

**Desarrollo de un modelo predictivo del tiempo de espera y demanda de usuarios en una  
electrificadora**

William Andrey Parra Becerra

Asesor

Rafael Gaitan Ospina

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería - ECBTI

Especialización en Ciencia de Datos y Analítica

2025

## Resumen

En el sector energético, la atención presencial a clientes presenta desafíos operativos debido a la alta variabilidad en los tiempos de espera y la afluencia diaria de usuarios. En la oficina principal de una electrificadora, se identificaron tiempos de espera de hasta cuatro horas y una incapacidad para anticipar la demanda, lo cual afecta la satisfacción del cliente y la eficiencia en la asignación de recursos. Este estudio abordó la problemática mediante el análisis de más de 100.000 registros históricos y la aplicación de la metodología CRISP-DM para desarrollar dos modelos predictivos: uno basado en XGBoost para estimar el tiempo de espera de cada usuario, y otro mediante el modelo Prophet para prever la demanda diaria. Los resultados mostraron que XGBoost alcanzó un buen equilibrio entre precisión y generalización (RMSE  $\approx$  13 minutos), mientras que Prophet logró un MAPE inferior al 20%, identificando patrones semanales y el efecto de días festivos. Estos hallazgos permiten establecer una base sólida para mejorar la planificación operativa, optimizar la distribución del personal y reducir la incertidumbre en la experiencia de atención al cliente.

**Palabras clave:** Modelo predictivo, Machine Learning, Metodología CRISP-DM, Tiempo de espera, Series temporales.

## Abstract

In the energy sector, in-person customer service presents operational challenges due to the high variability in waiting times and daily user demand. At the main office of an electric utility company, wait times of up to four hours were observed, along with a lack of demand forecasting, which negatively impacts customer satisfaction and resource allocation efficiency. This study addressed the issue through the analysis of over 100,000 historical records and the application of the CRISP-DM methodology to develop two predictive models: one using XGBoost to estimate individual waiting times, and another using Prophet to forecast daily demand. The results showed that XGBoost achieved a good balance between accuracy and generalization (RMSE  $\approx$  13 minutes), while Prophet achieved a MAPE below 20%, effectively capturing weekly patterns and the impact of holidays. These findings provide a solid foundation for improving operational planning, optimizing staff allocation, and reducing uncertainty in the customer service experience.

**Keywords:** Predictive model, Machine Learning, CRISP-DM methodology, Waiting time, Time series.

## Tabla de Contenido

Introducción .....	9
Descripción del Problema .....	11
Sistematización del Problema .....	13
Pregunta General.....	13
Preguntas Específicas.....	13
Justificación .....	14
Objetivos.....	17
Objetivo General .....	17
Objetivos Específicos.....	17
Marco de Referencia .....	18
Estado del Arte.....	18
Marco Contextual.....	20
Marco Teórico.....	23
Tiempos de Espera y Satisfacción del Cliente .....	23
Predicción de Tiempos de Espera con Machine Learning .....	24
Predicción de Demanda con Modelos de Series Temporales .....	25
Comparación de Algoritmos y Modelos .....	26
Relevancia de la Metodología CRISP-DM.....	26
Marco Conceptual .....	27
Marco Normativo.....	29
Metodología .....	32
Método .....	33

Tipo de Estudio .....	34
Recolección de Datos .....	34
Resultados .....	36
Análisis Exploratorio de los Datos (ADE).....	36
Modelado del Tiempo de Espera .....	45
Modelado de la Demanda .....	50
Modelado con ARIMA .....	52
Modelado con Prophet .....	53
Modelado con Redes Neuronales Recurrentes (RNN) .....	55
Validación Cruzada.....	58
Conclusiones .....	60
Recomendaciones .....	62
Referencias Bibliográficas .....	64

## Lista de Tablas

<b>Tabla 1</b> <i>Resultados Iniciales de Modelos Predictivos Según Tipo de Codificación</i> .....	47
<b>Tabla 2</b> <i>Desempeño del Modelo ARIMA para Diferentes Configuraciones</i> .....	52
<b>Tabla 3</b> <i>Comparación del Desempeño del Modelo Prophet con y sin Festivos</i> .....	54
<b>Tabla 4</b> <i>Desempeño de Modelos RNN con Diferentes Ciclos de Entrada</i> .....	56
<b>Tabla 5</b> <i>Comparación General de Modelos para la Demanda Diaria</i> .....	58

## Lista de Figuras

<b>Figura 1</b> <i>Matriz de Correlaciones Entre Variables Numéricas del Dataset Depurado</i> .....	38
<b>Figura 2</b> <i>Evolución Diaria de la Cantidad de Turnos Generados (Enero 2023 - Febrero 2025)</i>	39
<b>Figura 3</b> <i>Distribución de la Cantidad de Turnos Generados por Hora del día</i> .....	39
<b>Figura 4</b> <i>Distribución Diaria de Turnos por día del mes</i> .....	40
<b>Figura 5</b> <i>Distribución del Tiempo de Espera por Grupo de Atención</i> .....	41
<b>Figura 6</b> <i>Distribución del Tiempo de Atención por Grupo de Atención</i> .....	42
<b>Figura 7</b> <i>Distribución del Tiempo de Espera en los Turnos Abandonados, con Referencia a Cuartiles</i> .....	43
<b>Figura 8</b> <i>Distribución del Tiempo de Espera con Referencia a Mediana y Media</i> .....	43
<b>Figura 9</b> <i>Distribución del Tiempo de Espera por Hora del día con Promedio Suavizado</i> .....	44
<b>Figura 10</b> <i>Cantidad de Turnos Atendidos y Tiempo Promedio de Atención por Taquilla</i> .....	45
<b>Figura 11</b> <i>Curva de Aprendizaje de XGBoost (Modelo Base)</i> .....	47
<b>Figura 12</b> <i>Curva de Aprendizaje del Modelo Random Forest Optimizado</i> .....	48
<b>Figura 13</b> <i>Curva de Aprendizaje del Modelo XGBoost Optimizado</i> .....	49
<b>Figura 14</b> <i>Curva de Aprendizaje del Modelo de Regresión Lineal</i> .....	50
<b>Figura 15</b> <i>Evolución Diaria de la Demanda de Turnos (2023–2025)</i> .....	51
<b>Figura 16</b> <i>Descomposición de la Serie Temporal Diaria de Turnos</i> .....	51
<b>Figura 17</b> <i>Funciones de Autocorrelación (ACF) y Autocorrelación Parcial (PACF)</i> .....	52
<b>Figura 18</b> <i>Pronóstico de Demanda con Prophet: Ajuste al Conjunto de Entrenamiento</i> .....	53
<b>Figura 19</b> <i>Pronóstico de Demanda con Prophet: Teniendo en Cuenta los días Hábiles</i> .....	54
<b>Figura 20</b> <i>Componentes del Modelo Prophet: Tendencia, Festivos y Patrones Estacionales</i> ....	55
<b>Figura 21</b> <i>Pronóstico de Demanda con RNN: Comparación Real vs Predicción</i> .....	56

<b>Figura 22</b> <i>MAPE por Horizonte de Predicción con Modelo RNN</i> .....	59
---	----

## Introducción

Los tiempos de espera son un factor determinante en la experiencia del cliente y en la percepción de calidad del servicio en entornos de atención presencial. Diversos estudios han evidenciado que esperas prolongadas o inciertas generan insatisfacción, afectan la fidelización del usuario y deterioran la imagen institucional (Caruelle et al., 2023). En respuesta a esta problemática, numerosas organizaciones han implementado modelos predictivos que permiten anticipar flujos de usuarios, optimizar recursos y mejorar la planificación operativa (Kyritsis & Deriaz, 2019; Fernández López et al., 2020).

En el contexto de una empresa del sector energético, se identificó una alta variabilidad tanto en los tiempos de espera como en la afluencia diaria de usuarios, que oscila entre 200 y 300 personas, en el periodo comprendido entre enero del 2023 hasta febrero del 2025. En casos extremos, los tiempos de espera superan las cuatro horas, generando afectaciones tanto en la experiencia del usuario como en la eficiencia del servicio. Esta situación se ve agravada por la ausencia de herramientas que permitan anticipar la demanda y gestionar adecuadamente el recurso humano.

Para abordar esta situación, se desarrolló una propuesta basada en la metodología CRISP-DM, orientada al diseño de dos modelos predictivos: uno de regresión para estimar el tiempo de espera individual, y otro de series temporales para anticipar la demanda diaria. La elección de los algoritmos se fundamentó en su efectividad comprobada en contextos similares. Por ejemplo, Random Forest y XGBoost han mostrado un desempeño superior en la predicción de tiempos de atención en entornos clínicos y de transporte (Joseph et al., 2023; Tai et al., 2023), mientras que Prophet y ARIMA han sido ampliamente utilizados en la estimación de patrones de demanda en sectores con fuerte estacionalidad (Menculini et al., 2021; Fierro Torres et al., 2022). Así mismo,

se exploró el uso de redes neuronales recurrentes (RNN) para capturar relaciones temporales complejas, particularmente útiles en la predicción de series con alta variabilidad.

Los modelos fueron evaluados mediante métricas de precisión como RMSE, MAE y  $R^2$ , las cuales cuantifican el desempeño del modelo desde distintas perspectivas, lo que permitió identificar las soluciones más robustas para cada tarea. Los resultados obtenidos no solo permiten mejorar la asignación de recursos en la oficina principal de atención, sino que constituyen un aporte replicable para organizaciones con dinámicas operativas similares en el sector público o privado.

## Descripción del Problema

En los entornos de atención al cliente, los tiempos de espera son un factor crítico que incide directamente en la satisfacción del usuario y en la percepción del servicio ofrecido. En la oficina principal de una electrificadora, se ha evidenciado una alta variabilidad en dichos tiempos, con registros que oscilan entre un minuto y más de cuatro horas. Este fenómeno genera incomodidad, insatisfacción y dificultades operativas para la empresa, que no cuenta con mecanismos para anticipar el comportamiento diario de la demanda.

A este panorama se suma la imprevisibilidad en la afluencia de usuarios, la cual varía significativamente según el día, la hora o el ciclo de facturación. Esta falta de visibilidad impide una adecuada planificación del recurso humano y la distribución eficiente de los puntos de atención (Caruelle et al., 2023). Sin un sistema que permita estimar la cantidad de usuarios esperados y su tiempo probable de espera, tanto la experiencia del cliente como la eficiencia institucional se ven afectadas.

La literatura reciente ha mostrado que el uso de modelos predictivos, en particular aquellos basados en técnicas de aprendizaje automático y análisis de series temporales, puede ofrecer soluciones eficaces ante estos desafíos. Herramientas como Random Forest y XGBoost han sido empleadas en el sector salud y servicios para predecir tiempos de atención con resultados satisfactorios (Joseph et al., 2023), mientras que modelos como Prophet y ARIMA han demostrado utilidad en la proyección de demanda bajo condiciones de estacionalidad (Fernández López et al., 2020).

En este contexto, la presente investigación plantea el diseño de dos modelos predictivos complementarios: uno para estimar los tiempos de espera individuales a partir de variables operativas, y otro para prever la cantidad diaria de usuarios en la oficina. La implementación de

estos modelos permitirá fortalecer la toma de decisiones operativas y reducir la incertidumbre tanto para los usuarios como para la empresa.

No obstante, este desafío implica resolver aspectos técnicos fundamentales como la calidad y limpieza de los datos históricos, la selección de variables relevantes, y la evaluación comparativa de algoritmos. Para ello, se adopta la metodología CRISP-DM, reconocida por su enfoque estructurado en proyectos de ciencia de datos, que guiará todo el proceso desde la comprensión del problema hasta la validación de los modelos.

¿Cómo se puede desarrollar un modelo predictivo basado en Machine Learning, utilizando la metodología CRISP-DM, para estimar con precisión los tiempos de espera y la demanda de usuarios en la oficina principal de una electrificadora?

## Sistematización del Problema

### Pregunta General

¿Cómo desarrollar un modelo predictivo del tiempo de espera y de la demanda de los usuarios en la oficina principal de una electrificadora?

### Preguntas Específicas

¿Qué patrones y tendencias pueden inferirse a partir del análisis exploratorio de los datos históricos de atención en la oficina?

¿Cuáles son las variables operativas más significativas que afectan los tiempos de espera en la atención al cliente de la electrificadora?

¿Qué técnicas de Machine Learning pueden utilizarse para desarrollar un modelo predictivo del tiempo de espera en este contexto?

¿Cuál de los modelos evaluados presenta la mejor precisión para la predicción de tiempos de espera?

¿Cómo puede evaluarse la efectividad del modelo predictivo en términos de métricas de precisión y validación cruzada?

¿Cuáles son los patrones temporales más relevantes en la demanda de usuarios de la oficina principal de la empresa de energía?

¿Qué modelos de series temporales pueden aplicarse para predecir la cantidad de usuarios en distintos periodos de tiempo?

## Justificación

En el entorno actual, Thiongane et al. (2022) mencionan que garantizar una experiencia de atención al cliente eficiente y predecible es un desafío clave para las empresas de servicios públicos. En la oficina principal de la electrificadora, los tiempos de espera de los usuarios son altamente variables, oscilando desde pocos minutos hasta casos extremos de varias horas, esta situación genera insatisfacción y afecta tanto la percepción del servicio como la fidelización de los clientes. Además, la incapacidad de prever estos tiempos y la afluencia de usuarios limita la capacidad operativa de la empresa para gestionar sus recursos de manera óptima. Por tanto, existe una necesidad urgente de desarrollar soluciones tecnológicas que permitan anticipar estas dinámicas.

El desarrollo de un modelo predictivo basado en Machine Learning es una respuesta innovadora y práctica a esta problemática. Este estudio propone dos enfoques: (1) la estimación de los tiempos de espera en función de variables operativas y (2) la predicción de la cantidad de usuarios que asistirán en determinados periodos. La combinación de ambos modelos permitirá optimizar la experiencia del cliente y mejorar la gestión operativa de la electrificadora, según Loureiro et al. (2023) a través del análisis de datos históricos y del uso de técnicas avanzadas de aprendizaje automático, estos modelos pueden proporcionar estimaciones precisas, permitiendo anticiparse a escenarios de alta demanda y mejorar la asignación de recursos.

La implementación de estos modelos tiene un impacto significativo tanto para la empresa como para sus usuarios. Para los clientes, contar con información confiable sobre los tiempos de espera reducirá la frustración asociada a la incertidumbre y mejorará su experiencia general. Así mismo, la predicción de la demanda permitirá que la empresa realice ajustes en su planificación operativa, asegurando una distribución eficiente del personal y evitando congestiones en los

horarios de mayor afluencia. Dado que la variabilidad en la cantidad de usuarios puede estar influenciada por factores externos como días de pago, condiciones climáticas o eventos locales, la combinación de distintos enfoques de modelado permitirá capturar tanto patrones estacionales como fluctuaciones imprevistas. En sectores como el energético, donde el acceso a la información y la agilidad en la atención son esenciales, estas mejoras pueden traducirse en una mayor confianza y satisfacción del usuario.

Desde el punto de vista académico, este proyecto tiene un alto valor formativo. Su desarrollo se basa en la metodología CRISP-DM, ampliamente utilizada en proyectos de ciencia de datos, garantizando un enfoque estructurado desde la comprensión del problema hasta la evaluación del modelo. Además, integra técnicas avanzadas de aprendizaje automático con análisis exploratorio de datos, fortaleciendo el conocimiento en modelos de regresión, clasificación y series temporales. Se evaluarán diversos algoritmos, incluyendo Random Forest, XGBoost, Regresión Lineal, ARIMA, Prophet y RNN, lo que permitirá comparar el desempeño de metodologías tradicionales y avanzadas en predicción de tiempos de espera y demanda. Este análisis contribuirá a la identificación del enfoque más adecuado para la electrificadora, facilitando futuras implementaciones y mejoras en la gestión operativa.

Finalmente, el desarrollo de un modelo predictivo para los tiempos de espera y la demanda de usuarios en la electrificadora es una solución alineada con las necesidades tanto de la empresa como de sus clientes. Su impacto potencial incluye una mejora sustancial en la satisfacción del usuario, una mayor eficiencia operativa y el fortalecimiento de la imagen institucional. Al mismo tiempo, este estudio representa una aplicación real de herramientas analíticas avanzadas, contribuyendo tanto al ámbito académico como empresarial, y ofreciendo

un marco de referencia para futuras implementaciones en organizaciones con dinámicas de atención similares.

## **Objetivos**

### **Objetivo General**

Desarrollar un modelo predictivo del tiempo de espera y demanda de usuarios en una electrificadora.

### **Objetivos Específicos**

Analizar los datos históricos de atención al cliente en la oficina principal de la electrificadora, utilizando técnicas de análisis exploratorio de datos, e infiriendo patrones y tendencias que influyan en los tiempos de espera y la demanda de usuarios.

Construir modelos predictivos utilizando técnicas de Machine Learning, siguiendo la metodología CRISP-DM, determinando el de mayor precisión.

Evaluar el desempeño del modelo predictivo seleccionado mediante métricas de precisión y validación cruzada.

## Marco de Referencia

### Estado del Arte

El análisis de antecedentes para este estudio se enfoca en investigaciones previas que han abordado la predicción de tiempos de espera y la estimación de la demanda en contextos de atención al cliente, aplicando técnicas de Machine Learning y modelos de series temporales. A continuación, se presentan los aportes más relevantes, agrupados según la problemática abordada y los métodos utilizados.

1. Predicción de tiempos de espera en la atención al cliente: Diversos estudios han evidenciado que la incertidumbre sobre el tiempo de espera afecta negativamente la experiencia del usuario. Caruelle et al. (2023) demostraron que la discrepancia entre el tiempo esperado y el tiempo real incide en la satisfacción del cliente, siendo las esperas superiores a lo previsto especialmente perjudiciales. En el mismo sentido, Kyritsis y Deriaz (2019) desarrollaron un modelo basado en redes neuronales para predecir tiempos de espera en sistemas multicanal, logrando reducir la incertidumbre percibida.

En contextos similares, Joseph et al. (2023) aplicaron los algoritmos Random Forest y XGBoost para estimar tiempos de espera en clínicas ambulatorias, obteniendo una mejora del 15% en la precisión frente a métodos tradicionales. Tai et al. (2023), por su parte, compararon Decision Trees, Random Forest y XGBoost en entornos de soporte técnico, encontrando que XGBoost era más robusto frente a múltiples factores dinámicos.

Estos estudios justifican el uso de modelos como Random Forest y XGBoost en este proyecto para predecir el tiempo de espera. Sin embargo, también se incluye regresión lineal como modelo de base comparativa, por su simplicidad y facilidad de interpretación, lo que permite establecer una línea de referencia clara en la evaluación de modelos predictivos. Este enfoque ha

sido validado en estudios donde la regresión lineal, aunque con menor precisión, se emplea como punto de partida para comparar modelos más complejos como XGBoost y Random Forest (Ayyalasomayajula et al., 2021).

2. Modelos de predicción de demanda: La literatura también ha abordado la estimación de la demanda en servicios de atención al cliente. Fernández López et al., (2020) demostraron que el uso de modelos híbridos, combinando aprendizaje profundo y métodos estadísticos, puede superar a los enfoques tradicionales en escenarios con estacionalidad.

Modelos como ARIMA han sido ampliamente utilizados para capturar patrones lineales y estacionales en series temporales (Ayala & Bucio, 2020), mientras que Prophet, desarrollado por Facebook, ha destacado por su flexibilidad para manejar irregularidades y detectar tendencias con mayor precisión (Fierro Torres et al., 2022). Las redes neuronales recurrentes (RNN), especialmente las variantes LSTM, han mostrado un alto desempeño en series con patrones no lineales y ruidos aleatorios.

3. Aplicación en el sector energético y evaluación comparativa: García et al. (2023) desarrollaron un modelo para estimar los tiempos de atención a fallas eléctricas mediante técnicas de regresión multivariable y Random Forest, logrando una mejora del 18% en la eficiencia operativa. Aunque el estudio no se centró en atención al cliente, su enfoque destaca la aplicabilidad de modelos predictivos en el contexto energético.

Asimismo, Thiongane et al. (2022) utilizaron distribuciones condicionales de tiempo en call centers multihabilidades, resaltando la importancia de variables como la urgencia del servicio o la cantidad de agentes disponibles. Su estudio refuerza el papel de la ingeniería de variables en la mejora del rendimiento predictivo.

4. Enfoques metodológicos en ciencia de datos: En cuanto a la estructura del proceso analítico, Loureiro et al. (2023) destacaron la utilidad de la metodología CRISP-DM para proyectos predictivos, dado su enfoque progresivo desde la comprensión del negocio hasta la evaluación de modelos. Patel et al. (2023) complementan esta visión al mostrar cómo la automatización en el análisis exploratorio y la selección de variables acelera el desarrollo de soluciones aplicadas.

En resumen, la literatura revisada respalda la aplicación de modelos de aprendizaje automático y series temporales para mejorar la planificación operativa en entornos de atención al cliente. Este estudio se alinea con estos enfoques, adaptándolos al contexto particular de una electrificadora en Colombia, caracterizada por una alta variabilidad en la afluencia de usuarios y tiempos de espera. Al integrar estas metodologías, se busca desarrollar soluciones predictivas que optimicen la asignación de recursos y mejoren la experiencia del usuario.

### **Marco Contextual**

Las empresas del sector energético desempeñan un papel fundamental en la distribución y comercialización de energía eléctrica, garantizando la prestación del servicio con altos estándares de calidad y eficiencia. Para atender a sus clientes, estas compañías cuentan con múltiples oficinas de atención, donde se gestionan solicitudes relacionadas con facturación, conexiones, financiación de deudas y reclamos, entre otros trámites. En la sede principales de la electrificadora en estudio, se registra un alto flujo de usuarios diariamente, convirtiéndola en un punto clave para la interacción con los clientes.

Esta oficina atiende entre 200 y 300 personas por día, lo que representa un desafío significativo en términos de operatividad y asignación de recursos. La demanda de atención varía en función de múltiples factores, incluyendo la hora del día, la cantidad de usuarios en cola y la

disponibilidad del personal. Además, ciertos trámites requieren tiempos de gestión más prolongados que otros, lo que contribuye a la variabilidad en los tiempos de espera. En algunos casos, los clientes pueden ser atendidos en pocos minutos, mientras que en otros la espera puede extenderse hasta cuatro horas, generando insatisfacción y estrés.

Actualmente, la empresa no cuenta con una herramienta que permita predecir con precisión los tiempos de espera ni anticipar la afluencia de usuarios. Esto dificulta la optimización de la distribución del personal y la planificación operativa, afectando la capacidad de respuesta ante picos de demanda. La falta de información sobre el tiempo estimado de atención también limita la capacidad de los clientes para tomar decisiones informadas sobre si permanecer en la sede o regresar en otro momento.

Ante esta problemática, el presente estudio propone el desarrollo de un modelo predictivo basado en Machine Learning para estimar los tiempos de espera y anticipar la cantidad de usuarios que acudirán en distintos periodos. Al analizar datos históricos y variables operativas clave, esta herramienta permitirá optimizar la gestión de recursos sin necesidad de incrementar costos operativos. Para los clientes, la posibilidad de conocer tiempos estimados de espera reducirá la incertidumbre y mejorará su experiencia de servicio. Para la empresa, contar con un modelo de predicción facilitará la asignación estratégica de personal, evitando congestiones y mejorando la eficiencia operativa.

El uso de técnicas de Machine Learning en la gestión de la atención al cliente ha demostrado ser una estrategia efectiva en distintos sectores, permitiendo a las organizaciones anticiparse a fluctuaciones en la demanda y optimizar la distribución de sus recursos. Empresas en diversas industrias han implementado soluciones similares para gestionar la afluencia de usuarios

y mejorar la asignación de personal, logrando una reducción significativa en los tiempos de espera y un aumento en la satisfacción del cliente.

Este estudio se enmarca dentro de los esfuerzos por mejorar la eficiencia operativa de la electrificadora y la experiencia del usuario mediante el uso de herramientas analíticas avanzadas. Los hallazgos obtenidos podrán servir como base para futuras estrategias de optimización en la planificación de atención al cliente, proporcionando un modelo replicable en otros contextos con dinámicas de servicio similares.

## **Marco Teórico**

El presente estudio se fundamenta en diversas teorías y enfoques relacionados con los tiempos de espera, la satisfacción del cliente, la predicción de demanda, los modelos de Machine Learning y series temporales, así como la metodología CRISP-DM. A continuación, se presentan los principales aspectos teóricos que sustentan esta investigación.

### ***Tiempos de Espera y Satisfacción del Cliente***

En el contexto de la gestión de servicios, el tiempo de espera se define como el intervalo transcurrido entre la solicitud del servicio por parte del usuario y el momento en que este es atendido. Este puede clasificarse en tiempo real (medido objetivamente) y tiempo percibido, el cual depende de la subjetividad del usuario y sus expectativas. La literatura ha identificado que estos dos componentes pueden diferir notablemente, impactando de manera significativa la satisfacción del cliente.

Los estudios más recientes coinciden en que la percepción del tiempo influye tanto como el tiempo medido. Las organizaciones que logran gestionar adecuadamente las expectativas del usuario durante la espera tienden a mejorar sus niveles de satisfacción y fidelización. En este sentido, comprender la relación entre tiempos de espera y percepción del servicio es esencial para diseñar estrategias de mejora en la atención presencial, especialmente en sectores con alta carga operativa como el energético.

Este efecto es aún más notable en entornos de atención al cliente con alta variabilidad en la demanda. Konrad y Liu (2023) analizaron sistemas multicanal de atención al cliente y concluyeron que la incertidumbre en los tiempos de espera afecta la confianza del usuario en la organización. De manera similar, Correa (2023) destaca que el uso de algoritmos como Random

Forest y XGBoost permite anticipar patrones en los tiempos de atención, lo que facilita la planificación operativa y reduce la insatisfacción del cliente.

En contextos como el de la salud Shin et al. (2024) estudiaron el impacto de los tiempos de espera en la atención médica, demostrando que la implementación de modelos de predicción basados en Machine Learning permite mejorar la experiencia del cliente al reducir la incertidumbre y minimizar las quejas.

### ***Predicción de Tiempos de Espera con Machine Learning***

El uso de modelos predictivos basados en Machine Learning ha mostrado resultados prometedores en diversos sectores. Sanit-in y Runapongsa Saikaew (2019) aplicaron Random Forest en servicios centralizados, obteniendo un desempeño superior a los métodos tradicionales en la predicción de tiempos de espera.

Por otro lado, Mariño (2022) comparó modelos estadísticos con técnicas avanzadas de Machine Learning, concluyendo que redes neuronales como LSTM pueden ser altamente eficaces en la predicción de series temporales. Castro Pinzón (2023) evaluó la efectividad de Random Forest y XGBoost en entornos industriales, encontrando que estos algoritmos permiten manejar múltiples variables y realizar estimaciones precisas.

Xiao (2023) implementó Random Forest, KNN y XGBoost para la predicción de la duración de incendios forestales, resaltando la capacidad de estos modelos para manejar alta variabilidad e incertidumbre, un aspecto clave para su aplicación en la predicción de tiempos de espera en atención al cliente.

En el sector del transporte, Vassallo et al. (2024) aplicaron Random Forest y KNN para estimar tiempos de espera, resaltando la importancia del análisis de datos históricos para mejorar

la precisión del modelo. Estos enfoques pueden adaptarse a la atención al cliente, donde la disponibilidad de datos históricos permite la construcción de modelos predictivos robustos.

### ***Predicción de Demanda con Modelos de Series Temporales***

Además de la estimación de tiempos de espera, la literatura ha abordado la predicción de la demanda de usuarios, lo que permite a las empresas anticiparse a los picos de atención y mejorar su planificación operativa. Fernández López et al. (2020) realizaron un análisis comparativo entre modelos tradicionales y modernos para la predicción de la demanda, destacando que los modelos híbridos (combinación de modelos estadísticos y aprendizaje profundo) logran un mejor desempeño que los métodos convencionales.

Para la predicción de demanda, se han utilizado diversos modelos en función de la naturaleza de los datos y la necesidad de captura de patrones. Los modelos estadísticos, como ARIMA, han sido ampliamente aplicados en distintos sectores, destacándose por su eficacia en la identificación de tendencias y estacionalidades en series temporales (Ayala & Bucio, 2020). En el ámbito del Machine Learning, Prophet, se ha consolidado como una alternativa flexible y robusta a ARIMA, permitiendo una mejor detección de patrones estacionales y la predicción de tendencias de demanda (Fierro Torres et al., 2022). Por otro lado, las RNN, en particular LSTM (Long Short-Term Memory), han demostrado ser altamente eficaces en la predicción de datos secuenciales, ofreciendo mejoras significativas en la precisión en escenarios con alta variabilidad, como la estimación de la demanda de usuarios en sectores turísticos (Fernández López et al., 2020).

Dada la importancia de contar con estimaciones precisas sobre la cantidad de usuarios que asistirán en distintos momentos del día y la semana, este estudio implementará y comparará ARIMA, Prophet y RNN (LSTM) para determinar cuál de estos modelos ofrece la mejor capacidad predictiva en el contexto de la electrificadora.

### ***Comparación de Algoritmos y Modelos***

La literatura especializada destaca la importancia de comparar distintos algoritmos para seleccionar el más adecuado en un contexto específico. Tai et al. (2023) analizaron Decision Trees, Random Forest y XGBoost en la predicción de tiempos de resolución de tickets, concluyendo que XGBoost ofrece la mayor precisión en escenarios con múltiples variables dinámicas.

De manera similar, Mariño (2022) identificó que modelos como LSTM y Random Forest superan los métodos tradicionales en la predicción de series multivariadas. Castro Pinzón (2023) destacó que XGBoost y Random Forest son herramientas clave en la predicción de eventos industriales debido a su capacidad para manejar múltiples factores operativos.

En entornos de call centers multihabilidades, Thiongane et al. (2022) utilizaron modelos de aprendizaje automático basados en distribuciones de tiempo condicional, lo que permitió ajustar las predicciones a contextos con alta variabilidad en la demanda de atención.

### ***Relevancia de la Metodología CRISP-DM***

La metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) es ampliamente reconocida en proyectos de ciencia de datos por su enfoque estructurado y su aplicabilidad en múltiples industrias. Loureiro et al. (2023) aplicaron esta metodología en la predicción de tiempos de espera en distintos sectores, destacando la importancia de la fase de comprensión y preparación de los datos para garantizar la calidad del modelo final.

Correa (2023) enfatiza que una metodología estructurada, como CRISP-DM, es esencial para asegurar que los datos sean procesados de manera eficiente, especialmente cuando se manejan grandes volúmenes de información y múltiples variables operativas.

## Marco Conceptual

**Tiempos de Espera** :Se refiere al tiempo que transcurre entre la llegada de un usuario a un servicio y el momento en que recibe atención. En entornos de atención al cliente, la gestión de los tiempos de espera es un factor determinante en la percepción del servicio y la satisfacción del usuario.

**Satisfacción del Cliente**: Concepto que mide la percepción del usuario respecto a la calidad del servicio recibido. Según Caruelle et al. (2023), la relación entre los tiempos de espera y la satisfacción del cliente es clave en la evaluación de la eficiencia del servicio.

**Machine Learning**: Área de la inteligencia artificial que permite a los sistemas aprender automáticamente a partir de datos sin ser explícitamente programados. Se utiliza en este estudio para desarrollar modelos predictivos de tiempos de espera.

**CRISP-DM**: Metodología estándar utilizada en la minería de datos, estructurada en fases que incluyen comprensión del negocio, análisis y preparación de datos, modelado, evaluación y despliegue. Es ampliamente empleada en proyectos de ciencia de datos para garantizar un enfoque riguroso en el desarrollo de modelos.

**Random Forest**: Algoritmo de aprendizaje supervisado basado en la construcción de múltiples árboles de decisión para mejorar la precisión y reducir el sobreajuste. Se ha utilizado en la predicción de tiempos de espera en diversos sectores debido a su robustez en el manejo de datos heterogéneos (IBM, s.f.).

**XGBoost**: Extensión del algoritmo de boosting que mejora iterativamente la precisión de predicciones a partir de modelos previos. Según Bentéjac et al. (2020), XGBoost es reconocido por su eficiencia en tareas de regresión y clasificación, superando a otros algoritmos en competiciones de ciencia de datos.

ARIMA (AutoRegressive Integrated Moving Average): Modelo estadístico utilizado para el análisis y predicción de series temporales. Se basa en la combinación de autoregresión (AR), diferenciación integrada (I) y media móvil (MA) para identificar patrones en datos secuenciales y proyectar tendencias futuras.

Prophet: Modelo desarrollado por Facebook para la predicción de series temporales. Se caracteriza por su flexibilidad en la detección de patrones estacionales y tendencias, permitiendo realizar ajustes automáticos para mejorar la precisión de las predicciones. Es ampliamente utilizado en aplicaciones comerciales por su facilidad de implementación y capacidad para manejar datos con irregularidades.

Redes Neuronales Recurrentes (RNN): Arquitectura de redes neuronales diseñada para procesar datos secuenciales, como series temporales. A diferencia de las redes neuronales tradicionales, las RNN tienen conexiones cíclicas que les permiten mantener información sobre patrones previos, lo que resulta útil para la predicción de demanda y la identificación de tendencias a lo largo del tiempo.

LSTM (Long Short-Term Memory): Variante de las RNN que incorpora mecanismos de memoria a largo plazo, lo que permite retener información relevante durante períodos prolongados y evitar problemas como el desvanecimiento del gradiente. Este modelo es particularmente eficaz en la predicción de datos secuenciales con alta variabilidad, como la demanda de usuarios en un servicio de atención.

Métricas de Evaluación (RMSE, MAE y  $R^2$ ): Indicadores utilizados para medir la precisión de los modelos predictivos.

RMSE (Root Mean Squared Error): Mide la diferencia entre los valores predichos y los valores reales, penalizando errores grandes.

MAE (Mean Absolute Error): Representa el error promedio en términos absolutos, sin considerar la dirección del error.

R<sup>2</sup> (Coeficiente de determinación): Indica qué proporción de la variabilidad de los datos es explicada por el modelo.

### **Marco Normativo**

El Marco Normativo establece las disposiciones legales y reglamentarias que rigen la prestación del servicio de atención al cliente en las empresas del sector energético en Colombia, así como aquellas que regulan el manejo de datos y la calidad del servicio. Estas normativas son fundamentales para garantizar que la implementación de modelos predictivos de tiempos de espera y demanda de usuarios cumpla con los principios de eficiencia, transparencia y protección de datos personales.

La Ley 142 de 1994, que regula los servicios públicos domiciliarios en Colombia, establece en su Artículo 133 la obligación de las empresas prestadoras de garantizar un servicio eficiente y brindar información clara y oportuna a los usuarios sobre los procedimientos y tiempos de atención. En este sentido, la implementación de modelos predictivos para estimar los tiempos de espera y la demanda de usuarios en el servicio de atención al cliente está alineada con este marco normativo, ya que proporcionaría a los usuarios información anticipada sobre la gestión de sus turnos, optimizando la transparencia y percepción del servicio.

La Resolución CREG 108 de 1997, emitida por la Comisión de Regulación de Energía y Gas (CREG), establece estándares de calidad en la prestación del servicio de energía eléctrica, incluyendo indicadores sobre la atención al usuario y tiempos de respuesta en las solicitudes de servicio. Un modelo predictivo de tiempos de espera y demanda contribuiría a mejorar el cumplimiento de estos estándares, permitiendo a la empresa optimizar sus procesos operativos,

anticiparse a los momentos de alta afluencia y gestionar eficientemente la distribución del personal de atención.

A nivel internacional, la Norma ISO 9001:2015 establece estándares para la gestión de la calidad en los procesos organizacionales, promoviendo la mejora continua en la atención al cliente y la eficiencia operativa. La aplicación de modelos de Machine Learning en la predicción de tiempos de espera y demanda se encuentra alineada con este enfoque, ya que permite optimizar los procesos de atención mediante la identificación de patrones en la afluencia de usuarios y la gestión de recursos en función de las tendencias históricas.

En el ámbito de la protección de datos personales, la Ley 1581 de 2012 establece el marco legal en Colombia para la recolección, tratamiento y almacenamiento de información de los ciudadanos. Dado que los modelos predictivos se basarán en datos históricos de atención al cliente, es fundamental asegurar que la información utilizada cumpla con los principios de confidencialidad, anonimización y uso legítimo de los datos. Para garantizar el cumplimiento de esta normativa, el desarrollo del modelo se enfocará en la exclusión de datos personales y en la utilización de información anonimizada, asegurando que el análisis de tiempos de espera y demanda se realice sin comprometer la privacidad de los usuarios.

El Código de Procedimiento Administrativo y de lo Contencioso Administrativo – Ley 1437 de 2011 regula la relación entre los ciudadanos y las entidades prestadoras de servicios públicos, estableciendo derechos y deberes de los usuarios, incluyendo el derecho a recibir información clara y precisa sobre tiempos de atención y procesos administrativos. La implementación de un modelo predictivo contribuiría a garantizar este derecho, proporcionando a los usuarios una herramienta que les permita conocer con mayor precisión los tiempos de espera y facilitando la toma de decisiones sobre su permanencia en el establecimiento.

El desarrollo de modelos predictivos para la estimación de tiempos de espera y demanda en el servicio de atención al cliente de una empresa del sector energético se encuentra alineado con diversas normativas vigentes en Colombia. Desde la Ley 142 de 1994, que regula la prestación de servicios públicos domiciliarios, hasta estándares internacionales como la ISO 9001:2015, estas regulaciones enfatizan la importancia de mejorar la calidad en la atención al usuario y garantizar la eficiencia operativa. Además, el cumplimiento de la Ley 1581 de 2012 sobre Protección de Datos Personales será un aspecto clave en la implementación del modelo, garantizando que la información utilizada respete los principios de privacidad y confidencialidad. En este contexto, la integración de herramientas tecnológicas basadas en Machine Learning y modelos de predicción de series temporales no solo representa una innovación en la gestión operativa, sino que también contribuye al cumplimiento de los estándares regulatorios en el sector de servicios públicos.

## Metodología

El desarrollo del presente estudio se enmarca dentro de la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), ampliamente adoptada en proyectos de ciencia de datos por su enfoque estructurado y progresivo. Esta metodología permite abordar desde la comprensión del problema hasta la validación de los modelos, asegurando una alineación constante con los objetivos definidos: estimar los tiempos de espera y predecir la demanda de usuarios en la oficina principal de una electrificadora.

En la fase inicial de comprensión del negocio, se identificaron las principales necesidades operativas de la empresa, enfocadas en mejorar la planificación del recurso humano y reducir la incertidumbre en la atención presencial. Se establecieron las preguntas clave del proyecto, incluyendo la identificación de variables que influyen en los tiempos de espera y los patrones de comportamiento diario en la afluencia de usuarios.

Posteriormente, se desarrolló la comprensión de los datos disponibles a través del análisis de los registros históricos suministrados por la electrificadora, correspondientes al periodo enero de 2023 a febrero de 2025. Para la predicción del tiempo de espera, se consideraron variables como la hora del día, el tipo de servicio y la cantidad de turnos en sala. En el caso de la demanda, se utilizó la fecha como variable temporal, con el fin de modelar el número de usuarios atendidos por día.

En la fase de preparación, se aplicaron técnicas de limpieza, transformación y codificación de variables, se eliminaron registros con valores nulos o inconsistentes, y se excluyeron días atípicos relacionados con caídas de sistema que afectaban la distribución general. Las variables categóricas se codificaron utilizando One-Hot Encoding y Label Encoding, según las

características del modelo, y se realizó un análisis de correlación para seleccionar las variables predictoras más relevantes.

El proceso de modelado incluyó el entrenamiento de diferentes algoritmos de aprendizaje supervisado para la predicción del tiempo de espera, como Regresión Lineal, Random Forest y XGBoost. Aunque inicialmente se consideró el uso de redes neuronales recurrentes (RNN) para esta tarea, su aplicación fue descartada al no ajustarse a la naturaleza no secuencial de la variable dependiente. Para la predicción de la demanda de usuarios, se optó por modelos de series temporales como ARIMA, Prophet y RNN (en su variante LSTM), dada su capacidad para capturar patrones estacionales, fluctuaciones no lineales y efectos de días festivos en los datos.

Durante la fase de evaluación, se aplicaron métricas de desempeño adecuadas para cada tipo de modelo. En el caso del tiempo de espera, se utilizaron RMSE, MAE y  $R^2$ . Para la demanda, se incluyeron además métricas como el Mean Absolute Scaled Error (MASE) y el Mean Squared Logarithmic Error (MSLE), lo que permitió evaluar el comportamiento de los modelos incluso en presencia de picos o caídas abruptas en la afluencia de usuarios.

Finalmente, los modelos desarrollados fueron documentados en un entorno reproducible mediante Jupyter Notebook, incluyendo visualizaciones interpretables y código estructurado. Aunque este prototipo no será implementado en producción, se deja como base para futuras pruebas piloto dentro de la empresa, con potencial para fortalecer la toma de decisiones operativas.

## **Método**

La investigación adopta un enfoque estructurado basado en la metodología CRISP-DM, ampliamente validada en proyectos de ciencia de datos. Esta metodología permite abordar de manera ordenada la comprensión del problema, la exploración y transformación de los datos, el desarrollo de modelos predictivos y su evaluación mediante métricas estandarizadas. A través de

este proceso, se garantiza una alineación constante con los objetivos específicos del proyecto y se asegura la reproducibilidad del análisis.

### **Tipo de Estudio**

El presente trabajo se clasifica como una investigación de tipo aplicada, con un enfoque cuantitativo y predictivo. Su propósito es construir modelos que permitan anticipar comportamientos operativos en un entorno real, específicamente en el proceso de atención al cliente de una empresa del sector energético. Dado que no busca intervenir directamente en el sistema estudiado, sino generar insumos analíticos para futuras decisiones, se enmarca dentro de un diseño experimental no aplicado con orientación práctica.

### **Recolección de Datos**

Los datos analizados provienen de los sistemas de gestión interna de la electrificadora y abarcan el periodo comprendido entre enero de 2023 y febrero de 2025. Tras aplicar procesos rigurosos de limpieza, filtrado y transformación, se consolidó una base final compuesta por 106.389 registros, correspondiente a los turnos generados por usuarios que ingresaron a la sala de espera en la oficina principal de atención.

El dataset incluye variables operativas relevantes como zona, sede, grupo o tipo de turno, tipo de servicio solicitado, canal de atención, hora de impresión del turno, tiempo de espera y tiempo de atención. Además, se dispone de variables temporales como fecha y hora de impresión, llamado y atención, lo que permite reconstruir el flujo de usuarios. Para el modelo de predicción del tiempo de espera, la variable objetivo se definió como el tiempo transcurrido entre la generación del turno y el inicio efectivo de la atención, en el caso del modelo de demanda, la variable objetivo corresponde al número total de turnos generados por día.

El procesamiento incluyó la eliminación de registros nulos o inconsistentes, así como la exclusión de eventos atípicos como días con interrupciones operativas, las variables categóricas fueron codificadas mediante técnicas como One-Hot Encoding y Label Encoding, y se derivaron variables adicionales como “turnos esperando” para enriquecer la capacidad predictiva de los modelos. Cabe resaltar que durante todo el tratamiento de la información se garantizó el cumplimiento de lo dispuesto en la Ley 1581 de 2012, asegurando la anonimización de los datos personales y la protección de la información sensible.

## **Resultados**

El análisis de los datos permitió identificar patrones relevantes en la atención al cliente, tanto en los tiempos de espera individuales como en la demanda diaria. A partir de los 106.389 registros consolidados, se realizó un diagnóstico exploratorio que sirvió como base para el desarrollo de los modelos predictivos. En las siguientes secciones se presentan los hallazgos obtenidos en la fase de análisis exploratorio, así como el comportamiento de los modelos entrenados para cada una de las variables objetivo.

### **Análisis Exploratorio de los Datos (ADE)**

La base de datos utilizada contiene 106.389 registros (filas), correspondientes a turnos generados entre enero de 2023 y febrero de 2025 en la oficina principal de la electrificadora. Estos registros fueron depurados a partir de una base original de 164.299 filas, eliminando aquellos que no correspondían a servicios del sector eléctrico o que presentaban inconsistencias, valores nulos o formatos erróneos.

Es importante señalar que cada fila representa un identificador único de turno-servicio, ya que un mismo turno puede incluir múltiples servicios solicitados por el usuario, esto permite distintos niveles de análisis: individual por servicio o consolidado por turno.

Se identificó que de estos 106.389, 26.267 (24,69%) correspondían a turnos donde se solicitaron más de un servicio, lo que indica que al menos uno de cada cuatro usuarios realiza múltiples gestiones por visita y donde los turnos con dos servicios alcanzan una cifra notable de 21.335 casos, siendo la segunda frecuencia más alta después de los turnos con un único servicio.

Durante esta fase se construyó una variable derivada clave denominada “turnos esperando”, la cual representa cuántos turnos estaban activos en la sala al momento en que un nuevo usuario generaba el suyo. Esta variable no estaba disponible originalmente en el sistema,

por lo que fue calculada a partir de las marcas temporales registradas. Para cada turno, se identificaron todos aquellos generados previamente que aún no habían sido llamados, restringiendo la comparación al mismo día de atención. Esta lógica permitió estimar de manera precisa la saturación operativa al momento de llegada del usuario, siendo una de las variables con mayor correlación con el tiempo de espera observado.

La distribución de esta variable evidenció una alta concentración de observaciones por debajo de los 40 turnos esperando. Sin embargo, se identificaron valores extremos en días específicos, como el 9 de enero de 2024, en los que se registraron más de 80 turnos simultáneamente en sala, posiblemente debido a caídas del sistema u otras contingencias operativas. Para mitigar el impacto de estos eventos atípicos, se revisaron los casos y se excluyeron del conjunto de entrenamiento aquellos días que presentaban desviaciones anómalas sostenidas, con el fin de mejorar la calidad del modelo y su capacidad de generalización.

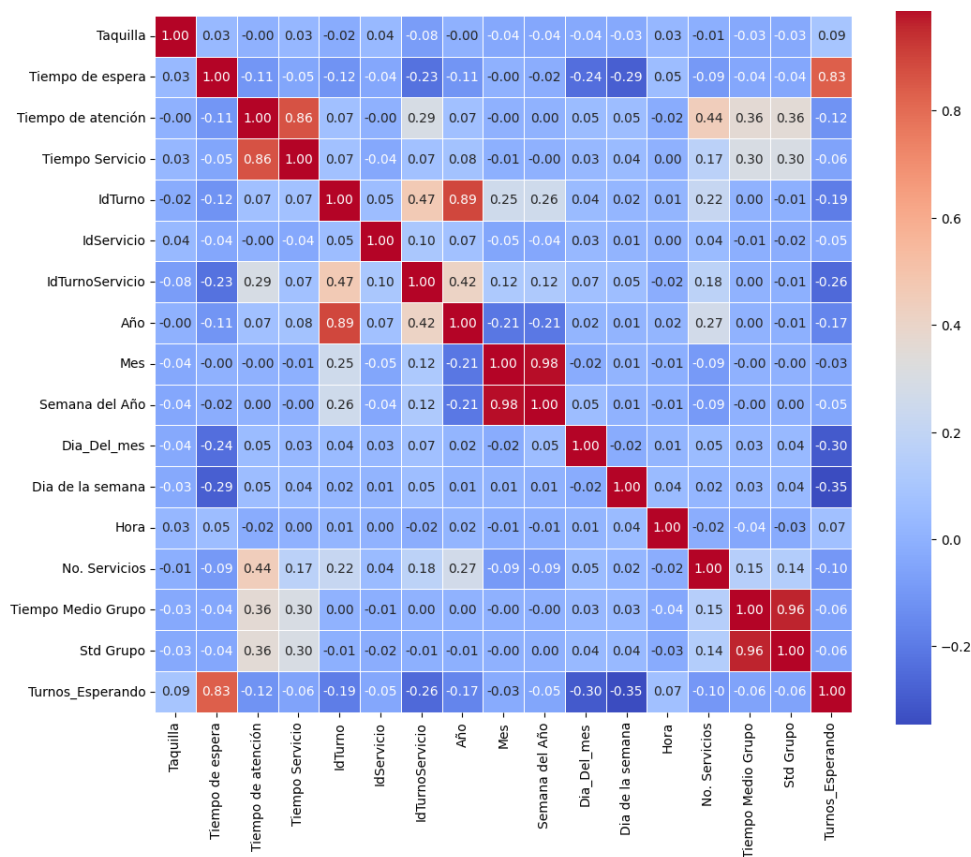
En cuanto al análisis de correlación entre variables, se destacan las relaciones que se pueden ver en la **Figura 1**.

La **Figura 1** muestra la matriz de correlaciones entre las principales variables numéricas del conjunto de datos. Se destaca una fuerte correlación positiva entre la variable derivada “turnos esperando” y el tiempo de espera ( $r = 0.83$ ), lo que confirma que la saturación en sala al momento de llegada del usuario es un factor determinante en la duración de su espera. Asimismo, se observa una alta correlación entre el tiempo de atención y el tiempo de servicio ( $r = 0.86$ ), lo cual es coherente dado que ambas variables representan distintas fases del mismo proceso operativo. También se identifica una relación moderada entre el número de servicios solicitados y el tiempo de atención ( $r = 0.47$ ), lo que sugiere que los usuarios que realizan múltiples gestiones tienden a requerir más tiempo en ventanilla. Estas relaciones fueron tenidas

en cuenta durante la selección de variables para el entrenamiento del modelo predictivo del tiempo de espera.

**Figura 1**

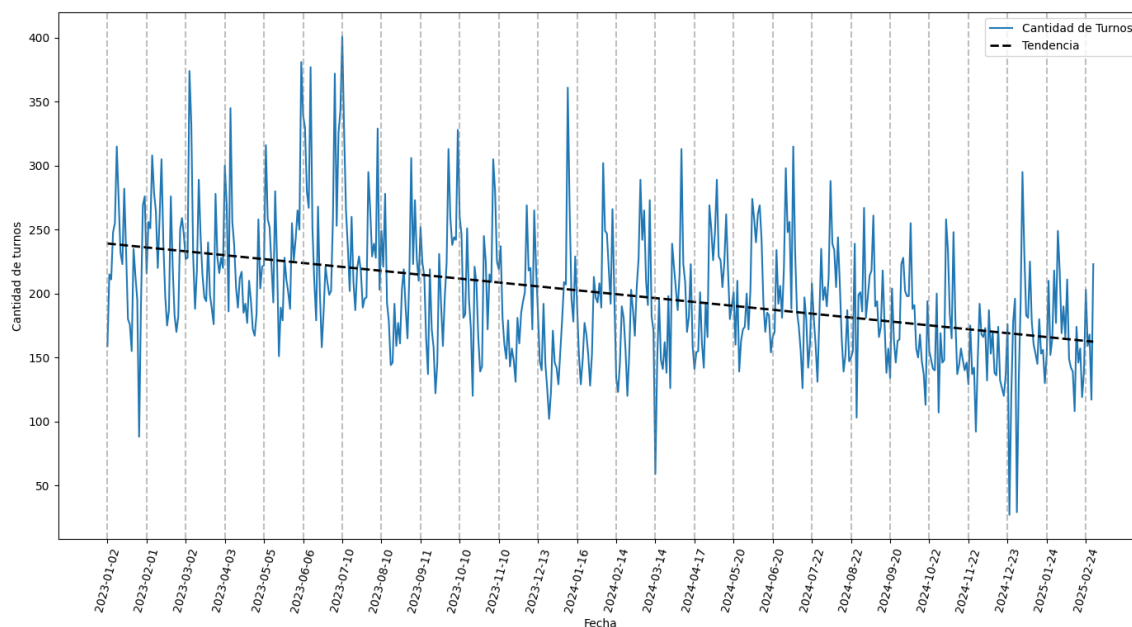
*Matriz de Correlaciones Entre Variables Numéricas del Dataset Depurado*



Respecto al análisis de arribos, la **Figura 2** presenta la evolución de la cantidad diaria de turnos durante el periodo 2023-2025. Se evidencia una tendencia general decreciente, aunque con alta variabilidad. Existen épocas con mayor intensidad operativa, especialmente al inicio de cada mes, relacionadas con los ciclos de facturación.

**Figura 2**

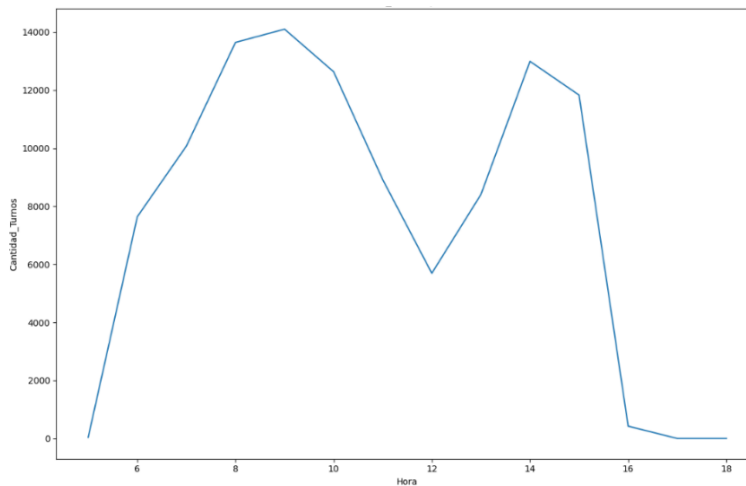
*Evolución Diaria de la Cantidad de Turnos Generados (Enero 2023 - Febrero 2025)*



En este sentido, la **Figura 3** muestra el patrón horario de llegada de usuarios, destacando dos picos: uno en la mañana entre las 8 y 10 a.m. y otro más moderado hacia las 2 p.m., indicando un comportamiento claramente bimodal.

**Figura 3**

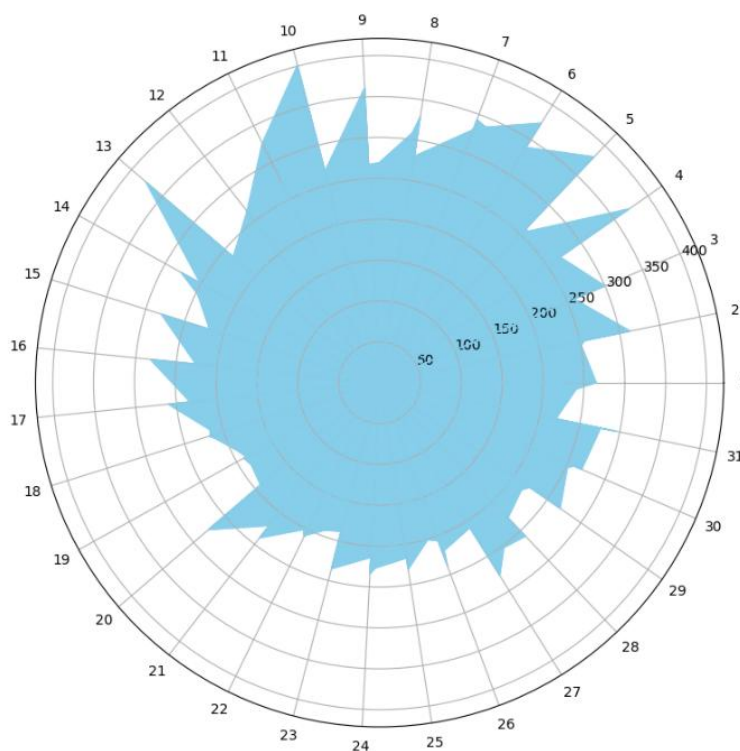
Distribución de la Cantidad de Turnos Generados por Hora del día



Complementariamente, la **Figura 4** representa la distribución diaria de turnos por mes, evidenciando que los mayores volúmenes de atención se concentran en los primeros 15 días de cada mes, especialmente entre los días 4 y 10. Este patrón respalda la hipótesis del impacto de los ciclos de facturación en la demanda del servicio.

#### **Figura 4**

*Distribución Diaria de Turnos por día del mes*

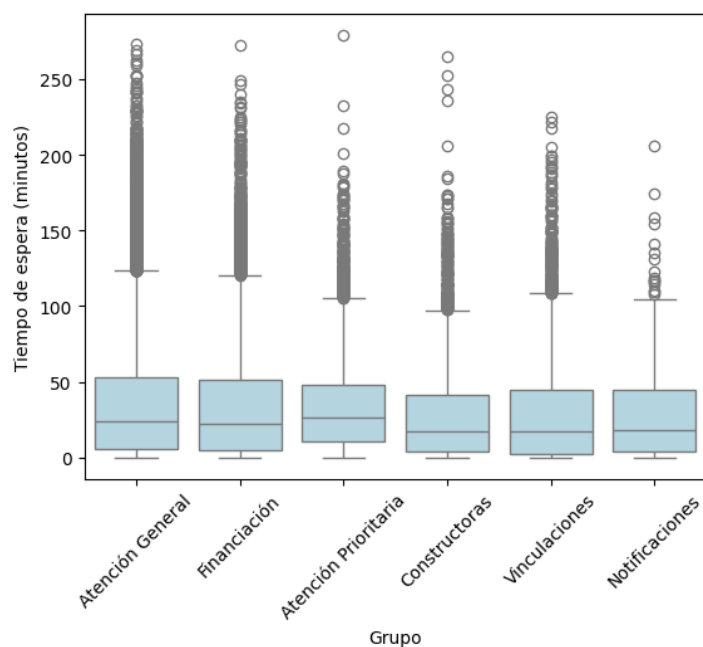


En nuestro contexto, los grupos de atención corresponden a categorías predefinidas que clasifican el motivo principal por el cual un usuario solicita un turno. Estos grupos son visibles para el cliente al momento de registrarse, y su elección permite dirigir la solicitud hacia las ventanillas adecuadas según la especialidad del trámite. Entre los grupos más frecuentes se encuentran: Atención General, Financiación, Atención Prioritaria, Constructoras, entre otras.

Desde esta perspectiva, “el grupo “Atención General” acumula la mayoría de los turnos registrados (54.305), seguido por “Financiación” con 19.335 y “Atención Prioritaria” con 11.261. Esta distribución podría explicarse, en parte, por una posible falta de claridad del usuario al clasificar su trámite, optando por la categoría más genérica. La **Figura 5** presenta la distribución del tiempo de espera por grupo, destacándose que “Atención Prioritaria” exhibe la mediana más alta. Este comportamiento sugiere que, podría no estar recibiendo una gestión diferencial, lo que obliga a los usuarios de este grupo a esperar en condiciones similares al resto.

### Figura 5

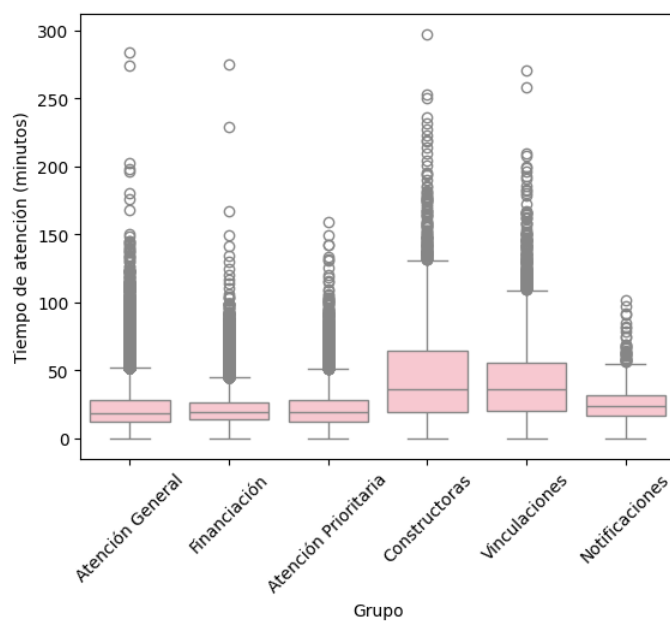
*Distribución del Tiempo de Espera por Grupo de Atención*



En contraste, en la **Figura 6**, se visualiza que los grupos “Constructoras” y “Vinculaciones” presentan los tiempos de atención más altos, con medianas que superan los 35 minutos. Esto es comprensible, dado que sus trámites tienden a ser más complejos y extensos, requiriendo una mayor dedicación por parte del personal.

**Figura 6**

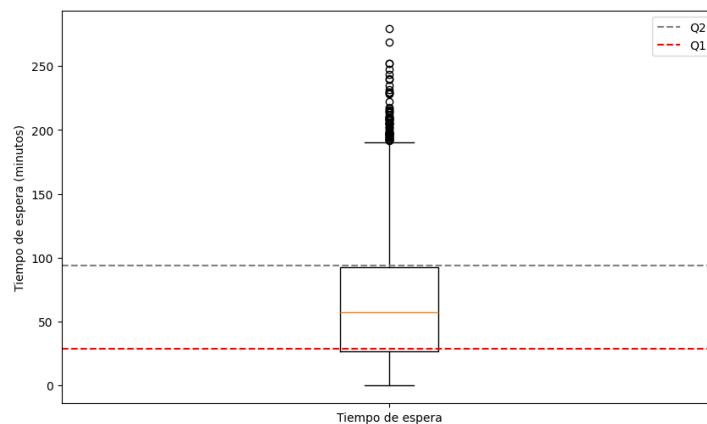
*Distribución del Tiempo de Atención por Grupo de Atención*



En términos de eficacia en la atención, se identificó un 5,5% de turnos abandonados frente a un 94,5% atendidos. Este porcentaje de abandono se analiza con mayor detalle en la **Figura 7**, donde se muestra la distribución del tiempo de espera de los turnos abandonados. La mediana se sitúa en torno a los 60 minutos, mientras que los valores extremos alcanzan hasta 270 minutos (Linea roja corresponde  $y=29$  min, y la gris  $y=94$  min). Algunos registros con tiempos cercanos a cero podrían corresponder a errores de sistema o cancelaciones inmediatas por parte del usuario.

### Figura 7

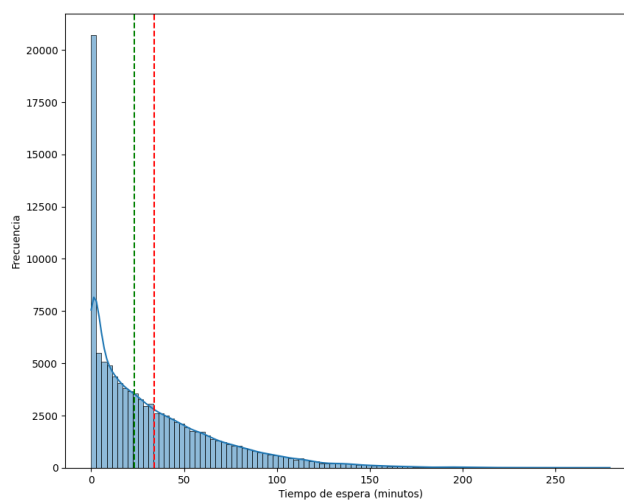
*Distribución del Tiempo de Espera en los Turnos Abandonados, con Referencia a Cuartiles*



El comportamiento del tiempo de espera se representa en la **Figura 8**, mediante un histograma que muestra una distribución asimétrica con sesgo a la derecha. La mayoría de los turnos tienen tiempos de espera inferiores a 50 minutos. La línea verde indica la mediana, y la roja la media, desplazada hacia la derecha por valores extremos. Esta diferencia confirma un sesgo positivo, típico en sistemas donde algunos usuarios experimentan esperas notablemente superiores al promedio.

### Figura 8

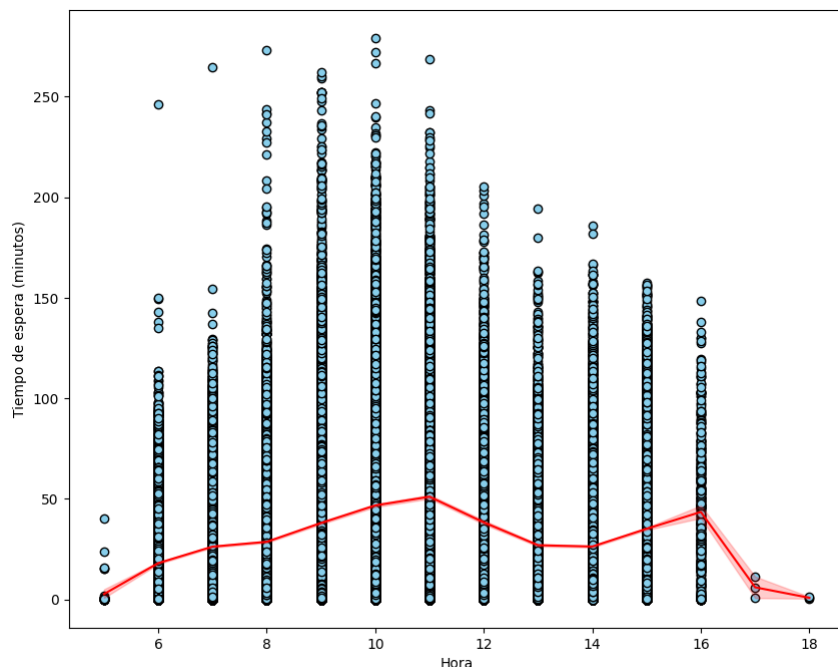
*Distribución del Tiempo de Espera con Referencia a Mediana y Media*



La **Figura 9** profundiza en cómo varía el tiempo de espera a lo largo del día. Se observa que las horas entre las 9:00 a.m. y 12:00 m. tienden a registrar los mayores tiempos de espera promedio, lo que puede relacionarse con los picos de afluencia mostrados previamente.

### Figura 9

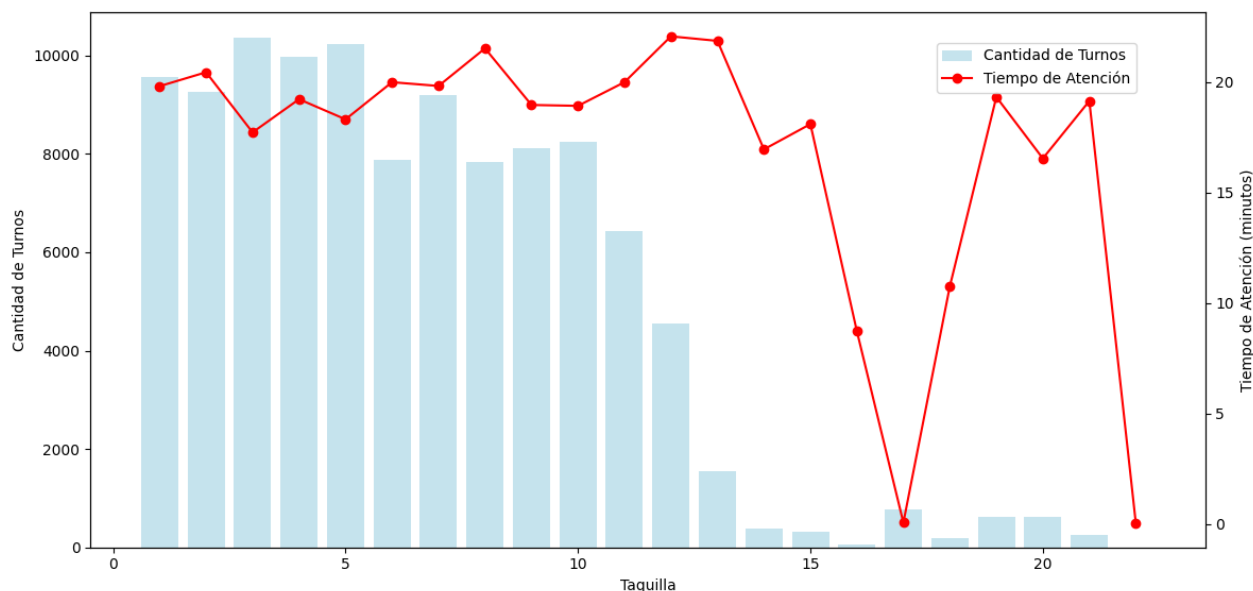
*Distribución del Tiempo de Espera por Hora del día con Promedio Suavizado*



Finalmente, la **Figura 10** analiza la cantidad de turnos atendidos por cada taquilla junto con su tiempo de atención promedio. Las taquillas 1 y 2, que atienden trámites de constructoras, presentan los tiempos más altos debido a la complejidad de sus gestiones. Asimismo, la taquilla 10, asignada a atención prioritaria, muestra un comportamiento diferenciado. Es importante destacar que las taquillas 11 y 12, aunque habilitadas, han permanecido inactivas durante varios meses, y que el personal suele rotar entre ventanillas, lo que diluye la posibilidad de hacer un análisis por agente individual.

**Figura 10**

*Cantidad de Turnos Atendidos y Tiempo Promedio de Atención por Taquilla*



En resumen, el análisis exploratorio permitió identificar patrones relevantes en la dinámica operativa del servicio de atención al cliente, destacando variaciones por hora del día, grupo de atención, taquilla y comportamiento de abandono. La construcción de variables derivadas como turnos esperando y la identificación de eventos atípicos fortalecieron la comprensión del sistema y permitieron preparar un conjunto de datos robusto para el entrenamiento de modelos predictivos. A continuación, se presentan los resultados obtenidos en el proceso de modelado, comenzando por la estimación del tiempo de espera a partir de variables explicativas observadas al momento del ingreso del usuario.

### **Modelado del Tiempo de Espera**

Previo al entrenamiento de los modelos predictivos, se realizó un análisis exhaustivo para identificar las variables con mayor capacidad explicativa sobre el tiempo de espera. La selección se basó en el análisis de correlaciones (como se mostró en la **Figura 1**), en observaciones de

comportamiento empírico durante la fase exploratoria, así como en criterios operativos basados en la lógica del proceso de atención. Dado el tipo de datos y la naturaleza del problema, no se aplicaron pruebas como ANOVA o chi-cuadrado, priorizándose un enfoque basado en transformación y evaluación empírica del desempeño de los modelos con distintas combinaciones de variables.

En el modelo final, se incluyeron variables numéricas como: hora de llegada, número de servicios, tiempo medio del grupo, desviación estándar del grupo, número de turnos esperando, tiempo desde la apertura, flujo previo, año, mes, día del mes, día de la semana y semana del año. A su vez, se codificaron variables cualitativas como el grupo de atención utilizando técnicas de One-Hot Encoding, lo cual generó columnas binarias para categorías como Atención General, Financiación, Prioritaria, entre otras. Este conjunto de variables resultó ser el más eficaz en términos predictivos, permitiendo capturar tanto patrones temporales como características operativas del entorno de atención.

Tras la selección y transformación de variables explicativas, se procedió a la construcción y evaluación de los modelos para predecir el tiempo de espera. Inicialmente se entrenaron tres algoritmos de referencia: regresión lineal, Random Forest y XGBoost. El propósito fue establecer una línea base de comparación, observar su comportamiento sin ajustes y seleccionar los modelos más prometedores para optimización posterior.

Los resultados iniciales, obtenidos tras aplicar ambos métodos de codificación (One-Hot Encoding y Label Encoding), se resumen en la **Tabla 1**. En general, se observó que el tipo de codificación no generó diferencias significativas en el rendimiento de los modelos. No obstante, en algunos casos, como en Random Forest y XGBoost, la codificación por etiquetas mostró ligeras mejoras en la métrica MAE.

**Tabla 1**

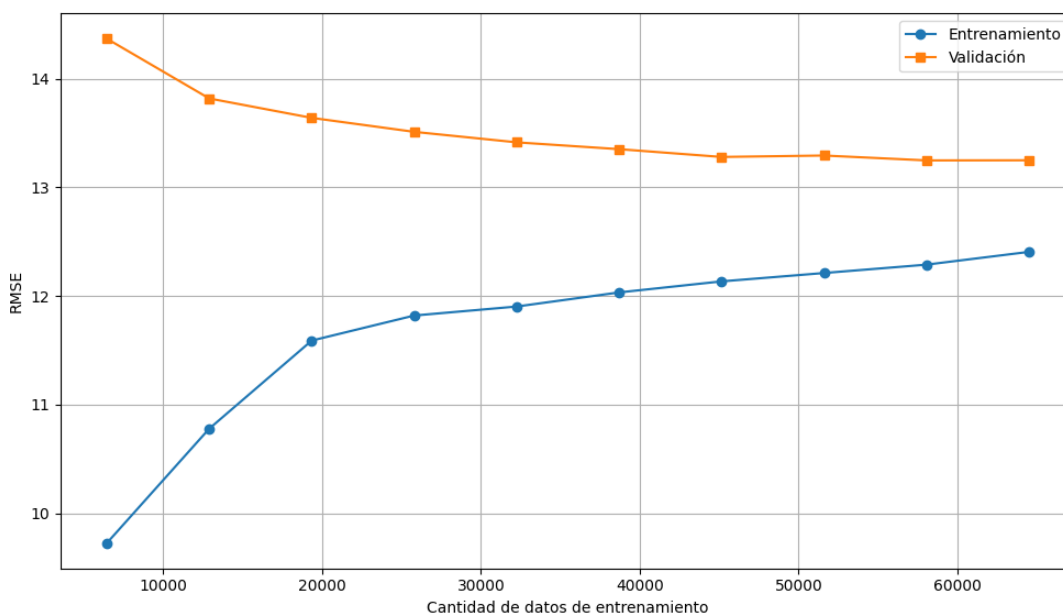
*Resultados Iniciales de Modelos Predictivos Según Tipo de Codificación*

Modelo	Codificación	RMSE	MAE	R <sup>2</sup>
Random Forest	One-Hot Encoding	11.99	7.33	0.8759
Random Forest	Label Encoding	12.01	7.37	0.8754
Regresión Lineal	One-Hot Encoding	18.41	12.58	0.7178
Regresión Lineal	Label Encoding	18.37	12.14	0.7084
XGBoost	One-Hot Encoding	13.78	8.86	0.8449
XGBoost	Label Encoding	13.45	8.56	0.8438

La **Figura 11** muestra la curva de aprendizaje del modelo XGBoost sin ajuste de hiperparámetros. Se observa una tendencia decreciente del error de validación conforme aumenta el tamaño del conjunto de entrenamiento, con una diferencia moderada frente al error de entrenamiento, lo que sugiere un buen equilibrio entre capacidad de aprendizaje y generalización.

**Figura 11**

*Curva de Aprendizaje de XGBoost (Modelo Base)*



Posteriormente, se aplicó una búsqueda aleatoria de hiperparámetros sobre los modelos más prometedores, particularmente Random Forest y XGBoost. Para el caso de Random Forest, los mejores parámetros encontrados fueron:  $n\_estimators = 200$ ,  $max\_depth = 30$ ,  $min\_samples\_split = 2$ ,  $min\_samples\_leaf = 1$ ,  $max\_features = 'log2'$ . Con estos ajustes, el modelo obtuvo:

RMSE: 11.45 minutos

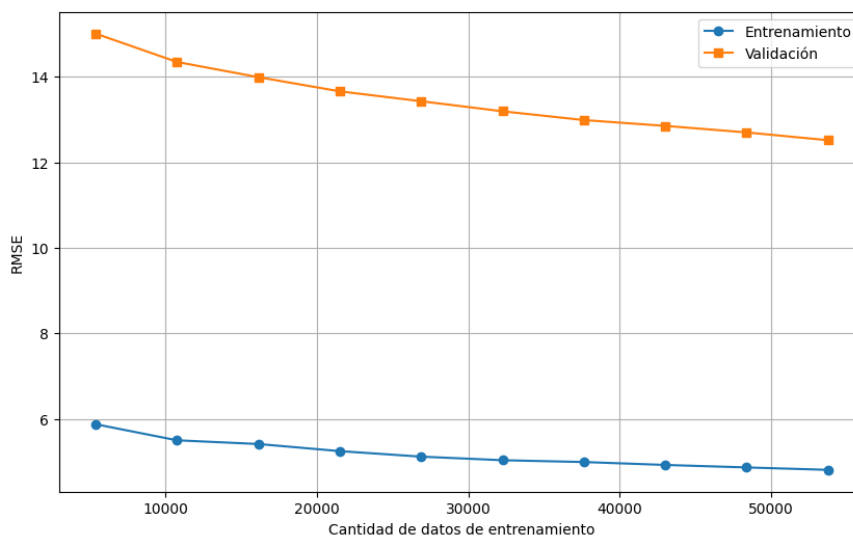
MAE: 6.96 minutos

$R^2$ : 0.89

No obstante, como se aprecia en la **Figura 12**, la curva de aprendizaje de Random Forest muestra una marcada diferencia entre los errores de entrenamiento y validación, indicando posible sobreajuste.

### Figura 12

*Curva de Aprendizaje del Modelo Random Forest Optimizado*



En el caso de XGBoost, la optimización mediante RandomizedSearchCV arrojó los siguientes hiperparámetros:  $n\_estimators = 200$ ,  $learning\_rate = 0.3$ ,  $max\_depth = 9$ ,  $gamma = 5$ ,

subsample = 1.0, colsample\_bytree = 1.0, reg\_alpha = 1, reg\_lambda = 2. Con esta configuración, el modelo logró:

RMSE: 11.62 minutos

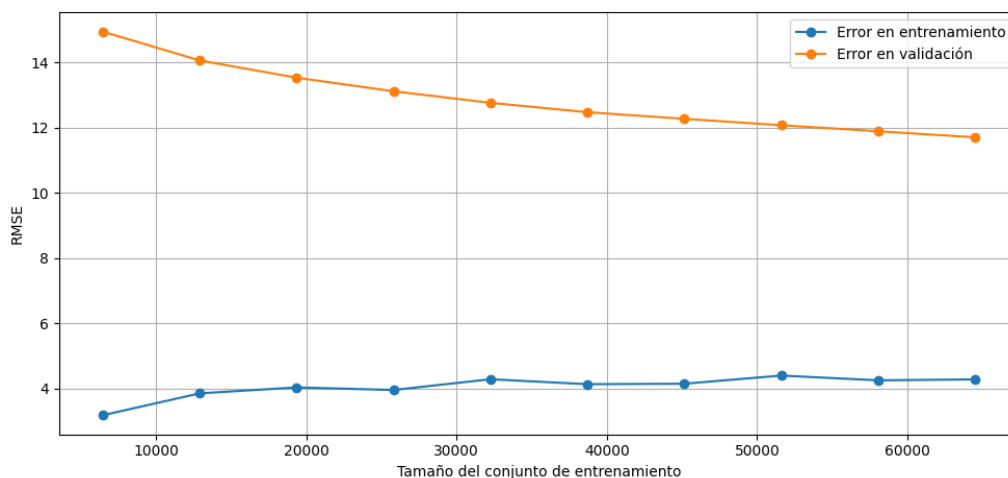
MAE: 7.21 minutos

R<sup>2</sup>: 0.8834

Sin embargo, tal como lo evidencia la **Figura 13**, este modelo también mostró una mayor varianza entre el error de entrenamiento y el de validación, lo que compromete parcialmente su capacidad de generalización.

### Figura 13

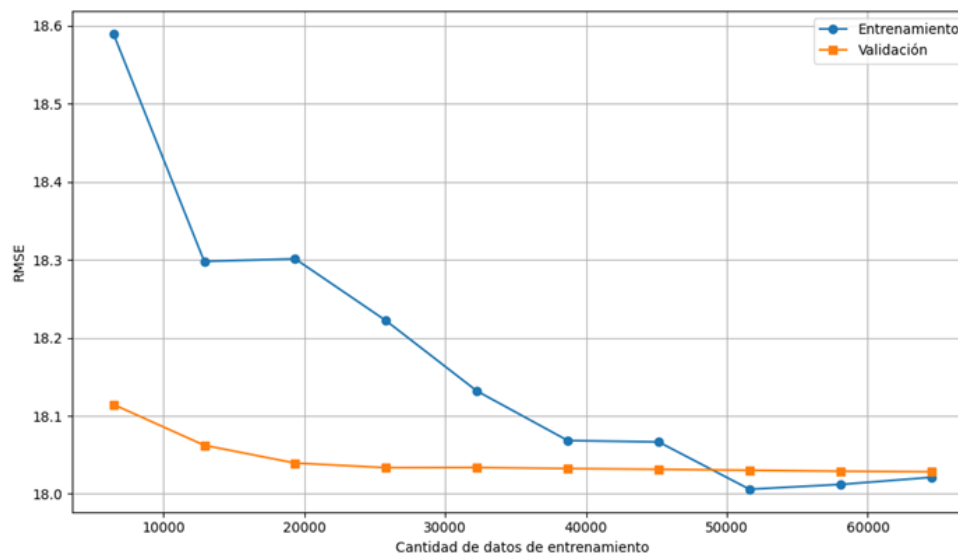
*Curva de Aprendizaje del Modelo XGBoost Optimizado*



Por comparación, la regresión lineal fue descartada como modelo viable. Su curva de aprendizaje, ilustrada en la **Figura 14**, evidenció errores constantes y relativamente altos tanto en entrenamiento como en validación, con escasa capacidad de mejora incluso al aumentar el tamaño del conjunto de entrenamiento.

**Figura 14**

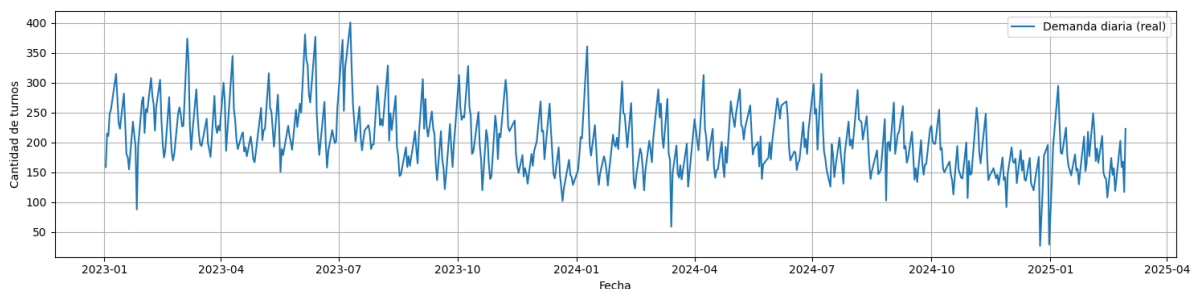
*Curva de Aprendizaje del Modelo de Regresión Lineal*



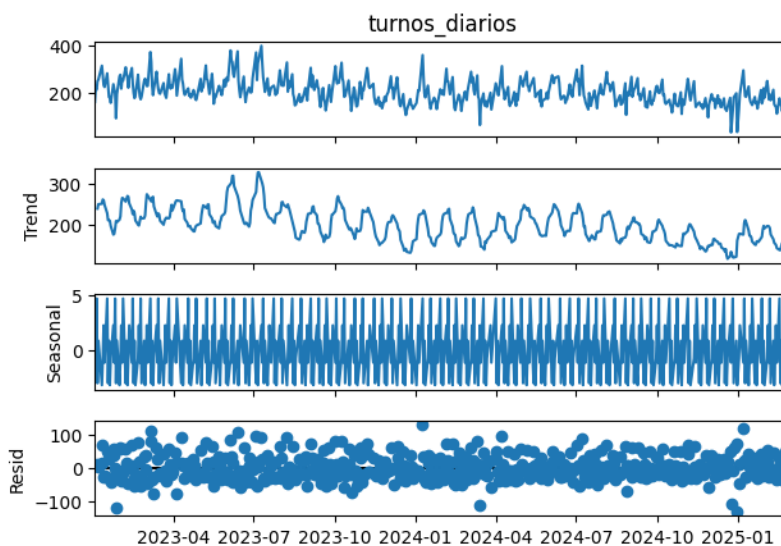
En síntesis, a pesar de que Random Forest obtuvo las mejores métricas de ajuste, su tendencia al sobreajuste lo convierte en una opción menos confiable. XGBoost, en cambio, ofreció un desempeño competitivo con una curva de aprendizaje más estable, lo que lo posiciona como el modelo más adecuado para la predicción del tiempo de espera en este caso de estudio.

### **Modelado de la Demanda**

Para el modelado de la demanda diaria de usuarios, fue necesario transformar la base de datos a una serie temporal continua con frecuencia diaria. En la **Figura 15** se observa la evolución de la demanda a lo largo del tiempo, donde se evidencia una tendencia general a la baja y una variabilidad consistente con los ciclos mensuales observados en el análisis exploratorio.

**Figura 15***Evolución Diaria de la Demanda de Turnos (2023–2025)*

Con el fin de explorar la viabilidad de aplicar modelos clásicos de series temporales, se realizó una descomposición estacional (**Figura 16**). Esta confirmó la presencia de una tendencia decreciente, estacionalidad semanal pronunciada (dado que el servicio opera en días hábiles) y residuos aparentemente aleatorios. Además, se aplicó la prueba de Dickey-Fuller aumentada para validar estacionariedad, arrojando un valor  $p > 0.05$ , lo que confirmó la necesidad de diferenciar la serie para estabilizarla. Tras una primera diferenciación ( $d = 1$ ), la serie resultó estacionaria con un  $p < 0.01$ .

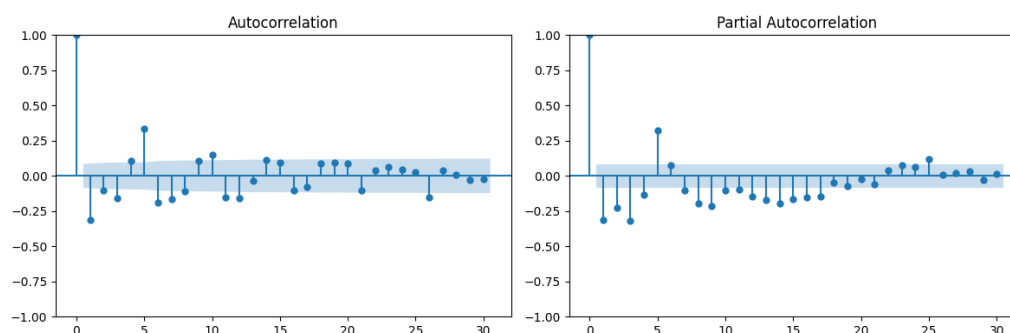
**Figura 16***Descomposición de la Serie Temporal Diaria de Turnos*

### Modelado con ARIMA

Posteriormente, se aplicó el modelo ARIMA. En la **Figura 17** se presentan los gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF), que guiaron la selección de los parámetros del modelo. El mejor desempeño fue obtenido con ARIMA(5,1,2), donde se compara el pronóstico con los datos reales. Las métricas obtenidas para diferentes combinaciones de parámetros se resumen en la **Tabla 2**:

#### Figura 17

*Funciones de Autocorrelación (ACF) y Autocorrelación Parcial (PACF)*



**Tabla 2**

*Desempeño del Modelo ARIMA para Diferentes Configuraciones*

Modelo	RMSE	MAE	MAPE(%)	R <sup>2</sup> Score
ARIMA(5,1,2)	40.40	30.14	25.95	0.0218
ARIMA(2,1,2)	40.62	30.45	26.28	0.0110
ARIMA(2,1,1)	48.32	39.06	35.11	-0.3995
ARIMA(1,1,1)	48.32	39.06	35.11	-0.3995

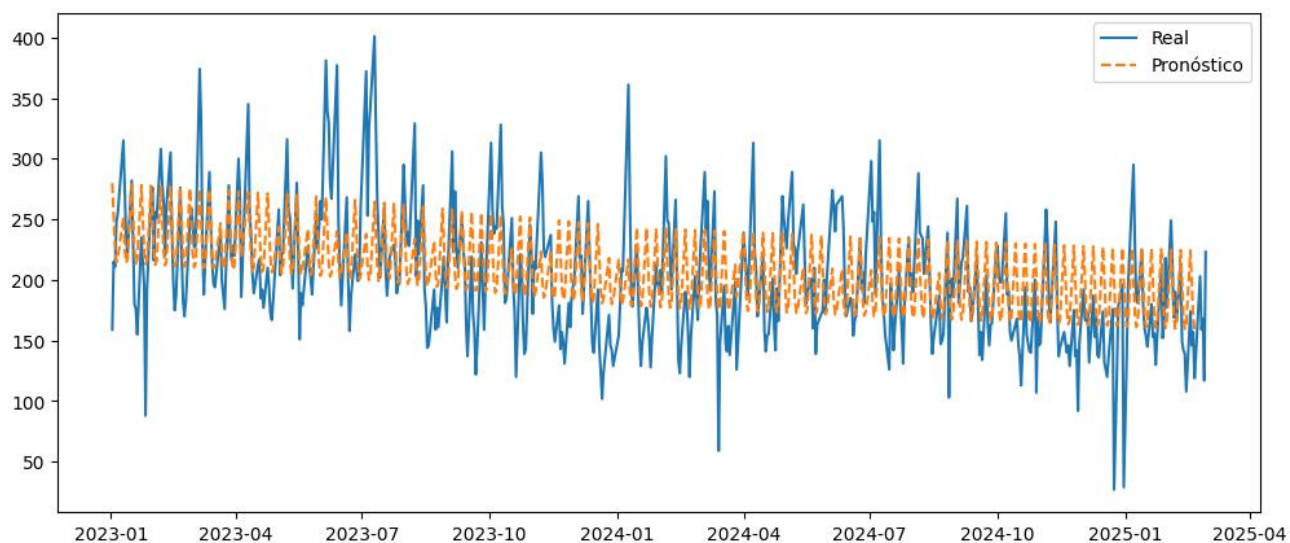
Los resultados muestran que el modelo ARIMA presentó limitaciones importantes para capturar la variabilidad de la demanda diaria. Aunque el RMSE se mantuvo por debajo de 41, el coeficiente de determinación ( $R^2$ ) apenas superó el 0.02, lo cual evidencia que el modelo explica muy poco de la variabilidad total de la serie. Las métricas MAPE también fueron elevadas (cercanas o superiores al 25%), indicando errores relativos considerables.

### ***Modelado con Prophet***

A continuación, se aplicó el modelo Prophet, primero sin ajustar festivos (**Figura 18**), y luego incorporando una lista de días no laborales como variable de entrada (**Figura 19**). Esto permitió ajustar mejor la estacionalidad semanal, ya que los días festivos suelen reducir la afluencia de usuarios.

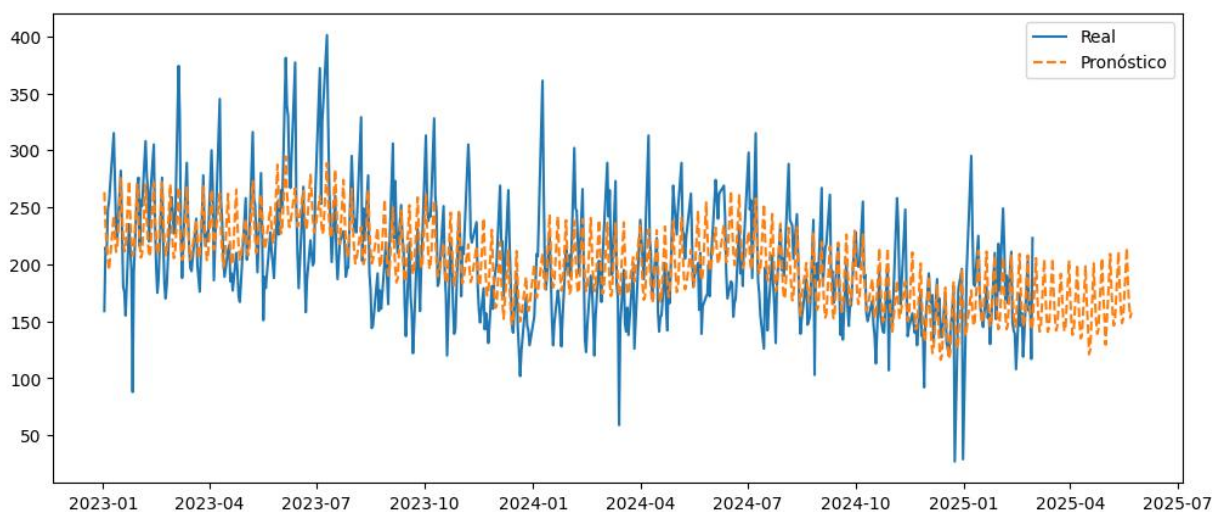
### **Figura 18**

*Pronóstico de Demanda con Prophet: Ajuste al Conjunto de Entrenamiento*



**Figura 19**

*Pronóstico de Demanda con Prophet: Teniendo en Cuenta los días Hábiles*



Los resultados se comparan en la **Tabla 3**:

**Tabla 3**

*Comparación del Desempeño del Modelo Prophet con y sin Festivos*

Modelo	RMSE	MAE	MAPE(%)	R <sup>2</sup> Score
Inicial	43.26	34.13	20.20	0.3319
Festivos	40.13	31.86	18.16	0.4240

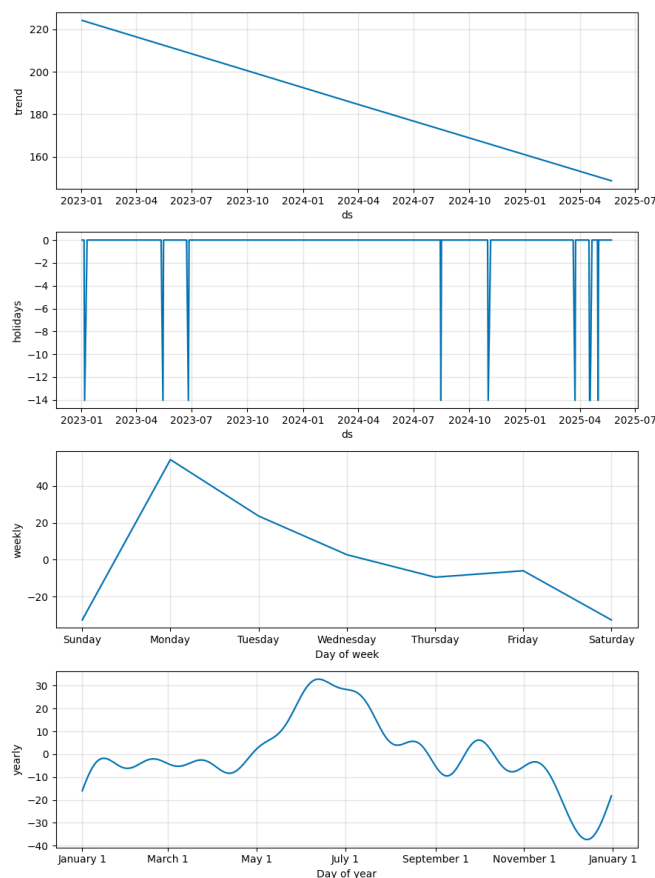
La incorporación de los festivos mejoró de forma clara el desempeño del modelo. Se logró reducir tanto el RMSE como el MAE, y el MAPE descendió por debajo del 20%, un umbral considerado muy bueno para predicciones relativas. El R<sup>2</sup> aumentó hasta 0.424, lo cual indica que Prophet es capaz de capturar una porción significativa de la variabilidad de la serie.

Finalmente, los componentes aprendidos por Prophet (**Figura 20**) permiten interpretar las principales influencias del modelo: una tendencia decreciente (trend), impacto de los festivos

(holidays), patrón semanal (weekly) donde los lunes presentan mayor demanda y sábados la menor, y una estacionalidad anual moderada con picos a mitad de año.

## Figura 20

*Componentes del Modelo Prophet: Tendencia, Festivos y Patrones Estacionales*



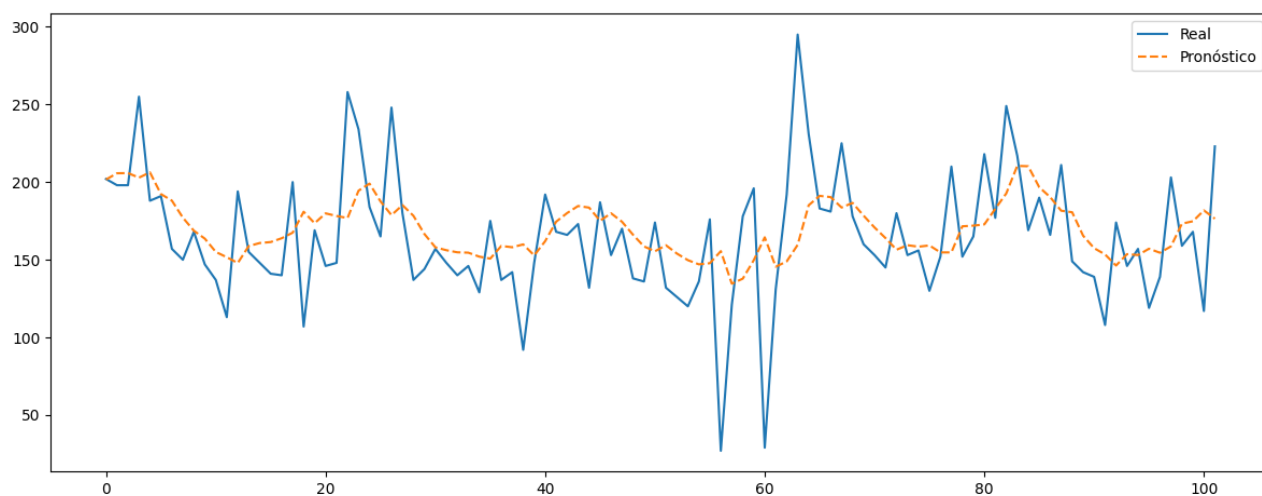
## *Modelado con Redes Neuronales Recurrentes (RNN)*

Con el fin de explorar métodos de aprendizaje profundo para la predicción de la demanda diaria, se implementaron modelos RNN utilizando distintas secuencias de entrada (ventanas temporales) para evaluar su impacto en la precisión de los resultados. Se probaron ciclos de 7, 15, 21 y 30 días. La siguiente **Tabla 4** resume los resultados obtenidos:

**Tabla 4***Desempeño de Modelos RNN con Diferentes Ciclos de Entrada*

Modelo	RMSE	MAE	MAPE (%)	R <sup>2</sup> Score
RNN 7 días	45.95	37.45	33.34	-0.2533
RNN 15 días	40.95	32.04	28.67	0.0044
RNN 21 días	37.60	27.60	24.62	0.1493
RNN 30 días	40.19	31.63	28.50	0.0321

La configuración de 21 días demostró ser la más eficaz, con el menor RMSE y MAE, así como el mejor R<sup>2</sup> de todos los modelos RNN evaluados. En la **Figura 21** se presenta la comparación entre los valores reales y las predicciones generadas por este modelo, donde puede observarse una mejor adaptación a la forma de la serie respecto a los otros experimentos.

**Figura 21***Pronóstico de Demanda con RNN: Comparación Real vs Predicción*

La gráfica ilustra cómo el modelo de RNN logra seguir parcialmente la tendencia de la demanda diaria, aunque presenta dificultades para capturar los picos más abruptos. Aun así, su comportamiento es claramente superior a los otros ciclos evaluados, haciendo de la configuración de 21 días la más favorable dentro de esta familia de modelos.

Con el propósito de mejorar los resultados obtenidos con RNN, se exploraron arquitecturas más complejas. Se aplicó una red LSTM de dos capas con función de activación tanh, Dropout para evitar sobreajuste y el optimizador RMSprop. Los resultados obtenidos fueron:

- RMSE: 41.07
- MAE: 32.82
- MAPE: 28.75%
- R<sup>2</sup> Score: -0.0107

Posteriormente, se implementó una red LSTM bidireccional con ventana de 15 y 21 días.

La variante con ventana de 15 días arrojó mejores resultados:

- RMSE: 39.55
- MAE: 29.26
- MAPE: 25.75%
- R<sup>2</sup> Score: 0.0715

A continuación se presenta la **Tabla 5** comparativa consolidada con todos los modelos de predicción de demanda diaria:

**Tabla 5***Comparación General de Modelos para la Demanda Diaria*

Modelo	RMSE	MAE	MAPE(%)	R <sup>2</sup> Score
ARIMA (5,1,2)	40.40	30.14	25.95	0.0218
Prophet (sin festivos)	43.26	34.13	23.19	0.3319
Prophet (con festivos)	40.13	31.86	18.16	0.4240
RNN 7 días	45.95	37.45	33.34	-0.2533
RNN 15 días	40.95	32.04	28.67	0.0044
RNN 21 días	37.60	27.60	24.62	0.1493
RNN 30 días	40.19	31.63	28.50	0.0321
Bidireccional 15 días	39.55	29.26	25.75	0.0715
Bidireccional 21 días	40.61	32.31	28.46	0.0077
ARIMA (2,1,2)	40.62	30.45	26.28	0.0110

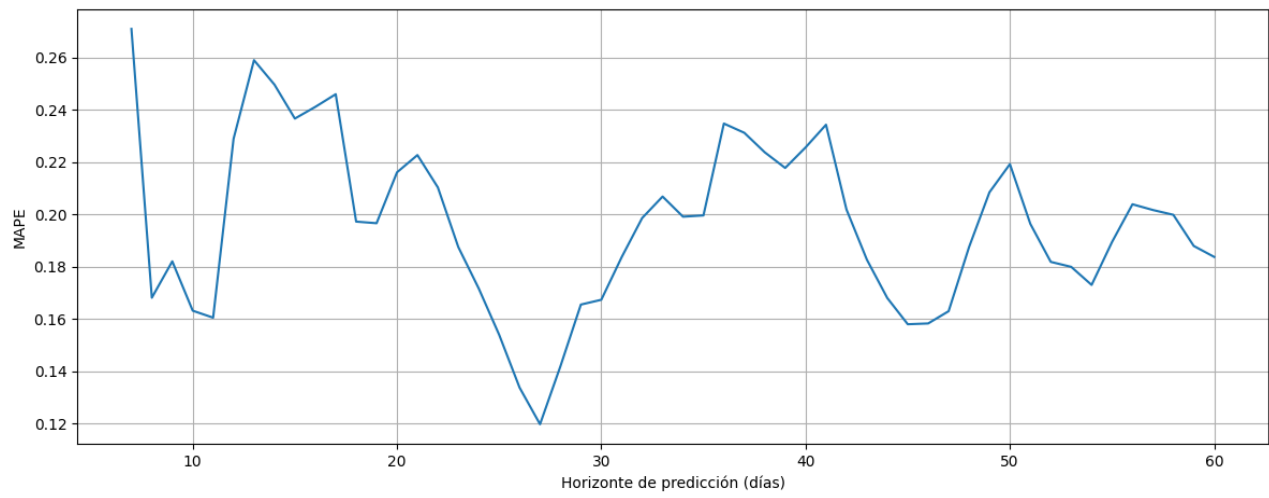
De todos los modelos aplicados, el modelo Prophet con ajuste por días festivos resultó ser el más preciso y robusto. No solo alcanzó el menor MAPE (18.16%), sino que logró un R<sup>2</sup> superior a 0.42, superando al resto de alternativas, incluidas las redes neuronales. Además, es el único modelo que mantuvo un equilibrio adecuado entre la capacidad explicativa y la estabilidad ante nuevas observaciones, sin necesidad de gran complejidad computacional.

**Validación Cruzada**

Tanto en el modelo de Prophet como en el de XGBoost se aplicaron técnicas de validación cruzada para medir el desempeño generalizado de los algoritmos. Para Prophet, se definieron ventanas de entrenamiento y predicción con periodos de 60 días de horizonte y 90 de intervalo. El resultado mostró que el MAPE (y las demás métricas) mínimo se obtuvo en el día 27 de horizonte (Ver **Figura 22**), respaldando su utilidad operativa. Para el modelo XGBoost, se utilizó validación cruzada con 5 particiones, obteniendo un RMSE promedio estable de 13.21 minutos. Estos métodos confirmaron la consistencia de los modelos elegidos.

**Figura 22**

*MAPE por Horizonte de Predicción con Modelo Prophet*



## Conclusiones

Este proyecto permitió desarrollar modelos predictivos para anticipar el tiempo de espera individual de los usuarios y la demanda diaria de atención en una oficina principal de una electrificadora. A partir de un conjunto de datos históricos con más de 100.000 registros, se aplicaron técnicas avanzadas de análisis exploratorio, transformación de variables y múltiples metodologías de machine learning y series temporales, siguiendo la metodología CRISP-DM.

En cuanto al tiempo de espera, se identificaron variables con alta correlación como el número de turnos esperando, el número de servicios solicitados y el grupo de atención. Estas fueron fundamentales en la construcción de los modelos. Entre los algoritmos evaluados Regresión Lineal, Random Forest y XGBoost, se destacó el modelo XGBoost por su equilibrio entre precisión y estabilidad, alcanzando un RMSE de 13.21 minutos en validación cruzada. Aunque el modelo optimizado de Random Forest presentó menor error (RMSE = 11.45), su curva de aprendizaje evidenció un sobreajuste considerable. En cambio, XGBoost mostró una curva más estable, lo que motivó su elección como el modelo más robusto para esta tarea.

Respecto a la predicción de la demanda diaria, se evaluaron distintos enfoques: modelos ARIMA, Prophet y redes neuronales recurrentes (RNN). El desempeño general reflejó la complejidad inherente de capturar los patrones de afluencia diaria, influenciados por factores exógenos como ciclos de facturación, eventos climáticos, festivos y posibles interrupciones del servicio. En este contexto, el modelo Prophet que incorporó la estacionalidad semanal y los días festivos logró el mejor desempeño global (RMSE = 40.13, MAPE = 18.16%,  $R^2 = 0.4240$ ), superando ampliamente a los demás.

Si bien el coeficiente de determinación  $R^2$  no alcanzó valores cercanos a 1, es importante resaltar que en modelos de series temporales donde el comportamiento de la variable objetivo

puede estar influenciado por ruidos, estacionalidades abruptas o factores no observados, métricas como RMSE y MAPE adquieren mayor relevancia práctica. Un MAPE por debajo del 20%, como el alcanzado por Prophet, se considera un resultado bastante aceptable en el ámbito de predicciones operativas.

Por su parte, los modelos RNN y LSTM mostraron progresos significativos al aumentar el tamaño de las secuencias (ventanas de tiempo), alcanzando sus mejores métricas con ventanas de 21 días (RMSE = 37.60, MAPE = 24.62%), aunque sin superar en consistencia al modelo Prophet. El modelo ARIMA, a pesar de sus fundamentos estadísticos, presentó limitaciones tanto en precisión como en capacidad de capturar estacionalidades relevantes.

Finalmente, la validación cruzada implementada en ambos frentes (regresión y series temporales) permitió evaluar la estabilidad de los modelos y su capacidad de generalización, reforzando la confiabilidad de los resultados obtenidos.

## Recomendaciones

Desde el punto de vista metodológico, se recomienda que futuras investigaciones sobre la predicción del tiempo de espera y la demanda de usuarios en entornos de atención al cliente consideren el uso de metodologías mixtas o híbridas. Aunque el presente estudio logró resultados satisfactorios empleando modelos como XGBoost y Prophet, existe potencial en combinar enfoques como redes neuronales recurrentes (RNN) con métodos estadísticos tradicionales. Asimismo, sería valioso explorar la integración de variables exógenas o contextuales (como el clima, fallas en el sistema o campañas institucionales) para fortalecer la capacidad explicativa de los modelos.

Por otra parte, en el ámbito académico, se recomienda fomentar el desarrollo de proyectos aplicados que integren problemáticas reales con herramientas de análisis predictivo, como se hizo en este estudio. La experiencia desarrollada demostró que modelos como Random Forest y XGBoost, al igual que técnicas de series temporales como Prophet y RNN, son capaces de predecir con alta precisión variables operativas como el tiempo de espera y la demanda diaria de usuarios. Este tipo de ejercicios permite a los estudiantes comprender la aplicación concreta de algoritmos de Machine Learning en contextos reales, especialmente en sectores como el de servicios públicos donde la optimización de recursos y la mejora de la atención son objetivos críticos. Por tanto, se propone utilizar este caso como insumo didáctico en cursos de modelado predictivo, análisis de series temporales y gestión de proyectos analíticos, incentivando una formación orientada a la solución de problemas del entorno.

Asimismo, desde el enfoque práctico, se recomienda a la electrificadora evaluar la posibilidad de implementar progresivamente el modelo XGBoost como herramienta de apoyo para la gestión operativa del flujo de usuarios en sala. Este modelo ha demostrado una alta

capacidad de predicción del tiempo de espera, lo cual podría traducirse en una mejora de la experiencia del usuario mediante la anticipación de congestiones y la redistribución estratégica de recursos. Así mismo, se sugiere incorporar el modelo Prophet con ajuste de días festivos como guía para prever la demanda diaria de atención, especialmente en periodos de facturación o cierre de mes. Estas recomendaciones prácticas pueden apoyar la planificación del talento humano, la apertura de taquillas y la preparación de infraestructura para atender con eficiencia.

Finalmente, se recomienda que tanto la institución como los desarrolladores del modelo mantengan procesos continuos de validación, retroalimentación y ajuste de los algoritmos, considerando que el comportamiento del usuario puede cambiar con el tiempo. En ese sentido, se sugiere la institucionalización de un sistema de monitoreo de predicciones y su comparación con los datos reales como parte del proceso de mejora continua en la gestión de atención al cliente.

### Referencias Bibliográficas

- Ayyalasomayajula, M. M. T., Bussa, S., & Ayyalasomayajula, S. (2021). Forecasting home prices employing machine learning algorithms: XGBoost, Random Forest, and Linear Regression. *ESP Journal of Engineering & Technology Advancements*, 1(1), 125–133.  
<https://doi.org/10.56472/25832646/JETA-V1I1P114>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2020). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 1937–1967.  
<https://doi.org/10.1007/s10462-020-09896-5>
- Caruelle, D., Lervik-Olsen, L., & Gustafsson, A. (2023). The clock is ticking—Or is it? Customer satisfaction response to waiting shorter vs. longer than expected during a service encounter. *Journal of Retailing*, 99(2), 247–264.  
<https://doi.org/10.1016/j.jretai.2023.03.003>
- Castro Pinzón, S. Y. (2023). Técnicas de Machine Learning aplicadas en modelos de mantenimiento predictivo en empresas industriales.  
<https://noesis.uis.edu.co/handle/20.500.14071/42369>
- Correa Loaiza, A. (2023). Análisis de modelos basados en Machine Learning para la predicción de la demanda de productos en la empresa Dyna & Cía. S.A.  
<https://hdl.handle.net/10495/35745>
- Fernández López, R., Díaz González, L. R., Alfonso Alemán, J. C., & Barrio Padrón, O. (2020). Modelo de predicción de series temporales para la demanda turística de la Cadena Hotelera Cubanacán. *COODES*, 8(3), 538–551.  
<http://coodes.upr.edu.cu/index.php/coodes/article/view/334>
- IBM. (s. f.). Random forest. IBM. <https://www.ibm.com/mx-es/topics/random-forest>

- Joseph, J., Senitha, S., Kirubaraja, A. A., & Ramson, J. S. R. (2023). Aprendizaje automático para la predicción de tiempos de espera en clínicas ambulatorias. *Procedia Computer Science*, 215, 230–239. <https://doi.org/10.1016/j.procs.2022.12.026>
- Konrad, K., & Liu, Y. (2023). Real-Time Estimations for the Waiting-Time Distribution in Time-Varying Queues. In *Proceedings of the 2023 Winter Simulation Conference*. <https://doi.org/10.1109/WSC60868.2023.10407204>
- Kyritsis, A. I., & Deriaz, M. (2019). A Machine Learning Approach to Waiting Time Prediction in Queueing Scenarios. *2019 Second International Conference on Artificial Intelligence for Industries (AI4I)*, 1–5. <https://doi.org/10.1109/AI4I46381.2019.00013>
- Loureiro, C., Pereira, P. J., Cortez, P., Guimarães, P., Moreira, C., & Pinho, A. (2019). Predicting multiple domain queue waiting time via machine learning. *International Journal of Machine Learning and Computing*, 9(3), 322–327. <https://doi.org/10.18178/ijmlc.2019.9.3.805>
- Mariño Villalba, J. A. (2023). Una comparación entre modelos estadísticos y de Machine Learning para la predicción de series de tiempo multivariadas. <https://repositorio.unal.edu.co/handle/unal/84522>
- Fierro Torres, C. Á., Castillo Pérez, V. H., & Torres Saucedo, C. I. (2022). Análisis comparativo de modelos tradicionales y modernos para pronóstico de la demanda: enfoques y características. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 12(24), e354. <https://doi.org/10.23913/ride.v12i24.1203>
- Menculini, L., Marini, A., Proietti, M., Garinei, A., Bozza, A., Moretti, C., & Marconi, M. (2021). Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices. <https://doi.org/10.48550/arXiv.2107.12770>

- Ayala Castrejón, R. F., & Bucio Pacheco, C. (2020). Modelo ARIMA aplicado al tipo de cambio peso-dólar en el periodo 2016–2017 mediante ventanas temporales deslizantes. *Revista Mexicana de Economía y Finanzas*, 15(3), 331–354.  
<https://doi.org/10.21919/remef.v15i3.466>
- Patel, H., Guttula, S., Gupta, N., Hans, S., Mittal, R. S., & Lokesh, N. (2023). A data-centric AI framework for automating exploratory data analysis and data quality tasks. *ACM Journal of Data and Information Quality*, 15(4), 44. <https://doi.org/10.1145/3603709>
- Sanit-in, Y., & Runapongsa Saikaew, K. (2019). Prediction of waiting time in one-stop service. *International Journal of Machine Learning and Computing*, 9(3), 322–327.  
<https://doi.org/10.18178/ijmlc.2019.9.3.805>
- Shin, J., Lee, D. A., Kim, J., Lim, C., & Choi, B. K. (2024). Dissatisfaction-considered waiting time prediction for outpatients with interpretable machine learning. *Health Care Management Science*, 27(3), 370–390. <https://doi.org/10.1007/s10729-024-09676-5>
- Tai, T. E., Haw, S. C., Ng, K. W., Naveen, P., & Al-Tarawneh, M. (2023). Performance evaluation on resolution time prediction using decision tree, random forest and extreme gradient boosting. 2023 International Conference on Computer Applications Technology (CCAT). <https://doi.org/10.1109/CCAT59108.2023.00021>
- Thiongane, M., Chan, W., & L'Ecuyer, P. (2022). Learning-based prediction of conditional wait time distributions in multiskill call centers. In *Operations Research and Enterprise Systems* (pp. 139–151). Springer. [https://doi.org/10.1007/978-3-031-10725-2\\_5](https://doi.org/10.1007/978-3-031-10725-2_5)
- Vassallo, M., Prakash, V., Garg, L., & Singh, P. (2024). Using historical trip information to determine the waiting time required for taxi services. In H. Sharma (Ed.), *Artificial*

Intelligence: Theory and Applications (pp. 139–151). Springer.

[https://doi.org/10.1007/978-981-99-8479-4\\_11](https://doi.org/10.1007/978-981-99-8479-4_11)

Xiao, H. (2023). Estimating wildfire duration using regression methods. arXiv preprint

arXiv:2308.08936. <https://doi.org/10.48550/arXiv.2308.08936>