en la distribución de eventos durante la semana. Pues en la madrugada de los días sábado y domingo, franja de 0:00 a 5:59 am, se observa un incremento notable en los incidentes.

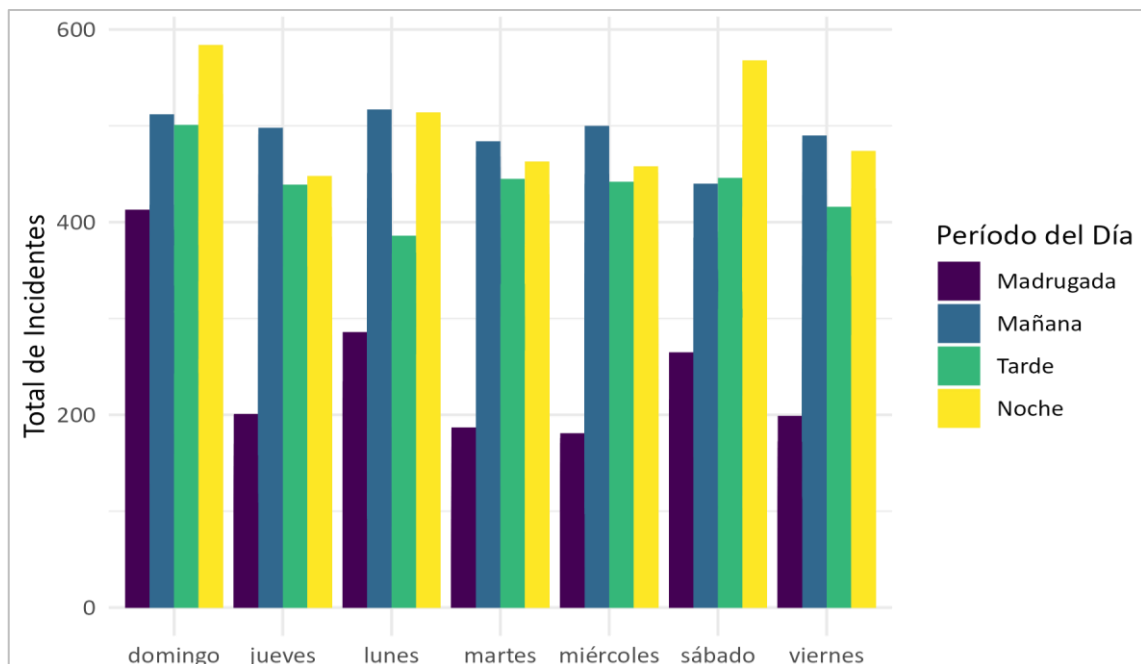


Figura 14. Total de Incidentes por Día de la Semana y Período del Día.

Por esta razón, las variables relacionadas con el día de la semana y el rango horario son cruciales en los modelos predictivos, ya que permiten capturar con mayor precisión los patrones de incidencia. Incorporar esta información mejora las predicciones de la demanda de servicios de emergencia, lo que a su vez optimiza la planificación y asignación de recursos, ajustándose al comportamiento social de la población y sus variaciones a lo largo de la semana.

- **Mapa de población por UPZ:** En la figura 15, se presenta un mapa que ilustra la distribución de la población por grupos etarios en las diferentes UPZs.

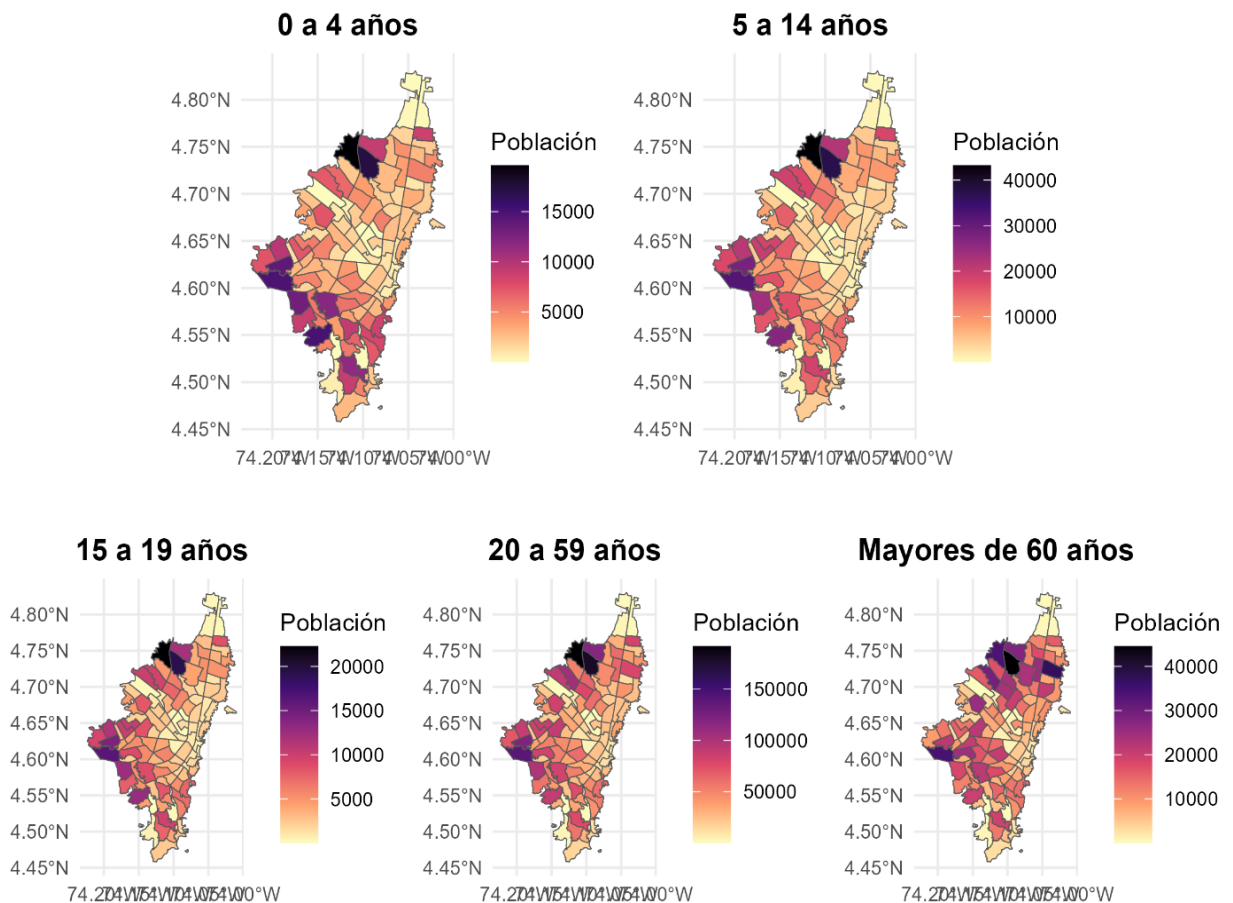


Figura 15. Mapa de distribución Demográfica por Grupos Etarios en UPZ.

Las UPZs El Rincón y Tibabuyes, en la localidad de Suba, registran la mayor concentración poblacional en todos los grupos etarios; este mismo patrón se identifica en las UPZs Bosa Central y Occidental en la localidad de Bosa, y en Patio Bonito en la localidad de Kennedy. La alta densidad poblacional en estas áreas podría correlacionarse con el número elevado de incidentes reportados por UPZ, como se evidencia en la figura 7.

- **Análisis de correlación:** Para el análisis de correlación, se calculó una matriz de correlación utilizando el coeficiente de Pearson. Por otra parte, para las variables categóricas, se empleó el análisis de varianza (ANOVA) con el fin de determinar si existían diferencias estadísticamente significativas en la cantidad de incidentes en función de las categorías.

En la Figura 16, se evidencia que las variables numéricas más altamente correlacionadas son los grupos etarios de población, lo cual es de esperarse ya que estas categorías representan subconjuntos de la población total.

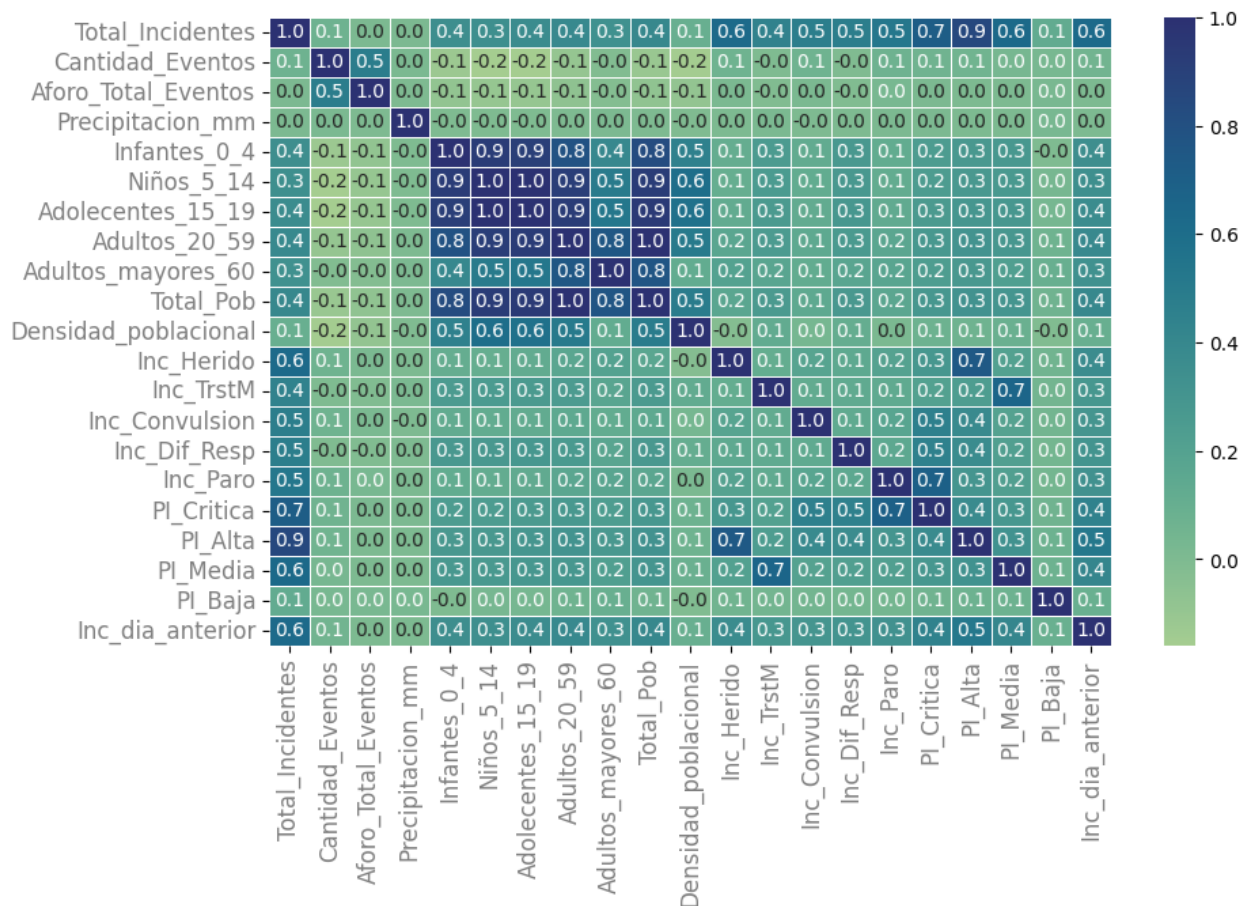


Figura 16. Mapa de correlación entre las variables.

Esta alta correlación sugiere que los cambios en la población total están directamente influenciados por las variaciones en estos grupos etarios, reflejando la distribución demográfica. Además, esta correlación indica que las diferentes edades tienden a moverse de manera conjunta, lo que podría estar relacionado con factores sociodemográficos o la estructura de la población en Bogotá. Esto también implica que, en un análisis predictivo, modelar el

comportamiento de la población total o de un grupo etario específico podría proporcionar información similar, dado que estos grupos están fuertemente entrelazados en términos de correlación.

La tabla 3 muestra los resultados del análisis de varianza realizado.

Tabla 3. Resultados ANOVA.

Variable	F- Estadístico	p-Valor
Año	47,74	0,00
Mes	20,91	0,00
Días_Mes	1,05	0,38
Días_del_año	1,73	0,00
Día_semana	21,03	0,00
Cod_UPL	1297	0,00

Como se puede observar en la Tabla 3, la mayoría de los p-valores son menores a 0.05, lo que indica que las variables categóricas consideradas tienen una relación estadísticamente significativa con la cantidad de incidentes, excepto por los días del mes. Esto sugiere que cada una de estas variables influye de manera significativa en la variabilidad de los incidentes, lo que justifica su inclusión en el análisis y modelado posterior.

El análisis descriptivo realizado en este capítulo fue esencial para comprender en profundidad las características y patrones de los datos, lo que permitió establecer una base sólida para las etapas posteriores del estudio. Este análisis facilitó la identificación de relaciones significativas entre las variables, así como la detección de valores atípicos y posibles inconsistencias en los datos. A partir de esta comprensión, se seleccionaron 24 variables, las cuales se presentan en la Figura 16, para ser incluidas y evaluadas en el modelo. La selección de estas variables se realizó bajo criterios de relevancia y aplicabilidad al contexto del problema, priorizando aquellas que ofrecieran información útil y predictiva en tiempo real.

Por otro lado, se descartaron variables cuya información se recolecta en etapas posteriores al evento, como la prioridad final, ya que esta solo está disponible después de que el incidente ha sido atendido, lo que limita su utilidad para fines predictivos. Asimismo, los tipos de incidentes con baja frecuencia no fueron considerados, dado que su limitada representatividad podría afectar la estabilidad y precisión del modelo. Este proceso no solo ayudó a depurar el conjunto de datos, sino que también aseguró que las variables seleccionadas tuvieran un impacto significativo en el desempeño del modelo.

En conclusión, el análisis descriptivo no solo contribuyó al entendimiento integral de los datos, sino que también guió decisiones críticas para optimizar la selección de variables, garantizando un enfoque metodológico sólido y alineado con los objetivos del estudio.

5 EVALUACIÓN DE MODELOS PREDICTIVOS

En los capítulos anteriores se abordaron las tres primeras etapas de la metodología CRISP-DM, avanzando en la comprensión del negocio, el análisis exploratorio y la preparación de los datos. A partir de este capítulo, se detallarán las actividades realizadas en las tres etapas finales de este flujo metodológico, las cuales se ilustran en la Figura 17, permitiendo concluir el proceso de modelado y análisis planteado.

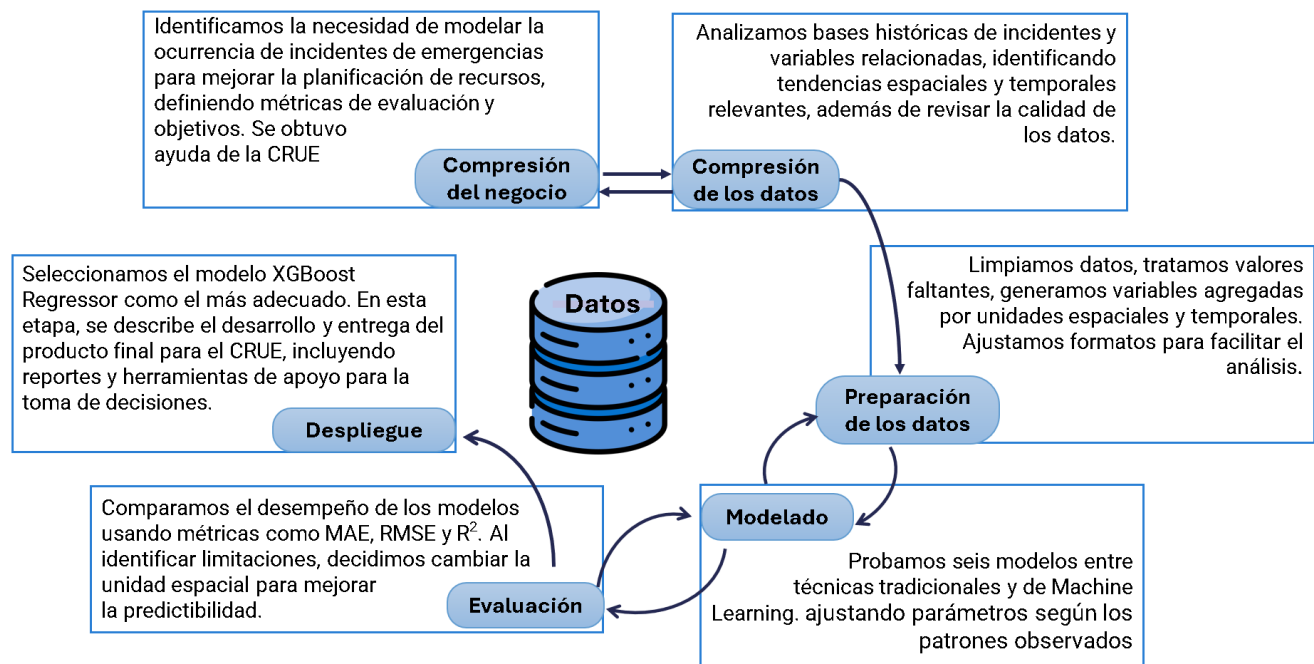


Figura 17. Flujo metodológico aplicado

Como lo sugieren Karthik R. y Abhishek S. “la evaluación de los modelos es el paso más importante en el desarrollo de cualquier solución analítica” [13], pues en esta etapa es donde medimos el rendimiento del modelo y decidimos si continuar con su desarrollo o devolvemos a revisar todos los pasos previos a este punto.

La evaluación de los modelos predictivos se realizó con el objetivo de determinar el rendimiento y la aplicabilidad de diferentes enfoques en la predicción de la demanda de emergencias médicas en Bogotá. En esta sección se presentan los pasos metodológicos clave, los criterios de selección de variables y parámetros, y los resultados de las evaluaciones bajo métricas estándar.

5.1 SELECCIÓN DE LA UNIDAD ESPACIAL Y TEMPORAL

Para este estudio, se seleccionó la UPZ como unidad espacial de análisis y un intervalo de 12 horas como unidad temporal. La elección de las 117 UPZs responde a su carácter administrativo y funcional en la planeación urbana de Bogotá, ya que agrupa varios barrios dentro de una misma zona, lo que permite una representación coherente y escalable de las dinámicas locales. Esta unidad facilita la agregación de datos espaciales.

En cuanto a la unidad temporal, se optó por intervalos de 12 horas debido a la estructura operativa del CRUE, que organiza sus turnos, distribuye el personal y los automotores en ciclos de 12 horas comenzando a la 9 a.m. Este enfoque temporal permite una alineación directa con las operaciones del sistema, garantizando que el modelo predictivo esté adaptado a las condiciones reales de trabajo y facilite su implementación en la planificación diaria y la asignación de recursos.

La combinación de estas unidades espaciales y temporales responde a un balance entre la granularidad de los datos, la viabilidad operativa y la relevancia práctica de los resultados esperados para mejorar la gestión del sistema de emergencias médicas en Bogotá.

5.2 DIVISIÓN DE DATOS PARA ENTRENAMIENTO Y PRUEBA

Para garantizar una evaluación robusta de los modelos y evitar problemas de sobre/sub-entrenamiento, los datos se dividieron de la siguiente manera:

- **Proporción de entrenamiento y prueba:** Se asignó el 80% de los datos para el entrenamiento del modelo, y el 20% restante para probar su rendimiento, esto es 111.125 y 27.782 registros respectivamente. Esta división asegura que los modelos se entrenen con un conjunto suficientemente amplio de datos y que las métricas de evaluación se calculen sobre datos no vistos durante el entrenamiento.
- **Método de División:** La división se realizó de manera estratificada para garantizar que las distribuciones de las variables clave, como el día de la semana, rango horario y las UPZs, fueran consistentes entre los conjuntos de entrenamiento y prueba.

5.3 METODOLOGÍA DE EVALUACIÓN

La evaluación metodológica se estructuró para garantizar que todos los modelos, tanto clásicos como de Machine Learning, fueran comparados bajo las mismas condiciones:

- **Métricas de Evaluación:**
 - **Error Cuadrático Medio (RMSE):** Mide la magnitud promedio de los errores de

predicción, penalizando más los errores grandes.

- **Error Absoluto Medio (MAE):** Ofrece una medida más intuitiva del error promedio.
 - **Coefficiente de determinación (R^2):** Evalúa que tan bien el modelo explica la variabilidad de los datos.
- **Procesos de Optimización:**
 - En los modelos de Machine Learning, se realizó una búsqueda de hiperparámetros utilizando Grid Search (búsqueda por grilla) con validación cruzada, evaluando combinaciones de hiperparámetros como el número de árboles (`n_estimators`), la tasa de aprendizaje (`learning_rate`) y la profundidad máxima (`max_depth`). Por otro lado, para las redes neuronales se exploraron diferentes arquitecturas y combinaciones de hiperparámetros mediante un enfoque iterativo. Entre los parámetros optimizados se incluyeron el número de capas y de neuronas, la tasa de aprendizaje, la tasa de regularización dropout, los tamaños del batch y la cantidad de épocas.
 - En los modelos clásicos como ARIMA y SARIMA, la selección de parámetros óptimos se realizó utilizando los criterios Akaike Information Criterion (AIC) y Bayesian Information Criterion (BIC). Estos métodos equilibran la calidad del ajuste y la simplicidad del modelo, penalizando la complejidad excesiva. Mientras que el AIC minimiza la pérdida de información, el BIC favorece modelos más parsimoniosos al considerar el tamaño de la muestra. Este enfoque permitió identificar configuraciones que maximizan el desempeño predictivo y garantizan robustez ante nuevos datos.
 - **Procesos de selección de variables:** La selección de variables es un paso fundamental en el desarrollo adecuado de los modelos, ya que influye directamente en su precisión, interpretabilidad y capacidad de generalización. Teniendo en cuenta lo anterior, se describen a continuación los criterios utilizados para la selección de las variables.
 - **Análisis Exploratorio de Datos:** se realizó un análisis de las características disponibles en el conjunto de datos, considerando tanto su relevancia teórica como empírica. Se emplearon diferentes gráficos para identificar relaciones entre variables predictoras y la demanda de incidentes. De igual forma, se analizaron patrones temporales (días de la semana, meses, y rangos horarios) en los datos históricos de incidentes.

- **Técnicas estadísticas:** Se emplearon métodos estadísticos para evaluar relaciones lineales entre las variables independientes y la variable objetivo (cantidad de incidentes). Esto incluyó el cálculo de correlaciones de Pearson. Además, se construyó una matriz de correlación que permitió visualizar y cuantificar estas relaciones, ayudando a identificar variables con mayor relevancia para el modelo. Este análisis también permitió detectar posibles problemas de multicolinealidad entre variables y por ende sobreajuste de los modelos, lo que ayudó a decidir qué características mantener o descartar para optimizar el desempeño de los modelos.
- **Relevancia Operativa:** Además de criterios estadísticos, se priorizó la inclusión de variables con significado práctico y relevancia operativa para el CRUE. Variables como los rangos horarios, los días de la semana, el tipo de incidente y la prioridad inicial fueron seleccionadas por su capacidad para reflejar patrones consistentes en la demanda de emergencias médicas y por su utilidad directa en la planificación y asignación de recursos (Tabla 6).

5.4 SELECCIÓN DE MODELOS

La selección de modelos para evaluar en el conjunto de datos se fundamentó en la revisión bibliográfica y en los antecedentes teóricos presentados en el capítulo 3. Los modelos seleccionados cumplen con los requerimientos estructurales para el análisis de variables espaciotemporales y fueron aplicados únicamente cuando el conjunto de datos satisfacía los supuestos mínimos necesarios para cada enfoque metodológico. En este estudio se evaluaron dos enfoques de modelado: modelos tradicionales y modelos de Machine Learning, con el propósito de establecer una base de comparación entre ambos. Este enfoque permitió contrastar el desempeño de técnicas clásicas con métodos más avanzados. A continuación, se describen los modelos evaluados y la metodología utilizada para analizar el desempeño de cada uno.

5.4.1 Modelos Tradicionales

A partir de la revisión de la literatura, se observa un consenso general respecto a que los datos de incidentes de emergencia suelen exhibir autocorrelación y patrones estacionales. Por esta razón, se evalúan los modelos temporales ARIMA y SARIMA, junto con sus supuestos. Estos modelos son teóricamente sólidos para capturar las dinámicas subyacentes y realizar predicciones precisas [29]. Además, se evaluó la regresión binomial dado que se detectó sobre-dispersión, con una varianza que triplica a la media en los datos de incidentes evaluados solo de forma temporal. Por lo tanto, el uso de un modelo de regresión binomial negativa está

teóricamente justificado, ya que permite capturar de manera más efectiva la naturaleza y la variabilidad de los conteos, dado que se basa en un modelo de mezcla Poisson-gamma [30].

Para evaluar los modelos temporales ARIMA y SARIMA, se utilizó una serie temporal que representa la cantidad de incidentes ocurridos en intervalos de 12 horas, comenzando a las 9 a.m. Los parámetros de los modelos fueron seleccionados mediante un proceso iterativo en Python, en el cual se buscaron aquellos que minimizaran la métrica AIC y maximizaran el ajuste del modelo a los datos; Obteniendo como mejores parámetros ARIMA (2,0,2) y SARIMA (0,0,2) (2,0,1)₁₂.

Por otra parte, se ajustó un modelo de regresión binomial negativa utilizando la función GLM de la biblioteca statsmodels. Este modelo se especificó con la variable dependiente siendo el conteo de incidentes (y_{train}) y las variables independientes representadas en la matriz de diseño (X_{train}); se utilizaron las variables descritas en la Tabla 2, y se usó la función de enlace logarítmica.

5.4.2 Modelos De Aprendizaje Automático

Los modelos probados, reconocidos por su importancia en la literatura y aplicación en problemas similares [20], fueron los siguientes:

- *Gradient Boosting Regressor (GBR)*: Modelo de ensamble que construye árboles de decisión de manera secuencial donde cada árbol intenta corregir los errores del árbol anterior, mejorando así el rendimiento en cada iteración [31]. Además, es reconocido por capturar interacciones no lineales de los datos. Para la ejecución de este modelo se toman 200 estimadores, profundidad máxima igual a 3 y tasa de aprendizaje de 0.05.
- *LightGBM (Light Gradient Boosting Machine)*: Es una variante del GBR con más alto rendimiento, ya que está optimizado para reducir el número de particiones y el crecimiento de hojas de los árboles para acelerar el proceso de entrenamiento manteniendo una buena precisión. La arquitectura escogida para este modelo es: tasa de aprendizaje de 0.03, profundidad máxima igual a 10, Boosting type igual a 'gbdt' (Gradient Boosting Decision Tree) y 1.000 iteraciones.
- *XGBoost (eXtreme Gradient Boosting)*: Es otro modelo de GBR optimizado para el rendimiento y velocidad en la fase de entrenamiento, ya que construye árboles en procesamiento paralelo [32]. Incluye mejoras en la regularización para evitar el sobreajuste. Para este modelo luego de varias pruebas se definió una tasa de aprendizaje de 0.02, profundidad máxima igual a 7 y una tasa de submuestras de 0.8.

- *MLP (Multilayer Perceptron)*: Esta es una arquitectura de red neuronal fully-connected, que hace parte de los modelos de aprendizaje profundo y es eficiente para capturar relaciones no lineales en los datos. Ha probado ser más eficiente que los modelos tradicionales para la predicción de demanda de ambulancias en diversos estudios [20]. La estructura utilizada se compone de 3 capas ocultas con función de activación ReLU, con regularización dropout y optimizador Adam.
- *RNN (Recurrent Neural Networks)*: Las redes neuronales recurrentes son diseñadas para manejar datos secuenciales, ya que tienen una memoria interna que les permite mantener información de pasos anteriores en la secuencia, lo que puede ser útil en este caso donde se tiene una dependencia temporal [33]. La estructura definida es de dos capas GRU (Gated Recurrent Unit) con activación ReLU, que introduce no linealidad a la red y le ayuda aprender patrones complejos, y dos capas densas también con activación ReLU más regularización dropout para prevenir el sobreajuste y mejorar la generalización.

Cada modelo fue ajustado utilizando las variables seleccionadas (Tabla 2), y los hiperparámetros fueron optimizados para maximizar el rendimiento en el conjunto de prueba.

5.5 COMPARACIÓN DE RESULTADOS

Con el objetivo de que la comparación fuera coherente se utilizaron las mismas variables para la evaluación, los resultados de los modelos se resumen a continuación:

Tabla 4. Comparación de resultados

Modelo	MAE	RMSE	R-squared
ARIMA (2,0,2)	20.05	24.30	
SARIMA (0,0,2) (2,0,1) ₁₂	28.30	36.35	
Regresión Binomial – Negativa	41.40	46.10	0.005
Gradient Boosting	1.122	1.545	0.411
LightGBM	1.061	1.465	0.476
XGBoost	1.058	1.457	0.478
MLP	1.161	2.329	0.317
RNN	1.130	2.282	0.325

Se concluye que los modelos de Boosting han demostrado un desempeño superior en comparación con los demás enfoques. En particular, el modelo XGBoost tiene una precisión ligeramente superior con los valores más bajos de MAE (1.058) y RMSE (1.457), así como el R^2 más alto (0.478), lo que sugiere que se ajusta mejor a los datos en comparación con los otros

modelos, pero sigue habiendo oportunidad de mejora para lograr un desempeño más eficiente del modelo.

Por otro lado, los modelos tradicionales muestran un desempeño considerablemente inferior en términos de MAE y RMSE, reflejando limitaciones para capturar las dinámicas espaciotemporales y variabilidad de los datos. En cuanto a las redes neuronales, si bien tienen el potencial de manejar relaciones no lineales complejas, es necesario tener presente que su arquitectura y el tamaño de los datos influyen significativamente en los resultados, y que el costo computacional de entrenar redes muy sofisticadas puede representar un desafío para la implementación práctica del modelo por parte del CRUE, lo que refuerza la ventaja de los modelos de boosting en este contexto. Aunque estos modelos lograron capturar ciertos patrones en los datos, su desempeño general no fue satisfactorio para cumplir con los objetivos establecidos, debido a limitaciones en su capacidad predictiva o en la representación de las características espaciales de los incidentes. Ante esta situación, se consideró necesario explorar nuevas alternativas metodológicas, incluyendo el ajuste de modelos más robustos y la redefinición de la unidad espacial de análisis, optando por un nivel de agregación más amplio que pudiera mejorar la calidad y precisión de las predicciones.

5.6 EXPLORACIÓN DE OTROS MODELOS

Se exploró el uso de modelos de procesos de Poisson con estructura de correlación espacial, específicamente el modelo de Cox log-gaussiano (LGCP), para capturar la dinámica de los incidentes de emergencia. Sin embargo, debido a la extensión de Bogotá y a la configuración de los datos a nivel de longitud y latitud, como es indispensable en este tipo de modelos tal como lo describen Bayisa y e.t. [21], las estimaciones obtenidas no lograron mostrar una correlación espacial suficiente. Esta falta de correlación impidió que el modelo capturara adecuadamente la variabilidad observada en los incidentes. Como resultado, se optó por enfoques alternativos que se ajustaran mejor a la naturaleza de los datos y a las condiciones del análisis.

Adicionalmente, con el fin de explorar otros tipos de técnicas y evaluar los datos desde diferentes perspectivas, se intentó analizar la distribución de incidentes en Bogotá a través de un modelo de proceso puntual espaciotemporal, inspirado en la técnica sugerida en “*Estimación no paramétrica y semiparamétrica*” de la interacción en patrones de puntos no homogéneos, donde se explica que “es posible realizar un estudio completamente no paramétrico de las interacciones utilizando un análogo de la función K” [34]. Este enfoque buscaba identificar patrones significativos en el tiempo y el espacio. Sin embargo, al emplear la distribución espacial en términos de puntos (latitud y longitud), en lugar de considerar variables como UPZ o UPL, los resultados mostraron que la distribución de los incidentes se tornó prácticamente aleatoria.

Aunque se lograron identificar ciertas concentraciones, la mayoría de los eventos no evidenciaron una tendencia clara hacia áreas específicas de la ciudad, ya que también se observó una notable dispersión.

Esta experiencia refuerza la idea de que es esencial abordar la espacialidad de Bogotá desde otro agrupamiento más amplio que permita tener más datos por unidad espacial y así los modelos puedan capturar mejor esta correlación.

5.7 CAMBIO DE UNIDAD ESPACIAL: DE UPZ A UPL

Durante el desarrollo del proyecto, se identificó la necesidad de ampliar la unidad espacial utilizada, ya que el desempeño de los modelos al predecir los incidentes no fue satisfactorio, mostrando un rendimiento apenas superior al azar. Específicamente, el modelo XGBoost alcanzó una capacidad predictiva máxima de solo el 48%. Debido a esto, se decidió agrupar los datos por las 31 UPLs en lugar de las 117 UPZs.

Este cambio responde, además, a necesidades operativas del CRUE, ya que las UPL representan un nivel de agregación más funcional para la planificación y asignación de recursos. Las UPL integran varias UPZs bajo un criterio de organización territorial definido en el nuevo Plan de Ordenamiento Territorial (POT) de Bogotá, permitiendo una mejor alineación con las dinámicas reales de atención médica de emergencia. Desde una perspectiva operativa, la gestión basada en UPL simplifica la asignación de ambulancias al reducir la fragmentación territorial, facilitando la planificación de rutas y la concentración de recursos en zonas críticas.

Según el POT Bogotá Reverdece 2022-2035, “Las UPL son 31 nuevas delimitaciones que redistribuyen las oportunidades y servicios esenciales para los habitantes de Bogotá. A través de ellas, se planean, organizan y gestionan las necesidades locales, permitiendo una administración más equitativa y efectiva” [35]. Estas unidades fueron distribuidas respetando las dinámicas sociales, culturales y económicas de cada territorio, garantizando que todas cuenten con un número de habitantes y una extensión territorial equilibrada. Las UPL surgieron como parte de una apuesta de justicia y equilibrio territorial y poblacional.

Además, trabajar con UPL permite considerar un mayor volumen de datos agregados por unidad espacial, lo que mejora la capacidad de los modelos predictivos para identificar patrones significativos de demanda. Para el CRUE, este nivel de agregación ofrece una visión más estratégica y operativa, ya que se priorizan áreas de intervención más amplias, con mayor coherencia en términos de tráfico, densidad poblacional y características urbanas, elementos clave en la toma de decisiones durante emergencias. En contraste, las UPZs, al ser más

pequeñas, generaban una dispersión de los datos que dificultaba la identificación de tendencias relevantes y complicaba la coordinación efectiva en incidentes.

Mientras que la UPZ es una división orientada principalmente a la planificación urbanística, las UPL ofrecen un nivel de granularidad más adecuado para la distribución y administración de recursos médicos. Además, cabe resaltar que este cambio beneficia al usuario final que es el CRUE, ya que las UPL se alinean mejor con el nuevo POT de la ciudad que se encuentra en implementación. En resumen, el cambio a las UPL no solo responde a la necesidad de mejorar el desempeño de los modelos predictivos, sino que también alinea los resultados del análisis con las necesidades operativas del CRUE, optimizando la gestión de recursos y mejorando la capacidad de respuesta ante emergencias en Bogotá.

Impacto en los datos

Se reasignaron los incidentes registrados en la base de datos a su respectiva UPL utilizando las coordenadas geográficas de cada evento. Esta transformación espacial implicó adaptar los datos, que originalmente estaban clasificados por UPZ, a la nueva división geográfica basada en UPL. Como resultado, el número de registros pasó de 138.907 agrupados por UPZ a 52.850 por UPL. Este proceso permitió recalcular la distribución de incidentes de manera más precisa, facilitando la detección de patrones espaciales que podrían haber pasado desapercibidos con la clasificación original. En la Figura 18, se presenta la dinámica de incidentes por UPL, destacando como la nueva delimitación espacial para Bogotá.

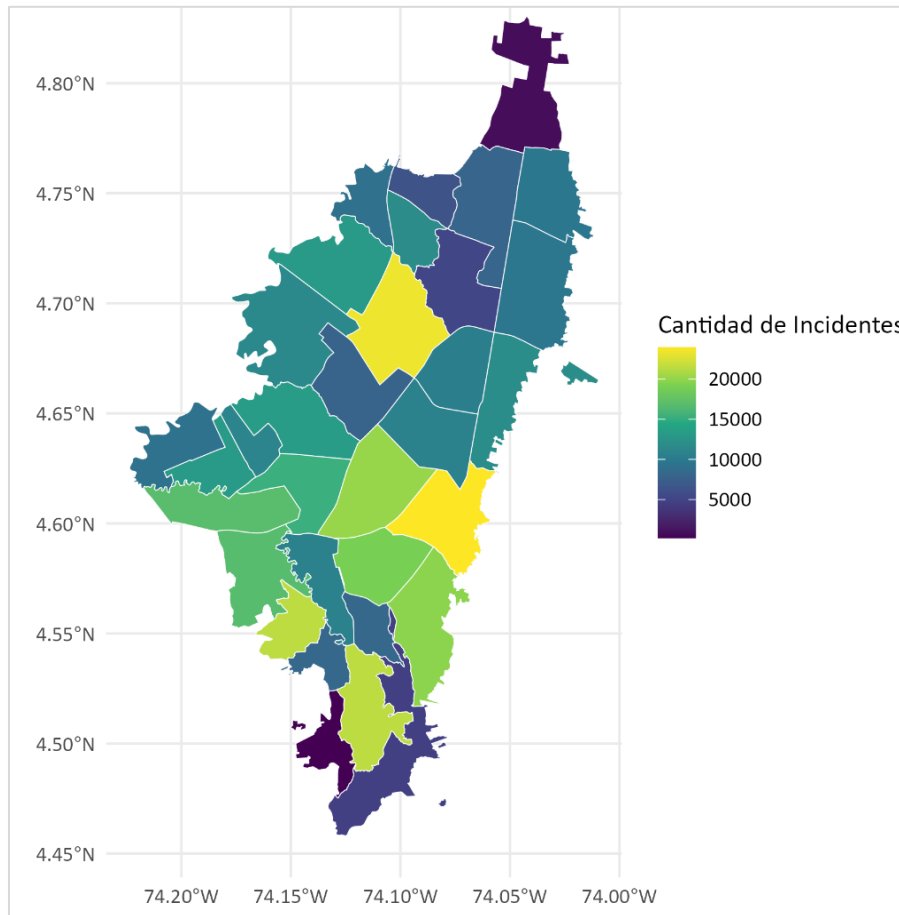


Figura 18. Cantidad de Incidentes por UPL.

5.7.1 Evaluación de Modelos Tradiciones

Se procedió a ejecutar nuevamente los modelos previamente implementados (ARIMA, SARIMA y Regresión Binomial Negativa) utilizando la nueva estructura de datos. En esta ocasión, los incidentes se agregaron a nivel de UPL y en los mismos intervalos de 12 horas, lo que permitió un análisis espaciotemporal que se adaptara a la necesidad de pronóstico por unidad de espacio y tiempo a la vez. En el caso de la Regresión Binomial Negativa. Inicialmente, la ejecución se realizó con las mismas variables utilizadas en el análisis mostrado en la Tabla 2, con el fin de mantener la consistencia en la comparación de los resultados. Esta reevaluación con los nuevos datos permitió una mayor precisión en la estimación de los incidentes y una mejor comprensión de las dinámicas temporales en las UPL, obteniendo los resultados de la Es importante resaltar que, para garantizar una comparación coherente, esta segunda evaluación de modelos se llevó a cabo utilizando el mismo conjunto de variables previamente seleccionado, garantizando que cada modelo se evaluara con las mismas variables.

Tabla 5.

5.7.2 Evaluación de Modelos de Aprendizaje Automático

En este nuevo escenario se probaron los modelos XGBoost, MLP y RNN, separando los datos de entrenamiento y prueba de la misma forma que en la ejecución anterior. A continuación, se describe la arquitectura usada en cada uno:

- *XGBoost*: Se define una tasa de aprendizaje igual a 0.05, número de árboles paralelos construidos durante cada iteración ($n_{estimators}$) igual a 100, y una semilla para replicar los resultados.
- *MLP*: Compuesto por tres capas ocultas con tamaños de 64, 32 y 16 neuronas respectivamente, cada una utilizando la función de activación ReLU, regularizaciones L2 y Dropout, Batch Normalization y la capa de salida.
- *RNN*: Este modelo utiliza una arquitectura híbrida, es decir, procesa las variables temporales utilizando una capa LSTM con 128 unidades, que captura relaciones secuenciales, mientras que para las variables de ubicación y contexto se utilizan capas densas con 16 y 64 neuronas, respectivamente. Luego se integran estas salidas mediante una capa de concatenación y se generan las predicciones a través de una capa densa final utilizando regularización L2 y Dropout para evitar el sobreajuste.

5.7.3 Comparación Final de Resultados

Es importante resaltar que, para garantizar una comparación coherente, esta segunda evaluación de modelos se llevó a cabo utilizando el mismo conjunto de variables previamente seleccionado, garantizando que cada modelo se evaluara con las mismas variables.

Tabla 5. Desempeño de los Modelos

Modelo	MAE	RMSE	R-squared
ARIMA (2,0,2)	4.174	3.341	
SARIMA (0,0,2) (2,0,1) ₁₂	4.168	3.312	
Regresión Binomial – Negativa	4.112	1.441	0.269
XGBoost	1.528	1.982	0.797
MLP	2.963	3.927	0.227
RNN	3.057	2.282	0.231

Representación visual de las métricas de evaluación:

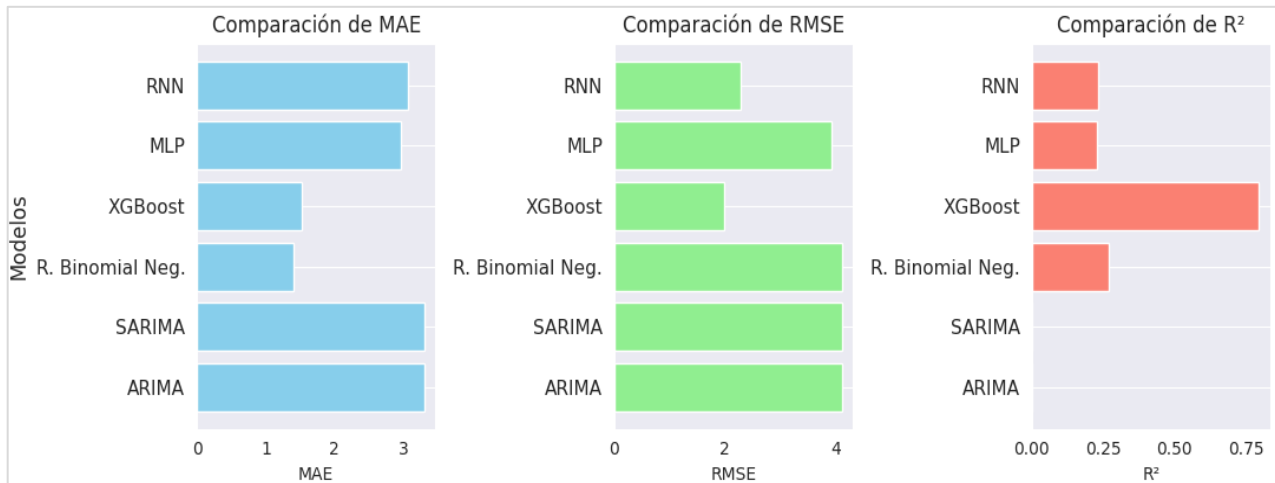


Figura 19. Comparación de las métricas de desempeño de los modelos.

Los resultados obtenidos en la comparación de los modelos evaluados con la nueva unidad espacial, tanto clásicos como de aprendizaje automático, destacan que el modelo XGBoost sigue mostrando el mejor desempeño general en términos de precisión, con un MAE de 1.528, un RMSE de 1.982 y un R^2 de 0.797. Esto sugiere que este modelo logra capturar patrones espaciotemporales complejos de manera más eficiente en comparación con los otros enfoques como lo muestra la Figura 19.

Desde una perspectiva operativa, los resultados refuerzan la elección de XGBoost como el mejor modelo para la implementación en el CRUE. Este modelo no solo ofrece una precisión superior en la predicción de incidentes, sino que también se adapta mejor al cambio de unidad espacial de UPZ a UPL, facilitando la planificación estratégica y la toma de decisiones.

El desempeño inferior de las redes neuronales destaca la importancia de seguir recopilando datos más extensos y consistentes, lo que permitirá en el futuro aprovechar al máximo su potencial. Adicionalmente, la comparación con modelos clásicos subraya la importancia de contar con enfoques más sofisticados como XGBoost, especialmente en un entorno operativo donde se manejan datos con alta dimensionalidad y variabilidad espaciotemporal.

6 IMPLEMENTACIÓN DEL MODELO PREDICTIVO ESPACIOTEMPORAL

En esta sección se describe el desarrollo y entrega del producto final para el CRUE: un modelo predictivo basado en aprendizaje automático diseñado para proyectar la demanda de servicios de emergencias médicas. Este modelo, desarrollado como parte del objetivo específico tres, no solo demuestra la aplicabilidad de técnicas avanzadas de predicción, sino que también resalta su rendimiento en un contexto real, y se detalla cómo el CRUE puede hacer uso del modelo pre-entrenado, integrándolo en sus operaciones diarias para tomar decisiones basadas en datos.

6.1 CRITERIOS DE SELECCIÓN

Para la implementación del modelo predictivo espaciotemporal se seleccionó el algoritmo XGBoost Regressor, debido a su rendimiento superior en comparación con los otros modelos evaluados. La selección del modelo se basó en los criterios estándar para evaluar modelos predictivos, los cuales incluyen aspectos como la precisión, interpretabilidad, capacidad de generalización y complejidad computacional [36]. A continuación, se detallan algunos de estos criterios:

- **Precisión y Desempeño:** El objetivo principal en la selección de un modelo es maximizar la precisión de las predicciones, y para problemas de regresión como este, una de las medidas más utilizadas es el RMSE que mide el error cuadrático medio y brinda una idea sobre que tanto error está teniendo el modelo en las predicciones, asignando un peso mayor a errores más grandes [37]. Sin embargo, dependiendo del contexto, en algunos casos especialmente cuando existen valores atípicos, es adecuado usar otras métricas como el MAE. RMSE y MAE son formas de medir la distancia entre el vector de valores reales y el vector de valores predichos. Por otro lado, el R^2 indica qué tan bien se ajustan las predicciones a los valores reales, con valores cercanos a 1 reflejando mejor ajuste.
- **Manejo Eficiente de Datos Desbalanceados:** En el contexto de la predicción de emergencias médicas, la distribución de los incidentes es desbalanceada, como se observó en el análisis exploratorio, con eventos que varían significativamente según la zona geográfica y el momento temporal. Modelos como XGBoost, gracias a su capacidad para incorporar regularización y ajustes finos en la construcción de árboles de decisión, han mostrado ser particularmente efectivos en manejar estos escenarios, logrando una buena precisión y evitando el sobreajuste.
- **Complejidad Computacional:** Es necesario considerar la complejidad computacional del modelo, ya que modelos con arquitecturas complejas como las redes neuronales pueden requerir un mayor tiempo de entrenamiento y mayores recursos computacionales, mientras que modelos como el XGBoost destacan por su eficiencia computacional.

- **Interpretabilidad:** Al momento de generar conclusiones sobre los resultados, es importante que el modelo permita tener una interpretación razonable sobre su desarrollo, por esto, el modelo XGBoost también resulta bastante adecuado ya que permite extraer la importancia de características lo que facilita la comprensión de cómo las diferentes variables afectan las predicciones.

6.2 VARIABLES FINALES

Luego de realizar diferentes pruebas y análisis, como el de correlación de variables y la evaluación del impacto en el rendimiento del modelo, se logró determinar que las características más representativas para el buen desempeño del modelo son las que se muestran en la Tabla 6:

Tabla 6. Variables del modelo final.

Variable	Tipo Variable
Cod_UPL	Explicativa
Rango_Hora	Explicativa
Prioridad_Inicial_Moda_Crítica	Explicativa
Tipoincidente_Herido	Explicativa
día_semana	Explicativa
día_del_año	Explicativa
Festivo	Explicativa
Infantes_0_4	Explicativa
Adultos_mayores_60	Explicativa
Densidad_poblacional	Explicativa
Cant_eventos_esp	Explicativa
Incidentes_día_anterior	Explicativa
Total_Incidentes	Dependiente

Estas variables fueron seleccionadas no solo por su correlación con la variable dependiente, sino también por su capacidad para capturar patrones complejos de la demanda de emergencias médicas en función del tiempo y la localización.

6.3 IMPLEMENTACION DEL MODELO XGBOOST REGRESSOR

Teniendo en cuenta las métricas de evaluación presentadas (Tabla 5), el modelo XGBoost mostró ser el más robusto y eficaz para predecir la demanda espaciotemporal de emergencias médicas por UPL en Bogotá, superando a otros algoritmos probados tanto tradicionales como de aprendizaje de máquina. Su capacidad para manejar datos desbalanceados y capturar patrones complejos en los datos, resalta su potencial para realizar predicciones precisas en este contexto donde variabilidad espacial y temporal son muy influyentes. El modelo logró captar el comportamiento y generar una buena predicción como lo muestra la Figura 20.

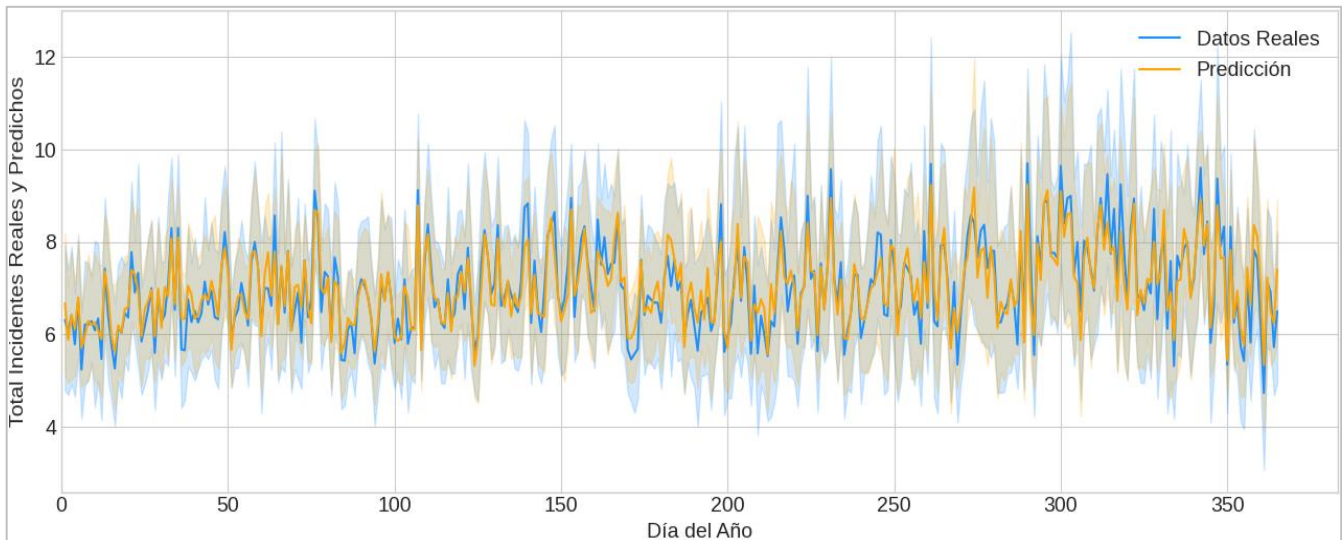


Figura 20. Datos reales Vs pronostico del modelo.

En la figura anterior (20) se presenta una comparación visual entre los valores reales de la demanda de emergencias médicas y las predicciones realizadas por el modelo XGBoost en el conjunto de prueba a lo largo de un año, representando los días del año en el eje horizontal y el número de incidentes en el eje vertical. Las líneas azules representan los datos reales, mientras que las líneas naranjas muestran los valores predichos. Además, las bandas sombreadas reflejan la dispersión o incertidumbre en ambos casos, proporcionando una indicación de la variabilidad en los datos y las predicciones.

El modelo muestra un ajuste cercano entre los valores reales y predichos, con una buena capacidad para capturar las fluctuaciones a lo largo del tiempo, especialmente en zonas de alta demanda. Aunque hay momentos en los que las predicciones tienden a suavizar los picos presentes en los datos reales, en general, el modelo logra seguir adecuadamente los patrones observados. Esta visualización refuerza los resultados de las métricas de evaluación, como el MAE y el RMSE, que indicaron un buen desempeño en términos de precisión y capacidad predictiva.

De otra parte, las variables más importantes para predecir la demanda de emergencias médicas según los resultados del modelo seleccionado son las que se muestran en la Figura 21. Este enfoque sigue las mejores prácticas recomendadas para la selección de características y validación orientada al objetivo, como se sugiere en estudios recientes sobre la predicción espaciotemporal en modelos de aprendizaje automático, que destacan la importancia de implementar estrategias de validación orientadas a evitar el sobreajuste y mejorar la precisión del modelo [38].

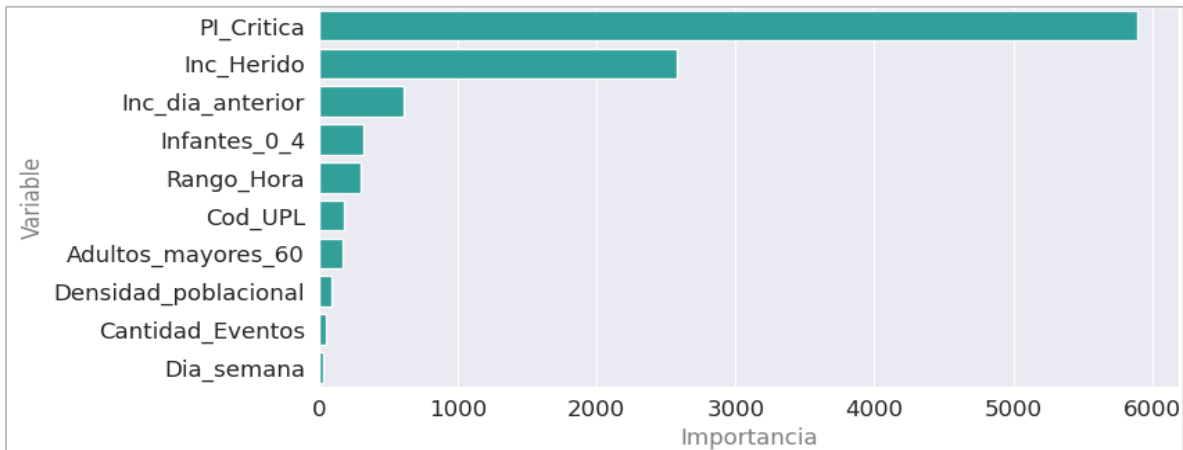


Figura 21. Importancia de variables.

Dentro de los factores más influyentes para la predicción de la demanda de emergencias médicas, se encuentra el tipo de prioridad inicial crítica y los incidentes relacionados con heridos, lo cual es consistente con la naturaleza de la atención prioritaria en situaciones críticas. Además, los incidentes del día anterior también tienen un papel relevante que sugiere una correlación temporal que afecta la demanda del día actual. Otras variables como el número de infantes, adultos mayores y el rango horario también contribuyen, aunque en menor medida, lo que indica que la demanda varía según los grupos etarios y momentos del día. De forma similar, la ubicación por UPL es significativa, demostrando que el componente espacial también es un factor a tener en cuenta, y que es importante diferenciar los sectores de la ciudad ya que las características de demanda pueden ser muy diferentes de una zona a otra.

Finalmente, variables como la densidad poblacional, eventos especiales y el día de la semana muestran una menor importancia en el modelo, lo que indica que, aunque influyen en la predicción, su impacto es menor en comparación con las demás variables.

El repositorio alojado en GitHub documenta de manera detallada cada etapa del desarrollo del modelo predictivo. Incluye desde el procesamiento y limpieza de datos, el análisis exploratorio y la selección de variables, hasta la implementación y evaluación de los modelos predictivos, tanto tradicionales como basados en aprendizaje automático. Además, contiene los scripts utilizados para generar las visualizaciones geoespaciales presentadas en este documento. Este repositorio ofrece una guía reproducible, útil tanto para la implementación del modelo por parte del CRUE, como para estudios futuros o para la implementación del modelo en otros contextos. Desde el siguiente enlace se puede acceder al repositorio: <https://github.com/AndreaAmaya19/PUJC-Tesis>

6.4 DESPLIEGUE DEL MODELO

El despliegue del modelo predictivo está diseñado para ser utilizado por el CRUE con un enfoque práctico y accesible para operadores con o sin experiencia en Machine Learning. A continuación, se describe cómo se puede implementar y usar la herramienta:

1. Entrega del Modelo Pre-entrenado

El modelo entrenado se entrega en formato serializado (modelo_xgb.pkl), que puede cargarse fácilmente en un entorno Python, Jupyter o en Google Colab el notebook proporcionado.

2. Notebook de Implementación

El notebook 'Implementación del Modelo Predictivo', disponible en el repositorio de GitHub, incluye instrucciones detalladas para:

- a) Preparar el entorno de trabajo y cargar las librerías necesarias.

1. Preparación del Entorno

Antes de comenzar, asegúrese de tener los siguientes elementos:

1. Acceso a este notebook desde Google Colab.
2. El archivo CSV que contenga los datos necesarios, como incidentes históricos, variables climáticas, y demográficas.
3. Cuenta de Google conectada para guardar y cargar datos.

2. Librerías Necesarias

Primero, se cargan las librerías que nos ayudarán a ejecutar el modelo.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import xgboost as xgb
import joblib
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from sklearn.preprocessing import StandardScaler
from google.colab import drive, files
```

b) Validar la estructura de los datos e introducir datos nuevos.

3. Cargar los Datos

Cargaremos los datos históricos que utilizaremos para realizar las predicciones. Asegúrese de que el archivo esté correctamente cargado en su entorno de Colab, y que siga la estructura descrita en `data_structure/estructura de datos para implementación del modelo.txt`.

```
# Definir ruta del archivo
drive.mount('/content/drive')
# Subir el archivo CSV
Datos = pd.read_csv(f"/content/drive/MyDrive/Proyecto de grado/Datos_sds/datos_pred_julio.csv", delimiter=',', encoding='')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
# Visualizar las primeras filas
Datos.head()
```

	cod_incidente	hora_fecha_inicial	Anio	Mes	Dia_mes	Infantes_0_4	Niños_5_14	Adolescentes_15_19	Adultos_20_59	Adultos_mayor
0	CRU-0100000-24	7/1/2024 7:30	2024	7	1	513	1754	883	6402	

c) Preparar los datos antes de ejecutar el modelo.

4. Preparación de los Datos

El modelo necesita que los datos estén organizados y limpios. Este paso se encarga de preparar las columnas necesarias.

Inicialmente se agrupan los datos por las variables fecha del incidente, código de UPL y rango de hora del incidente.

```
df_agrupado = Datos.groupby(['Fecha', 'Cod_UPL', 'Rango_Hora']).agg(
    Total_Incidentes=('cod_incidente', lambda x: x.nunique()),
    Anio = ('Anio', 'max'),
    Mes = ('Mes', 'max'),
    Dia_mes = ('Dia_mes', 'max'),
    Dia_del_anio = ('dia_del_anio', 'max'),
    Dia_semana = ('dia_semana', 'max'),
    Festivo = ('Festivo', 'max'),
    Cantidad_Eventos = ('Cantidad Eventos', 'max'),
    Aforo_Total_Eventos = ('Aforo_total', 'max'),
    Precipitacion_mm = ('ValorPrecip', 'max'),
```

d) Cargar el modelo pre-entrenado.

5. Cargar el Modelo Preentrenado

El modelo ya ha sido entrenado previamente. Aquí cargaremos ese modelo para usarlo directamente.

```
modelo_xgb = joblib.load('/content/drive/MyDrive/Proyecto de grado/Datos_sds/modelo_xgb.pkl')  
print("Modelo cargado exitosamente")
```

Modelo cargado exitosamente

e) Generar predicciones utilizando el modelo.

En el ejemplo de implementación, se realizan predicciones para la demanda de la primera semana de julio de 2024, simulando algunos datos que el modelo no ha visto en su entrenamiento. De esta forma se garantiza que este esté generalizando adecuadamente a nueva información.

6. Realizar Predicciones

Utilizaremos el modelo cargado para predecir la demanda en los próximos días.

```
# Seleccionar las columnas necesarias para las predicciones  
X = df[['Cod_UPL', 'Dia_semana', 'Dia_del_anio', 'Rango_Hora', 'Densidad_poblacional', 'Adultos_mayores_60'],
```

```
# Realizar predicciones  
df['prediccion'] = modelo_xgb.predict(X)
```

```
# Visualizar resultados  
df[['Fecha', 'Rango_Hora', 'prediccion']].head(5)
```

	Fecha	Rango_Hora	prediccion
15	2024-07-02 11:56:00	1	3.946150
36	2024-07-03 06:50:00	0	2.477051

f) Visualizar las predicciones y exportar los resultados en un archivo csv.

7. Visualización de Resultados

Generamos gráficos para mostrar las predicciones de manera visual y comprensible.

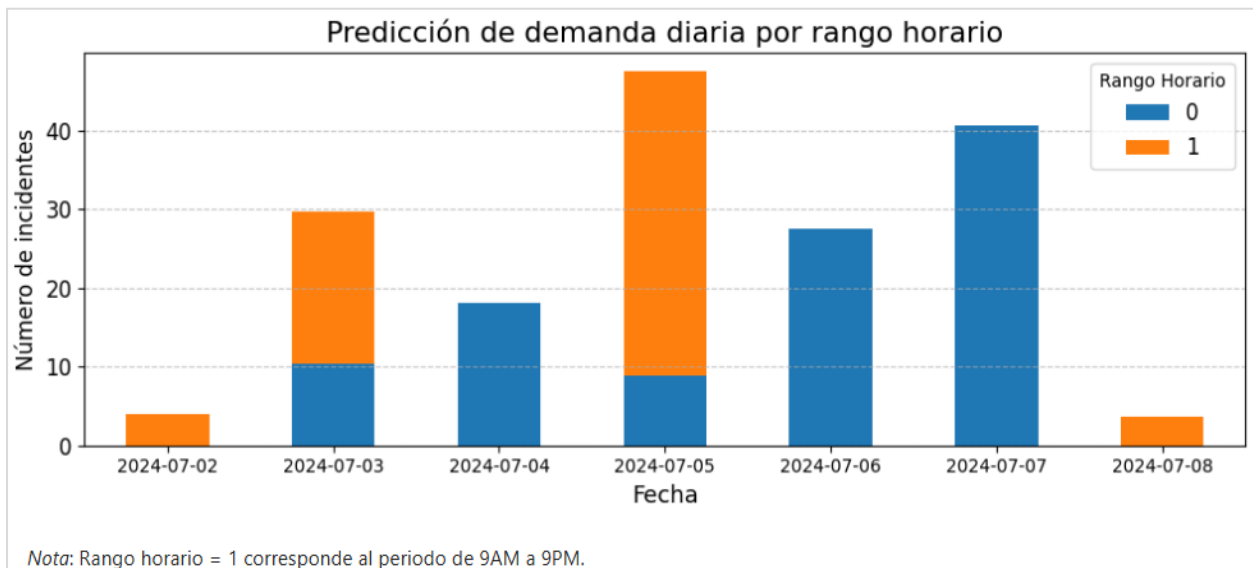
```

dia_hora = df.groupby([df['Fecha'].dt.date, 'Rango_Hora'])['prediccion'].sum().unstack()

plt.figure(figsize=(10, 4))
dia_hora.plot(kind='bar', stacked=True, figsize=(10, 4), color=['#1f77b4', '#ff7f0e', '#2ca02c'])
plt.title('Predicción de demanda diaria por rango horario', fontsize=16)
plt.xlabel('Fecha', fontsize=13)
plt.ylabel('Número de incidentes', fontsize=12)
plt.legend(title='Rango Horario', fontsize=12)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.xticks(rotation=0, fontsize=10)
plt.yticks(fontsize=12)
plt.tight_layout()
plt.show()

```

<Figure size 1000x400 with 0 Axes>



3. Uso Operacional

El modelo está preparado para integrarse en el flujo de trabajo diario del CRUE. Los pasos básicos para su uso incluyen:

- a) **Preparar Datos:** El equipo debe cargar datos actualizados en formato CSV siguiendo la estructura definida (descrita en el archivo 'estructura de datos para implementación del modelo.txt').

- b) **Ejecutar el Notebook:** Los operadores pueden seguir las instrucciones en el notebook para generar predicciones.
- c) **Interpretar Resultados:** Las predicciones podrán ser descargadas en un archivo CSV para su revisión e integración en la planificación operativa diaria desde el tablero de control entregado.

4. Requisitos Técnicos


- a) **Entorno:** Google Colab o cualquier equipo con Python 3.8+ instalado.
- b) **Dependencias:** Se incluye el archivo 'requirements.txt' con las librerías necesarias para ejecutar el modelo.
- c) **Guía de Ejecución:** En el archivo ReadMe del repositorio se brinda una Guía de Uso con el detalle necesario para ejecutar adecuadamente los códigos de implementación del modelo.

Guía de Uso

1. Requisitos Previos

1. Asegúrate de tener instalado Python 3.8 o superior.
2. Comprueba que la estructura de la base de datos en cada parte del procesamiento.
3. Instala las dependencias necesarias ejecutando:

```
pip install -r requirements.txt
```



Finalmente, el despliegue del modelo predictivo espaciotemporal brinda al CRUE una herramienta completa y accesible para anticipar la demanda de emergencias médicas diaria por rango de horas y por localidades, optimizando así la asignación de recursos y mejorando la capacidad de respuesta. Este modelo, diseñado para su fácil implementación mediante un notebook interactivo, permite a los operadores obtener predicciones precisas con pasos claros y definidos. Cabe resaltar que el éxito y la continuidad de esta solución dependerán del compromiso del CRUE en mantener el modelo actualizado con datos nuevos, revisar periódicamente su desempeño, y realizar ajustes según los cambios en los patrones de demanda.

7 DESARROLLO DEL TABLERO DE CONTROL INTERACTIVO

El avance tecnológico ha permitido que la visualización de datos se convierta en una herramienta clave para la toma de decisiones en diversas áreas, incluyendo la gestión de emergencias médicas. La capacidad de transformar grandes volúmenes de datos en información visualmente comprensible facilita la comprensión rápida y efectiva de los resultados obtenidos por los modelos predictivos. Como afirma Few (2012), “el objetivo principal de la visualización de datos es comunicar información de manera clara y eficiente a través de medios gráficos, permitiendo que los tomadores de decisiones detecten patrones, anomalías y tendencias de forma rápida” [39].

El tablero de control desarrollado en Power BI está diseñado para visualizar y analizar la demanda de servicios de emergencias médicas en Bogotá, segmentada por UPLs. Este instrumento busca entender las dinámicas y distribuciones geográficas de los incidentes en la ciudad, permitiendo una asignación más eficiente de recursos y una respuesta más rápida en situaciones críticas.

7.1 ESTRUCTURA DEL TABLERO

El tablero de control interactivo está compuesto por dos hojas principales, cada una con un enfoque particular en los datos y las predicciones del modelo. A continuación, se describen las secciones y funcionalidades de cada hoja:

7.1.1 Hoja No. 1: Visualización de Incidentes por UPL y Parámetros Temporales

- **Mapa de Incidentes por UPL**

Muestra la distribución geográfica de los incidentes médicos en Bogotá utilizando un código de colores para representar la cantidad de incidentes en cada UPL, facilitando así la identificación de las áreas con mayor o menor demanda de servicios de emergencias médicas, como se observa en la Figura 22. Además, el mapa incluye tres botones que permiten realizar zoom (acercar con el botón "+" y alejar con el botón "-") y seleccionar un área específica utilizando un rectángulo. La escala de colores va desde el rojo, que indica las UPL con mayor concentración de incidentes, pasando por tonos amarillos para niveles intermedios, y finalizando en verde para las áreas con menor cantidad de incidentes. Esta información visual se refuerza con el tamaño de los círculos, donde los más grandes representan un mayor número de incidentes.

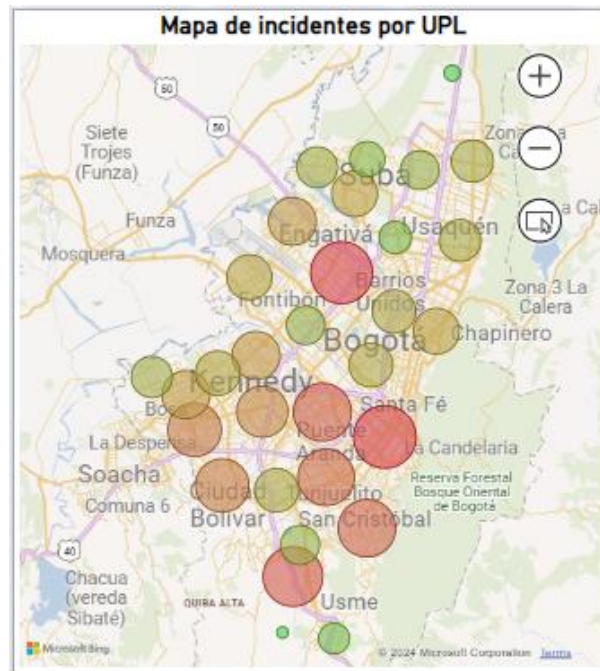


Figura 22. Mapa de incidentes – Tablero de control.

- **Gráfico de Incidentes por Fecha**

Representa la evolución temporal de los incidentes mediante un gráfico de líneas, lo que permite visualizar la variación en la ocurrencia de estos a lo largo del tiempo. Esta representación es especialmente útil para identificar tendencias, como picos de actividad o patrones estacionales recurrentes.

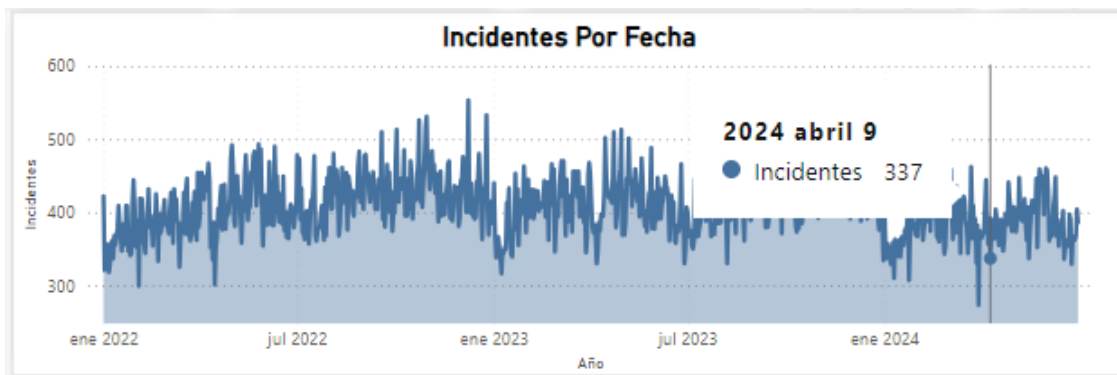


Figura 23. Incidentes por fecha – Tablero de control.

Además, el gráfico (Figura 23) incorpora una funcionalidad interactiva que permite, al desplazar el cursor sobre la línea, obtener información precisa sobre la cantidad exacta de incidentes registrados en cada día específico, facilitando un análisis detallado en puntos temporales clave.

- **Gráfico de Incidentes por Hora del Día**

Desglosa los incidentes por hora del día mediante un gráfico de barras, permitiendo identificar con precisión los periodos de mayor y menor demanda de servicios médicos. Esta información es fundamental para optimizar la planificación de recursos y ajustar la disponibilidad del personal según las horas de mayor carga operativa.

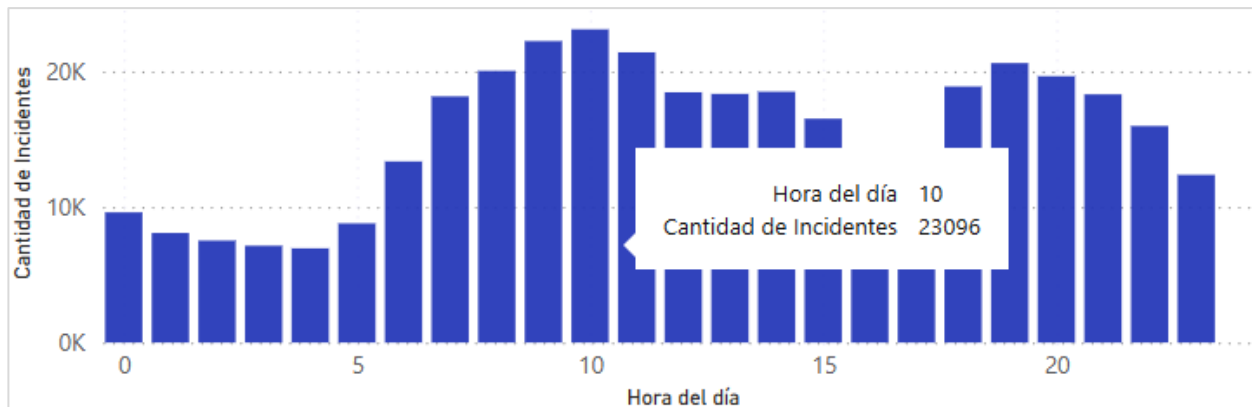


Figura 24. Incidentes por Hora del Día – Tablero de Control.

El gráfico (Figura 24) incluye una funcionalidad interactiva que permite, al posicionar el cursor sobre cada barra, visualizar de forma detallada el número exacto de incidentes registrados en cada franja horaria, facilitando un análisis más granular,

- **Gráfico de Incidentes por UPL**

Muestra la cantidad total de incidentes por cada UPL en un gráfico de barras horizontales. Este gráfico complementa el mapa, ofreciendo un desglose cuantitativo que permite comparar la carga de incidentes entre diferentes UPL como se puede observar en la Figura 25.

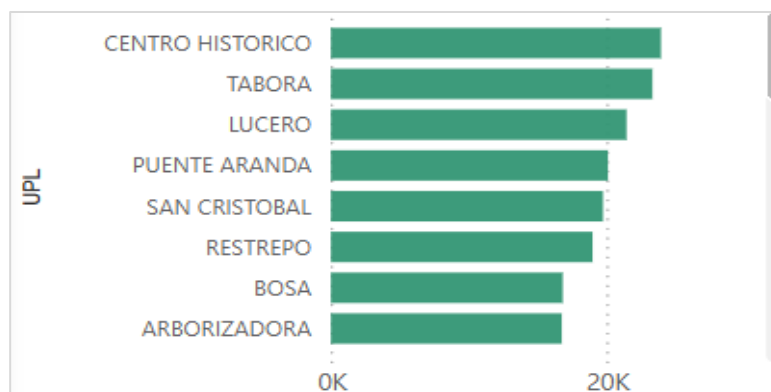


Figura 25. Incidentes por UPL – Tablero de Control.

7.1.2 Hoja No. 2: Comparación de Datos históricos y las predicciones del modelo.

- **Incidentes Reales vs. Incidentes Predichos por UPL**

Desglosa los incidentes observados y los pronósticos del modelo por UPL en un gráfico de barras (Figura 26). Permite una rápida visualización de las diferencias entre los valores reales y los estimados, evaluando la precisión del modelo predictivo.

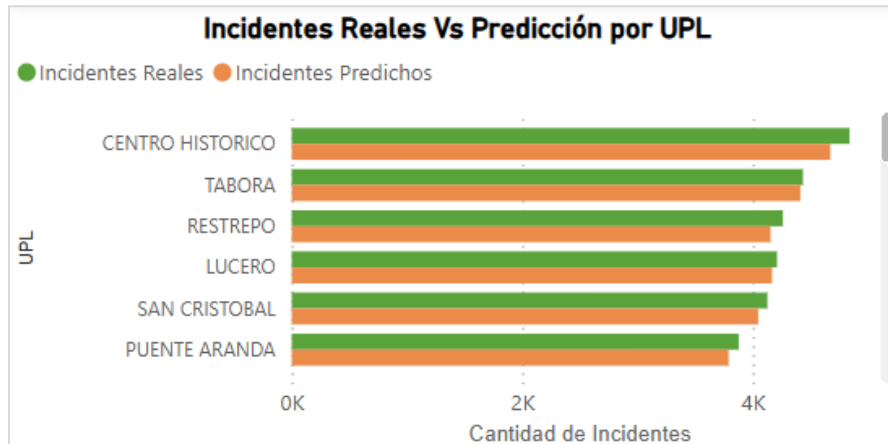


Figura 26. Comparativo Incidentes - Tablero de control.

- **Gráfico de Incidentes Reales vs. Predichos por día**

El gráfico de líneas superpuestas (Figura 27) compara de manera simultánea los incidentes observados y los predichos para cada día del año, proporcionando una evaluación visual de la precisión del modelo predictivo. Esta representación es fundamental para analizar el desempeño del modelo en diferentes días del año, ya que permite identificar de forma clara las divergencias entre los valores reales y los estimados.

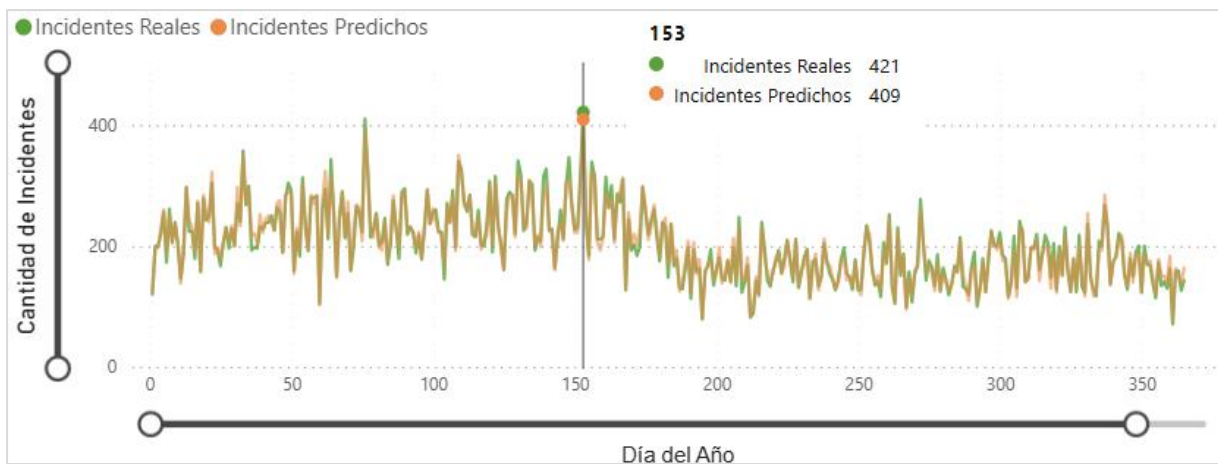


Figura 27. Incidentes Reales Vs Predicción - Tablero de Control.

La cercanía entre ambas líneas (real y predicción) refleja la capacidad del modelo para capturar adecuadamente la dinámica de los incidentes.

- **Mapa de Incidentes con las Predicciones**

Este mapa observado en la Figura 28 representa la distribución geoespacial de los incidentes predichos por el modelo, utilizando una escala de colores para indicar la magnitud de los incidentes proyectados en cada UPL. Siguiendo el mismo esquema que en el primer mapa, los colores van desde el rojo para las UPL con mayor concentración de incidentes predichos, pasando por tonos intermedios en amarillo, hasta el verde para aquellas con menor cantidad de eventos estimados.

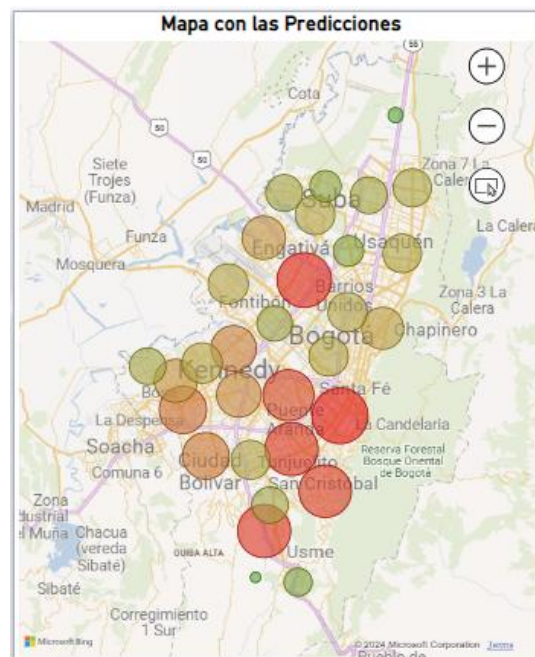


Figura 28. Mapa predicción - Tablero de Control.

Esta codificación visual facilita una rápida identificación de las áreas más vulnerables o críticas en términos de demanda proyectada. Además, el mapa permite una comparación directa con los incidentes observados, lo que ayuda a detectar discrepancias geográficas y ajustar el modelo predictivo según las necesidades operativas.

7.2 FUNCIONALIDADES DINÁMICAS DEL TABLERO

El tablero está diseñado para ser altamente interactivo, permitiendo al usuario filtrar los datos según varios criterios como la UPL, el año, el mes, el tipo de incidente, la prioridad de la emergencia y el día de la semana. Esto posibilita a los usuarios personalizar la visualización de

los datos según sus necesidades específicas. Además, la integración de mapas interactivos permite que los tomadores de decisiones visualicen de manera rápida la distribución espacial de los incidentes y sus predicciones, ofreciendo una ventaja estratégica para la asignación de recursos.

La herramienta también incluye gráficos dinámicos que permiten identificar tendencias a lo largo del tiempo, patrones según la hora del día y una comparación directa entre los incidentes reales y los pronósticos, lo que facilita la evaluación de la efectividad del modelo predictivo. Este enfoque basado en la interactividad y en el análisis visual acelera el proceso de interpretación y reduce la carga cognitiva en la toma de decisiones.

En el siguiente enlace se puede acceder al tablero de control elaborado a modo de consulta:

[Tablero de Control](#)

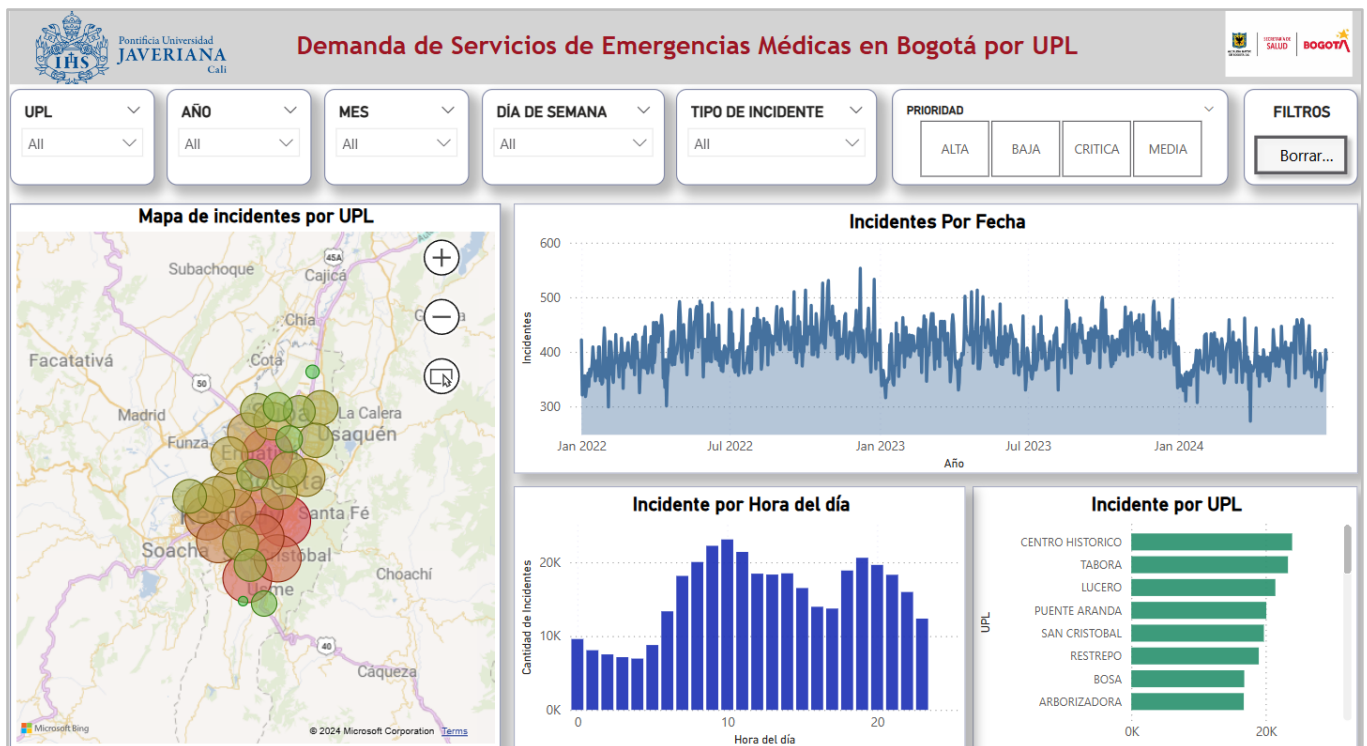


Figura 29. Vista descriptiva – Tablero de control.

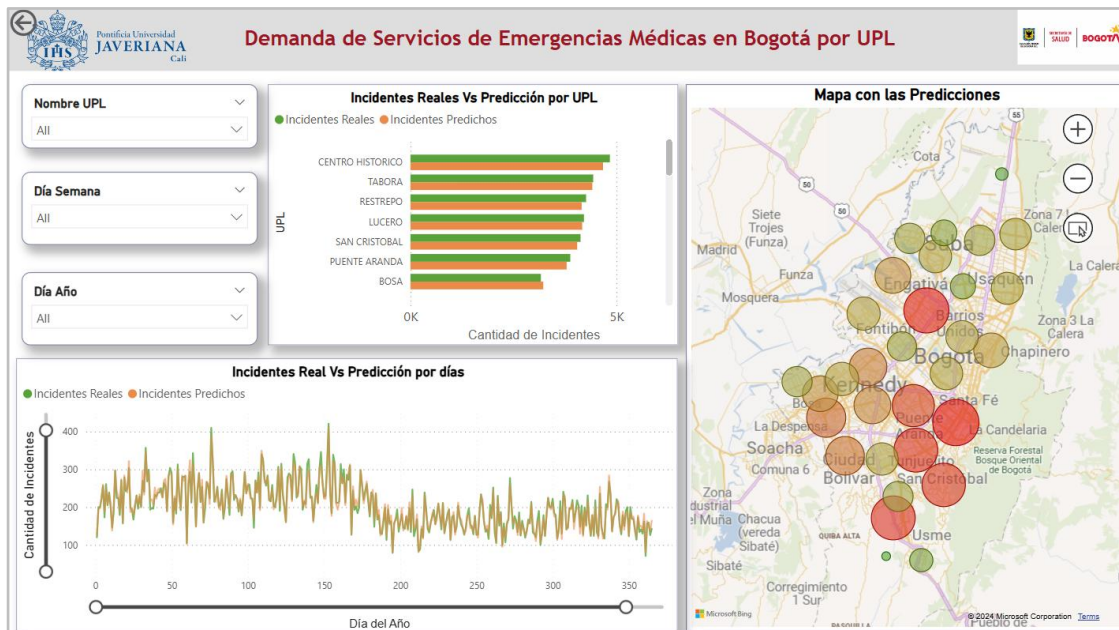


Figura 30. Vista predictiva – Tablero de control.

La Figura 29 y Figura 30 muestran el resultado final del dashboard realizado en Power BI. Es importante destacar que el uso de herramientas de visualización, como las presentadas en este tablero de control, no solo facilita la interpretación de los datos, sino que también potencia la toma de decisiones estratégicas en tiempo real. La integración de modelos predictivos y su representación gráfica, en combinación con herramientas interactivas, permite a los usuarios explorar patrones complejos y obtener una comprensión más profunda de la demanda de servicios médicos en diferentes UPL. Esto es especialmente valioso en el contexto de la gestión de emergencias, donde la optimización de recursos puede tener un impacto directo en la atención y respuesta ante situaciones críticas.

En este sentido, el concepto de Business Intelligence (BI) juega un papel fundamental al permitir que los usuarios no solo accedan a una visualización estática de los datos, sino que también puedan interactuar dinámicamente con ellos, obteniendo insights valiosos de manera intuitiva y eficiente. El BI se refiere a la capacidad de transformar datos en información procesable y útil, facilitando la toma de decisiones informadas mediante herramientas como gráficos interactivos y mapas. Según Turban "el BI combina una serie de metodologías, procesos y herramientas que permiten analizar grandes volúmenes de datos para generar informes que apoyen la toma de decisiones" [40]. En este tablero, los gráficos y mapas no solo presentan los datos, sino que también permiten la exploración activa, lo que amplía su utilidad al brindar la posibilidad de ajustar estrategias según los cambios en la demanda y ofrecer una visión integral y adaptable del panorama de emergencias médicas en Bogotá.

8 CONCLUSIONES Y TRABAJOS FUTUROS

8.1 CONCLUSIONES

En la consecución del objetivo de identificar las variables que impactan la demanda de emergencias médicas en Bogotá, se logró la construcción sólida de la base de datos a partir de fuentes primarias y secundarias, complementada con un análisis exploratorio detallado. La selección de variables revela que variables como la fecha, el rango horario, la UPL, y la cantidad de eventos especiales son relevantes en la variabilidad de los incidentes. Sin embargo, el durante análisis realizado se determinó que los eventos especiales y características demográficas, como la densidad poblacional y grupos etarios específicos (infantes de 0-4 años y adultos mayores de 60), tienen un impacto significativo para entender patrones específicos de demanda, especialmente en su interacción con factores temporales como días festivos y la ocurrencia de incidentes previos. En este sentido, la incorporación de estas variables no solo enriquece el modelo predictivo, sino que resulta esencial para abordar variaciones espaciales y temporales.

En cuanto a la evaluación de los modelos de predicción, el algoritmo XGBoost mostró el mejor desempeño, superando de manera significativa a modelos tradicionales como ARIMA, SARIMA y la Regresión Binomial Negativa. Los resultados indicaron que XGBoost presentó un RMSE de 1.9, un MAE de 1.5 y un R^2 de 0.7, lo que evidencia su capacidad para capturar de manera más precisa la variabilidad de los datos espaciotemporales en comparación con los modelos estadísticos tradicionales. Asimismo, los modelos de redes neuronales, como MLP y RNN, aunque mostraron resultados aceptables, no lograron superar la precisión y menor error de XGBoost, lo que se puede aludir a la importancia en la definición de la arquitectura de las redes neuronales para su óptimo rendimiento. Esto sugiere que los enfoques basados en aprendizaje automático son más adecuados para predecir la demanda de emergencias médicas en Bogotá.

El modelo predictivo se implementó a nivel de UPL, lo cual resultó ser una decisión acertada, dado que esta unidad espacial permitió capturar con mayor precisión los patrones de demanda y la variabilidad espacial en la ciudad. La transición de UPZ a UPL se justificó debido a que la UPL proporcionó un mejor ajuste a los datos disponibles y mejoró el desempeño del modelo. Este cambio pone de relieve la importancia de no establecer de manera rígida la unidad espacial de análisis, sino evaluar previamente la escala más adecuada según la disponibilidad y calidad de los datos. Por otra parte, se evidenció que la prioridad inicial crítica, los incidentes del día anterior y aquellos relacionados con heridos son los principales factores que influyen en la demanda, mientras que otras variables, como el rango horario, la demografía y la ubicación espacial, también aportan información relevante.

El tablero de control desarrollado en Power BI facilita la visualización y el análisis de la demanda de servicios de emergencias médicas en Bogotá, segmentada por Unidades de Planeamiento Local. Instrumentos de visualización de resultados son claves para los usuarios finales, ya que permiten una comprensión más clara de los patrones y tendencias en la demanda de emergencias. Esta herramienta no solo mejora la capacidad de respuesta ante incidentes, sino que también optimiza la asignación de recursos y la planificación estratégica, lo que es fundamental para una gestión eficiente en situaciones críticas. La interactividad y el acceso a datos en tiempo real garantizan que los tomadores de decisiones puedan reaccionar de manera informada y oportuna, contribuyendo así a mejorar la atención médica y la seguridad en la ciudad.

8.2 TRABAJOS FUTUROS

Para futuros trabajos, se recomienda explorar el desarrollo de modelos predictivos que puedan estimar la demanda de llamadas de incidentes trasladadas al centro de llamadas del CRUE. Utilizando la base de datos actual, que incluye información detallada sobre el tipo de incidentes, horarios y localizaciones, se podría implementar un modelo que no solo pronostique el volumen de llamadas, sino también identifique patrones específicos en la demanda. Este enfoque permitiría optimizar los tiempos de respuesta y mejorar la organización interna del CRUE para prestar un servicio más oportuno y eficiente.

Asimismo, sería relevante investigar la aplicación de modelos más avanzados, como redes neuronales profundas o modelos de ensamblaje, que puedan adaptarse de manera más precisa a la complejidad de los datos espaciotemporales. El uso de técnicas de aprendizaje profundo, como las redes neuronales recurrentes (RNN) o las redes neuronales convolucionales (CNN), podría aportar mejoras significativas en la precisión de las predicciones si se desarrollan arquitecturas específicas para este tipo de problemas. Estas técnicas podrían complementar el enfoque actual y abrir nuevas posibilidades para optimizar la gestión de emergencias médicas en Bogotá.

9 REFERENCIAS BIBLIOGRÁFICAS

- [1] Alcaldía de Bogotá, «Alcaldía de Bogotá,» 28 Enero 2021. [En línea]. Available: <https://bogota.gov.co/mi-ciudad/salud/que-es-una-upz>. [Último acceso: Noviembre 2023].
- [2] Secretaria de Movilidad, «Encuesta de Movilidad 2023 Bogotá - Región,» 05 Mayo 2024. [En línea]. Available: <https://bogota.gov.co/mi-ciudad/movilidad/movilidad-en-bogota-asi-se-movilizan-los-capitalinos-segun-encuesta#:~:text=Actualmente%2C%20en%20Bogot%C3%A1%20se%20registran,transporte%20p%C3%ABblico%20o%20en%20bicicleta..> [Último acceso: 03 Septiembre 2024].
- [3] INRIX, «Cuadro de indicadores de tráfico mundial INRIX 2022,» 10 Enero 2023. [En línea]. Available: <https://inrix.com/press-releases/2022-global-traffic-scorecard-uk/>. [Último acceso: Octubre 2024].
- [4] S. d. S. -. I. C. 2020, «Secretaría de Salud - saludcapital.gov.co,» 2 Marzo 2021. [En línea]. Available: http://www.saludcapital.gov.co/Informes%20de%20Gestin/Planes%20de%20Gesti%C3%B3n/Plan_de_Accion_2020/Informe_CBN_1045_2020.pdf. [Último acceso: 11 2023].
- [5] C. d. Bogotá, «Más de una hora se demora la atención de ambulancias en Bogotá, Diana Diago revela la cifra.,» Julio 2024. [En línea]. Available: <https://concejodebogota.gov.co/mas-de-una-hora-se-demora-la-atencion-de-ambulancias-en-bogota-diana/cbogota/2024-07-24/113443.php>. [Último acceso: Oct 2024].
- [6] Secretaría de Salud, «Secretaría Distrital de Salud,» [En línea]. Available: <http://www.saludcapital.gov.co/DDS/Paginas/Ofer-ta-y-demanda-servicios-salud.aspx#:~:text=Demanda%20de%20servicios%20de%20salud%3A%20la%20demanda%20de%20servicios%20de,consumo%20de%20servicios%20de%20salud..> [Último acceso: 30 11 2023].
- [7] K. S. P. A. F. J. E. T. R. P. F. D. P. Gina C. Prichard, «Differential health responses to climate change projections in three UK cities as measured by ambulance dispatch data,» *Environmental Advances*, vol. 7, nº 100146, p. 6, 2022.
- [8] F. Villarreal, «Introducción a los Modelos de Pronósticos,» Septiembre 2016. [En línea]. Available: https://www.matematica.uns.edu.ar/uma2016/material/Introduccion_a_los_Modelos_de_Pronosticos.pdf. [Último acceso: 03 Diciembre 2023].
- [9] J. M. Izar, «Modelos de Pronóstico,» de *Contabilidad Gerencial*, Trillas, 2007, p. 32.
- [10] N. Morales , «Algunas Consideraciones para la Organización de Servicios de Salud para Emergencias y Desastres,» *Anales de la Facultad de Medicina*, vol. 62, nº 1, pp. 44-55, 2001.
- [11] Subsecretaría de Servicios de Salud y Aseguramiento, Dirección de Urgencias y Emergencias en Salud, «BOLETÍN ESTADÍSTICO Y EPIDEMIOLOGICO MENSUAL,»

- Secretaría Distrital de Salud, Bogotá, 2021.
- [12] C. Véliz Capuñay, «Capítulo 1 Introducción,» de *Aprendizaje automático. Introducción al aprendizaje profundo*, Lima, Perú, Fondo editorial, 2020, pp. 1- 6.
- [13] A. S. Karthik Ramasubramanian, *Machine Learning Using R*, New Delhi, Delhi, India: Apress, 2019.
- [14] IBM, «El modelo de redes neuronales,» 17 Agosto 2021. [En línea]. Available: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=networks-neural-model>. [Último acceso: 03 Diciembre 2023].
- [15] I. S. S. A. Halima Bousqaoui, «Comparative analysis of short-term demand predicting models using ARIMA and Deep learning.,» *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, nº 4, p. 10, 2021.
- [16] S. Dorado Gonzalez y V. Velásquez Robayo, *Pronóstico De Demanda Utilizando Inteligencia Artificial*, Cali, 2020.
- [17] P. Abreu, D. Santos y A. Barbosa-Póvoa, «Data-driven forecasting for operational planning of emergency medical services,» *Socio-Economic Planning Sciences*, 2023.
- [18] K. S. Chinchay Velasco, R. d. P. Franco Suarez y J. T. Rosas Angulo, *Análisis de Factores que Influyen en la Cobertura de Atención Prehospitalaria en Lima metropolitana 2013 – 2019*, Lima, 2020.
- [19] J. O. Barrera Gutiérrez, «Cantidad y distribución de ambulancias para responder con oportunidad y eficacia a los incidentes médicos y de atención pre hospitalaria que se suscitan en la ciudad de Bogotá durante el año 2018,» Pontificia Universidad Javeriana, Bogotá, 2018.
- [20] A. F. W. H. K. H. C. Adrian Xi Lin, «Leveraging Machine Learning Techniques and Engineering of Multi-Nature Features for National Daily Regional Ambulance Demand Prediction.,» *International Journal of Environmental Research and Public Health.*, Singapur, 2020.
- [21] F. L. Bayisa, M. Ådahla, P. Rydén y O. Cronie, «Large-scale modelling and forecasting of ambulance calls in northern Sweden using spatio-temporal log-Gaussian Cox processes,» Department of Mathematics and Mathematical Statistics, Umeå University, Västerbotten, 2020.
- [22] Sistema Único de Gestión de Aglomeraciones de Público Idiger, «Datos Abiertos,» 30 junio 2024. [En línea]. Available: <https://datosabiertos.bogota.gov.co/dataset/suga-sistema-unico-de-gestion-de-aglomeraciones-de-publico/resource/1d80743e-00ad-4a17-bc3d-411e813df7c8>. [Último acceso: Julio 2024].
- [23] Instituto de Hidrología, Meteorología y Estudios Ambientales, «Datos Abiertos,» Junio 2024. [En línea]. Available: https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaci-n/s54a-sgyg/about_data. [Último acceso: Julio 2024].
- [24] Departamento Administrativo Nacional de Estadísticas , «Proyecciones de población desagregadas por localidades 2018-2035 y UPZ 2018-2024,» 26 Marzo 2021. [En línea].

- Available: <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion/proyecciones-de-poblacion-bogota>. [Último acceso: Junio 2024].
- [25] Laboratorio Urbano, «Laboratorio Urbano - Bogotá portal de datos,» opendatasoft, [En línea]. Available: https://bogota-laburbano.opendatasoft.com/explore/dataset/poblacion-upz-bogota/table/?flg=es&sort=-cod_loc&dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6InBvYmxhY2lvbi11cHotYm9nb3Rhlwib3B0aW9ucyI6eyJmbGciOiJlcyJ9fSwiY2hhcnRzIjpbeyJhbGlnbk1vbnoRolj. [Último acceso: 15 Agosto 2024].
- [26] Secretaría Distrital de Salud, «Infraestructura de Datos Espaciales de Bogotá,» 30 Noviembre 2020. [En línea]. Available: <https://www.ideca.gov.co/recursos/mapas/territorios-redes-salud-por-upz>. [Último acceso: Junio 2024].
- [27] «Corporación Para EL Control Social,» 15 Marzo 2016. [En línea]. Available: <https://contrial.co/mapa-de-upzs-de-bogota-y-listado-de-barrios-para-consulta-en-foros-control-social-encuentre-a-su-representante/>. [Último acceso: Junio 2024].
- [28] J. L. ,. F. H. Salvador García, *Data Preprocessing in Data Mining*, Switzerland: Springer, Cham., 2015.
- [29] P. Congdon, «A spatio-temporal autoregressive model for monitoring and predicting COVID infection rates.,» *Journal of Geographical Systems*, vol. 24, pp. 583 - 610, 2022.
- [30] J. M. Hilbe, *Negative Binomial Regression*, Arizona: Cambridge University Press, 2012.
- [31] K. P. Murphy, *Machine Learning A Probabilistic Perspective*, Massachusetts: The MIT Press, 2012.
- [32] R. Z. F. L. O. S. S. F. S. V. C. M. L. d. S. C. Marcos Yamasaki, «Optimized hybrid ensemble learning approaches applied to very short-term load forecasting,» *International Journal of Electrical Power & Energy Systems*, Vols. %1 de %2155, Part B, nº 109579, p. 17, 2024.
- [33] U. W. H. Muhammad Waqas, «A critical review of RNN and LSTM variants in hydrological time series predictions,» *MethodsX*, vol. 13, nº 102946, p. 19, 2024.
- [34] J. A. Baddeley, J. Waagepetersen y J. Møller, «Non- and semi-parametric estimation of interaction in inhomogeneous point patterns,» *Statistica Neerlandica*, vol. 54, nº 3, pp. 265-382, 2001.
- [35] S. D. d. Planeación, «SECRETARÍA DE PLANEACIÓN,» [En línea]. Available: <https://www.sdp.gov.co/micrositios/pot/upl>. [Último acceso: 20 Septiembre 2024].
- [36] Y. H. W. Y. Xiaodi Wang, «Novel wind power ensemble forecasting system based on mixed-frequency modeling and interpretable base model selection strategy,» *Energy*, vol. 297, nº ISSN 0360-5442, p. 15, 2024.
- [37] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, O'Reilly, 2019.

- [38] C. R. T. H. M. K. T. N. Hanna Meyer, «Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation,» *Environmental Modelling & Software*, vol. 101, nº ISSN 1364-8152, pp. 1-9, 2018.
- [39] S. Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Oakland: Analytics Press, 2012.
- [40] E. Turban, R. Sharda y D. Delen, *Decision Support and Business Intelligence Systems*, New Jersey: Prentice Hall Press, 2010.