

Maestría en

INTELIGENCIA ARTIFICIAL APLICADA

**Trabajo previo a la obtención de título de Magister en
Inteligencia Artificial Aplicada**

AUTOR/ES:

JORDAN ALEXIS AYALA LOPEZ

OSCAR DANIEL RIERA GUAMUSHIG

ANGEL ESTEBAN JUMBO VICENTE

TUTOR/ES:

JUAN FERNANDO LÓPEZ TITO

ANDRÉS ROBERTO NAVAS CASTELLANOS

TEMA:

Desarrollo de un modelo analítico basado en inteligencia artificial y análisis de series temporales para la identificación de patrones, factores asociados y escenarios de saturación hospitalaria en entidades hospitalarias del Distrito Metropolitano de Quito.

Certificación de autoría

Nosotros, **Jordan Alexis Ayala López, Ángel Esteban Jumbo Vicente y Oscar Daniel Riera Guamushig**, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.



Firma
Jordan Alexis Ayala López



Firma
Ángel Esteban Jumbo Vicente



Firma
Oscar Daniel Riera Guamushig

Autorización de Derechos de Propiedad Intelectual

Nosotros, **Jordan Alexis Ayala López, Ángel Esteban Jumbo Vicente y Oscar Daniel Riera Guamushig**, en calidad de autores del trabajo de investigación titulado ***Desarrollo de un modelo analítico basado en inteligencia artificial y análisis de series temporales para la identificación de patrones, factores asociados y escenarios de saturación hospitalaria en entidades hospitalarias del Distrito Metropolitano de Quito***, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, Agosto 2026



Validar únicamente en FirmatC.
Firmado electrónicamente por:
**JORDAN ALEXIS AYALA
LOPEZ**

Firma
Jordan Alexis Ayala López



**Angel Esteban
Jumbo Vicente**
Time Stamping
Security Data

Firma
Ángel Esteban Jumbo Vicente

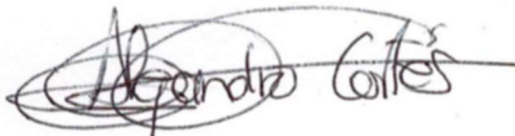


**Oscar Daniel Riera
Guamushig**
Time Stamping
Security Data

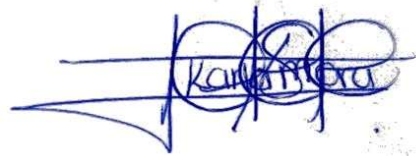
Firma
Oscar Daniel Riera Guamushig

Aprobación de dirección y coordinación del programa

Nosotros, **Alejandro Cortés López** Director EIG y **Karla Estefanía Mora Cajas** Coordinadora UIDE, declaramos que: **Jordán Alexis Ayala López, Ángel Esteban Jumbo Vicente y Oscar Daniel Riera Guamushig** son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés López Director de la
Maestría en Inteligencia Artificial
Aplicada



Karla Estefanía Mora Cajas
Coordinadora de la
Maestría en Inteligencia Artificial Aplicada

DEDICATORIA

Dedico este trabajo de titulación a mis padres, quienes han sido un pilar fundamental a lo largo de mi formación académica y personal. Su apoyo constante, sus enseñanzas y la confianza que han depositado en mí han sido esenciales para alcanzar esta meta.

A ellos, por acompañarme en cada etapa, por motivarme a continuar aun en los momentos de dificultad y por enseñarme, con su ejemplo, el valor del esfuerzo, la responsabilidad y la perseverancia.

Este logro también les pertenece, porque detrás de cada paso recorrido siempre estuvo su respaldo y el deseo de verme alcanzar mis objetivos.

Oscar Daniel Riera Guamushig

El presente trabajo de titulación lo dedico a mis padres, hermanos y sobrinos, quienes han sido base fundamental para lograr mis metas tanto en lo académico como en lo personal. Su apoyo constante, consejos y la confianza que han depositado en mí han inspirado para mejorar y conseguir nuevas metas.

A cada uno de ellos, por acompañarme en cada etapa, por motivarme a continuar aun en los momentos de dificultad y por enseñarme, con su ejemplo, el valor del esfuerzo, la responsabilidad y la perseverancia.

De tal modo que este logro también les pertenece, porque detrás de cada paso recorrido siempre estuvo su respaldo y el deseo de verme triunfar.

Angel Esteban Jumbo Vicente

Este proyecto va dedicado a Dios por ser quien me ha guiado, cuidado y por haberme brindado todas las capacidades necesarias para tener una buena vida. A mis hermanas Krisstel, Sabine y Salma por estar siempre ahí independientemente de la situación, demostrando que siempre tengo con quien contar, motivándonos y apoyándonos mutuamente y por último a Marcia, por ser literalmente la mejor persona que pude haber encontrado en mi vida.

Jordan Alexis Ayala López

AGRADECIMIENTOS

Quiero expresar mi agradecimiento a todos los docentes que me acompañaron durante el desarrollo de esta maestría. Cada uno de ellos aportó, desde su experiencia y conocimientos, enseñanzas que fueron importantes para mi formación y que podré aplicar en mi vida profesional.

Agradezco especialmente el tiempo, la orientación y la disposición que tuvieron para compartir sus conocimientos durante las diferentes asignaturas. A lo largo de este proceso no solo adquirí nuevos aprendizajes, sino que también pude fortalecer habilidades y ampliar mi forma de comprender los retos que se presentan en el ámbito profesional. También agradezco a quienes me orientaron durante la elaboración de este trabajo de titulación, ya que sus recomendaciones y observaciones fueron de gran ayuda para poder desarrollarlo y llegar hasta esta etapa.

Finalmente, me quedo con los conocimientos, las experiencias y los aprendizajes obtenidos durante estos años de estudio, y con la satisfacción de haber culminado una etapa muy importante de mi formación profesional.

Oscar Daniel Riera Guamushig

Agradezco a todos los docentes que me acompañaron durante la maestría, aportándome sus conocimientos que fueron importantes para mi formación académica y su experiencia que me ayudará en mi vida profesional. Agradezco por su paciencia y la disposición que tuvieron para compartir sus conocimientos durante las diferentes asignaturas y a lo largo de este proceso no solo adquirí nuevos aprendizajes, sino que también pude fortalecer habilidades y ampliar mi forma de comprender los retos que se presentan en el ámbito profesional.

Agradezco a mi Familia que siempre me han apoyado de manera incondicional directa o indirectamente en todo mi periodo de estudio y proceso de titulación

Finalmente agradezco a quienes me orientaron durante la elaboración de este trabajo de titulación, ya que sus recomendaciones y observaciones fueron de gran ayuda para poder desarrollarlo y llegar hasta esta etapa.

Angel Esteban Jumbo Vicente

Agradezco a mis padres y abuelos por apoyarme siempre no solo en esta etapa de mi vida, siendo mi soporte en todo ámbito y decisión tomada, sabiéndome brindar todo su esfuerzo en pro de mi bienestar.

A Adrián Criollo y Marcia Chancusig por ser mis grandes y mejores amigos, compañeros e incluso familia, estando conmigo desde el inicio, apoyándome y ayudándome en todo tipo de situación demostrando ser las mejores personas con las que uno se podría encontrar.

Finalmente a todos mis maestros que, a pesar de no poder nombrarlos a todos, quiero que sepan que les agradezco con todo mi ser haberme enseñado todo lo que me ha llevado a estar en este lugar, haberme formado como un profesional responsable y a entender que todo lo bueno conlleva un esfuerzo.

Jordan Alexis Ayala López

RESUMEN

La gestión de la demanda y de la capacidad hospitalaria requiere herramientas que permitan anticipar variaciones en el flujo de pacientes e identificar periodos de elevada presión sobre los recursos disponibles. En este contexto, la presente investigación tuvo como objetivo desarrollar un modelo analítico basado en inteligencia artificial y análisis de series temporales para predecir la demanda hospitalaria, identificar patrones y factores asociados, y caracterizar la ocupación y la presión hospitalaria relativa a la capacidad en los establecimientos de salud del Distrito Metropolitano de Quito. Para ello, se utilizaron los Registros Estadísticos de Camas y Egresos Hospitalarios del Instituto Nacional de Estadística y Censos correspondientes al periodo 2017–2024. La demanda hospitalaria se representó mediante los ingresos diarios y se evaluaron modelos Naive, Seasonal Naive, LSTM y Prophet mediante validación temporal fuera de muestra. Adicionalmente, se analizó la ocupación hospitalaria y se reconstruyó un proxy diario de presión relativa a la capacidad para el periodo 2017–2023. Los resultados mostraron que Seasonal Naive $t-7$ presentó el menor MAE promedio durante la validación 2020–2023, con 50,07 ingresos diarios, mientras que LSTM-28 obtuvo el mejor desempeño observado en el conjunto independiente de 2024, con un MAE de 39,77. La ocupación agregada del Distrito Metropolitano de Quito se mantuvo por debajo del 85 % durante todo el periodo; sin embargo, el análisis sectorial permitió identificar episodios de alta presión concentrados principalmente en el sector público. Para la predicción operacional del proxy diario, XGBoost alcanzó un MAE de 1,5260 puntos porcentuales y redujo el error en 42,81 % frente al baseline de persistencia. El análisis de interpretabilidad evidenció que la presión reciente constituye el principal componente predictivo, mientras que el análisis exploratorio de factores asociados mostró una jerarquía consistente de sector, demanda reciente, calendario y composición demográfica. En conjunto, los resultados demuestran la utilidad de integrar predicción de demanda, capacidad hospitalaria e interpretabilidad para generar evidencia complementaria orientada a la planificación y gestión de los recursos hospitalarios.

Palabras Claves: demanda hospitalaria, ocupación hospitalaria, inteligencia artificial, series temporales, aprendizaje automático, presión hospitalaria.

ABSTRACT

Hospital demand and capacity management requires tools capable of anticipating variations in patient flow and identifying periods of high pressure on available resources. In this context, this study aimed to develop an analytical model based on artificial intelligence and time-series analysis to forecast hospital demand, identify associated patterns and factors, and characterize occupancy and capacity-related hospital pressure in healthcare facilities located in the Metropolitan District of Quito. Statistical Records of Hospital Beds and Hospital Discharges from Ecuador's National Institute of Statistics and Censuses for the 2017–2024 period were used. Hospital demand was represented by daily admissions, and Naive, Seasonal Naive, LSTM, and Prophet models were evaluated through out-of-sample temporal validation. In addition, hospital occupancy was analyzed, and a daily proxy of capacity-related hospital pressure was reconstructed for the 2017–2023 period. The results showed that Seasonal Naive $t-7$ achieved the lowest average MAE during the 2020–2023 validation period, with 50.07 daily admissions, whereas LSTM-28 achieved the best observed performance on the independent 2024 test set, with an MAE of 39.77. Aggregate hospital occupancy in the Metropolitan District of Quito remained below 85% throughout the study period; however, the sector-level analysis identified high-pressure episodes concentrated mainly in the public sector. For operational prediction of the daily pressure proxy, XGBoost achieved an MAE of 1.5260 percentage points and reduced the error by 42.81% compared with the persistence baseline. The interpretability analysis showed that recent hospital pressure was the main predictive component, while the exploratory analysis of associated factors revealed a consistent hierarchy comprising hospital sector, recent demand, calendar-related variables, and demographic composition. Overall, the findings demonstrate the usefulness of integrating demand forecasting, hospital capacity analysis, and model interpretability to provide complementary evidence for hospital resource planning and management.

Keywords: hospital demand, hospital occupancy, artificial intelligence, time series, machine learning, hospital pressure.

TABLA DE CONTENIDOS

CAPITULO 1:	1
INTRODUCCIÓN.....	1
Definición del proyecto	1
1.1. Justificación e importancia del trabajo de investigación	2
1.2. Alcance	3
1.3. Objetivos.....	4
CAPITULO 2:	6
2. REVISIÓN DE LITERATURA	6
2.1. Estado del Arte	6
2.2. Dataset o fuente de datos.	12
2.3. Marco Teórico.	14
CAPITULO 3:	20
3. DESARROLLO.....	20
3.1. Diseño Metodológico de la investigación	20
3.2. Roles	21
3.3. Product Backlog	21
3.4. Preparación e integración de los datos	22
3.4.1. Integración de los Registros de Egresos Hospitalarios.....	23
3.4.2. Integración de los Registros de Camas Hospitalarias.....	24
3.4.3. Homologación, depuración y control de calidad	25
3.4.4. Construcción de variables derivadas	26
3.5. Construcción y análisis de la demanda hospitalaria	27
3.5.1. Definición de la demanda hospitalaria	27
3.5.2. Construcción de la serie temporal diaria	28
3.5.3. Análisis exploratorio del comportamiento temporal de la demanda	29
3.5.4. Diferenciación entre demanda y ocupación hospitalaria.....	30
3.6. Desarrollo de modelos predictivos de demanda.....	31
3.6.1. Diseño de la validación temporal	31
3.6.2. Modelos de referencia	32
3.6.3. Modelo Long Short-Term Memory	33
3.6.4. Modelo Prophet	34
3.6.5. Métricas de evaluación	35

3.6.6. Criterio de selección y evaluación independiente	36
3.7. Desarrollo del análisis de ocupación hospitalaria.....	36
3.7.1. Definición y cálculo de la ocupación hospitalaria.....	37
3.7.2. Variables de capacidad hospitalaria	38
3.7.3. Caracterización de los niveles de ocupación hospitalaria	39
3.7.4. Análisis agregado y por sector hospitalario.....	40
3.7.5. Alcance del indicador anual de ocupación	40
3.8.1. Reconstrucción de hospitalizaciones activas por día.....	41
3.8.2. Estimación de la capacidad media diaria disponible	42
3.8.3. Construcción del proxy diario de presión hospitalaria	42
3.8.4. Validación frente a la ocupación anual oficial	43
3.8.5. Reconstrucción por sector hospitalario.....	44
3.8.6. Identificación de episodios de alta presión hospitalaria	44
3.8.7. Delimitación temporal	45
3.9. Desarrollo del modelo de factores asociados	45
3.9.1. Construcción de la base analítica sector-día.....	46
3.9.2. Selección y construcción de variables predictoras	46
3.9.3. Modelo A: enfoque operacional y predictivo	47
3.9.4. Modelo B: análisis exploratorio de factores asociados	48
3.9.5. Diseño de validación temporal	48
3.10. Interpretabilidad de los modelos.....	49
3.10.1.	SHAP
49	
3.10.2.	Permutation Importance
50	
3.10.3.	Importancia agrupada de variables
50	
3.10.4.	Análisis de sensibilidad
51	
CAPITULO 4:	53
4. ANÁLISIS DE RESULTADOS.....	53
4.1. Caracterización de los registros hospitalarios	53
4.1.1. Características demográficas de los pacientes.....	53
4.1.2. Caracterización según sector hospitalario y parroquia	55
4.1.3. Duración de la estancia hospitalaria	57

4.1.4. Síntesis de la caracterización hospitalaria	58
4.2. Evolución de la demanda hospitalaria 2017–2024	59
4.2.1. Evolución temporal de la demanda hospitalaria.....	59
4.2.2. Caracterización de la serie diaria de demanda.....	61
4.3. Evaluación de los modelos predictivos	63
4.3.1. Evaluación de los modelos predictivos	63
4.3.2. Resultados del modelo LSTM	64
4.3.3. Resultados del modelo Prophet	65
4.3.4. Comparación de resultados por periodo de validación.....	65
4.3.5. Desempeño promedio y selección del modelo	67
4.3.6. Análisis de sensibilidad 2021–2023	68
4.3.7. Síntesis de la evaluación y selección del modelo	69
4.4. Evaluación independiente en 2024	69
4.4.1. Resultados del test independiente.....	70
4.4.2. Comportamiento del error al cierre de 2024.....	71
4.5. Capacidad y ocupación hospitalaria 2017–2024	72
4.5.1. Evolución de la capacidad y ocupación hospitalaria	73
4.5.2. Diferencias entre establecimientos y sectores	74
4.6. Presión hospitalaria diaria y episodios de alta ocupación	76
4.6.1. Comportamiento diario de la presión hospitalaria.....	76
4.6.2. Episodios de alta ocupación	77
4.7. Resultados del modelo operacional de presión — Modelo A	78
4.8. Factores asociados — Modelo B	79
4.9. Interpretabilidad y robustez	81
4.9.1. Interpretabilidad del Modelo A	81
4.9.2. Interpretabilidad del Modelo B.....	83
4.9.3. Análisis de robustez.....	85
4.10. Contraste de hipótesis	85
4.10.1.Respuesta a la pregunta de investigación	86
4.10.2.Contraste de hipótesis	87
CAPITULO 5:	88
5. CONCLUSIONES Y RECOMENDACIONES	88
5.1. Conclusiones.....	88

5.2. Recomendaciones	90
Bibliografía.....	92
Apéndices	95

LISTA DE TABLAS

Tabla 1	23
Tabla 2	24
Tabla 3	32
Tabla 4	39
Tabla 5	39
Tabla 6	43
Tabla 7	44
Tabla 8	48
Tabla 9	51
Tabla 10	54
Tabla 11	57
Tabla 12	59
Tabla 13	62
Tabla 14	65
Tabla 15	67
Tabla 16	68
Tabla 17	70
Tabla 18	71
Tabla 19	73
Tabla 20	78
Tabla 21	79
Tabla 22	80
Tabla 23	83
Tabla 24	85

LISTA DE FIGURAS

Figura 1	54
Figura 2	55
Figura 3	58
Figura 4	60
Figura 5	66
Figura 6	67
Figura 7	71
Figura 8	72
Figura 9	74
Figura 10	75
Figura 11	76
Figura 12	82
Figura 13	84

CAPITULO 1:

INTRODUCCIÓN

Definición del proyecto

La capacidad y ocupación hospitalaria representan elementos fundamentales para la adecuada gestión de los servicios de salud, debido a que influyen directamente en la respuesta del sistema sanitario, la calidad de atención y la disponibilidad de recursos. En este contexto, la saturación hospitalaria ocurre cuando la demanda de atención supera temporalmente la capacidad disponible de los establecimientos de salud, generando dificultades relacionadas con la disponibilidad de camas, el flujo de pacientes, la duración de las estancias hospitalarias y la organización de los recursos institucionales.

La congestión de los servicios hospitalarios constituye un desafío importante para los sistemas sanitarios, ya que puede provocar retrasos en la atención, aumento del tiempo de permanencia de los pacientes, mayor carga laboral para el personal sanitario y posibles afectaciones en la calidad y seguridad de los servicios (Pearce et al., 2023; Youssef et al., 2024). En el Distrito Metropolitano de Quito, esta situación adquiere relevancia debido a la alta demanda de atención médica, la concentración poblacional y la llegada de pacientes provenientes de diferentes provincias del país, factores que pueden generar mayor presión sobre la infraestructura y los recursos hospitalarios.

En Ecuador, los registros de camas y egresos hospitalarios constituyen fuentes oficiales que permiten analizar el comportamiento de la atención sanitaria, la utilización de los servicios y las principales características de la actividad hospitalaria. Durante el año 2024 se registraron 1.132.667 egresos hospitalarios y 23.726 camas disponibles en 626 establecimientos de salud a nivel nacional, lo que evidencia la magnitud de la demanda atendida por el sistema sanitario ecuatoriano (Instituto Nacional de Estadística y Censos [INEC], 2025). En este sentido, el análisis de la información correspondiente al Distrito Metropolitano de Quito permite identificar patrones de comportamiento, tendencias y periodos con mayor presión asistencial.

Actualmente, diversos estudios han aplicado modelos estadísticos e inteligencia artificial para predecir la ocupación hospitalaria. Sin embargo, gran parte de estas investigaciones se enfocan principalmente en la generación de pronósticos, dejando en segundo plano la interpretación de los resultados y el análisis de los factores que pueden

influir en la aparición de escenarios críticos. Esta limitación puede dificultar el uso de estos modelos como herramientas de apoyo para la planificación preventiva y la toma de decisiones.

Ante esta problemática, el presente proyecto plantea el desarrollo de un modelo analítico basado en inteligencia artificial y análisis de series temporales, orientado no solo a predecir el comportamiento de la demanda hospitalaria, sino también a identificar patrones temporales, tendencias y condiciones asociadas a posibles escenarios de saturación en las entidades de salud del Distrito Metropolitano de Quito.

Mediante el análisis de datos históricos correspondientes al periodo 2017–2024, se busca generar información que permita comprender mejor el comportamiento de la demanda hospitalaria y aportar elementos para fortalecer la planificación y gestión de los recursos sanitarios. Además, el uso de técnicas de análisis y visualización de datos permitirá presentar los resultados de manera comprensible y proporcionar evidencia local sobre los periodos de mayor presión asistencial.

1.1. Justificación e importancia del trabajo de investigación

La saturación hospitalaria representa un desafío importante para los sistemas de salud, debido al impacto que genera sobre la capacidad operativa de los establecimientos sanitarios. Cuando la demanda de atención supera la capacidad disponible, pueden presentarse dificultades en la asignación de camas, retrasos en los procesos de atención, aumento de la carga laboral del personal sanitario y limitaciones para responder de manera adecuada ante periodos de alta demanda.

En ciudades con una alta concentración poblacional, como el Distrito Metropolitano de Quito, resulta fundamental contar con herramientas que permitan analizar el comportamiento histórico de la demanda hospitalaria e identificar patrones que contribuyan a una mejor planificación de los recursos disponibles. La capacidad de anticipar posibles escenarios de presión asistencial puede servir como apoyo en la gestión de camas, organización del personal y distribución eficiente de los recursos institucionales.

Los avances en inteligencia artificial, aprendizaje automático y análisis de series temporales han permitido desarrollar modelos capaces de generar estimaciones sobre la demanda hospitalaria y la ocupación de camas (Seo et al., 2024; Tello et al., 2022). Sin embargo, la utilidad de estas herramientas no depende únicamente de su precisión predictiva,

sino también de la posibilidad de interpretar sus resultados y evaluar su aplicabilidad dentro del contexto específico donde serán implementadas.

Por ello, esta investigación propone un enfoque analítico que integra inteligencia artificial, análisis de series temporales y visualización de datos para estudiar el comportamiento histórico de la demanda hospitalaria en el Distrito Metropolitano de Quito durante el periodo 2017–2024. Su relevancia se centra en generar evidencia local que permita identificar patrones, tendencias y periodos asociados con mayores niveles de presión asistencial.

Además, el proyecto busca complementar los modelos predictivos tradicionales mediante una perspectiva orientada al análisis e interpretación de los resultados. La identificación de comportamientos recurrentes y condiciones asociadas a la saturación hospitalaria puede contribuir al desarrollo de estrategias de planificación más preventivas y eficientes. Los resultados obtenidos no buscan reemplazar los registros directos de ocupación de camas ni los criterios clínicos y administrativos de las instituciones de salud, sino proporcionar una herramienta adicional para apoyar la toma de decisiones basada en información.

Finalmente, la investigación presenta viabilidad debido a la disponibilidad de registros históricos de egresos y camas hospitalarias proporcionados por el Instituto Nacional de Estadística y Censos (INEC), así como al acceso a herramientas tecnológicas y librerías especializadas para el procesamiento de datos, análisis temporal y desarrollo de modelos basados en inteligencia artificial.

1.2. Alcance

Incluye

- Análisis exploratorio de los datos históricos de camas y egresos hospitalarios correspondientes al periodo 2017–2024.
- Identificación de patrones temporales, tendencias y estacionalidades relacionadas con la demanda hospitalaria.
- Análisis de escenarios de alta ocupación y presión hospitalaria relativa a la capacidad mediante técnicas de inteligencia artificial y análisis de series temporales.
- Implementación y comparación de modelos predictivos y analíticos aplicados a datos

hospitalarios.

- Identificación de factores asociados al comportamiento de la presión hospitalaria relativa a la capacidad.
- Segmentación de escenarios hospitalarios de acuerdo con niveles de ocupación y presión relativa a la capacidad disponible.
- Generación de resultados orientados al apoyo en la toma de decisiones y planificación hospitalaria.
- Aplicación del estudio en las entidades hospitalarias del Distrito Metropolitano de Quito.

No incluye

- Predicción o análisis clínico individual de pacientes, con el propósito de preservar la confidencialidad de la información individual y observar la protección especial aplicable a los datos relativos a la salud (Ley de Estadística, 1976, art. 21; Ley Orgánica de Protección de Datos Personales, 2021).
- Uso de información personal sensible o identificación individual de pacientes.
- Implementación de sistemas hospitalarios en tiempo real.
- Determinación directa de una condición operativa de saturación hospitalaria, debido a que esta requiere información adicional sobre recursos y condiciones de funcionamiento de los establecimientos.
- Evaluación financiera o presupuestaria de las entidades hospitalarias.
- Diferenciación específica entre zonas urbanas y rurales.

1.3. Objetivos

Objetivo general

Desarrollar un modelo analítico basado en inteligencia artificial y análisis de series temporales para predecir la demanda hospitalaria, identificar patrones y factores asociados, y caracterizar la ocupación y la presión hospitalaria relativa a la capacidad en los establecimientos de salud del Distrito Metropolitano de Quito, utilizando datos históricos de camas y egresos hospitalarios correspondientes al periodo 2017–2024.

Hipótesis

Hipótesis nula (H_0):

Los modelos de inteligencia artificial y series temporales evaluados no presentan una reducción del error de predicción de la demanda hospitalaria diaria frente a un modelo de referencia estacional, considerando su desempeño mediante validación temporal fuera de muestra.

Hipótesis alternativa (H_1):

Al menos uno de los modelos de inteligencia artificial o series temporales evaluados presenta una reducción del error de predicción de la demanda hospitalaria diaria frente a un modelo de referencia estacional, considerando su desempeño mediante validación temporal fuera de muestra.

Pregunta de investigación

¿En qué medida los modelos de series temporales e inteligencia artificial permiten predecir la demanda hospitalaria y qué patrones y factores se encuentran asociados al comportamiento de la ocupación y de la presión hospitalaria relativa a la capacidad en los establecimientos de salud del Distrito Metropolitano de Quito durante el periodo 2017–2024?

1.3.1. Objetivos específicos.

- Realizar un análisis exploratorio de los datos históricos de camas y egresos hospitalarios correspondientes al periodo 2017–2024 en los establecimientos de salud del Distrito Metropolitano de Quito.
- Identificar patrones temporales, tendencias y comportamientos estacionales asociados a la demanda hospitalaria.
- Implementar y comparar modelos de inteligencia artificial y análisis de series temporales para la predicción de la demanda hospitalaria diaria.
- Caracterizar la ocupación y la presión hospitalaria relativa a la capacidad, identificando periodos de alta ocupación a partir de la información disponible de capacidad hospitalaria.
- Identificar factores y variables asociados al comportamiento de la presión hospitalaria relativa a la capacidad mediante técnicas de análisis e interpretabilidad de datos.
- Generar resultados que constituyan un insumo complementario para la planificación y gestión de recursos hospitalarios.

CAPITULO 2:

2. REVISIÓN DE LITERATURA

2.1.Estado del Arte

En los últimos años, el tema de la ocupación hospitalaria ha tomado bastante importancia dentro del área de la salud pública. Esto se hizo aún más evidente a raíz de la pandemia de COVID-19, donde muchos hospitales colapsaron por la alta demanda de pacientes. A partir de este tipo de situaciones, surgió la necesidad de desarrollar herramientas que permitan anticipar estos escenarios y así mejorar la planificación y el uso de recursos en los hospitales.

Uno de los enfoques más utilizados para este tipo de problemas es el análisis de series temporales, especialmente con modelos como SARIMA. Según Rob J. Hyndman y Athanasopoulos (2018), estos modelos permiten trabajar con datos históricos para identificar patrones como tendencias o comportamientos estacionales, lo cual resulta bastante útil cuando se quiere hacer predicciones. Por esta razón, han sido ampliamente utilizados en diferentes áreas, incluyendo el sector salud.

En algunos estudios se ha aplicado este tipo de modelos para analizar la demanda hospitalaria. Por ejemplo, Rodríguez et al. (2021) utilizaron SARIMA en hospitales públicos de España y lograron identificar ciertos patrones que se repetían en determinadas épocas del año, especialmente relacionados con enfermedades respiratorias. Aunque los resultados fueron buenos, también se observó que la precisión del modelo depende mucho de la calidad de los datos disponibles.

Sin embargo, no todo se puede resolver con modelos estadísticos tradicionales. En muchos casos, los datos presentan comportamientos más complejos que no siempre pueden ser capturados de forma adecuada con este tipo de métodos. Por eso, en los últimos años se ha empezado a utilizar con más frecuencia la inteligencia artificial y el aprendizaje automático. De acuerdo con Ian Goodfellow et al. (2016), estos modelos permiten trabajar mejor con relaciones no lineales y patrones más difíciles de identificar, lo que puede mejorar la capacidad de predicción.

Dentro de estas técnicas, uno de los modelos más utilizados son las redes neuronales recurrentes, especialmente los modelos LSTM. En el estudio de Shahid et al. (2020), se compararon modelos tradicionales como SARIMA con modelos LSTM para predecir datos relacionados con salud, y se encontró que los modelos basados en inteligencia artificial tenían mejor rendimiento en escenarios donde los datos cambiaban constantemente. Esto demuestra que pueden adaptarse mejor a situaciones más dinámicas.

De igual manera, Rustam et al. (2020) analizaron distintos modelos de aprendizaje automático como Random Forest y Support Vector Machines para predecir la demanda hospitalaria durante la pandemia. Los resultados mostraron que no existe un único modelo perfecto, sino que en muchos casos es mejor combinar varios enfoques para obtener resultados más precisos. Esto ha dado paso a lo que se conoce como modelos híbridos.

Según Adadi y Berrada (2018), la inteligencia artificial explicable busca proporcionar mecanismos que permitan interpretar las decisiones generadas por modelos complejos, facilitando la identificación de variables relevantes y patrones críticos dentro de los sistemas analizados. Este enfoque resulta especialmente importante en el ámbito de la salud, donde la interpretación de los resultados constituye un elemento fundamental para apoyar procesos de toma de decisiones clínicas y administrativas.

En esta línea, diferentes investigaciones recientes han comenzado a incorporar técnicas de interpretabilidad en modelos hospitalarios. Mahmoudian et al. (2023) desarrollaron un enfoque basado en aprendizaje automático y aprendizaje profundo para analizar la planificación de capacidad hospitalaria en un departamento de cardiología. Los autores integraron modelos SARIMA, regresión lineal y redes LSTM para estimar demanda hospitalaria y duración de estancia, demostrando que la incorporación de variables explicativas y factores asociados permite generar estimaciones más realistas y útiles para la gestión hospitalaria.

Por otra parte, Seo et al. (2024) analizaron la predicción de tasas de ocupación hospitalaria mediante modelos LSTM y Bi-LSTM incorporando simultáneamente información dinámica y estática de las unidades hospitalarias. Los resultados demostraron que la integración de variables explicativas específicas mejoraba significativamente el

rendimiento y la capacidad interpretativa de los modelos. Además, el estudio permitió desarrollar herramientas de apoyo operativo para la gestión hospitalaria.

Asimismo, investigaciones recientes han incorporado técnicas de segmentación y análisis de patrones para identificar escenarios críticos de ocupación hospitalaria. Tello et al. desarrollaron un modelo híbrido basado en algoritmos k-means y Support Vector Regression (K-SVR) para analizar la demanda de camas hospitalarias. Los resultados evidenciaron bajos errores de predicción y permitieron identificar agrupamientos de comportamiento hospitalario asociados a diferentes niveles de demanda, demostrando la utilidad de las técnicas de segmentación para el análisis de capacidad hospitalaria.

En el contexto latinoamericano, algunos estudios realizados en países como Brasil, México y Chile han aplicado modelos predictivos y técnicas de inteligencia artificial para mejorar la planificación hospitalaria y optimizar el uso de recursos sanitarios. Sin embargo, gran parte de estos trabajos presentan limitaciones relacionadas con la disponibilidad de datos, la interpretación de resultados y la escasa incorporación de enfoques explicativos e interpretables dentro del análisis hospitalario.

En el contexto ecuatoriano y particularmente en el Distrito Metropolitano de Quito, la disponibilidad de registros administrativos oficiales permite abordar conjuntamente estas dimensiones. No obstante, el aporte de la presente investigación no consiste únicamente en aplicar un modelo predictivo, sino en integrar el análisis de la demanda hospitalaria, la información de capacidad y ocupación, la evaluación temporal de modelos predictivos y técnicas de inteligencia artificial explicable dentro de un mismo marco analítico.

Tello et al. (2022) desarrollaron una estrategia de aprendizaje automático para pronosticar la demanda de camas hospitalarias, con el objetivo de optimizar la planificación de recursos hospitalarios y la utilización de la capacidad instalada hospitalaria. Para ello, usaron cinco años de datos históricos de pacientes de un centro médico y construyeron un modelo que combina algoritmos k-means y regresión de máquinas de vectores de soporte (K-SVR). Los resultados mostraron bajos errores de predicción, con valores de MAPE que oscilaron entre el 0,49 % y el 4,10 %, lo cual demuestra que estas técnicas pueden generar

pronósticos estables y precisos. Este estudio es un precursor directo de la presente investigación, debido a que confirma la utilidad de los modelos predictivos para anticipar las tasas de ocupación o la demanda de camas y fortalecer la predictibilidad para la gestión hospitalaria.

Avinash et al. (2025) aplicaron modelos de aprendizaje automático para predecir las tasas de ocupación de camas en un hospital psiquiátrico en la India, demostrando que estas técnicas pueden capturar variaciones temporales y estacionales relevantes para la gestión hospitalaria. Este estudio tuvo como fuente de datos a la ocupación semanal de 2008 a 2024 y comparó seis modelos predictivos: Regresión de Vectores de Soporte (SVR), XGBoost, Bosque Aleatorio, K-Vecinos Más Cercanos, Impulso de Gradiente y Árboles de Decisión. Los resultados mostraron que el modelo de Bosque Aleatorio tuvo el mejor desempeño en los datos de prueba, con un error porcentual absoluto medio (MAPE) del 3,57 %, lo que demuestra una alta capacidad para predecir con precisión las tasas de ocupación de camas. Si bien la investigación se realizó en el campo de la salud mental, sus hallazgos son relevantes para respaldar la aplicación de modelos predictivos a la gestión de la ocupación de camas hospitalarias.

Además, Mahmoudian et al. (2023) abordaron la planificación de la capacidad hospitalaria mediante un enfoque predictivo basado en aprendizaje automático y aprendizaje profundo. Este estudio integró la clasificación de la duración de la estancia y la previsión de pacientes, para esto se usó modelos como SARIMA, regresión lineal y redes neuronales LSTM, esto fue realizado en base a un estudio de caso realizado en el departamento de cardiología de un hospital público, estos autores demostraron que la integración de variables como la duración de la estancia y el número de pacientes por paciente produce estimaciones más realistas de la capacidad requerida. Por otro lado, los resultados mostraron que la demanda futura del departamento podría aumentar significativamente, lo que subraya la importancia de los modelos predictivos para anticipar las necesidades de infraestructura y fortalecer la planificación hospitalaria en entornos complejos.

Siguiendo esta línea, Bravo Silva (2024) en su tesis sobre la predicción de la demanda de camas pediátricas relacionadas con infecciones respiratorias agudas analizó el problema de la ocupación estacional de camas durante el invierno y propuso tres enfoques de modelado:

un modelo epidemiológico, una red neuronal recurrente de tipo LSTM y un modelo híbrido que combina ambos. Los resultados mostraron que la integración del modelado epidemiológico y el aprendizaje profundo proporciona una alternativa eficaz para anticipar la ocupación máxima de camas, con un margen de error de predicción inferior a 12 días. Si bien la investigación se centró en un hospital pediátrico y específicamente en patologías respiratorias, sus hallazgos son relevantes para el uso de modelos predictivos en contextos de fuerte demanda estacional y alta presión sobre los servicios de salud.

Por último, Seo et al. (2024) analizaron la predicción de las tasas de ocupación hospitalaria a nivel departamento y habitación mediante el uso de modelos de series temporales basados en aprendizaje profundo, mismos que incorporaban simultáneamente información dinámica y estática. Este estudio empleó arquitecturas LSTM y Bi-LSTM para predecir las tasas de ocupación de camas por unidad hospitalaria y demostró que la inclusión de variables estáticas específicas de cada habitación mejoraba el rendimiento del modelo, especialmente para las predicciones a nivel de habitación. Se demostró que los modelos basados en Bi-LSTM lograron el mejor rendimiento, con indicadores favorables independientemente del nivel. Esto también permitió el desarrollo de un panel de control web para facilitar la planificación operativa del hospital. Si bien la investigación se centró en la ocupación de departamentos y habitaciones en un solo hospital, sus hallazgos son relevantes para respaldar el uso de modelos predictivos de series temporales para la optimización de recursos y la gestión proactiva de la ocupación hospitalaria.

En este contexto, el presente proyecto busca contribuir mediante el desarrollo de un modelo analítico basado en inteligencia artificial y análisis de series temporales orientado no solamente a la predicción de ocupación hospitalaria, sino también a la identificación de patrones temporales, tendencias, factores asociados y escenarios críticos de saturación hospitalaria en las entidades hospitalarias del Distrito Metropolitano de Quito.

A diferencia de investigaciones previas centradas exclusivamente en el pronóstico, la presente propuesta incorpora un enfoque analítico e interpretativo apoyado en técnicas de inteligencia artificial explicable y análisis temporal, con el objetivo de generar conocimiento útil para fortalecer la planificación hospitalaria y apoyar la toma de decisiones dentro del sistema de salud ecuatoriano.

2.2.1 Analisis Critico y vacios identificados.

La revisión de los estudios previos demuestra que los modelos basados en inteligencia artificial y análisis de series temporales tienen potencial para anticipar el comportamiento de la demanda y ocupación hospitalaria. Sin embargo, los resultados obtenidos en diferentes investigaciones no pueden compararse de manera directa, debido a las diferencias existentes en las poblaciones analizadas, la frecuencia de los datos, las variables consideradas, los periodos de predicción y las métricas utilizadas para evaluar el desempeño de los modelos.

Por ejemplo, Tello et al. (2022) emplearon información histórica de un único centro médico para realizar pronósticos semanales de demanda de camas, mientras que Avinash et al. (2025) analizaron la ocupación semanal en un hospital psiquiátrico. Aunque ambos estudios presentaron resultados favorables, su aplicabilidad depende de las características particulares de cada institución, servicio médico y población atendida.

De igual manera, Mahmoudian et al. (2023) enfocaron su investigación en el área de cardiología de un hospital público, mientras que Bravo Silva (2024) estudió la demanda pediátrica relacionada con infecciones respiratorias. Estos trabajos evidencian la importancia de incorporar variables clínicas y epidemiológicas específicas; sin embargo, su nivel de especialización puede limitar la aplicación de sus resultados en otros sistemas hospitalarios con diferentes características.

Respecto a las técnicas utilizadas, algunos estudios han reportado un mejor desempeño de modelos de aprendizaje automático y aprendizaje profundo frente a métodos estadísticos tradicionales. No obstante, esta ventaja depende de diversos factores, como la calidad y cantidad de datos disponibles, el horizonte de predicción, la configuración del modelo y la métrica utilizada. Por esta razón, la evaluación de un modelo no debe basarse únicamente en la reducción del error, sino también en aspectos como la estabilidad de las predicciones, su capacidad de generalización y la posibilidad de interpretar sus resultados.

Otro aspecto relevante identificado en la literatura corresponde al equilibrio entre precisión e interpretación. Modelos como LSTM y Bi-LSTM permiten capturar relaciones temporales complejas, aunque presentan una menor transparencia en comparación con métodos estadísticos tradicionales. Si bien la inteligencia artificial explicable busca mejorar esta limitación (Adadi & Berrada, 2018), su aplicación en estudios relacionados con demanda hospitalaria todavía es limitada. Seo et al. (2024), por ejemplo, lograron mejorar la capacidad predictiva mediante la combinación de información dinámica y estática; sin embargo, su

estudio fue desarrollado en un único hospital, lo que limita la generalización de sus resultados.

En términos generales, la mayoría de investigaciones revisadas han sido desarrolladas en contextos hospitalarios, especialidades médicas o sistemas sanitarios diferentes al ecuatoriano. Existe poca evidencia sobre la aplicación conjunta de modelos predictivos, análisis temporal, identificación de factores asociados y clasificación de niveles de presión asistencial en el Distrito Metropolitano de Quito.

Este vacío fundamenta la presente investigación, la cual utiliza datos históricos del periodo 2017–2024 para comparar los modelos LSTM y Prophet, identificar patrones temporales y diferencias entre sectores hospitalarios, y transformar los resultados obtenidos en indicadores de apoyo para la planificación sanitaria. Sin embargo, debido a que los datos utilizados corresponden a información agregada, los niveles establecidos deben interpretarse como indicadores relativos de presión asistencial y no como mediciones directas de ocupación o saturación física de camas.

A partir de la revisión realizada se identificaron cuatro principales vacíos de investigación: la limitada evidencia aplicada al contexto hospitalario ecuatoriano; la escasa comparación entre modelos estadísticos y de aprendizaje profundo utilizando una misma fuente de datos; la falta de integración entre predicción, análisis de factores asociados y clasificación de escenarios de presión hospitalaria; y la limitada incorporación de herramientas visuales que faciliten la interpretación de resultados para la gestión sanitaria.

Frente a estos desafíos, la presente investigación propone un enfoque analítico que integra la exploración de registros hospitalarios, la identificación de patrones temporales, la comparación entre modelos LSTM y Prophet, el análisis de factores asociados y la clasificación estadística de escenarios de presión asistencial. Además, incorpora un visualizador como herramienta de apoyo para facilitar la interpretación de los resultados, considerando que las predicciones no reemplazan los criterios clínicos, administrativos ni institucionales, sino que complementan la toma de decisiones basada en datos.

2.2. Dataset o fuente de datos.

Fuente de los datos y justificación de la elección:

Registros Estadísticos de Camas y Egresos Hospitalarios.

Los Registros Estadísticos de Camas y Egresos Hospitalarios presenta información de morbilidad hospitalaria, y la utilización de camas censables y camas no censables. La información se recolecta de los establecimientos de salud que prestan el servicio de internación hospitalaria, de la Red Pública Integral de Salud (RPIS) y Red Complementaria (RC).

Características del dataset.

Tipo de datos y variables disponibles.

Según la ficha metodológica de la operación estadística las principales variables de las camas y egresos hospitalarios son:

Camas Hospitalarias: Provincia de ubicación del establecimiento de salud, camas de dotación normal, camas disponibles, camas no censables, entre otros.

Egresos Hospitalarios: Provincia, cantón, parroquia, clase del establecimiento, tipo de establecimiento, entidad del establecimiento, sector del establecimiento, mes de registro, sexo paciente, edad paciente, condición de edad, etnia paciente, causa morbilidad 10ª revisión internacional detallada a 3 dígitos, entre otros.

Cabe recalcar que existen otras variables sin embargo se analizará la inclusión de las mismas en función de la necesidad.

Volumen de información

Egresos

- 1.143.765 (2017)
- 1.164.659 (2018)
- 1.195.311 (2019)
- 907.515 (2020)
- 1.038.235 (2021)
- 1.130.603 (2022)
- 1.170.813 (2023)

- 1.132.667 (2024)

Número de camas

- 23.452 (2017)
- 23.803 (2018)
- 24.054 (2019)
- 23.212 (2020)
- 23.196 (2021)
- 23.395 (2022)
- 23.645 (2023)
- 23.726 (2024)

Fuente de origen

Para el desarrollo de la investigación se utilizaron las bases de datos de Camas y Egresos Hospitalarios correspondientes a los años 2022, 2023 y 2024, publicadas por el Instituto Nacional de Estadística y Censos (INEC, s. f.).

2.3. Marco Teórico.

El presente estudio se sustenta en la relación entre gestión hospitalaria, análisis de series temporales, inteligencia artificial aplicada a la salud y análisis de la presión hospitalaria relativa a la capacidad disponible. Esta integración permite abordar la demanda hospitalaria como un fenómeno dinámico, relacionado con el comportamiento temporal de los ingresos hospitalarios, la disponibilidad de recursos, la duración de la estancia y las características de los establecimientos y de la población atendida.

La ocupación hospitalaria es un indicador utilizado para analizar el grado de utilización de la capacidad instalada de los establecimientos de salud. De manera general, relaciona el uso de las camas hospitalarias durante un periodo determinado con la capacidad disponible en ese mismo periodo. En este sentido, el porcentaje de ocupación puede expresarse como la relación entre los días-paciente y los días-cama disponibles. Sin embargo, la presión hospitalaria representa un concepto más amplio, ya que no depende únicamente de la disponibilidad de camas, sino también de factores como la demanda asistencial, el flujo de pacientes, la duración de la estancia, la disponibilidad del personal sanitario y la organización interna de los servicios. De acuerdo con la literatura, la congestión hospitalaria puede estar

influenciada por múltiples factores relacionados con los pacientes, los recursos humanos y el funcionamiento general del sistema de salud (Pearce et al., 2023).

$$\text{Ocupación Hospitalaria} = \frac{\text{días} - \text{paciente}}{\text{días} - \text{camas disponibles}} * 100$$

La ocupación hospitalaria debe diferenciarse de la demanda hospitalaria. Mientras la demanda representa el flujo de pacientes que requieren o reciben atención, la ocupación refleja el nivel de utilización de la capacidad disponible. Asimismo, una alta demanda o una alta ocupación no constituyen, por sí solas, evidencia suficiente de saturación hospitalaria, debido a que esta última corresponde a una condición operativa más compleja en la que también intervienen factores como la disponibilidad de personal sanitario, recursos físicos, organización de los servicios, insumos y capacidad de respuesta institucional.

En esta investigación, el concepto de presión hospitalaria relativa a la capacidad se utiliza para analizar situaciones en las que la utilización de los recursos hospitalarios alcanza niveles elevados respecto de la capacidad disponible. Este enfoque permite caracterizar escenarios de alta ocupación sin asumir automáticamente la existencia de saturación institucional. En consecuencia, el estudio no pretende realizar un diagnóstico clínico individual ni determinar una condición operativa de saturación en cada establecimiento, sino analizar patrones agregados de demanda, ocupación y presión relativa a la capacidad.

La presión hospitalaria constituye un aspecto importante dentro de la planificación sanitaria, ya que permite identificar periodos en los que los servicios de salud pueden presentar mayores exigencias operativas. Diversos estudios relacionan la congestión hospitalaria con retrasos en la atención, incremento de la estancia de los pacientes, mayor carga laboral del personal sanitario y dificultades en el flujo asistencial; sin embargo, sus efectos pueden variar según las características propias de cada establecimiento y del sistema de salud en el que se desarrolla (Pearce et al., 2023; Youssef et al., 2024).

En este contexto, el análisis de datos históricos representa una herramienta útil para comprender el comportamiento de la demanda hospitalaria, identificar periodos de mayor presión y aportar información que facilite la planificación de recursos. No obstante, sus resultados deben considerarse como elementos de apoyo para la toma de decisiones y no

como sustitutos de los criterios clínicos, administrativos y operativos establecidos por las instituciones de salud.

Los Registros Estadísticos de Camas y Egresos Hospitalarios del Instituto Nacional de Estadística y Censos (INEC) constituyen la principal fuente de información utilizada en esta investigación. Esta operación estadística recopila datos relacionados con la morbilidad hospitalaria, la disponibilidad y utilización de camas, así como las características de los egresos registrados en los establecimientos de salud que brindan servicios de internación dentro del sistema público y complementario (Instituto Nacional de Estadística y Censos [INEC], 2025).

Para el desarrollo del estudio se analizaron los registros correspondientes a los años 2022, 2023 y 2024 del Distrito Metropolitano de Quito. La base de datos incluyó variables asociadas a la ubicación del establecimiento, sector hospitalario, causas de hospitalización, especialidad médica, sexo del paciente, condición de egreso y días de estancia hospitalaria, permitiendo caracterizar el comportamiento histórico de la demanda asistencial.

El análisis de series temporales permite estudiar la evolución de una variable a lo largo del tiempo. En el ámbito hospitalario, este enfoque es útil porque la demanda de atención no se comporta de manera constante, sino que puede variar según meses, semanas, temporadas epidemiológicas, condiciones sociales o eventos extraordinarios.

Una serie temporal corresponde a un conjunto de observaciones registradas de manera consecutiva a lo largo del tiempo, cuya estructura puede presentar componentes como tendencia, estacionalidad, ciclos y variaciones aleatorias (Hyndman & Athanasopoulos, 2018). La tendencia refleja el comportamiento general de la serie, mientras que la estacionalidad hace referencia a patrones que se repiten en periodos determinados. Por su parte, las variaciones irregulares representan cambios no predecibles que no siguen una dinámica constante.

En el presente proyecto, el análisis de series temporales se utiliza para identificar patrones temporales en los registros hospitalarios, especialmente en el volumen de ingresos o egresos diarios y mensuales. Esta perspectiva permite reconocer periodos de mayor presión hospitalaria y construir una base cuantitativa para estimar escenarios de alerta o saturación.

De esta manera, los datos administrativos dejan de ser únicamente registros históricos y se transforman en insumos para la planificación y toma de decisiones.

La inteligencia artificial se refiere al desarrollo de sistemas capaces de analizar información, aprender a partir de los datos, reconocer patrones y generar respuestas o decisiones basadas en el conocimiento adquirido (Russell & Norvig, 2021). Dentro de este campo, el aprendizaje automático utiliza algoritmos que permiten identificar relaciones entre variables y mejorar su rendimiento mediante la experiencia obtenida de los datos, sin necesidad de establecer previamente todas las reglas de funcionamiento.

En el ámbito hospitalario, estas técnicas permiten analizar comportamientos complejos, identificar relaciones no lineales y generar estimaciones sobre la evolución de la demanda asistencial. De esta manera, pueden constituirse en herramientas de apoyo para la planificación de recursos y la toma de decisiones dentro de los servicios de salud (Tello et al., 2022).

Dentro de los modelos de aprendizaje profundo, las redes neuronales recurrentes son especialmente útiles para trabajar con datos secuenciales. Las redes LSTM, o Long Short-Term Memory, fueron diseñadas para conservar información relevante a través del tiempo y reducir problemas asociados a la pérdida de dependencia temporal en secuencias largas (Hochreiter & Schmidhuber, 1997). En el contexto de esta investigación, el uso de LSTM permite analizar la evolución diaria de la demanda hospitalaria a partir de una ventana temporal previa, con el objetivo de estimar comportamientos posteriores.

Además de los modelos de aprendizaje profundo, existen herramientas específicas para pronóstico de series temporales, como Prophet. Este modelo permite analizar datos con tendencia y estacionalidad, y resulta útil cuando se requiere una aproximación interpretable al comportamiento temporal de una variable (Taylor & Letham, 2018). La comparación entre LSTM y Prophet es pertinente porque ambos modelos abordan el comportamiento temporal desde enfoques distintos: LSTM se orienta al aprendizaje de patrones secuenciales complejos, mientras que Prophet representa explícitamente componentes como la tendencia y la estacionalidad, lo que facilita la interpretación de la estructura temporal modelada (Taylor & Letham, 2018).

No obstante, en investigaciones aplicadas a salud no basta con obtener predicciones numéricas. Los modelos deben ser comprensibles y útiles para apoyar decisiones. En este punto adquiere importancia la inteligencia artificial explicable, entendida como el conjunto de métodos orientados a facilitar la interpretación de los resultados producidos por modelos complejos. La interpretabilidad permite identificar qué variables o condiciones se asocian con escenarios críticos y evita que el modelo funcione como una “caja negra” sin utilidad práctica para la gestión hospitalaria (Adadi & Berrada, 2018).

La identificación de factores asociados representa un aspecto importante dentro de esta investigación, debido a que la presión hospitalaria no está determinada únicamente por el volumen de pacientes atendidos, sino también por elementos relacionados con el flujo asistencial, la disponibilidad de camas, la duración de la estancia y las características operativas de los servicios de salud (Pearce et al., 2023).

En este estudio se consideran variables como los días de estancia hospitalaria, el sector del establecimiento, las causas de hospitalización, la especialidad médica y la condición de egreso, con el propósito de analizar su comportamiento en relación con la demanda hospitalaria. Estas asociaciones se interpretan desde una perspectiva agregada y descriptiva, por lo que no representan relaciones causales ni predicciones clínicas individuales.

La segmentación de escenarios permite clasificar los periodos analizados según niveles de presión hospitalaria. Para ello, pueden emplearse umbrales estadísticos basados en percentiles, los cuales permiten diferenciar días de comportamiento normal, días de alerta y días de saturación crítica. En esta investigación, el uso de percentiles se plantea como una aproximación analítica para identificar valores superiores al comportamiento habitual de la serie histórica. Esta clasificación no reemplaza criterios administrativos o clínicos propios de cada institución, pero ofrece una referencia objetiva para interpretar el riesgo relativo de saturación.

En conjunto, el marco teórico de esta investigación articula cuatro dimensiones: la gestión de la presión hospitalaria, el análisis temporal de la demanda, el uso de inteligencia artificial para modelar patrones y la interpretación de factores asociados a escenarios críticos.

Esta articulación permite que el proyecto no se limite a predecir valores futuros, sino que también contribuya a comprender el comportamiento de la demanda hospitalaria en el Distrito Metropolitano de Quito.

Por tanto, el desarrollo de un modelo analítico basado en inteligencia artificial y análisis de series temporales resulta pertinente para identificar patrones, factores asociados y escenarios de saturación hospitalaria. La finalidad del modelo no es sustituir la decisión institucional, sino proporcionar evidencia cuantitativa que fortalezca la planificación, la gestión de recursos y la capacidad de respuesta de las entidades hospitalarias frente a periodos de alta demanda.

CAPITULO 3:

3. DESARROLLO

3.1. Diseño Metodológico de la investigación

La presente investigación se desarrolla bajo un enfoque cuantitativo, debido a que se basa en el análisis de datos numéricos provenientes de registros administrativos de salud y en la aplicación de métricas estadísticas y de aprendizaje automático, como MAE y RMSE, para evaluar el desempeño predictivo de los modelos implementados.

El estudio es de tipo aplicado y presenta un alcance predictivo, ya que busca estimar el comportamiento de la demanda hospitalaria a partir de patrones identificados en datos históricos. La investigación no contempla la manipulación de las variables analizadas, sino el estudio de su comportamiento mediante técnicas de procesamiento y modelamiento de datos.

El diseño de la investigación es no experimental, longitudinal y retrospectivo, debido a que utiliza registros históricos correspondientes a los años 2022, 2023 y 2024 del Distrito Metropolitano de Quito. Estos datos fueron analizados sin intervención directa del investigador sobre las unidades de estudio, permitiendo observar su comportamiento a través del tiempo.

La investigación adopta un enfoque cuantitativo, un alcance predictivo y un diseño no experimental, longitudinal y retrospectivo, de acuerdo con las características de los datos y los objetivos planteados (Hernández-Sampieri & Mendoza, 2018).

La población de estudio está conformada por la totalidad de los registros de ingresos y egresos hospitalarios reportados por las entidades de salud del Distrito Metropolitano de Quito durante el periodo 2017–2024. Al trabajar con el conjunto completo de registros administrativos proporcionados por el INEC, no fue necesario aplicar técnicas de muestreo probabilístico.

Técnicas e instrumentos para el desarrollo de la investigación se aplicaron técnicas de análisis documental sobre bases de datos secundarias, específicamente microdatos en formato .sav proporcionados por el INEC. El procesamiento de la información se realizó mediante herramientas computacionales basadas en Python, utilizando librerías para análisis de datos, modelamiento de series temporales y aprendizaje profundo. Posteriormente, se efectuó una evaluación comparativa de los modelos predictivos mediante métricas de error.

Es importante señalar que el diseño metodológico descrito corresponde a la estrategia científica utilizada para responder los objetivos de investigación, mientras que Scrum fue empleado únicamente como marco de gestión del desarrollo técnico del proyecto.

En este sentido, Scrum no representa un diseño de investigación, sino una metodología de organización del trabajo utilizada para estructurar de manera iterativa las actividades de procesamiento de datos, construcción de modelos, validación y desarrollo del visualizador. La aplicación de ciclos de trabajo o sprints permitió realizar ajustes progresivos y mejorar los componentes técnicos durante las diferentes etapas del proyecto. Scrum se utilizó como un marco de gestión iterativa para organizar las actividades técnicas del proyecto mediante ciclos de trabajo denominados sprints (Schwaber & Sutherland, 2020).

3.2. Roles

Product Owner (PO): Equipo investigador, debido a que las decisiones relacionadas con los objetivos, el alcance y la priorización de las actividades fueron tomadas conjuntamente por los autores de la investigación.

Scrum Master: Ángel Esteban Jumbo Vicente, responsable de coordinar los sprints, realizar el seguimiento de las actividades y facilitar la organización del equipo.

Equipo de desarrollo: Jordan Alexis Ayala López, Ángel Esteban Jumbo Vicente y Oscar Daniel Riera Guamushig, responsables de la preparación y análisis de los datos, la implementación y evaluación de los modelos, la generación de visualizaciones y el desarrollo del tablero interactivo.

Los roles de Scrum fueron adaptados al contexto académico del proyecto. Aunque se asignaron responsabilidades específicas de coordinación y priorización, los tres autores participaron activamente en las actividades técnicas y en la elaboración de la investigación.

3.3. Product Backlog

El Product Backlog reúne y prioriza las actividades necesarias para alcanzar los objetivos del proyecto (Schwaber & Sutherland, 2020). En esta investigación estuvo compuesto por:

- Preparación e integración de los registros hospitalarios del periodo 2017–2024.
- Construcción y análisis de la serie diaria de demanda hospitalaria.

- Identificación de tendencias, estacionalidades y patrones temporales.
- Implementación y evaluación de los modelos Naive, Seasonal Naive, LSTM y

Prophet.

- Análisis de la ocupación y reconstrucción de la presión hospitalaria.
- Identificación e interpretación de los factores asociados.
- Desarrollo del tablero interactivo y las visualizaciones.
- Validación y documentación de los resultados.

3.4. Preparación e integración de los datos

La preparación de los datos constituyó la primera etapa del desarrollo analítico. Para la investigación se utilizaron los Registros Estadísticos de Egresos Hospitalarios y Camas Hospitalarias correspondientes al periodo 2017–2024, obtenidos de las bases publicadas por el Instituto Nacional de Estadística y Censos (INEC). Para la investigación se utilizaron los Registros Estadísticos de Egresos Hospitalarios y Camas Hospitalarias correspondientes al periodo 2017–2024, publicados por el Instituto Nacional de Estadística y Censos (INEC, 2025).

Debido a que la información se encuentra disponible de forma independiente para cada año, fue necesario desarrollar un proceso de integración temporal que permitiera construir bases consolidadas y comparables para todo el periodo de estudio. Este procedimiento incluyó la selección geográfica de los establecimientos ubicados en el Distrito Metropolitano de Quito, la homologación de variables entre años, la revisión de tipos de datos, el procesamiento de fechas y la construcción de variables derivadas requeridas para los análisis posteriores.

Durante todo el proceso se preservaron las variables originales de las fuentes. Las transformaciones realizadas se incorporaron mediante nuevas variables o bases analíticas derivadas, evitando modificar los archivos fuente y manteniendo la trazabilidad de los procedimientos aplicados.

3.4.1. Integración de los Registros de Egresos Hospitalarios

Los archivos anuales del Registro Estadístico de Egresos Hospitalarios correspondientes a los años 2017 a 2024 fueron procesados individualmente y posteriormente integrados en una única base maestra.

El criterio geográfico utilizado correspondió a la ubicación del establecimiento de salud, seleccionándose aquellos registros pertenecientes al Distrito Metropolitano de Quito. Por tanto, la delimitación territorial de la investigación hace referencia al lugar donde se produjo la atención hospitalaria y no necesariamente al lugar de residencia habitual del paciente.

Una vez aplicado este criterio y consolidada la información de los ocho años, la base maestra quedó conformada por 1.814.236 registros de egresos hospitalarios, distribuidos de la siguiente manera:

Tabla 1

Distribución anual de los registros de egresos hospitalarios del Distrito Metropolitano de Quito, 2017–2024

Año	Registros
2017	226.735
2018	237.817
2019	241.611
2020	185.048
2021	210.098
2022	228.021
2023	244.332
2024	240.574
Total	1.814.236

Nota. Elaboración propia a partir de los Registros Estadísticos de Egresos Hospitalarios del INEC (2017–2024).

La integración permitió mantener información relevante relacionada con las características de los establecimientos y de los pacientes, entre ellas sector hospitalario, tipo de establecimiento, fecha de ingreso y egreso, edad, sexo, causa de morbilidad, condición de egreso y duración de la estancia.

Para los análisis posteriores se utilizaron subconjuntos específicos de esta base en función del objetivo metodológico. En particular, la construcción de la serie de demanda hospitalaria consideró los ingresos producidos dentro del periodo comprendido entre el 1 de enero de 2017 y el 31 de diciembre de 2024. Bajo este criterio se identificaron 1.811.587 ingresos hospitalarios.

La diferencia entre esta cifra y los 1.814.236 registros de la base maestra se debe principalmente a hospitalizaciones cuyo ingreso ocurrió antes del inicio de 2017, pero cuyo egreso fue registrado dentro del periodo analizado. Estos casos no fueron eliminados de la base general, debido a que contienen información relevante para la posterior reconstrucción de la ocupación hospitalaria al inicio del periodo de estudio.

3.4.2. Integración de los Registros de Camas Hospitalarias

De manera complementaria, se procesaron los archivos del Registro Estadístico de Camas Hospitalarias correspondientes al periodo 2017–2024. Se aplicó el mismo criterio geográfico utilizado en la base de egresos, conservando únicamente los registros correspondientes a establecimientos ubicados en el Distrito Metropolitano de Quito.

Como resultado, se obtuvo una base consolidada de 706 registros de establecimientos hospitalarios, distribuida anualmente de la siguiente manera:

Tabla 2

Distribución anual de los registros de camas hospitalarias del Distrito Metropolitano de Quito, 2017–2024

Año	Registros de establecimientos
2017	91
2018	89
2019	89
2020	88
2021	85
2022	86
2023	88
2024	90
Total	706

Fuente: Registro Estadístico de Camas Hospitalarias, INEC, 2017–2024.

Esta fuente permitió incorporar información relacionada con la capacidad hospitalaria, principalmente camas de dotación normal, camas disponibles, días-cama disponibles y días-paciente. Por otro lado, la incorporación del Registro de Camas fue fundamental para diferenciar el análisis de la demanda del análisis de ocupación. Mientras los registros de egresos permiten reconstruir el flujo de pacientes y las hospitalizaciones observadas, la información de camas proporciona una medida de la capacidad disponible contra la cual puede compararse la utilización hospitalaria.

3.4.3. Homologación, depuración y control de calidad

Una vez consolidadas las bases anuales, se realizó un proceso de revisión orientado a garantizar la consistencia temporal de las variables utilizadas. Debido a que determinados nombres, tipos de datos o formatos pueden presentar variaciones entre periodos, se homologaron las variables necesarias para mantener una estructura analítica común durante los ocho años de estudio.

Particular atención recibió el procesamiento de las fechas de ingreso y egreso, debido a que estas variables constituyen la base tanto para la construcción de la serie temporal de demanda como para la reconstrucción posterior de pacientes hospitalizados por día. Las fechas fueron estandarizadas a un formato temporal común y se verificó su consistencia cronológica.

A partir de estas variables se calculó la duración de la estancia hospitalaria. Para los pacientes cuyo ingreso y egreso ocurrieron el mismo día se consideró un día de estancia, debido a que existió utilización de recursos hospitalarios durante esa jornada. Asimismo, se realizaron verificaciones entre la estancia calculada a partir de las fechas y la información disponible en los registros originales.

También se revisaron los tipos de datos de variables categóricas y numéricas, evitando realizar conversiones que pudieran modificar el significado original de los códigos. Cuando fue necesario normalizar una variable para su análisis, se generó una versión derivada y se conservó paralelamente la variable proveniente de la fuente.

Los controles de calidad incluyeron la revisión de valores faltantes, rangos temporales, consistencia entre fechas, duplicados, categorías disponibles y valores extremos. Los registros considerados válidos fueron conservados aun cuando presentaran comportamientos poco frecuentes, siempre que no existiera evidencia suficiente para considerarlos errores de procesamiento. De esta forma, la depuración buscó corregir inconsistencias verificables sin eliminar observaciones únicamente por presentar valores atípicos.

3.4.4. Construcción de variables derivadas

Una vez integradas y verificadas las fuentes, se construyeron las variables requeridas para las diferentes etapas del análisis. Estas transformaciones se realizaron sobre copias analíticas de las bases consolidadas, preservando las variables originales.

La edad de los pacientes requirió un tratamiento específico debido a que el registro combina el valor de la edad con la unidad en la que esta se encuentra expresada. Para permitir comparaciones homogéneas, se generó una variable de edad equivalente en años, manteniendo los años como unidad directa y transformando, cuando correspondía, los meses, días y horas a su equivalente anual.

Las causas de morbilidad fueron conservadas de acuerdo con su codificación original basada en la Clasificación Internacional de Enfermedades, décima revisión (CIE-10). Para facilitar su interpretación se utilizó adicionalmente un catálogo externo de correspondencia entre los códigos y sus respectivas descripciones, sin reemplazar los códigos originales presentes en la fuente.

Asimismo, se construyeron variables temporales derivadas a partir de la fecha de ingreso, entre ellas año, mes, día, día de la semana y otras características requeridas posteriormente para el análisis de series temporales.

Para la reconstrucción de la utilización hospitalaria se generaron variables destinadas a determinar el intervalo comprendido entre la fecha de ingreso y la fecha de egreso de cada hospitalización. Estas variables permitieron posteriormente identificar qué pacientes permanecían hospitalizados en cada día del periodo analizado.

Finalmente, las bases consolidadas fueron almacenadas en formatos analíticos que facilitaron su reutilización durante las siguientes etapas del proyecto. A partir de estas bases maestras se generaron posteriormente las series temporales de demanda, los indicadores de ocupación, el proxy diario de presión hospitalaria y las bases específicas utilizadas en los modelos predictivos y de factores asociados.

3.5.Construcción y análisis de la demanda hospitalaria

Una vez integrada y depurada la información de los Registros Estadísticos de Egresos Hospitalarios, se procedió a construir la serie temporal utilizada para representar el comportamiento de la demanda hospitalaria en el Distrito Metropolitano de Quito. Esta etapa tuvo como finalidad transformar los registros individuales de hospitalización en una estructura temporal diaria que permitiera analizar su evolución y posteriormente desarrollar los modelos predictivos.

Para efectos de esta investigación, la demanda hospitalaria se analizó a partir del número de ingresos hospitalarios registrados diariamente. Esta definición permite estudiar el flujo de pacientes que requieren internación en los establecimientos de salud y diferenciarlo conceptualmente de la ocupación hospitalaria, que depende adicionalmente de la duración de la estancia y de la capacidad disponible.

El análisis comprendió el periodo entre el 1 de enero de 2017 y el 31 de diciembre de 2024. A partir de esta información se construyó una serie diaria continua, sobre la cual se efectuaron análisis exploratorios orientados a identificar tendencias, patrones recurrentes, variaciones asociadas al calendario y cambios extraordinarios en el comportamiento de la demanda.

3.5.1. Definición de la demanda hospitalaria

En el contexto de la investigación, se definió la demanda hospitalaria diaria como el número de hospitalizaciones cuya fecha de ingreso corresponde a un día determinado. De esta manera, para cada fecha t , la demanda se obtuvo mediante:

$$Dt = \sum_{i=1}^N I(F_i = t)$$

donde:

- Dt representa el número de ingresos hospitalarios registrados en el día t ;
- F_i corresponde a la fecha de ingreso de la hospitalización i ;
- $I(.)$ es una función indicadora que toma el valor de 1 cuando la fecha de ingreso coincide con el día analizado y 0 en caso contrario.

La elección de la fecha de ingreso, en lugar de la fecha de egreso, responde al objetivo de analizar el momento en que los pacientes entran al sistema hospitalario y generan nueva demanda de internación. La fecha de egreso, por su parte, fue utilizada posteriormente para determinar la permanencia del paciente y reconstruir la utilización diaria de la capacidad hospitalaria.

Esta diferenciación resulta metodológicamente importante, debido a que el número de ingresos diarios constituye una medida del flujo de demanda, pero no representa directamente el número de camas ocupadas ni permite determinar por sí mismo la existencia de saturación hospitalaria.

Para esta etapa se consideraron únicamente los registros cuya fecha de ingreso se encontraba comprendida entre el 1 de enero de 2017 y el 31 de diciembre de 2024. Como resultado, la base utilizada para construir la serie temporal estuvo constituida por 1.811.587 ingresos hospitalarios.

3.5.2. Construcción de la serie temporal diaria

A partir de las fechas de ingreso previamente estandarizadas, los registros individuales fueron agrupados por día, contabilizando el número total de ingresos hospitalarios ocurridos en cada fecha.

Posteriormente, se construyó un calendario continuo comprendido entre el 1 de enero de 2017 y el 31 de diciembre de 2024, con la finalidad de garantizar que cada día del periodo estuviera representado en la serie, independientemente del número de ingresos registrados.

Este procedimiento permitió evitar discontinuidades temporales que pudieran afectar el análisis y entrenamiento posterior de los modelos.

La serie resultante quedó conformada por 2.922 observaciones diarias consecutivas, correspondientes a los ocho años analizados. Se verificó la ausencia de fechas duplicadas y discontinuidades dentro del calendario construido. Formalmente, la serie utilizada puede expresarse como:

$$Y = \{D_1, D_2, \dots, D_{2922}\}$$

donde cada D_t corresponde al número total de ingresos hospitalarios observados en un día específico.

La utilización de periodicidad diaria fue seleccionada debido a que permite conservar una mayor cantidad de información temporal que una agregación mensual o anual y facilita la identificación de patrones de corto plazo, particularmente aquellos relacionados con el comportamiento semanal de la demanda. A partir de la fecha se generaron adicionalmente variables temporales auxiliares, como año, mes y día de la semana. Estas variables fueron utilizadas principalmente con fines exploratorios y, cuando correspondía, como parte de los análisis posteriores.

3.5.3. Análisis exploratorio del comportamiento temporal de la demanda

Sobre la serie diaria construida se realizó un análisis exploratorio orientado a comprender la evolución de la demanda hospitalaria antes de implementar los modelos predictivos.

En primer lugar, se examinó el comportamiento de los ingresos a diferentes niveles de agregación temporal. La agregación anual permitió observar cambios generales en el volumen de hospitalizaciones a lo largo del periodo 2017–2024, mientras que las agregaciones mensuales y semanales facilitaron el reconocimiento de variaciones de menor duración.

También se analizó el comportamiento de la demanda según el día de la semana, con el propósito de identificar la posible presencia de patrones semanales recurrentes. Este análisis resulta especialmente relevante para la selección de los modelos de referencia utilizados posteriormente, debido a que una serie hospitalaria puede presentar dependencia respecto de observaciones correspondientes al mismo día de semanas anteriores.

De manera complementaria, se utilizaron representaciones gráficas y estadísticas descriptivas para evaluar la distribución de los ingresos diarios, identificar valores extremos y observar modificaciones en el nivel o variabilidad de la serie.

El periodo 2020–2021 recibió un tratamiento interpretativo particular, debido a que coincide con la emergencia sanitaria provocada por la pandemia de COVID-19. En consecuencia, las variaciones observadas durante estos años no fueron asumidas automáticamente como patrones estacionales recurrentes, sino como posibles cambios estructurales asociados a un evento extraordinario que modificó el comportamiento habitual de los servicios hospitalarios.

Esta consideración resulta relevante debido a que la estacionalidad implica la repetición de patrones en intervalos regulares, mientras que un choque externo puede producir modificaciones temporales o estructurales que no necesariamente volverán a presentarse bajo condiciones normales.

3.5.4. Diferenciación entre demanda y ocupación hospitalaria

Aunque la demanda diaria constituye un componente relacionado con la utilización de los servicios hospitalarios, en la presente investigación no se empleó como sustituto de la ocupación hospitalaria.

Un día con un elevado número de ingresos puede generar mayor presión sobre los establecimientos, pero su efecto sobre la ocupación depende también de factores como la cantidad de pacientes que ya permanecían hospitalizados, la duración de sus estancias y la capacidad de camas disponible.

Por esta razón, el análisis desarrollado en esta sección se interpreta exclusivamente como un análisis de demanda hospitalaria. La ocupación y la presión hospitalaria relativa a la capacidad se abordan posteriormente mediante la integración de la información de camas, días-paciente y días-cama disponibles.

Esta separación conceptual permitió establecer dos componentes analíticos diferenciados: por una parte, la predicción del flujo futuro de ingresos hospitalarios y, por otra, el análisis de la utilización de la capacidad hospitalaria.

3.6.Desarrollo de modelos predictivos de demanda

Una vez construida la serie temporal diaria de ingresos hospitalarios, se desarrolló la etapa de modelado predictivo con el propósito de evaluar la capacidad de diferentes enfoques para estimar la demanda hospitalaria diaria en el Distrito Metropolitano de Quito.

El proceso de modelado se diseñó considerando la naturaleza temporal de los datos. A diferencia de una partición aleatoria convencional, las observaciones conservaron en todo momento su orden cronológico, de manera que cada predicción se generara utilizando únicamente información disponible con anterioridad al día que se pretendía estimar.

Para establecer una comparación metodológicamente consistente se evaluaron cuatro aproximaciones: un modelo ingenuo basado en la observación inmediatamente anterior, un modelo ingenuo estacional basado en el comportamiento de la semana previa, un modelo de aprendizaje profundo mediante redes Long Short-Term Memory (LSTM) y el modelo Prophet para pronóstico de series temporales.

El desarrollo se dividió en dos etapas. La primera correspondió a un proceso de validación temporal fuera de muestra durante 2020–2023, utilizado para comparar modelos y seleccionar sus configuraciones. La segunda consistió en una evaluación independiente sobre el año 2024, que permaneció separado durante el proceso de selección.

3.6.1. Diseño de la validación temporal

Debido a la dependencia cronológica presente en una serie temporal, no se aplicó una división aleatoria entre observaciones de entrenamiento y validación. Este tipo de partición

podría permitir que información cronológicamente posterior intervenga en el entrenamiento de modelos destinados a predecir observaciones anteriores, generando una estimación optimista de su capacidad predictiva.

En su lugar, se implementó un esquema de ventana expansiva de validación temporal, en el cual cada conjunto de entrenamiento incorporó progresivamente la información histórica disponible y el año inmediatamente posterior fue utilizado como periodo de validación. El diseño empleado se lo puede apreciar en la siguiente tabla:

Tabla 3

Esquema de validación temporal para la evaluación de los modelos predictivos de demanda hospitalaria

Iteración	Periodo de entrenamiento	Periodo de validación
1	2017–2019	2020
2	2017–2020	2021
3	2017–2021	2022
4	2017–2022	2023

Nota. Elaboración propia a partir de datos del Instituto Nacional de Estadística y Censos (INEC, 2017–2024).

Este esquema permitió evaluar los modelos frente a diferentes condiciones temporales. Su utilización fue particularmente relevante debido al comportamiento extraordinario observado durante 2020 y 2021, asociado al periodo de la pandemia de COVID-19. Por este motivo, dichos años fueron conservados dentro de la evaluación y no fueron considerados como una manifestación convencional de estacionalidad.

La comparación principal de los modelos se realizó a partir del desempeño promedio obtenido en los cuatro periodos de validación. Adicionalmente, se consideraron análisis de sensibilidad sobre los años posteriores a la mayor disrupción observada durante 2020, sin modificar el criterio principal de selección.

3.6.2. Modelos de referencia

Antes de evaluar modelos de mayor complejidad se implementaron dos modelos de referencia o baselines, cuya finalidad fue establecer un nivel mínimo de desempeño que los modelos avanzados debían ser capaces de superar.

El primer modelo correspondió al enfoque Naive, en el cual la demanda prevista para el día siguiente es igual a la demanda observada en el día inmediatamente anterior, mientras que el segundo modelo fue el Seasonal Naive, que utiliza como predicción el valor registrado siete días antes

Este segundo enfoque fue incorporado debido a la posible presencia de un patrón semanal en los ingresos hospitalarios, permitiendo evaluar si los modelos más complejos proporcionaban mejoras frente a una regla de predicción sencilla basada en el mismo día de la semana anterior.

La inclusión de estos modelos permitió evitar que el desempeño de LSTM y Prophet fuera analizado de manera aislada. Un modelo de mayor complejidad únicamente puede considerarse útil desde el punto de vista predictivo si su desempeño representa una mejora frente a métodos de referencia razonables para la estructura temporal de los datos.

3.6.3. Modelo Long Short-Term Memory

Las redes Long Short-Term Memory (LSTM) constituyen una variante de las redes neuronales recurrentes diseñada para aprender dependencias temporales dentro de datos secuenciales. Su utilización en esta investigación tuvo como finalidad determinar si un modelo de aprendizaje profundo podía capturar relaciones temporales en la evolución diaria de los ingresos hospitalarios que no fueran representadas adecuadamente mediante los modelos de referencia. (Hochreiter & Schmidhuber, 1997).

Para convertir la serie temporal en secuencias de entrada se evaluaron cuatro tamaños de ventana histórica o *lookback*: 7, 14, 21 y 28 días. Cabe mencionar que para una ventana de tamaño *L*, el modelo utiliza las *L* observaciones anteriores para estimar la demanda del día siguiente.

Cada tamaño de ventana fue evaluado utilizando exactamente los mismos periodos de validación temporal establecidos previamente. De esta manera, la selección de la longitud de la secuencia no se basó en el comportamiento del conjunto final de prueba.

Debido a que las redes neuronales son sensibles a la escala de los datos, se aplicó un proceso de normalización previo al entrenamiento. (Goodfellow et al., 2016). Para evitar fuga de información, los parámetros de escalamiento fueron estimados exclusivamente con las observaciones pertenecientes al conjunto de entrenamiento de cada iteración. Posteriormente, ese mismo escalador fue utilizado para transformar las observaciones del periodo de validación. Por tanto, ninguna característica estadística del conjunto de validación fue utilizada para determinar la transformación aplicada durante el entrenamiento.

La evaluación se efectuó mediante predicciones un paso adelante, manteniendo la secuencia cronológica de la serie. Asimismo, las configuraciones utilizadas durante el proceso de desarrollo fueron definidas y comparadas utilizando únicamente los periodos de validación 2020–2023.

La selección de la ventana temporal se realizó utilizando como criterio principal el desempeño promedio obtenido mediante el Error Absoluto Medio (MAE) en los periodos de validación. Una vez elegida la configuración, esta quedó establecida antes de realizar la evaluación sobre el conjunto independiente correspondiente al año 2024.

3.6.4. Modelo Prophet

Como segundo enfoque de modelado temporal se utilizó Prophet, un modelo diseñado para representar series temporales mediante componentes asociados a tendencia y estacionalidad. (Taylor & Letham, 2018). Su incorporación permitió comparar el desempeño de una arquitectura de aprendizaje profundo como LSTM con una metodología diseñada específicamente para pronóstico temporal que modela de manera explícita diferentes componentes de la serie.

Prophet fue evaluado bajo los mismos periodos de entrenamiento y validación utilizados para los demás modelos, garantizando que las comparaciones se realizaran sobre exactamente las mismas observaciones.

Las predicciones se generaron siguiendo un esquema de pronóstico un paso adelante, utilizando únicamente la información históricamente disponible hasta el momento de cada

estimación. De esta manera, el modelo no tuvo acceso a observaciones futuras del periodo que estaba siendo evaluado. Este procedimiento permitió mantener un criterio de comparación homogéneo entre Prophet, LSTM y los modelos de referencia.

3.6.5. Métricas de evaluación

El desempeño predictivo se evaluó utilizando tres métricas complementarias: Error Absoluto Medio (MAE), Raíz del Error Cuadrático Medio (RMSE) y Error Porcentual Absoluto Medio (MAPE).

El MAE representa la magnitud promedio del error expresada en las mismas unidades de la variable analizada, es decir, número de ingresos hospitalarios diarios. Por su interpretación directa, el MAE fue establecido como criterio principal para la comparación y selección de modelos.

El RMSE penaliza con mayor intensidad los errores de gran magnitud y fue utilizado como métrica complementaria para identificar modelos particularmente sensibles ante desviaciones elevadas.

Finalmente, el MAPE proporciona una representación relativa del error respecto del valor observado. No obstante, su interpretación se realizó de manera complementaria debido a que puede presentar mayor sensibilidad cuando los valores reales de demanda son excepcionalmente bajos.

Estas métricas permiten evaluar dimensiones complementarias del error de pronóstico: el MAE mantiene la unidad de la variable, el RMSE penaliza con mayor intensidad los errores grandes y el MAPE expresa el error en términos porcentuales, aunque puede presentar dificultades ante valores reales cercanos a cero (Hyndman & Athanasopoulos, 2018).

3.6.6. Criterio de selección y evaluación independiente

Para evitar seleccionar un modelo a partir de su desempeño sobre el mismo periodo utilizado como prueba final, se estableció una separación entre la fase de desarrollo y la fase de evaluación independiente.

Durante el desarrollo, cada modelo fue evaluado sobre los años 2020, 2021, 2022 y 2023. Para cada aproximación se calcularon las métricas correspondientes y posteriormente se obtuvo el desempeño promedio entre los periodos de validación.

El RMSE y el MAPE fueron considerados métricas complementarias para caracterizar el comportamiento de cada alternativa.

Una vez determinado el modelo y las configuraciones utilizadas durante la fase de desarrollo, estos criterios fueron congelados antes de examinar los resultados correspondientes al año 2024. De esta manera, el último año funcionó como una evaluación independiente del comportamiento de los modelos frente a información no utilizada durante su selección.

El conjunto de prueba quedó definido como: 1 de enero de 2024 – 31 de diciembre de 2024; correspondiente a 366 observaciones diarias, debido a que 2024 fue un año bisiesto.

En esta etapa se evaluaron nuevamente los modelos de referencia, LSTM y Prophet utilizando las mismas métricas establecidas durante el desarrollo. Los resultados obtenidos sobre 2024 fueron utilizados para analizar la capacidad de generalización de cada alternativa, pero no para modificar retrospectivamente el modelo seleccionado durante la etapa de validación.

3.7. Desarrollo del análisis de ocupación hospitalaria

Una vez analizado el comportamiento de la demanda hospitalaria, se incorporó la información del Registro Estadístico de Camas Hospitalarias con el propósito de estudiar la utilización de la capacidad disponible en los establecimientos de salud del Distrito Metropolitano de Quito.

A diferencia de la demanda, representada en esta investigación mediante el número diario de ingresos hospitalarios, la ocupación requiere relacionar la utilización efectiva de los servicios con la capacidad disponible. Por esta razón, este componente se desarrolló utilizando principalmente las variables días-paciente y días-cama disponibles, disponibles en los registros de Camas Hospitalarias correspondientes al periodo 2017–2024.

El análisis permitió obtener indicadores anuales de ocupación para el conjunto de establecimientos ubicados en el Distrito Metropolitano de Quito y posteriormente caracterizar diferentes niveles de utilización de la capacidad hospitalaria. Estos indicadores constituyeron, además, la referencia para el desarrollo posterior del proxy diario de presión hospitalaria.

3.7.1. Definición y cálculo de la ocupación hospitalaria

Para efectos de esta investigación, la ocupación hospitalaria se calculó mediante la relación entre los días-paciente registrados durante un periodo y los días-cama disponibles correspondientes al mismo periodo.

El indicador se expresa mediante: De acuerdo con la metodología del INEC (2025), el porcentaje de ocupación hospitalaria relaciona los días-paciente con los días-cama disponibles durante un periodo determinado.

$$\text{Ocupación Hospitalaria (\%)} = \frac{\text{Días - Paciente}}{\text{Días - camas disponibles}} * 100$$

Los días-paciente representan la utilización acumulada de camas por parte de los pacientes hospitalizados durante el periodo analizado, mientras que los días-cama disponibles representan la capacidad temporal de hospitalización efectivamente disponible.

Esta relación permite analizar el grado de utilización de la capacidad hospitalaria y evita considerar el número de ingresos o egresos como una medida directa de ocupación.

Para obtener el indicador agregado se utilizaron las sumas de días-paciente y días-cama disponibles de todos los registros correspondientes a cada año. Por tanto, la ocupación anual agregada se calculó de la siguiente manera:

$$O_y = \frac{\sum_{i=1}^{n_y} DP_{iy}}{\sum_{i=1}^{n_y} DC_{iy}} * 100$$

donde:

- O_y corresponde al porcentaje de ocupación hospitalaria del Distrito Metropolitano de Quito en el año y ;
- DP_{iy} representa los días-paciente del establecimiento o registro i durante el año y ;
- DC_{iy} representa los días-cama disponibles correspondientes al mismo registro;
- n_y corresponde al número de registros hospitalarios disponibles para ese año.

Es importante señalar que el indicador agregado no se obtuvo mediante el promedio simple de los porcentajes de ocupación individuales, debido a que los establecimientos presentan capacidades hospitalarias diferentes. El cálculo mediante el cociente de las sumas permite ponderar implícitamente cada registro de acuerdo con su volumen de días-cama disponibles.

3.7.2. Variables de capacidad hospitalaria

Además de los días-paciente y días-cama disponibles utilizados para el cálculo del porcentaje de ocupación, se examinaron otras variables relacionadas con la capacidad hospitalaria.

Entre ellas se consideraron principalmente las camas de dotación normal y las camas disponibles. Por otro lado, las camas de dotación normal corresponden a la capacidad establecida habitualmente para el funcionamiento del establecimiento, mientras que las camas disponibles representan aquellas que estuvieron efectivamente habilitadas para la atención durante el periodo registrado. Esta diferenciación resulta relevante, debido a que la capacidad teórica de un establecimiento no necesariamente coincide con la capacidad efectivamente disponible para hospitalización.

La evolución anual de estas variables fue analizada para contextualizar los cambios observados en la ocupación y determinar si las variaciones en la utilización hospitalaria estuvieron acompañadas por modificaciones en la capacidad disponible.

Tabla 4

Variables utilizadas para el análisis de capacidad y ocupación hospitalaria

Variable	Utilización en el estudio
Camas de dotación normal	Caracterización de la capacidad hospitalaria instalada
Camas disponibles	Caracterización de la capacidad efectivamente disponible
Días-cama disponibles	Denominador para el cálculo de la ocupación hospitalaria
Días-paciente	Numerador para el cálculo de la ocupación hospitalaria
Egresos hospitalarios	Variable complementaria para validación y caracterización

Nota. Elaboración propia a partir de las variables del Registro Estadístico de Camas Hospitalarias del Instituto Nacional de Estadística y Censos (INEC, 2017–2024).

3.7.3. Caracterización de los niveles de ocupación hospitalaria

Con el propósito de facilitar la interpretación de los indicadores obtenidos, se establecieron categorías referenciales de ocupación hospitalaria. La selección de los puntos de corte de 85 % y 90 % se sustentó en el estudio de Bagust et al. (1999), quienes, mediante un modelo de simulación estocástica aplicado a la utilización de camas hospitalarias, identificaron que los riesgos asociados a una disponibilidad insuficiente de camas comienzan a incrementarse cuando la ocupación promedio supera aproximadamente el 85 %, mientras que niveles iguales o superiores al 90 % se relacionan con una mayor probabilidad de presentar déficits recurrentes de capacidad. A partir de esto, y con fines exclusivamente analíticos, en la presente investigación los porcentajes de ocupación se clasificaron de la siguiente manera:

Tabla 5

Categorías referenciales utilizadas para caracterizar los niveles de ocupación hospitalaria

Porcentaje de ocupación	Categoría analítica
Menor a 85 %	Por debajo del nivel de alta ocupación
85 % a menos de 90 %	Alta ocupación
90 % a menos de 100 %	Ocupación muy alta
100 % o más	Sobreocupación observada

Nota. Elaboración propia a partir de las variables del Registro Estadístico de Camas Hospitalarias del Instituto Nacional de Estadística y Censos (INEC, 2017–2024).

3.7.4. Análisis agregado y por sector hospitalario

El cálculo de la ocupación se realizó inicialmente para el conjunto de establecimientos del Distrito Metropolitano de Quito, permitiendo obtener una medida agregada de la utilización hospitalaria durante cada año del periodo 2017–2024.

Sin embargo, un indicador agregado puede ocultar diferencias importantes entre grupos de establecimientos. Por esta razón, el análisis fue complementado mediante la desagregación según sector hospitalario, considerando los establecimientos públicos, privados con fines de lucro y privados sin fines de lucro. Para cada sector se aplicó el mismo principio de cálculo, utilizando el cociente entre la suma de días-paciente y la suma de días-cama disponibles correspondientes al grupo analizado. Esta desagregación permitió examinar si los niveles de utilización de la capacidad presentaban comportamientos diferentes entre sectores, evitando que las características de establecimientos con mayor capacidad dominaran la interpretación del indicador general del Distrito Metropolitano de Quito.

3.7.5. Alcance del indicador anual de ocupación

El indicador desarrollado en esta sección corresponde a una medida de ocupación hospitalaria con periodicidad anual, debido a la estructura temporal de la información disponible en el Registro Estadístico de Camas Hospitalarias. Por esta razón, los resultados permiten analizar la utilización global de la capacidad durante cada año, pero no permiten conocer directamente cuántas camas se encontraban disponibles u ocupadas en un día específico.

Esta limitación resulta especialmente importante para la presente investigación, debido a que un porcentaje anual puede ocultar episodios de elevada utilización ocurridos durante periodos más cortos. Dada esta novedad, en la siguiente etapa metodológica se desarrolló una aproximación complementaria basada en las fechas individuales de ingreso y egreso de los pacientes, con el propósito de reconstruir la cantidad de hospitalizaciones activas en cada día y relacionarla con una referencia de capacidad hospitalaria. El indicador resultante se interpreta como un proxy diario de presión hospitalaria relativa a la capacidad y no como una tasa oficial diaria de ocupación.

3.8.Reconstrucción de la presión hospitalaria diaria

El Registro Estadístico de Camas Hospitalarias permite calcular indicadores de ocupación a partir de los días-paciente y los días-cama disponibles; sin embargo, la información de capacidad utilizada en esta investigación posee una periodicidad anual. Esta característica limita la posibilidad de identificar directamente variaciones de corto plazo o episodios específicos de elevada utilización hospitalaria.

Con el propósito de complementar el análisis anual desarrollado en la sección anterior, se construyó un proxy diario de presión hospitalaria relativa a la capacidad. Para ello, se utilizaron las fechas individuales de ingreso y egreso disponibles en el Registro Estadístico de Egresos Hospitalarios para reconstruir el número de hospitalizaciones activas en cada día y relacionarlo con una estimación de la capacidad media diaria disponible durante el año correspondiente.

Esta aproximación se desarrolló para el periodo 2017–2023 y posteriormente se desagregó según sector hospitalario. El año 2024 no fue incorporado en el análisis principal del proxy debido a la existencia de censura a la derecha al finalizar la serie de egresos disponible.

3.8.1. Reconstrucción de hospitalizaciones activas por día

La reconstrucción diaria se realizó a partir de las fechas de ingreso y egreso de cada hospitalización. Para cada registro se determinó el intervalo durante el cual el paciente permaneció hospitalizado y, posteriormente, se contabilizó el número de hospitalizaciones activas en cada fecha del periodo de estudio.

La duración utilizada para este procedimiento respetó la regla establecida durante la preparación de los datos, según la cual toda hospitalización representa al menos un día de utilización. En consecuencia, los casos en los que el ingreso y el egreso ocurrieron durante la misma fecha fueron contabilizados como un día de hospitalización.

Para hospitalizaciones con una duración superior a cero, el intervalo de utilización se reconstruyó a partir de las fechas de ingreso y egreso. En los casos de estancia del mismo día se asignó una jornada de utilización, evitando que estas observaciones fueran excluidas por presentar una diferencia de cero días entre ambas fechas. La reconstrucción se implementó

mediante un procedimiento basado en eventos de entrada y salida, evitando expandir físicamente cada hospitalización en múltiples registros diarios. Este enfoque permitió procesar eficientemente el elevado volumen de información y obtener el número de hospitalizaciones activas para cada fecha.

3.8.2. Estimación de la capacidad media diaria disponible

Para relacionar el número de hospitalizaciones activas con la capacidad hospitalaria fue necesario obtener una referencia diaria de camas disponibles. Debido a que el Registro Estadístico de Camas Hospitalarias proporciona los días-cama disponibles a nivel anual, se calculó una capacidad media diaria para cada año mediante la siguiente fórmula:

$$C_y = \frac{DC_y}{N_y}$$

donde:

- C_y corresponde a la capacidad media diaria estimada para el año y ;
- DC_y representa el total anual de días-cama disponibles;
- N_y corresponde al número de días del año, es decir, 365 o 366 en los años bisiestos.

Este indicador representa el número promedio de camas disponibles por día durante el año y y permite disponer de un denominador común para relacionar la reconstrucción diaria de hospitalizaciones con la capacidad reportada en el Registro de Camas.

3.8.3. Construcción del proxy diario de presión hospitalaria

Una vez obtenido el número de hospitalizaciones activas y la capacidad media diaria, se construyó el indicador de presión hospitalaria relativa a la capacidad mediante:

$$P_t = \frac{H_t}{C_y} * 100$$

donde:

- P_t representa el proxy de presión hospitalaria correspondiente al día t ;
- H_t es el número de hospitalizaciones activas reconstruidas para ese día;
- C_y corresponde a la capacidad media diaria estimada para el año al que pertenece t .

El indicador expresa, en términos porcentuales, la relación entre las hospitalizaciones reconstruidas y una referencia de capacidad media disponible.

Es importante señalar que este indicador no constituye una tasa oficial diaria de ocupación hospitalaria. La ocupación oficial se calculó en la sección anterior utilizando directamente los días-paciente y días-cama disponibles reportados en el Registro Estadístico de Camas Hospitalarias. En cambio, el indicador diario combina una reconstrucción basada en los microdatos de egresos con una capacidad media derivada de información anual. Por esta razón, en toda la investigación se lo denomina proxy diario de presión hospitalaria relativa a la capacidad.

3.8.4. Validación frente a la ocupación anual oficial

Como procedimiento de control, la reconstrucción diaria fue agregada a nivel anual y comparada con los indicadores oficiales de ocupación calculados a partir del Registro Estadístico de Camas Hospitalarias. El propósito de esta comparación es determinar el grado de correspondencia entre la utilización reconstruida mediante las fechas de ingreso y egreso y la información agregada reportada en el Registro de Camas.

Para cada año, se compararon el cociente oficial entre días-paciente y días-cama disponibles contra el coeficiente reconstruido de las hospitalizaciones a partir de los registros individuales.

Tabla 6

Indicadores utilizados en la reconstrucción de la presión hospitalaria diaria

Indicador	Construcción	Interpretación
Hospitalizaciones activas diarias	Fechas de ingreso y egreso	Pacientes hospitalizados reconstruidos para cada fecha
Capacidad media diaria	Días-cama disponibles anuales / días del año	Referencia promedio de capacidad disponible
Proxy diario de presión	Hospitalizaciones activas / capacidad media diaria $\times 100$	Presión relativa a la capacidad media disponible
Ocupación anual oficial	Días-paciente / días-cama disponibles $\times 100$	Indicador oficial utilizado como referencia de validación

Nota. Elaboración propia a partir de las variables del Registro Estadístico de Camas Hospitalarias del Instituto Nacional de Estadística y Censos (INEC, 2017–2024).

3.8.5. Reconstrucción por sector hospitalario

Además del análisis agregado para el Distrito Metropolitano de Quito, el procedimiento fue aplicado independientemente según el sector al que pertenecía el establecimiento de salud. Se consideraron tres grupos:

- sector público;
- sector privado con fines de lucro;
- sector privado sin fines de lucro.

Para cada sector s, se reconstruyeron las hospitalizaciones activas y se estimó una capacidad media diaria utilizando exclusivamente los días-cama disponibles correspondientes al mismo grupo. Esta desagregación permitió evitar que el comportamiento agregado del Distrito Metropolitano de Quito ocultara episodios de mayor presión concentrados en determinados grupos de establecimientos.

La comparación entre la reconstrucción y la ocupación oficial también se efectuó a nivel sectorial. Las discrepancias identificadas se mantuvieron sin calibración y fueron consideradas al momento de interpretar posteriormente los resultados.

3.8.6. Identificación de episodios de alta presión hospitalaria

Una vez construida la serie diaria sectorial, se identificaron periodos en los que el proxy alcanzó niveles elevados respecto de la capacidad media disponible. Para esto, se utilizaron las mismas referencias establecidas previamente para el análisis de ocupación:

Tabla 7

Categorías utilizadas para la clasificación del proxy diario de presión hospitalaria

Proxy diario de presión	Categoría analítica
Menor a 85 %	Por debajo del nivel de alta ocupación
85 % a menos de 90 %	Alta presión relativa
90 % a menos de 100 %	Presión relativa muy alta
100 % o más	Sobreocupación observada según el proxy

Nota. Elaboración propia a partir de las variables del Registro Estadístico de Camas Hospitalarias del Instituto Nacional de Estadística y Censos (INEC, 2017–2024).

Adicionalmente, se definió un episodio de alta presión como uno o más días consecutivos durante los cuales el indicador permaneció igual o por encima del umbral analizado. Este procedimiento permitió distinguir entre aumentos aislados de un solo día y periodos persistentes de elevada presión relativa a capacidad. Para cada episodio se determinaron posteriormente características como su duración, nivel promedio y valor máximo observado.

3.8.7. Delimitación temporal

Aunque los registros de egresos utilizados en la investigación se encuentran disponibles hasta el 31 de diciembre de 2024, el análisis principal del proxy diario se restringió al periodo 2017–2023. La razón de esta decisión corresponde a un problema de limitación a la derecha.

Para reconstruir correctamente la utilización hospitalaria de un día determinado es necesario conocer tanto la fecha de ingreso como la posterior fecha de egreso del paciente. Sin embargo, un paciente que ingresó durante los últimos días de 2024 y permaneció hospitalizado durante 2025 no puede ser identificado completamente utilizando únicamente el archivo de egresos disponible hasta 2024. En consecuencia, a medida que se aproxima el final del último año disponible aumenta la posibilidad de subestimar el número de hospitalizaciones activas, debido a que no se observan todavía los egresos ocurridos en el periodo siguiente.

Si únicamente se dispone de los egresos hasta 2024, esa hospitalización no puede reconstruirse completamente. Por este motivo, 2024 fue conservado para el análisis y predicción de demanda hospitalaria, donde la fecha de ingreso sí se encuentra observada, pero fue excluido del análisis principal de presión diaria reconstruida. Esta decisión evita interpretar artificialmente la reducción de hospitalizaciones observada al final de 2024 como una disminución real de la presión hospitalaria.

3.9.Desarrollo del modelo de factores asociados

Una vez construido el proxy diario de presión hospitalaria relativa a la capacidad, se desarrolló un componente analítico orientado a identificar variables asociadas con su comportamiento. Este análisis se realizó utilizando una base con nivel de agregación sector-

día para el periodo 2017–2023 y diferenciando los sectores público, privado con fines de lucro y privado sin fines de lucro. La variable objetivo principal correspondió al proxy diario continuo de presión hospitalaria relativa a la capacidad. Se optó por utilizar el indicador de forma continua debido a que los días que alcanzaban niveles iguales o superiores al 85 % representaban una proporción reducida del total de observaciones, lo que generaba un fuerte desbalance para el desarrollo de un modelo de clasificación. Por esta razón, los episodios de alta presión identificados en la sección anterior se utilizaron principalmente con fines descriptivos, mientras que el análisis de factores asociados se desarrolló sobre la variación continua del indicador.

3.9.1. Construcción de la base analítica sector-día

Para el desarrollo de los modelos se construyó una base analítica cuya unidad de observación correspondió a la combinación de sector hospitalario y día. De esta manera, para cada fecha se generaron observaciones independientes para los sectores público, privado con fines de lucro y privado sin fines de lucro. La base integró el proxy diario de presión hospitalaria con información procedente de los registros de egresos, variables temporales y características agregadas de los pacientes atendidos.

El periodo de análisis comprendió los años 2017–2023, manteniendo la misma delimitación utilizada para la reconstrucción diaria de presión hospitalaria. Debido a que algunas variables predictoras requirieron información histórica de hasta 28 días anteriores, las primeras observaciones de cada serie sectorial no pudieron utilizarse en el modelado. La base resultante permitió evaluar de manera conjunta la evolución temporal de la presión hospitalaria y las diferencias existentes entre sectores.

3.9.2. Selección y construcción de variables predictoras

Las variables utilizadas en el análisis se organizaron en cuatro grupos principales: demanda reciente, calendario, sector hospitalario y características demográficas agregadas. Las variables de demanda reciente se construyeron a partir del número de ingresos hospitalarios observados en días anteriores. Se incluyeron valores rezagados y estadísticas móviles destinadas a representar tanto el nivel reciente de demanda como su comportamiento durante periodos previos.

Para representar los efectos relacionados con el calendario se incorporaron variables asociadas al día de la semana, mes y condición de fin de semana. Algunos componentes temporales fueron representados mediante transformaciones cíclicas, con el objetivo de conservar la naturaleza periódica del calendario. Por otro lado, el sector hospitalario se incorporó como una variable categórica para representar posibles diferencias estructurales entre establecimientos públicos, privados con fines de lucro y privados sin fines de lucro.

Finalmente, se generaron características demográficas agregadas a partir de los pacientes ingresados durante los días anteriores, entre ellas edad promedio y proporciones correspondientes a determinados grupos de edad y sexo.

Cabe mencionar que, todas las variables basadas en comportamiento histórico se calcularon utilizando exclusivamente información de fechas anteriores al día analizado. De esta manera, se evitó incorporar información futura dentro de las variables predictoras.

3.9.3. Modelo A: enfoque operacional y predictivo

El primer enfoque, denominado Modelo A, tuvo como finalidad evaluar la capacidad de predecir el comportamiento diario de la presión hospitalaria utilizando información disponible previamente. Este modelo incorporó variables relacionadas con demanda reciente, calendario, sector hospitalario y, adicionalmente, información histórica del propio proxy de presión hospitalaria mediante valores rezagados y estadísticas móviles.

La inclusión de estos antecedentes permitió evaluar hasta qué punto el comportamiento reciente de la presión aportaba información para estimar su evolución inmediata. Se compararon tres algoritmos:

- Regresión Lineal.
- Random Forest.
- XGBoost.

La regresión lineal se utilizó como referencia de comportamiento lineal. Random Forest combina múltiples árboles construidos mediante aleatorización y agrega sus resultados para generar la predicción final (Breiman, 2001). XGBoost implementa un sistema escalable de potenciación de árboles que construye modelos sucesivos orientados a reducir los errores

residuales (Chen & Guestrin, 2016). Las configuraciones de los modelos fueron establecidas antes de analizar los resultados finales y se mantuvieron sin modificaciones posteriores orientadas a mejorar artificialmente su desempeño sobre los periodos de validación.

3.9.4. Modelo B: análisis exploratorio de factores asociados

El segundo enfoque, denominado Modelo B, se desarrolló con una finalidad principalmente exploratoria e interpretativa. A diferencia del Modelo A, se excluyeron las variables correspondientes al historial del propio proxy de presión hospitalaria. Esta decisión permitió reducir la influencia directa del comportamiento pasado de la variable objetivo y examinar con mayor claridad la contribución de otros factores.

El Modelo B incorporó información relacionada con demanda reciente, calendario, sector hospitalario y características demográficas agregadas de los pacientes. Al igual que en el Modelo A, se evaluaron Regresión Lineal, Random Forest y XGBoost bajo un mismo esquema de validación.

Este segundo enfoque no fue planteado como un modelo causal. Por tanto, las relaciones identificadas entre las variables y el proxy de presión hospitalaria fueron interpretadas como asociaciones dentro del modelo, sin afirmar que dichas características provoquen directamente aumentos o reducciones de la presión hospitalaria.

3.9.5. Diseño de validación temporal

La evaluación de ambos enfoques se realizó mediante un esquema de validación temporal expansiva, conservando el orden cronológico de las observaciones. El diseño utilizado fue el siguiente:

Tabla 8

Esquema de validación temporal para los modelos de factores asociados

Iteración	Periodo de entrenamiento	Periodo de validación
1	2017–2020	2021
2	2017–2021	2022
3	2017–2022	2023

Nota. Elaboración propia a partir de las variables del Registro Estadístico de Camas Hospitalarias del Instituto Nacional de Estadística y Censos (INEC, 2017–2024).

El desempeño de los modelos se evaluó mediante el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el coeficiente de determinación R^2 . El MAE fue utilizado como criterio principal para comparar las alternativas.

También se incorporaron modelos de referencia sencillos, entre ellos el promedio global, el promedio histórico por sector y, cuando correspondía, la persistencia del valor observado en el día anterior. Estos modelos permitieron determinar si las técnicas de aprendizaje automático aportaban una mejora frente a estrategias de predicción más simples. La selección dentro de cada enfoque se realizó considerando el menor MAE promedio obtenido en los periodos de validación 2021–2023. Posteriormente, los modelos seleccionados fueron utilizados para el análisis de interpretabilidad desarrollado en la siguiente sección.

3.10. Interpretabilidad de los modelos

Una vez seleccionados los modelos de mejor desempeño dentro de los enfoques definidos en la sección anterior, se desarrolló una etapa de interpretabilidad orientada a comprender la contribución de las variables utilizadas en las predicciones.

El análisis se aplicó sobre el modelo seleccionado del Modelo A, correspondiente al enfoque predictivo operacional, y sobre el modelo seleccionado del Modelo B, orientado al análisis exploratorio de factores asociados. La interpretabilidad se realizó utilizando exclusivamente observaciones de los periodos de validación fuera de muestra, con el propósito de analizar el comportamiento de los modelos sobre información no utilizada durante su entrenamiento.

Para este análisis se utilizaron principalmente las técnicas SHAP (SHapley Additive exPlanations) y Permutation Importance, complementadas con un análisis de importancia agrupada y una evaluación de sensibilidad.

3.10.1. SHAP

SHAP (SHapley Additive exPlanations) es una técnica de interpretabilidad basada en los valores de Shapley que permite asignar a cada característica una medida de contribución sobre una predicción específica (Lundberg & Lee, 2017). Su fundamento se basa en los

valores de Shapley provenientes de la teoría de juegos, mediante los cuales la predicción puede descomponerse en contribuciones atribuibles a las diferentes características utilizadas. En esta investigación, SHAP se utilizó para identificar cuáles variables presentaban una mayor contribución promedio al comportamiento estimado del proxy de presión hospitalaria.

Para el Modelo A, seleccionado bajo el enfoque predictivo operacional, se aplicó SHAP sobre el modelo XGBoost. Este análisis permitió examinar la importancia de variables relacionadas con la historia reciente de la presión, la demanda hospitalaria y los componentes temporales. En cuanto al Modelo B, seleccionado para el análisis exploratorio de factores asociados, se aplicó SHAP sobre el modelo de Regresión Lineal. En este caso, la interpretación se complementó con los coeficientes estandarizados del modelo y con verificaciones adicionales de estabilidad.

Los valores SHAP fueron utilizados como medidas de contribución dentro de los modelos y no como evidencia de relaciones causales entre las variables y la presión hospitalaria.

3.10.2. Permutation Importance

Como técnica complementaria se utilizó Permutation Importance, procedimiento que evalúa la relevancia de una variable mediante el deterioro del desempeño predictivo producido al permutar aleatoriamente sus valores en el conjunto de evaluación (Breiman, 2001). El procedimiento consiste en permutar aleatoriamente los valores de una variable dentro del conjunto de validación, manteniendo el resto de las características sin modificaciones, y posteriormente calcular nuevamente el error predictivo.

La aplicación conjunta de SHAP y Permutation Importance permitió analizar la relevancia de las características desde perspectivas complementarias: SHAP evaluó su contribución dentro de las predicciones, mientras que la permutación permitió determinar cuánto se deterioraba la capacidad predictiva al alterar su información.

3.10.3. Importancia agrupada de variables

Debido a que algunas variables predictoras presentaban relaciones importantes entre sí, especialmente aquellas construidas a partir de rezagos y ventanas móviles de demanda, la

interpretación individual podía distribuir la importancia entre características altamente correlacionadas. Para complementar el análisis se realizó una Permutation Importance agrupada, en la que las variables fueron organizadas según su naturaleza conceptual. Los principales grupos considerados fueron:

Tabla 9

Grupos de variables considerados en el análisis de interpretabilidad

Grupo	Tipo de información
Sector	Sector hospitalario
Demanda reciente	Rezagos, medias móviles y variaciones recientes de ingresos
Calendario	Día de la semana, mes, fin de semana y componentes temporales
Composición demográfica	Edad y distribución agregada por grupos etarios y sexo
Periodo 2020–2021	Indicador correspondiente al periodo de mayor disrupción sanitaria

Nota. Elaboración propia. Los grupos reúnen variables conceptualmente relacionadas para evaluar su importancia conjunta mediante técnicas de interpretabilidad. El indicador 2020–2021 identifica el periodo de mayor alteración observado en la serie y no representa por sí mismo una relación causal con la presión hospitalaria.

En lugar de permutar una única variable, el procedimiento modificó simultáneamente las características pertenecientes a cada grupo y posteriormente evaluó el incremento producido en el MAE. Este análisis permitió obtener una interpretación más robusta de la relevancia de cada dimensión y reducir el riesgo de atribuir importancia excesiva a una variable específica cuando existían otras características estrechamente relacionadas con ella.

3.10.4. Análisis de sensibilidad

Finalmente, se realizó un análisis de sensibilidad con el propósito de evaluar la estabilidad de los resultados obtenidos para los factores asociados. Este procedimiento se concentró en los periodos posteriores a 2020, debido a que dicho año presentó una alteración extraordinaria en el comportamiento hospitalario asociada a la pandemia de COVID-19.

El análisis complementario consideró evaluaciones temporales posteriores, manteniendo el mismo enfoque de modelado y sin modificar las configuraciones seleccionadas previamente. La finalidad de esta etapa no fue seleccionar nuevamente los

modelos ni ajustar sus hiperparámetros, sino comprobar si la jerarquía general de los grupos de variables se mantenía al reducir la influencia del periodo de mayor disrupción.

Los resultados del análisis principal y de sensibilidad fueron posteriormente comparados para determinar qué factores presentaban una contribución relativamente estable a través de diferentes periodos temporales.

En todos los casos, las técnicas de interpretabilidad fueron utilizadas para analizar contribuciones y asociaciones dentro de los modelos. Por tanto, los resultados obtenidos mediante SHAP, Permutation Importance o sus variantes agrupadas no fueron interpretados como evidencia de causalidad.

CAPITULO 4:

4. ANÁLISIS DE RESULTADOS

4.1. Caracterización de los registros hospitalarios

La integración de los Registros Estadísticos de Egresos Hospitalarios permitió consolidar una base maestra de 1.814.236 registros correspondientes a establecimientos de salud ubicados en el Distrito Metropolitano de Quito durante el periodo 2017–2024. A partir de esta información se realizó una caracterización inicial orientada a conocer la estructura de los registros, las características demográficas de los pacientes, el sector de atención y la duración de las hospitalizaciones.

Para los análisis asociados a la demanda hospitalaria se consideraron 1.811.587 ingresos cuya fecha de admisión se encontraba comprendida entre el 1 de enero de 2017 y el 31 de diciembre de 2024. La diferencia respecto de la base maestra corresponde principalmente a hospitalizaciones iniciadas antes de 2017 y finalizadas dentro del periodo de estudio, las cuales fueron conservadas en la base general para otros componentes del análisis, particularmente para la reconstrucción de la utilización hospitalaria.

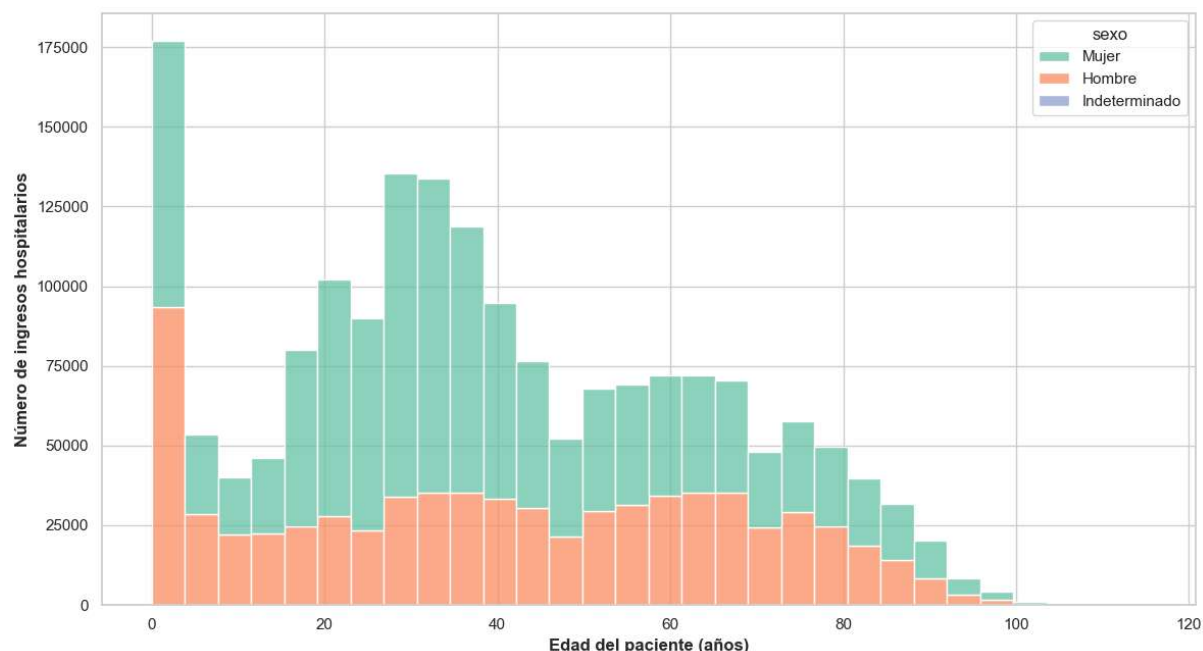
Es importante señalar que la delimitación territorial se realizó según la ubicación del establecimiento de salud, por lo que los resultados representan la actividad hospitalaria registrada en establecimientos localizados en el Distrito Metropolitano de Quito y no exclusivamente a pacientes residentes en el distrito.

4.1.1. Características demográficas de los pacientes

La población hospitalizada presentó una amplia distribución etaria. La edad promedio fue de 39,32 años, mientras que la mediana se ubicó en 36 años, registrándose edades comprendidas entre menos de un año y 115 años. La diferencia entre la media y la mediana evidencia una distribución ligeramente desplazada hacia edades superiores, característica esperable en una población hospitalaria que comprende desde pacientes pediátricos hasta adultos mayores.

Figura 1

Distribución de la edad de los pacientes hospitalizados en el Distrito Metropolitano de Quito, 2017–2024



Nota. Elaboración propia a partir de datos del INEC (2017–2024).

En relación con el sexo de los pacientes, se observó una mayor participación de mujeres, quienes representaron el 60,13 % de los ingresos analizados. Los hombres representaron aproximadamente el 39,87 %, mientras que únicamente cuatro registros presentaron la categoría indeterminado.

Tabla 10

Distribución de los ingresos hospitalarios según sexo del paciente, 2017–2024

Sexo	Número de registros	Porcentaje
Mujer	1.089.372	60,13 %
Hombre	722.211	39,87 %
Indeterminado	4	<0,01 %
Total	1.811.587	100,00 %

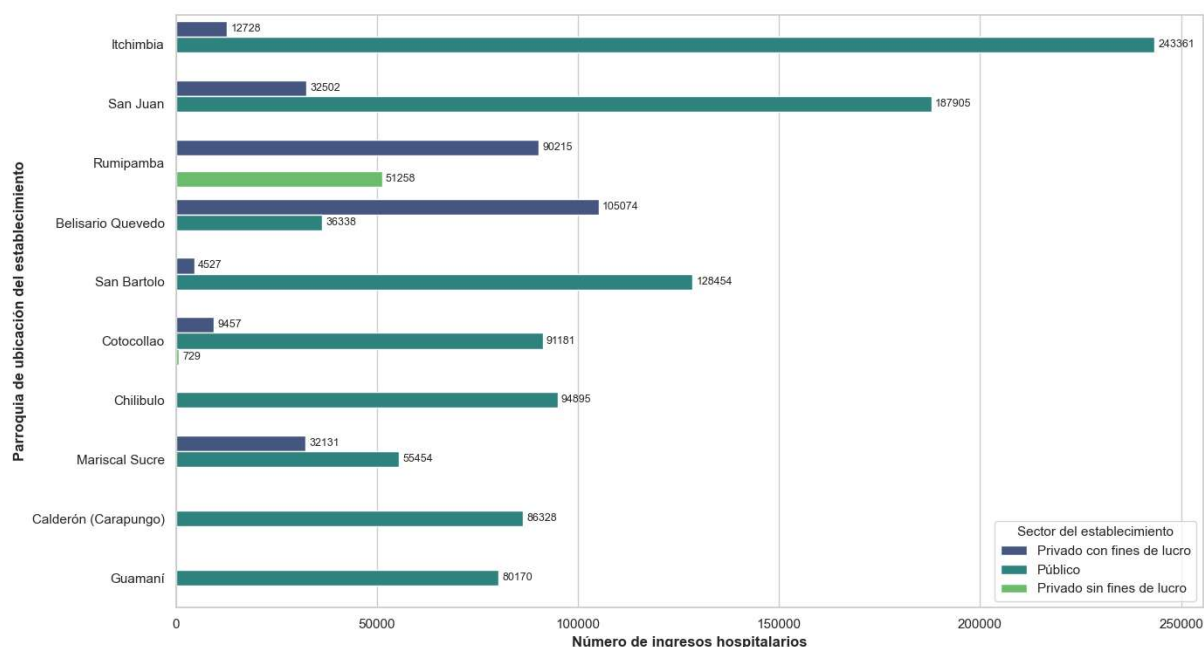
Nota. Elaboración propia a partir de los Registros Estadísticos de Egresos Hospitalarios del Instituto Nacional de Estadística y Censos (INEC, 2017–2024).

4.1.2. Caracterización según sector hospitalario y parroquia

La distribución de los ingresos hospitalarios evidencia una concentración importante de la actividad en determinadas parroquias del Distrito Metropolitano de Quito. La siguiente Figura presenta las diez parroquias con mayor número de ingresos hospitalarios durante el periodo 2017–2024, diferenciadas según el sector al que pertenece el establecimiento de salud.

Figura 2

Distribución de la demanda hospitalaria según parroquia y sector del establecimiento de salud en el Distrito Metropolitano de Quito, 2017–2024



Nota. La parroquia corresponde a la ubicación del establecimiento de salud y no al lugar de residencia del paciente.

A partir de los valores representados en la figura, Itchimbia registró el mayor volumen de ingresos hospitalarios, con 256.089 casos, equivalente aproximadamente al 14,14 % del total de ingresos analizados durante el periodo. Le siguió San Juan, con 220.407 ingresos (12,17 %). En conjunto, estas dos parroquias concentraron aproximadamente el 26,30 % de toda la demanda hospitalaria registrada en establecimientos del Distrito Metropolitano de Quito.

Un segundo grupo estuvo conformado por Rumipamba, con 141.473 ingresos; Belisario Quevedo, con 141.412; y San Bartolo, con 132.981. Posteriormente se ubicaron Cotocollao, Chilibulo, Mariscal Sucre, Calderón (Carapungo) y Guamaní. En conjunto, las diez parroquias representadas en la figura concentraron aproximadamente el 74,12 % de los ingresos hospitalarios considerados en el análisis, lo que evidencia que la actividad hospitalaria se encuentra territorialmente concentrada en un conjunto relativamente reducido de parroquias.

Es importante señalar que la concentración territorial no presenta la misma composición sectorial, y esto lo podemos notar en Itchimbía, por ejemplo, 243.361 de los 256.089 ingresos correspondieron al sector público, lo que representa aproximadamente el 95,03 % de la demanda registrada en esa parroquia. Una estructura similar se observa en San Juan, donde el sector público concentró cerca del 85,25 % de los ingresos, y en San Bartolo, donde alcanzó aproximadamente el 96,60 %.

Por el contrario, Rumipamba y Belisario Quevedo presentan una estructura marcadamente diferente. En Rumipamba, los establecimientos privados con fines de lucro concentraron aproximadamente el 63,77 % de los ingresos registrados, mientras que el restante 36,23 % correspondió a establecimientos privados sin fines de lucro. En la información representada para esta parroquia no se registró participación del sector público. En Belisario Quevedo, aproximadamente el 74,30 % de los ingresos correspondió a establecimientos privados con fines de lucro y el 25,70 % al sector público.

Por último, en la Mariscal Sucre se observó una distribución más mixta: aproximadamente el 63,31 % de los ingresos correspondió al sector público y el 36,69 % a establecimientos privados con fines de lucro.

Esta concentración tiene además una implicación importante para los análisis posteriores de la investigación. El comportamiento agregado de la demanda hospitalaria del Distrito Metropolitano de Quito combina establecimientos con estructuras institucionales considerablemente diferentes. Mientras determinadas parroquias presentan una concentración casi exclusiva de actividad pública, otras poseen una participación predominantemente privada. En consecuencia, el análisis podría ocultar diferencias relevantes en la utilización de

la capacidad hospitalaria entre sectores. Por este motivo, los análisis posteriores de ocupación y presión hospitalaria relativa a la capacidad mantienen la diferenciación entre el sector público, privado con fines de lucro y privado sin fines de lucro.

4.1.3. Duración de la estancia hospitalaria

La duración de la estancia constituye una característica especialmente relevante para el análisis posterior de utilización de la capacidad hospitalaria, debido a que dos periodos con un número similar de ingresos pueden generar niveles diferentes de ocupación dependiendo del tiempo durante el cual los pacientes permanezcan hospitalizados.

En los registros analizados, la duración promedio de la hospitalización fue de 5,39 días, mientras que la mediana fue de 3 días. El 75 % de las hospitalizaciones presentó una estancia de hasta 5 días y el 90 % de hasta 11 días. Por su parte, el percentil 95 se ubicó en 17 días y el percentil 99 en 42 días.

Tabla 11

Estadísticos de la duración de la estancia hospitalaria

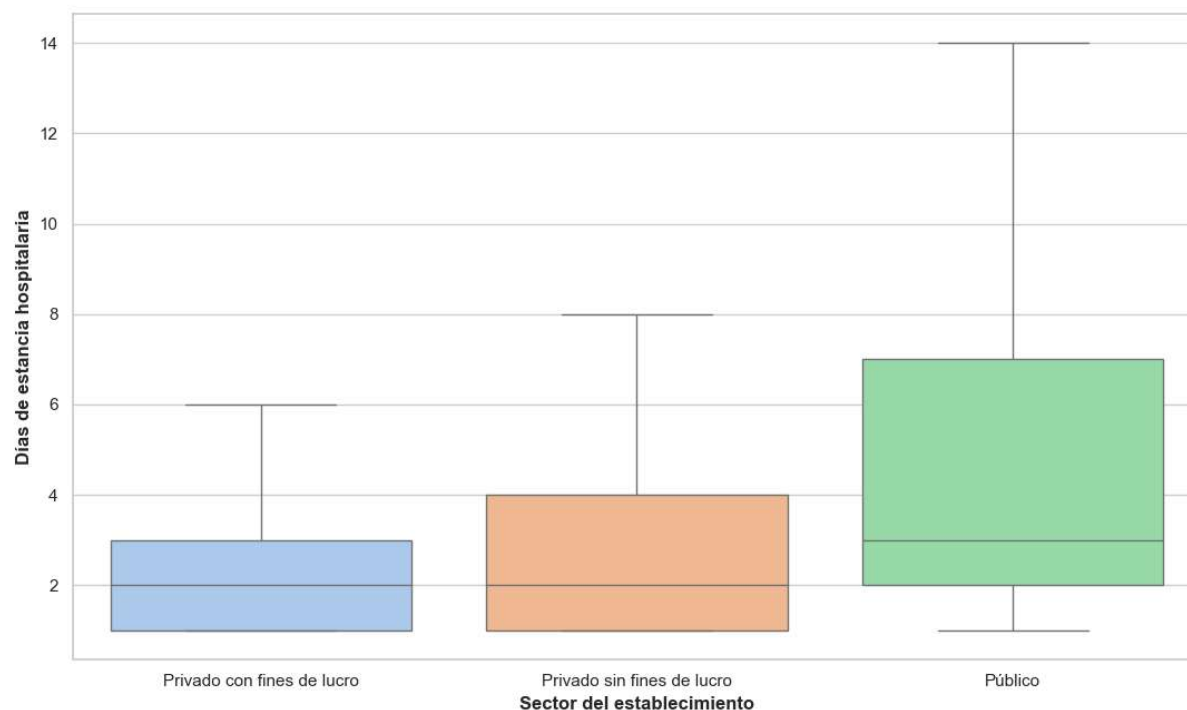
Indicador	Días
Media	5,39
Mediana	3
Percentil 75	5
Percentil 90	11
Percentil 95	17
Percentil 99	42

Nota. Elaboración propia las hospitalizaciones con ingreso y egreso durante la misma fecha fueron contabilizadas como un día de estancia.

Este resultado refuerza la necesidad de diferenciar entre demanda y ocupación hospitalaria. El número de nuevos ingresos permite caracterizar el flujo de pacientes que accede a los establecimientos, pero no refleja por sí solo la utilización acumulada de las camas, debido a que esta depende también del tiempo de permanencia de los pacientes.

Figura 3

Distribución de la duración de la estancia hospitalaria según sector del establecimiento en el Distrito Metropolitano de Quito, 2017–2024



Nota. Para facilitar la visualización y comparación entre sectores, se excluyen de la representación gráfica los valores extremos.

4.1.4. Síntesis de la caracterización hospitalaria

La caracterización inicial evidencia que los registros hospitalarios analizados corresponden a una población heterogénea en términos de edad, con una mayor participación de mujeres y un predominio de atención registrada en establecimientos del sector público. Asimismo, la duración de las hospitalizaciones presenta una distribución asimétrica, en la que la mayoría de las estancias se concentra en pocos días, mientras que una proporción reducida alcanza periodos considerablemente más prolongados.

Estos resultados son relevantes para las siguientes etapas del análisis. Por una parte, el volumen de nuevos ingresos permite estudiar la evolución de la demanda hospitalaria; por otra, la duración de las estancias y la disponibilidad de camas intervienen en la utilización de la capacidad hospitalaria. Esta diferenciación permite analizar posteriormente la demanda y la ocupación como dimensiones relacionadas, pero no equivalentes.

4.2. Evolución de la demanda hospitalaria 2017–2024

Para analizar el comportamiento temporal de la demanda hospitalaria se utilizaron los registros cuya fecha de ingreso se encontraba comprendida entre el 1 de enero de 2017 y el 31 de diciembre de 2024. En total, se identificaron 1.811.587 ingresos hospitalarios, a partir de los cuales se construyó una serie temporal diaria continua para todo el periodo de estudio.

El análisis permitió identificar cambios importantes en la demanda a lo largo de los ocho años, incluyendo un crecimiento previo a 2020, una ruptura pronunciada durante ese año y una recuperación progresiva en los años posteriores. Adicionalmente, la evaluación de la frecuencia diaria permitió observar patrones recurrentes de corto plazo que posteriormente resultaron relevantes para el desarrollo de los modelos predictivos.

4.2.1. Evolución temporal de la demanda hospitalaria

El análisis de los ingresos hospitalarios durante el periodo 2017–2024 evidencia cambios importantes en la evolución de la demanda hospitalaria del Distrito Metropolitano de Quito. A nivel anual, se identificó un crecimiento durante los años previos a la pandemia, seguido de una reducción considerable en 2020 y de un proceso progresivo de recuperación durante los años posteriores.

Tabla 12

Evolución anual de los ingresos hospitalarios en el Distrito Metropolitano de Quito, 2017–2024

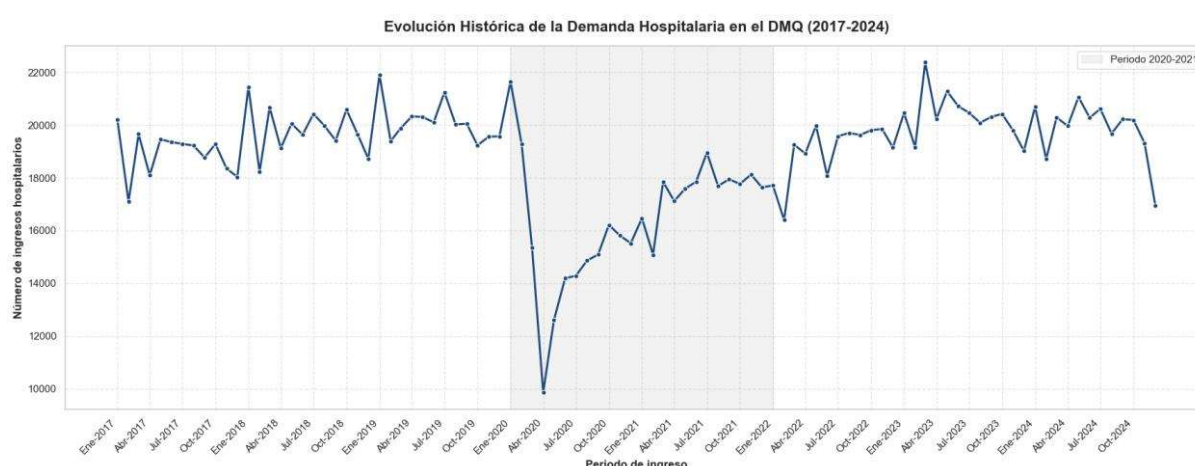
Año	Ingresos hospitalarios	Variación anual
2017	226.822	—
2018	237.947	4,90 %
2019	241.623	1,54 %
2020	184.706	-23,56 %
2021	210.085	13,74 %
2022	228.057	8,55 %
2023	244.344	7,14 %
2024	238.003	-2,60 %
Total	1.811.587	—

Nota. Elaboración propia a partir de datos del INEC (2017–2024). La variación corresponde al cambio respecto del año anterior.

La evolución anual muestra que entre 2017 y 2019 los ingresos aumentaron de 226.822 a 241.623, equivalente a un crecimiento acumulado aproximado del 6,53 %. Esta trayectoria cambia notablemente en 2020, cuando se registraron 184.706 ingresos, una reducción de 23,56 % respecto de 2019 y la mayor caída interanual de todo el periodo analizado.

Figura 4

Evolución mensual de la demanda hospitalaria en el Distrito Metropolitano de Quito, 2017–2024



Nota. Elaboración propia a partir de datos del INEC (2017–2024). Los valores corresponden al total mensual de ingresos hospitalarios.

La desagregación mensual permite observar que durante 2017–2019 la demanda se mantuvo en niveles relativamente estables, aunque con fluctuaciones entre meses. La mayor parte de los valores se concentra aproximadamente entre 18.000 y 21.000 ingresos mensuales, sin evidenciarse durante estos años una modificación persistente en el nivel general de la serie.

Esta dinámica cambia abruptamente durante los primeros meses de 2020. El gráfico muestra una reducción pronunciada de los ingresos hospitalarios, alcanzándose el punto más bajo de toda la serie durante este periodo. La caída observada explica en gran medida la reducción anual de 23,56 % registrada en 2020 y evidencia que esta no correspondió a una disminución uniforme durante todo el año, sino a una alteración particularmente intensa a partir del inicio de la emergencia sanitaria.

Posteriormente se observa una recuperación progresiva. Durante la segunda mitad de 2020 los ingresos comienzan a incrementarse, aunque todavía permanecen por debajo de los niveles habituales observados antes de la pandemia. Esta recuperación continúa durante 2021 y, hacia 2022, la demanda mensual retorna a valores comparables con los registrados entre 2017 y 2019.

Durante 2023 se alcanzaron nuevamente niveles elevados de demanda, consistentes con el hecho de que este año presentó el mayor número anual de ingresos del periodo analizado, con 244.344 registros. Durante buena parte de 2024 también se mantienen niveles elevados, aunque hacia el final del año se observa una reducción marcada. Esta disminución se debe a que la fuente corresponde a registros de egresos hospitalarios y algunas hospitalizaciones iniciadas durante los últimos días de 2024 y cuyo egreso ocurrió durante 2025 pueden no encontrarse representadas en la información disponible.

4.2.2. Caracterización de la serie diaria de demanda

Con el propósito de analizar la demanda hospitalaria con una mayor resolución temporal, los registros individuales fueron agregados según la fecha de ingreso. Como resultado, se obtuvo una serie temporal compuesta por 2.922 observaciones diarias consecutivas, correspondientes al periodo comprendido entre el 1 de enero de 2017 y el 31 de diciembre de 2024. La serie no presentó fechas faltantes ni duplicadas y acumuló un total de 1.811.587 ingresos hospitalarios, coincidente con el número de registros considerados para el análisis de demanda.

En términos generales, durante el periodo analizado se registró un promedio de 619,98 ingresos hospitalarios por día, con una mediana de 651 ingresos diarios. El 75 % de los días presentó hasta 738 ingresos, mientras que los percentiles 90 y 95 alcanzaron 783 y 808,95 ingresos, respectivamente. El máximo de la serie fue de 891 ingresos en un día, mientras que el valor mínimo fue de 17 ingresos.

La comparación anual de los estadísticos diarios evidencia diferencias relevantes en el comportamiento de la demanda. Durante el periodo previo a 2020, el promedio diario aumentó progresivamente desde 621,43 ingresos en 2017 hasta 661,98 en 2019. Esta

tendencia se interrumpió en 2020, cuando el promedio disminuyó a 504,66 ingresos diarios, acompañado de una reducción de la mediana hasta 491,5 ingresos. En 2021 se observó una recuperación parcial, con 575,58 ingresos diarios en promedio, mientras que en 2022 el indicador ascendió a 624,81.

Para 2023 se alcanzó el mayor promedio diario del periodo analizado, con 669,44 ingresos, además de una mediana de 716 ingresos. En 2024 el promedio se situó en 650,28 ingresos diarios, manteniéndose en niveles próximos a los observados antes de la pandemia. Sin embargo, el valor mínimo anual descendió hasta 17 ingresos, considerablemente inferior a los mínimos registrados en los años anteriores.

Tabla 13

Estadísticos descriptivos de la demanda hospitalaria según año en el Distrito Metropolitano de Quito, 2017–2024

Año	Días	Total de ingresos	Promedio diario	Mediana diaria	Mínimo	Máximo
2017	365	226.822	621,43	669,0	311	834
2018	365	237.947	651,91	707,0	336	865
2019	365	241.623	661,98	711,0	305	891
2020	366	184.706	504,66	491,5	221	874
2021	365	210.085	575,58	612,0	306	761
2022	365	228.057	624,81	655,0	325	838
2023	365	244.344	669,44	716,0	305	872
2024	366	238.003	650,28	697,0	17	858

Nota. Elaboración propia a partir de datos del INEC (2017–2024). Los estadísticos se calcularon con los ingresos hospitalarios diarios.

El comportamiento observado en 2020 confirma, también a escala diaria, la alteración identificada previamente en la evolución mensual. La disminución del promedio y de la mediana indica que durante este año se produjo un cambio generalizado en el nivel de la demanda hospitalaria y no únicamente una reducción concentrada en unos pocos meses o días. Posteriormente, la recuperación progresiva registrada entre 2021 y 2023 permitió retornar e incluso superar los niveles diarios observados antes de la pandemia.

El valor mínimo registrado en 2024 requiere una consideración particular. Durante los últimos días de diciembre se observa una reducción progresiva de los registros, desde 593

ingresos el 23 de diciembre hasta 17 ingresos el 31 de diciembre. Debido a que la fuente utilizada corresponde a registros de egresos hospitalarios, esta disminución puede estar influenciada por hospitalizaciones iniciadas al final de 2024 cuyo egreso ocurrió posteriormente, por lo que el extremo final de la serie debe interpretarse con cautela.

En conjunto, la caracterización diaria evidencia que la demanda hospitalaria presenta una variabilidad importante a lo largo del periodo estudiado y que su comportamiento se modifica de manera sustancial ante eventos extraordinarios. Esta característica resulta especialmente relevante para la evaluación de los modelos predictivos, ya que permite analizar su capacidad no solamente frente a periodos relativamente estables, sino también ante cambios importantes en el nivel histórico de la demanda.

4.3. Evaluación de los modelos predictivos

La evaluación de la capacidad predictiva de la demanda hospitalaria se realizó comparando cuatro alternativas: Naive $t-1$, Seasonal Naive $t-7$, LSTM con una ventana retrospectiva de 28 días y Prophet. Los resultados corresponden a los cuatro periodos de validación temporal comprendidos entre 2020 y 2023, manteniéndose el año 2024 completamente separado durante esta etapa para su posterior evaluación independiente. El criterio principal de comparación fue el menor MAE promedio obtenido en los cuatro periodos de validación.

4.3.1. Evaluación de los modelos predictivos

Los modelos de referencia permitieron establecer un punto de comparación mínimo antes de evaluar las alternativas de mayor complejidad. Los resultados evidenciaron una diferencia consistente entre utilizar únicamente el valor del día anterior y considerar el comportamiento observado una semana antes. El modelo Naive $t-1$ obtuvo un MAE promedio de 86,37 ingresos diarios, un RMSE de 119,83 y un MAPE de 14,97 %. En contraste, el modelo Seasonal Naive $t-7$ redujo estos valores a un MAE de 50,07, RMSE de 78,02 y MAPE de 9,11 %. Además, Seasonal Naive presentó un menor error que Naive $t-1$ en cada uno de los cuatro años de validación.

Este resultado evidencia que la demanda hospitalaria contiene una dependencia temporal de carácter semanal importante. Utilizar como referencia el número de ingresos

registrado siete días antes resultó considerablemente más efectivo que asumir que la demanda del día siguiente sería similar a la observada inmediatamente el día anterior. La estabilidad de este comportamiento también resulta relevante. El MAE de Seasonal Naive osciló entre 45,89 y 54,69, mientras que el error del modelo Naive aumentó progresivamente desde 71,64 en 2020 hasta 98,83 en 2023.

Por tanto, Seasonal Naive $t-7$ constituyó una referencia particularmente exigente para evaluar si los modelos LSTM y Prophet producían una mejora real sobre una estrategia predictiva sencilla.

4.3.2. Resultados del modelo LSTM

Durante la evaluación de LSTM se compararon ventanas retrospectivas de 7, 14, 21 y 28 días. Considerando el MAE promedio de los cuatro periodos de validación, la configuración con 28 días presentó el mejor desempeño dentro de las alternativas LSTM, con un MAE promedio de 61,57, RMSE de 79,04 y MAPE de 12,49 %.

Sin embargo, el comportamiento de esta configuración fue marcadamente diferente entre los años evaluados. Durante 2020, LSTM-28 alcanzó un MAE de 111,04, considerablemente superior al MAE de 49,04 obtenido por Seasonal Naive. Por el contrario, en 2021 el error de LSTM disminuyó hasta 40,91, en 2022 fue de 50,96 y en 2023 descendió nuevamente a 43,38. Este contraste muestra que el desempeño global de LSTM estuvo fuertemente condicionado por el primer periodo de validación. En 2020 el modelo había sido entrenado únicamente con información de 2017–2019, por lo que tuvo que realizar predicciones durante un año cuyo comportamiento presentó una ruptura considerable respecto de la dinámica histórica previa.

A partir de 2021, una vez que el proceso de entrenamiento incorporó progresivamente información correspondiente al periodo de alteración hospitalaria, el comportamiento del modelo mejoró de manera sustancial. De hecho, LSTM-28 obtuvo el menor MAE en 2021 y 2023, y también el menor RMSE y MAPE en 2021, 2022 y 2023. Es así que, los resultados de LSTM no muestran un desempeño uniformemente inferior al modelo estacional. Más bien evidencian una diferencia entre su capacidad de respuesta ante una ruptura no observada previamente, como la de 2020, y su desempeño durante los periodos posteriores.

4.3.3. Resultados del modelo Prophet

Prophet presentó un comportamiento similar en cuanto a la diferencia entre 2020 y los años posteriores. Durante el primer periodo de validación obtuvo un MAE de 95,84, RMSE de 125,62 y MAPE de 23,02 %. En 2021 el MAE disminuyó a 44,82, en 2022 fue de 49,89 y en 2023 alcanzó 45,97.

En términos globales, Prophet alcanzó un MAE promedio de 59,13, RMSE de 81,21 y MAPE de 11,98 %. Esto lo ubicó como el modelo no-baseline con menor MAE promedio durante el periodo completo 2020–2023, superando en esta métrica a LSTM-28, cuyo MAE promedio fue de 61,57.

No obstante, al igual que LSTM, Prophet no logró superar globalmente al Seasonal Naive. Su MAE promedio fue aproximadamente un 18,09 % superior al del modelo estacional de referencia. El comportamiento anual muestra nuevamente que esta conclusión está determinada en gran medida por 2020. Prophet superó a Seasonal Naive en MAE durante 2021, 2022 y 2023, pero presentó un deterioro considerable durante la ruptura observada en 2020.

4.3.4. Comparación de resultados por periodo de validación

La siguiente tabla resume los resultados de los cuatro modelos en cada periodo de validación.

Tabla 14

Desempeño de los modelos predictivos según periodo de validación, 2020–2023

Año	Modelo	MAE	RMSE	MAPE (%)
2020	Naive (t-1)	71,64	101,21	14,36
2020	Seasonal Naive (t-7)	49,04	83,74	10,39
2020	LSTM-28	111,04	130,43	26,82
2020	Prophet	95,84	125,62	23,02
2021	Naive (t-1)	84,02	115,88	15,07
2021	Seasonal Naive (t-7)	45,89	69,56	8,53
2021	LSTM-28	40,91	58,54	7,78
2021	Prophet	44,82	65,16	8,61
2022	Naive (t-1)	90,98	126,42	15,07
2022	Seasonal Naive (t-7)	54,69	82,69	9,38
2022	LSTM-28	50,96	67,83	8,54
2022	Prophet	49,89	70,75	8,65

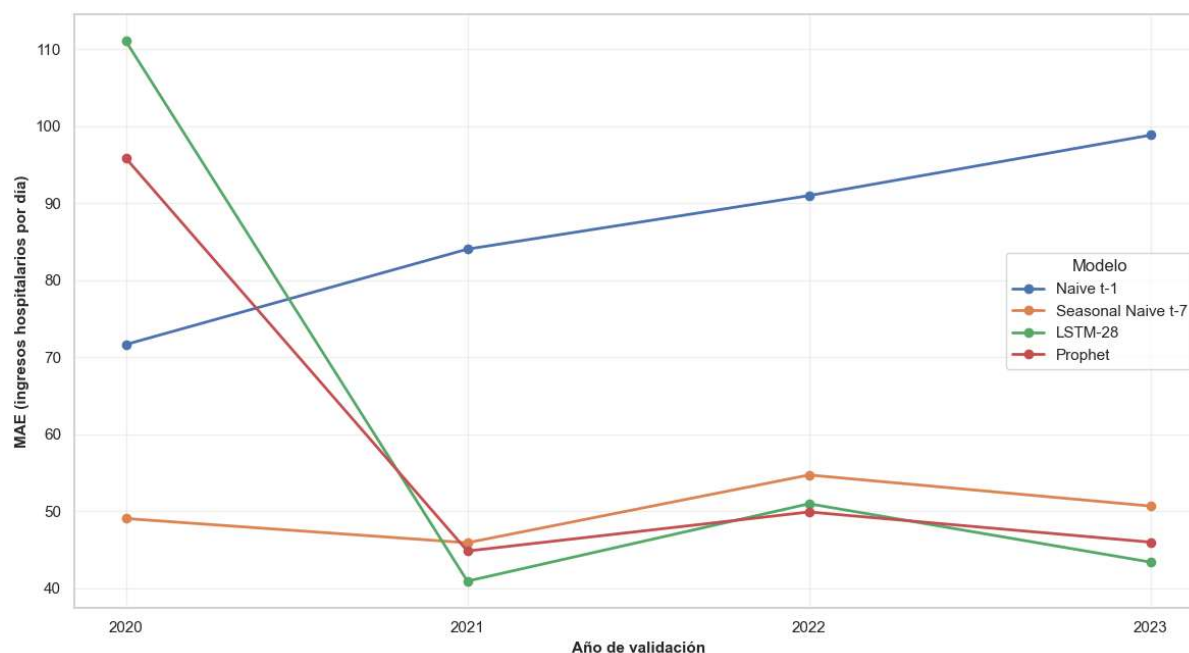
2023	Naive (t-1)	98,83	135,80	15,36
2023	Seasonal Naive (t-7)	50,66	76,10	8,15
2023	LSTM-28	43,38	59,36	6,84
2023	Prophet	45,97	63,32	7,66

Los resultados muestran que ningún modelo dominó todos los periodos de validación. En 2020, Seasonal Naive presentó el menor error en las tres métricas. En 2021, LSTM-28 obtuvo el mejor desempeño general. En 2022 Prophet registró el menor MAE, mientras que LSTM-28 presentó los menores RMSE y MAPE. Finalmente, en 2023 LSTM-28 volvió a obtener los menores valores en las tres métricas.

Esta variabilidad resulta más informativa que una comparación basada únicamente en promedios. Los modelos más complejos presentaron mejor capacidad predictiva durante tres de los cuatro periodos, pero su error aumentó sustancialmente ante la alteración extraordinaria de 2020. Seasonal Naive, por su parte, mostró una mayor robustez en dicho escenario.

Figura 5

Evolución del MAE de los modelos predictivos según periodo de validación, 2020–2023



Nota. Elaboración propia. Un menor MAE indica un mejor desempeño predictivo del modelo.

La figura compara el Error Absoluto Medio de Naive $t-1$, Seasonal Naive $t-7$, LSTM-28 y Prophet en cada uno de los cuatro periodos de validación temporal. Cabe mencionar que un menor MAE indica mayor precisión predictiva.

4.3.5. Desempeño promedio y selección del modelo

Al promediar los cuatro periodos de validación, el modelo con menor MAE fue Seasonal Naive $t-7$, con 50,07 ingresos diarios. Le siguieron Prophet con 59,13, LSTM-28 con 61,57 y Naive $t-1$ con 86,37.

Tabla 15

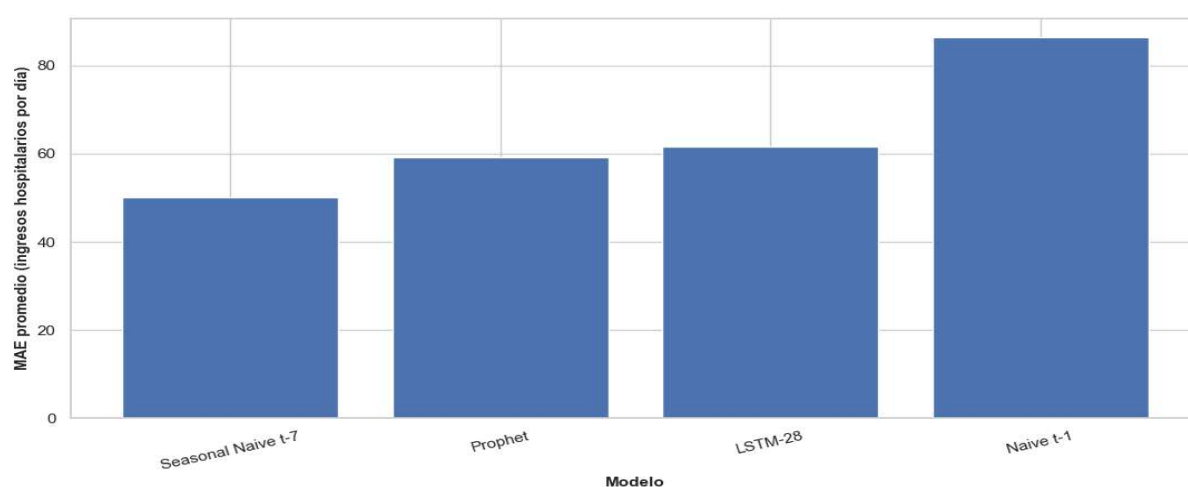
Desempeño promedio de los modelos predictivos durante la validación temporal 2020–2023

Modelo	MAE promedio	RMSE promedio	MAPE promedio
Seasonal Naive (t-7)	50,07	78,02	9,11 %
Prophet	59,13	81,21	11,98 %
LSTM-28	61,57	79,04	12,49 %
Naive (t-1)	86,37	119,83	14,97 %

Nota. Los valores corresponden al promedio de los resultados obtenidos en los cuatro periodos de validación temporal: 2020, 2021, 2022 y 2023. El MAE fue utilizado como criterio principal de selección.

Figura 6

Comparación del MAE promedio de los modelos predictivos durante la validación 2020–2023



Nota. Elaboración propia. Un menor MAE promedio indica un mejor desempeño durante la validación temporal.

De acuerdo con el criterio establecido, Seasonal Naive $t-7$ fue seleccionado como el modelo con mejor desempeño durante la etapa de desarrollo. Este resultado es relevante porque demuestra que una mayor complejidad algorítmica no garantiza necesariamente una mayor capacidad de generalización. En una serie con una estructura semanal marcada y una ruptura extraordinaria como la observada en 2020, una estrategia estacional sencilla presentó el menor error promedio global.

Esto no significa, sin embargo, que LSTM y Prophet carezcan de capacidad predictiva. La comparación por año muestra que las dos alternativas avanzadas superaron al modelo estacional en distintos periodos posteriores a 2020.

4.3.6. Análisis de sensibilidad 2021–2023

Debido al comportamiento excepcional observado durante 2020, se realizó adicionalmente una comparación considerando exclusivamente los periodos de validación 2021–2023. Este análisis no modificó el criterio ni la selección principal, sino que permitió examinar cuánto influía la ruptura de 2020 sobre la comparación global.

Tabla 16

Análisis de sensibilidad del desempeño predictivo durante el periodo 2021–2023

Modelo	MAE promedio	RMSE promedio	MAPE promedio
LSTM-28	45,08	61,91	7,72 %
Prophet	46,89	66,41	8,31 %
Seasonal Naive (t-7)	50,41	76,12	8,68 %
Naive (t-1)	91,28	126,03	15,17 %

Nota. Elaboración propia. MAE, RMSE y MAPE corresponden al desempeño promedio obtenido durante 2021–2023.

En este escenario, LSTM-28 pasó a ocupar el primer lugar, con un MAE promedio de 45,08, seguido por Prophet con 46,89 y Seasonal Naive con 50,41. Respecto del modelo estacional, LSTM redujo el MAE en 10,58 %, el RMSE en 18,67 % y el MAPE en 11,12 %. Prophet también presentó una mejora, reduciendo el MAE en 6,99 % y el RMSE en 12,75 %.

Este resultado confirma que la clasificación global de los modelos depende considerablemente del comportamiento registrado durante 2020. Cuando se analizan exclusivamente los años posteriores, los modelos de mayor complejidad presentan ventajas frente a la referencia estacional.

Sin embargo, no corresponde sustituir retrospectivamente el modelo seleccionado. La regla de selección había sido definida sobre los cuatro periodos 2020–2023 y, bajo dicho criterio, Seasonal Naive fue el ganador. La sensibilidad proporciona evidencia complementaria sobre el comportamiento de los modelos después de la ruptura, pero no constituye un nuevo proceso de selección. Esta decisión quedó expresamente fijada en la programación antes de evaluar 2024.

4.3.7. Síntesis de la evaluación y selección del modelo

La evaluación temporal permitió identificar dos resultados complementarios. En primer lugar, Seasonal Naive $t-7$ presentó el menor error promedio global durante 2020–2023, mostrando especial robustez frente a la alteración registrada en 2020. En segundo lugar, LSTM y Prophet mostraron una mejora considerable en los años posteriores, llegando a superar al modelo estacional durante 2021–2023.

En consecuencia, los resultados no sustentan la idea de que un único modelo sea superior bajo todas las condiciones temporales. La predicción de la demanda hospitalaria mostró sensibilidad a cambios en la dinámica histórica de la serie: el modelo estacional sencillo presentó mayor robustez ante una ruptura abrupta, mientras que LSTM y Prophet alcanzaron mejores resultados durante periodos posteriores caracterizados por una dinámica más cercana a la información incorporada progresivamente en su entrenamiento.

Con base en el criterio previamente definido, Seasonal Naive $t-7$ quedó seleccionado al finalizar la etapa de desarrollo, mientras que las configuraciones de los cuatro modelos permanecieron congeladas para su evaluación sobre el conjunto independiente de 2024.

4.4. Evaluación independiente en 2024

Una vez finalizado el proceso de desarrollo y selección de los modelos, el año 2024 se utilizó como conjunto independiente de prueba, sin haber intervenido previamente en la

selección de modelos. El periodo estuvo conformado por 366 observaciones diarias y permitió evaluar el comportamiento de Naive $t-1$, Seasonal Naive $t-7$, LSTM-28 y Prophet utilizando las configuraciones definidas durante las etapas anteriores. Antes de realizar esta evaluación, Seasonal Naive $t-7$ había sido seleccionado como el modelo de mejor desempeño durante la validación 2020–2023.

4.4.1. Resultados del test independiente

Los resultados obtenidos en 2024 muestran un comportamiento diferente al observado durante la etapa de validación. LSTM-28 presentó el menor error en las tres métricas evaluadas, con un MAE de 39,77 ingresos diarios, un RMSE de 61,26 y un MAPE de 13,86 %. Prophet ocupó el segundo lugar según MAE y RMSE, mientras que Seasonal Naive $t-7$ se ubicó en tercer lugar.

Tabla 17

Desempeño de los modelos predictivos en el conjunto independiente de prueba de 2024

Modelo	MAE	RMSE	MAPE
LSTM-28	39,77	61,26	13,86 %
Prophet	42,32	74,68	18,39 %
Seasonal Naive (t-7)	54,64	84,97	15,74 %
Naive (t-1)	93,11	128,32	17,59 %

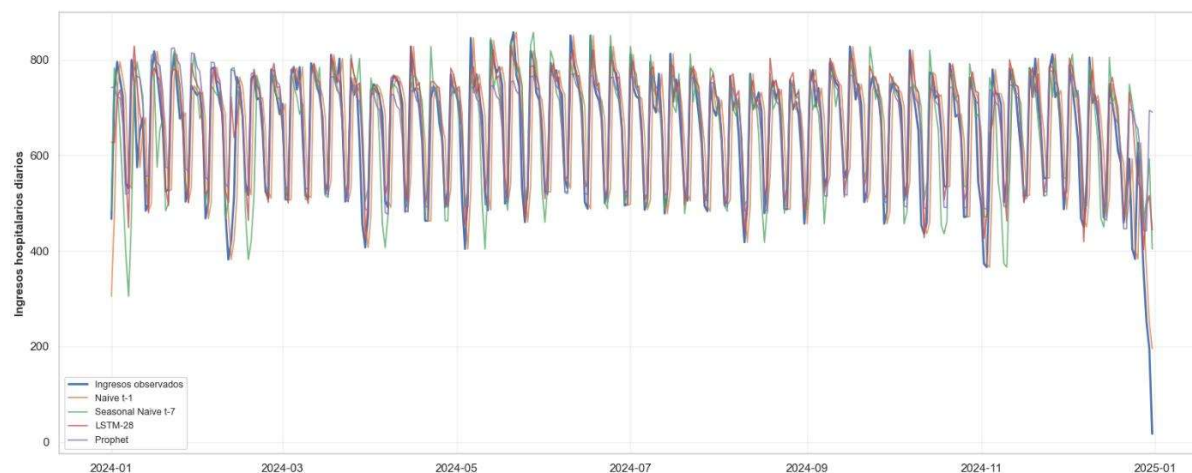
Nota. Elaboración propia. Las métricas corresponden a la evaluación independiente realizada con los datos de 2024.

El desempeño de LSTM-28 resulta especialmente relevante porque durante la validación 2020–2023 no había presentado el menor MAE promedio global. Sin embargo, su comportamiento favorable en 2024 es coherente con los resultados obtenidos en los años posteriores a 2020, donde había mostrado una mayor capacidad predictiva.

Este resultado no modifica retrospectivamente la selección realizada en la etapa de desarrollo. Seasonal Naive $t-7$ continúa siendo el modelo seleccionado bajo el criterio previamente establecido sobre 2020–2023; LSTM-28 se identifica únicamente como el modelo con mejor desempeño observado en el conjunto independiente de 2024. Esta distinción evita utilizar el conjunto de prueba como un nuevo mecanismo de selección.

Figura 7

Comparación entre los ingresos hospitalarios observados y las predicciones de los modelos durante 2024



Nota. Elaboración propia. Las líneas comparan los ingresos observados con las predicciones generadas para 2024.

La comparación gráfica permite observar que los modelos reproducen buena parte de las fluctuaciones de la demanda a lo largo del año, aunque las diferencias se incrementan hacia el cierre de 2024.

4.4.2. Comportamiento del error al cierre de 2024

Para examinar con mayor detalle este comportamiento se realizó un análisis descriptivo diferenciando enero–noviembre y diciembre. Entre enero y noviembre, LSTM-28 alcanzó un MAE de 35,15, seguido por Prophet con 36,87 y Seasonal Naive con 50,79. En diciembre, el error aumentó en los cuatro modelos y LSTM-28, aunque continuó presentando el menor MAE, alcanzó 89,76.

Tabla 18

Comparación del MAE de los modelos predictivos entre enero–noviembre y diciembre de 2024

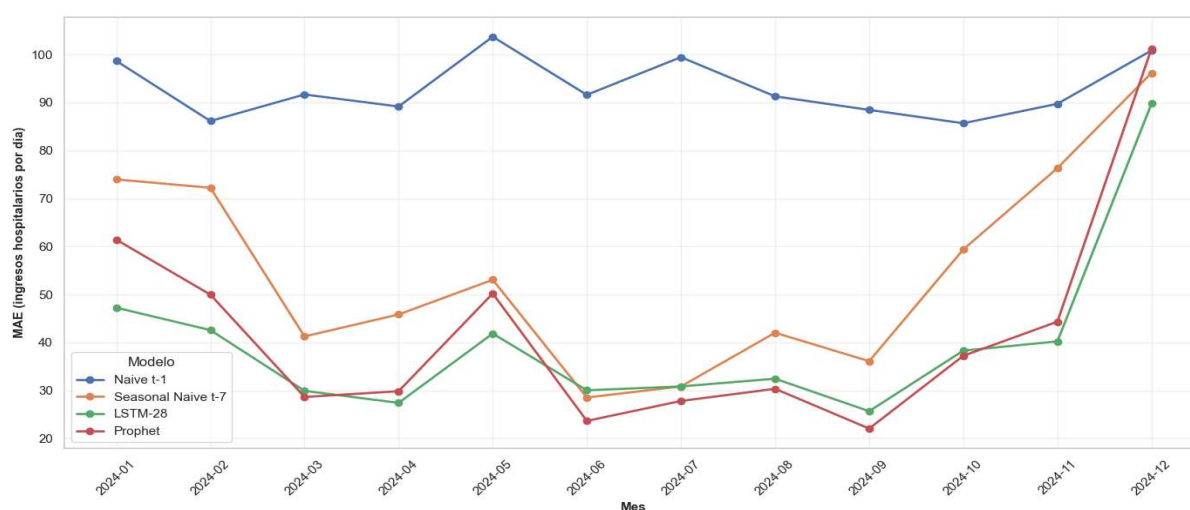
Periodo	Naive (t-1)	Seasonal Naive (t-7)	LSTM-28	Prophet
Enero–noviembre	92,39	50,79	35,15	36,87
Diciembre	100,84	96,16	89,76	101,22

Nota. Elaboración propia. El MAE se presenta por periodo para comparar el desempeño de los modelos durante 2024.

El deterioro de diciembre es particularmente visible en los modelos LSTM, Prophet y Seasonal Naive. En LSTM, por ejemplo, el MAE aumentó aproximadamente 155,38 % respecto al obtenido entre enero y noviembre.

Figura 8

Evolución mensual del MAE de los modelos predictivos durante el test independiente de 2024



Nota. Elaboración propia. El incremento del MAE en diciembre puede estar influido por la censura a la derecha de los registros disponibles.

El incremento del error durante diciembre debe interpretarse considerando las características de la fuente utilizada. Al provenir de registros de egresos hospitalarios, algunas hospitalizaciones iniciadas al final de 2024 cuyo egreso ocurrió durante 2025 pueden no estar representadas en la información disponible. Por tanto, el cierre de la serie puede encontrarse afectado por censura a la derecha. La evaluación oficial mantuvo las 366 observaciones sin eliminar, recortar o imputar los valores correspondientes al final del año.

4.5.Capacidad y ocupación hospitalaria 2017–2024

El análisis de la capacidad hospitalaria se realizó a partir del Registro Estadístico de Camas Hospitalarias para los establecimientos ubicados en el Distrito Metropolitano de Quito. Para cada año se consideraron la dotación normal de camas, las camas disponibles, los días-paciente y los días-cama disponibles. La ocupación hospitalaria agregada se calculó

como la razón entre la suma de días-paciente y la suma de días-cama disponibles de los registros válidos.

4.5.1. Evolución de la capacidad y ocupación hospitalaria

La siguiente tabla muestra un resumen de los principales indicadores de capacidad y utilización hospitalaria durante el periodo 2017–2024.

Tabla 19

Capacidad y ocupación hospitalaria en el Distrito Metropolitano de Quito, 2017–2024

Año	Dotación normal de camas	Camas disponibles	Días-paciente	Días-cama disponibles	Ocupación (%)
2017	5.568	5.212	1.339.916	1.901.638	70,46
2018	5.795	5.520	1.415.277	2.014.350	70,26
2019	5.652	5.552	1.347.245	2.023.162	66,59
2020	6.010	5.194	1.203.565	2.091.336	57,55
2021	5.983	5.048	1.275.435	1.862.273	68,49
2022	5.602	5.126	1.232.582	1.885.970	65,36
2023	5.718	5.318	1.374.226	2.060.387	66,70
2024	5.691	5.335	1.387.330	2.023.265	68,57

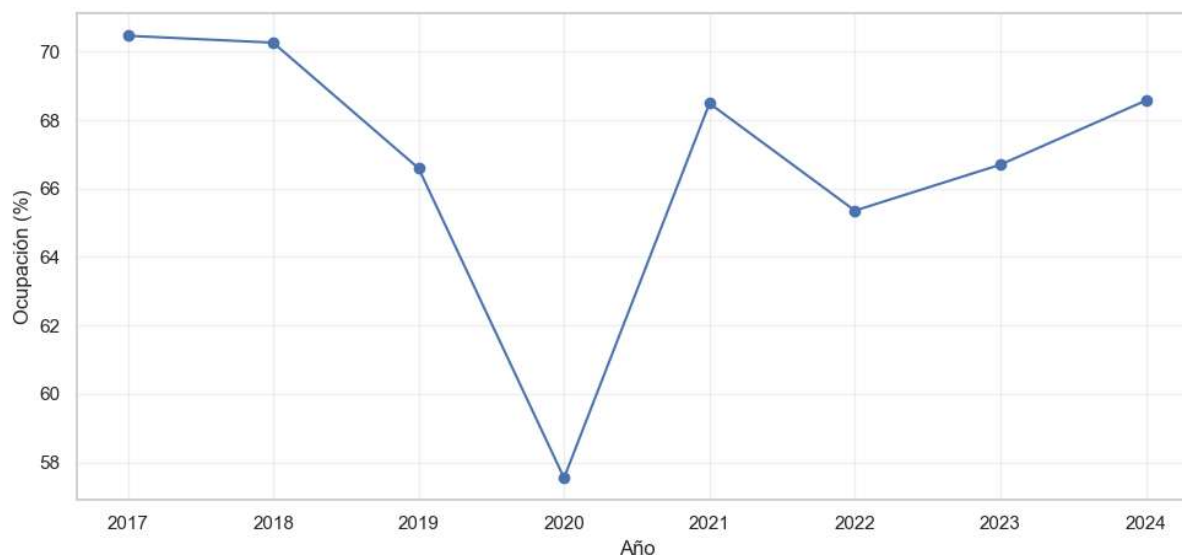
Nota. La ocupación agregada corresponde a la razón entre los días-paciente y los días-cama disponibles de los registros válidos.

La ocupación agregada se mantuvo alrededor del 70 % en 2017 y 2018, y descendió a 66,59 % en 2019. El principal cambio se produjo en 2020, cuando la ocupación cayó hasta 57,55 %, el valor más bajo del periodo. Esta reducción estuvo relacionada principalmente con la disminución de los días-paciente, que pasaron de 1,35 millones en 2019 a 1,20 millones en 2020, en concordancia con la reducción de la actividad hospitalaria observada previamente en los registros de demanda.

A partir de 2021 se produjo una recuperación de la ocupación, que alcanzó 68,49 %. En 2022 volvió a disminuir ligeramente hasta 65,36 %, mientras que en 2023 y 2024 ascendió a 66,70 % y 68,57 %, respectivamente. De esta forma, al final del periodo la ocupación agregada se aproximó nuevamente a los niveles observados al inicio de la serie.

Figura 9

Evolución de la ocupación hospitalaria agregada en el Distrito Metropolitano de Quito, 2017–2024



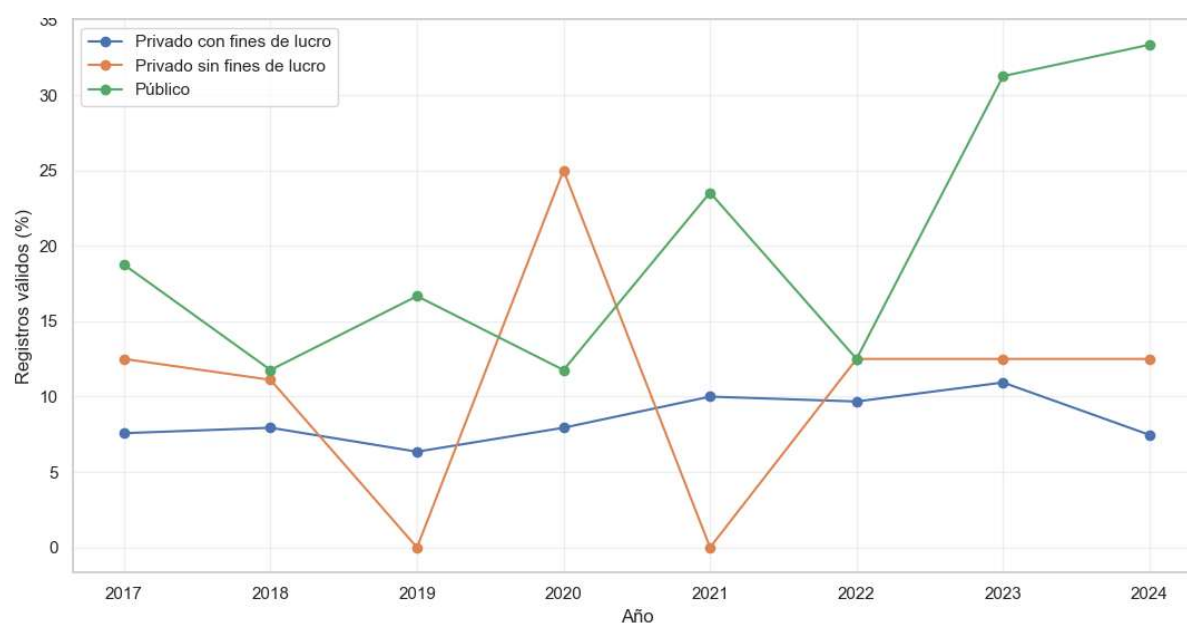
Nota. La figura propia presenta la ocupación hospitalaria agregada calculada a partir de la relación entre días-paciente y días-cama disponibles para los establecimientos de salud del Distrito Metropolitano de Quito.

4.5.2. Diferencias entre establecimientos y sectores

Aunque la ocupación agregada del DMQ se mantuvo por debajo del 85 % durante todos los años, este resultado no implica que todos los establecimientos presentaran niveles similares de utilización. En cada año se identificaron registros individuales con ocupaciones iguales o superiores al 85 % y también casos superiores al 100 %. Estos últimos oscilaron entre 3 y 6 registros anuales, alcanzando el mayor número en 2022.

Figura 10

Alta ocupación hospitalaria según sector del establecimiento, 2017–2024



Nota. La figura propia presenta la frecuencia de registros con niveles de ocupación iguales o superiores al 85 % según sector del establecimiento y año.

La coexistencia de una ocupación agregada moderada con establecimientos que superan el 100 % evidencia que el indicador global puede ocultar diferencias importantes dentro de la red hospitalaria. Por esta razón, los valores superiores al 100 % fueron conservados en el análisis y no tratados automáticamente como errores, siempre que sus componentes fueran válidos.

También se observaron diferencias según el sector del establecimiento. El sector público presentó la mayor ocupación agregada en siete de los ocho años analizados, con valores de 79,88 % en 2017, 79,89 % en 2018, 76,13 % en 2019, 77,98 % en 2021, 75,44 % en 2022, 81,77 % en 2023 y 83,37 % en 2024. La excepción fue 2020, cuando el sector privado sin fines de lucro alcanzó una ocupación agregada de 83,69 %.

En contraste, el sector privado con fines de lucro presentó generalmente los niveles agregados más bajos, incluyendo una reducción particularmente marcada hasta 34,28 % en 2020. Estas diferencias muestran que el comportamiento de la ocupación hospitalaria no fue

homogéneo entre sectores y justifican profundizar posteriormente en el análisis de la presión hospitalaria.

4.6. Presión hospitalaria diaria y episodios de alta ocupación

Para complementar el indicador anual de ocupación, se reconstruyó un proxy diario de presión hospitalaria para el periodo 2017–2023. Este indicador permitió analizar la utilización relativa de la capacidad con mayor detalle temporal y diferenciar el comportamiento entre los sectores público, privado con fines de lucro y privado sin fines de lucro.

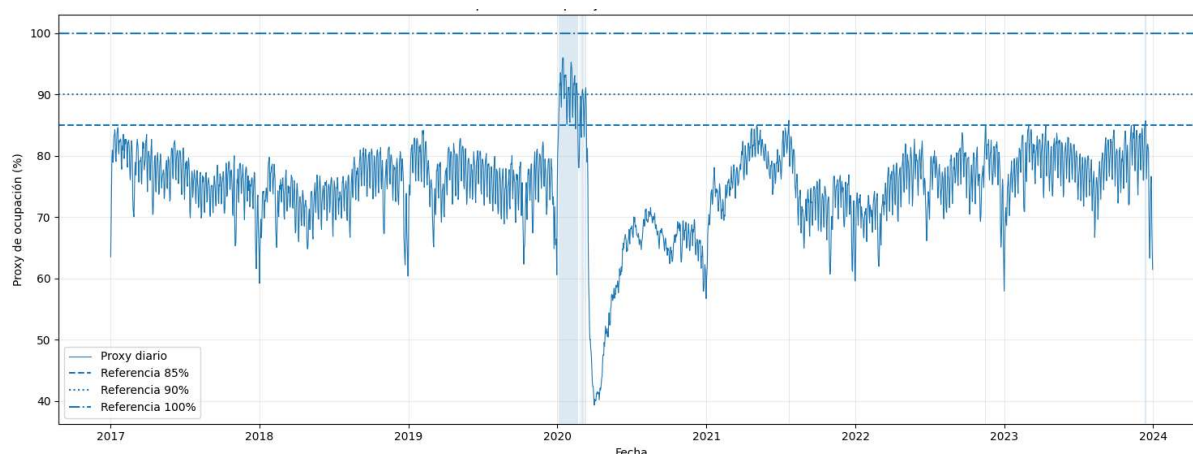
4.6.1. Comportamiento diario de la presión hospitalaria

A nivel agregado del Distrito Metropolitano de Quito, el proxy diario no alcanzó el umbral referencial del 85 % en ningún día del periodo analizado. Los valores máximos anuales oscilaron entre 69,34 % y 75,07 %, por lo que una lectura exclusivamente agregada sugeriría una presión relativamente moderada sobre la capacidad hospitalaria.

Sin embargo, este resultado cambia al desagregar la información por sector. El sector público concentró prácticamente todos los días de alta presión, mientras que el sector privado con fines de lucro permaneció por debajo del 85 % durante todo el periodo. La diferencia muestra que el indicador agregado del DMQ puede ocultar situaciones de presión localizadas en determinados componentes de la red hospitalaria.

Figura 11

Evolución diaria del proxy de presión hospitalaria en el sector público, 2017–2023



Nota. La figura propia presenta la evolución diaria del proxy de presión hospitalaria para el sector público y permite identificar los periodos en los que el indicador alcanzó o superó el umbral referencial del 85 %.

El comportamiento más destacado ocurrió en 2020. Durante ese año, el sector público presentó 59 días-sector con valores iguales o superiores al 85 %, equivalentes al 16,12 % de los días del año, y 35 días-sector iguales o superiores al 90 %. Su máximo diario alcanzó el 96,00 %. En los años posteriores los eventos fueron considerablemente menos frecuentes: dos días-sector en 2021, uno en 2022 y tres en 2023.

En el sector privado con fines de lucro no se identificaron días por encima del 85 %. En el sector privado sin fines de lucro se registró únicamente un día en 2020, con un valor de 85,45 %. Este último resultado debe interpretarse con cautela debido a que la comparación entre los días-paciente reconstruidos y los registrados oficialmente presentó para ese sector y año una diferencia relativa de aproximadamente 30 %.

4.6.2. Episodios de alta ocupación

Para analizar la persistencia de estos eventos, se definió como episodio una secuencia de uno o más días consecutivos en los que el proxy sectorial alcanzó o superó el 85 %. En total se identificaron 66 observaciones diarias sectoriales por encima del 85 %, agrupadas en 9 episodios. De estas observaciones, 35 alcanzaron o superaron el 90 %, mientras que ninguna llegó al 100 %.

La mayor parte de los episodios correspondió al sector público. El evento de mayor duración ocurrió entre el 5 de enero y el 14 de febrero de 2020, con 41 días consecutivos por encima del 85 %. Durante este episodio, el proxy presentó un promedio de 91,06 %, alcanzó un máximo de 96,00 % y permaneció 27 días por encima del 90 %.

Además de este episodio principal, durante 2020 se identificaron otros cuatro episodios en el sector público, con duraciones de 5, 2, 6 y 5 días. Los episodios registrados durante 2021, 2022 y 2023 fueron considerablemente más breves, con duraciones máximas de 2, 1 y 3 días, respectivamente.

Tabla 20

Episodios de alta presión hospitalaria identificados mediante el proxy diario, 2017–2023

Sector y año	Número de episodios	Días en episodios ≥ 85 %	Duración máxima	Máximo observado
Público, 2020	5	59	41 días	96,00 %
Privado sin fines de lucro, 2020	1	1	1 día	85,45 %
Público, 2021	1	2	2 días	85,78 %
Público, 2022	1	1	1 día	85,07 %
Público, 2023	1	3	3 días	85,72 %

Nota. Elaboración propia. Los episodios corresponden a días consecutivos con un proxy de presión hospitalaria igual o superior al 85 %.

Los resultados muestran que la presión hospitalaria diaria se concentró principalmente en el sector público y, particularmente, durante 2020. Asimismo, aunque se observaron episodios con niveles superiores al 90 %, ninguno de los valores diarios reconstruidos alcanzó o superó el 100 %. Esto diferencia los resultados del proxy diario de los indicadores anuales a nivel de establecimiento analizados anteriormente, donde sí se identificaron registros superiores al 100 %.

4.7.Resultados del modelo operacional de presión — Modelo A

El Modelo A fue desarrollado con un enfoque operacional y predictivo, utilizando información temporal disponible hasta el día anterior para estimar el proxy diario de presión hospitalaria por sector. Su desempeño fue evaluado mediante validación temporal para los años 2021, 2022 y 2023, comparando Regresión Lineal, Random Forest y XGBoost. Como referencia adicional se utilizó un baseline de persistencia $t-1$, que asume como predicción para el día actual el nivel de presión observado el día anterior.

Tabla 21

Desempeño promedio del Modelo A durante la validación temporal, 2021–2023

Método	MAE	RMSE	R²
XGBoost	1,5260	2,0278	0,9760
Random Forest	1,5929	2,1232	0,9738
Regresión Lineal	1,7635	2,3058	0,9687
Baseline persistencia (t-1)	2,6684	3,4614	0,9305

Nota. Elaboración propia. Las métricas corresponden al desempeño promedio del Modelo A durante la validación temporal de 2021–2023.

Los tres modelos evaluados superaron al baseline de persistencia en términos de MAE. Entre ellos, XGBoost presentó el mejor desempeño, con un MAE promedio de 1,5260 puntos porcentuales, seguido por Random Forest con 1,5929 y Regresión Lineal con 1,7635. XGBoost también alcanzó un R² promedio de 0,9760, mostrando una elevada capacidad para reproducir las variaciones del proxy de presión en los periodos evaluados.

Frente al baseline de persistencia, cuyo MAE fue de 2,6684 puntos porcentuales, XGBoost logró una reducción del error absoluto medio de 42,81 %, por lo que fue seleccionado como el algoritmo principal del Modelo A.

En términos prácticos, este resultado significa que utilizar únicamente la presión observada el día anterior proporciona una referencia predictiva relativamente fuerte; sin embargo, la incorporación conjunta de la información temporal considerada por XGBoost permitió estimar con mayor precisión el nivel de presión del día siguiente. Por tanto, el Modelo A aporta una mejora predictiva relevante frente a una estrategia basada exclusivamente en la persistencia reciente.

4.8. Factores asociados — Modelo B

El Modelo B fue planteado con un propósito exploratorio, orientado a analizar factores asociados al comportamiento de la presión hospitalaria sin utilizar rezagos del propio proxy como predictores. Se evaluaron Regresión Lineal, Random Forest y XGBoost mediante validación temporal, comparando su desempeño con un baseline basado en la media sectorial del periodo de entrenamiento.

Tabla 22

Desempeño promedio del Modelo B durante la validación temporal

Método	MAE	RMSE	R²
Baseline media sector	6,1621	7,5022	0,6462
Regresión Lineal	7,4343	8,7702	0,5009
XGBoost	7,8452	9,1769	0,4096
Random Forest	7,8550	9,3592	0,3550

Nota. Elaboración propia. Las métricas corresponden al desempeño promedio del Modelo B durante la validación temporal.

Entre los modelos evaluados, la Regresión Lineal presentó el menor MAE, con 7,4343 puntos porcentuales. Sin embargo, este valor fue superior al MAE de 6,1621 obtenido por el baseline sectorial, lo que representa un deterioro aproximado del 20,64 % respecto de esa referencia. En consecuencia, ninguno de los algoritmos del Modelo B demostró una mejora predictiva global frente a una estrategia sencilla basada únicamente en el nivel medio histórico de cada sector.

Por esta razón, el Modelo B no se interpreta como un modelo predictivo competitivo, sino como una herramienta exploratoria para estudiar asociaciones entre las características sectoriales, la demanda reciente, el calendario y la composición demográfica con el nivel de presión hospitalaria.

Como análisis complementario se evaluó el comportamiento posterior a 2020. En la validación de 2022, la Regresión Lineal obtuvo un MAE de 3,9348, frente a 5,2499 del baseline sectorial; para 2023, los valores fueron 5,4765 y 6,7678, respectivamente. Estos resultados indican que el desempeño del Modelo B mejoró cuando el análisis se restringió al periodo posterior a la alteración de 2020, aunque esta sensibilidad no modifica su carácter exploratorio ni los resultados principales.

Finalmente, el diagnóstico de correlación mostró una elevada redundancia entre algunos predictores del Modelo B. Se identificaron 30 pares de variables con correlaciones absolutas iguales o superiores a 0,85, especialmente entre distintas medidas rezagadas y medias móviles de la demanda. Esta multicolinealidad limita la interpretación aislada de coeficientes o importancias individuales, por lo que en el análisis posterior se priorizó la evaluación de grupos de variables en lugar de atribuir el resultado a un predictor específico.

4.9. Interpretabilidad y robustez

La interpretación de los modelos se realizó mediante SHAP y Permutation Importance, con el propósito de identificar las variables que contribuyeron en mayor medida a las estimaciones de presión hospitalaria. Debido a las diferencias entre el Modelo A y el Modelo B, los resultados fueron analizados por separado.

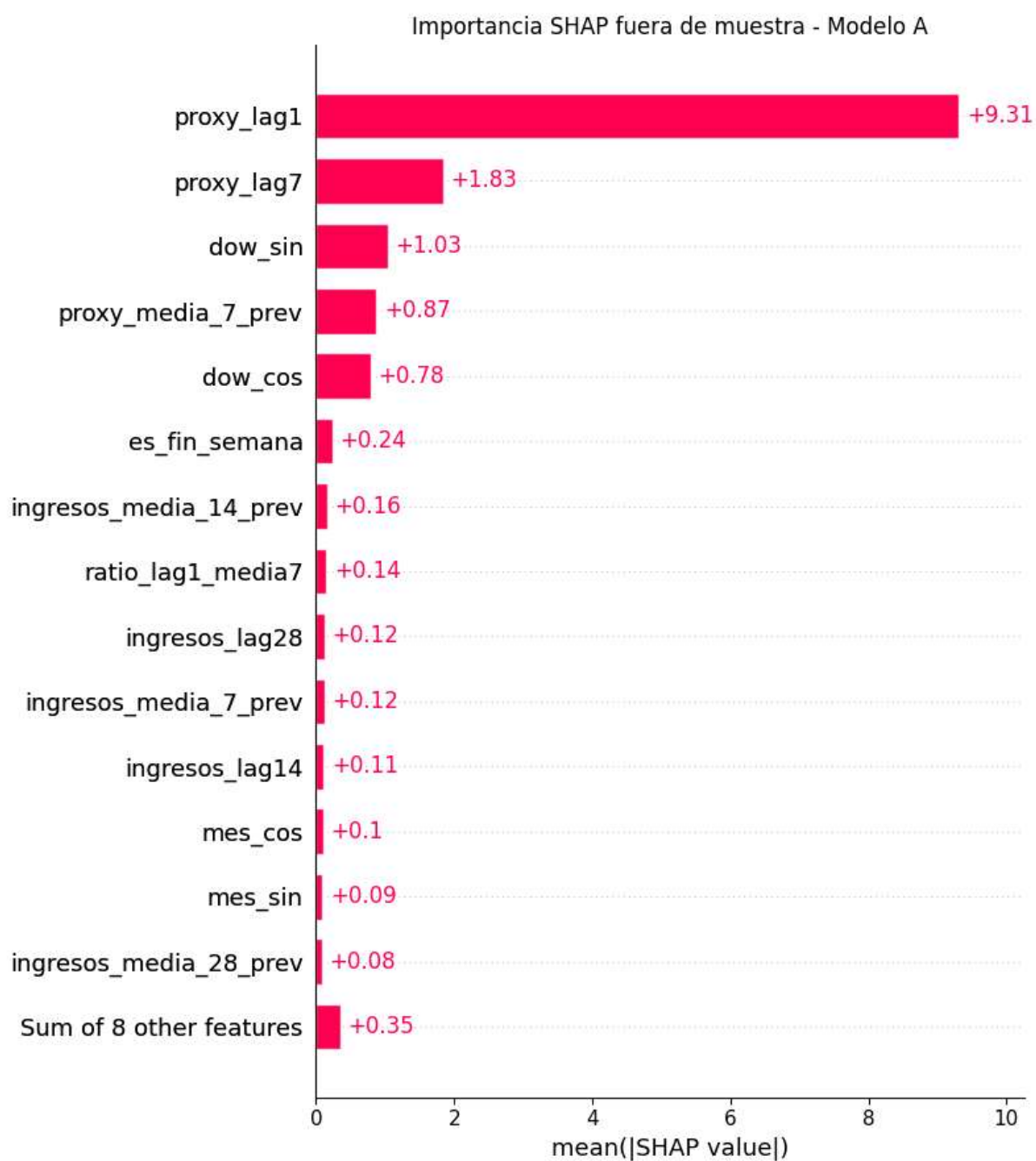
4.9.1. Interpretabilidad del Modelo A

En el Modelo A, seleccionado previamente con XGBoost, los valores SHAP mostraron un predominio marcado de la información histórica del propio indicador de presión. La variable de mayor importancia fue `proxy_lag1`, correspondiente al nivel de presión observado el día anterior, con una importancia SHAP absoluta media de 9,3072 puntos porcentuales. A considerable distancia se ubicaron `proxy_lag7`, con 1,8332 puntos, y las variables relacionadas con el patrón semanal y la media móvil reciente del proxy.

Este resultado indica que el comportamiento de la presión hospitalaria presenta un componente autorregresivo importante: conocer el nivel alcanzado el día anterior constituye la principal fuente de información para estimar el comportamiento del día siguiente. La presencia de `proxy_lag7` y de los componentes asociados al día de la semana muestra, además, que existe una estructura temporal semanal secundaria.

Figura 12

Importancia de las variables del Modelo A mediante valores SHAP



Nota. Elaboración Propia valores mayores representan una mayor contribución de la variable a las predicciones del modelo.

Los resultados fueron contrastados mediante Permutation Importance, obteniéndose una estructura similar. Al alterar aleatoriamente proxy_lag1, el MAE del modelo aumentó en promedio 11,1619 puntos porcentuales, mientras que la alteración de proxy_lag7 produjo un

incremento de 0,8467 puntos. Los componentes del día de la semana y la media móvil del proxy presentaron efectos menores, pero positivos.

La coincidencia entre SHAP y Permutation Importance refuerza la conclusión de que el buen desempeño predictivo del Modelo A depende principalmente de la persistencia reciente de la presión hospitalaria, complementada por información semanal. Esto también permite entender por qué el baseline de persistencia utilizado en la sección anterior constituía una referencia relativamente competitiva, aunque XGBoost consiguió mejorar su desempeño al combinar múltiples componentes temporales.

4.9.2. Interpretabilidad del Modelo B

La interpretación del Modelo B requiere mayor cautela. A nivel individual, los resultados de SHAP otorgaron una elevada importancia a variables como la media móvil de ingresos de los siete días previos, el sector y algunas características demográficas. Sin embargo, el diagnóstico de correlación identificó 30 pares de predictores con correlaciones absolutas iguales o superiores a 0,85, incluyendo correlaciones cercanas a 1 entre distintas medias móviles de la demanda.

Esta redundancia implica que la importancia atribuida a una variable individual puede depender de cómo el modelo distribuya información compartida entre predictores altamente correlacionados. Por esta razón, no resulta apropiado interpretar literalmente una variable específica como un factor independiente ni atribuirle un efecto causal. Para obtener resultados más robustos, las variables fueron agrupadas en categorías conceptuales y se aplicó Permutation Importance sobre cada grupo completo.

Tabla 23

Importancia agrupada de los factores considerados en el Modelo B

Grupo de variables	Δ MAE promedio (pp)	Posición
Sector	2,8965	1
Demanda reciente	0,3550	2
Calendario	0,2687	3
Composición demográfica	0,0099	4

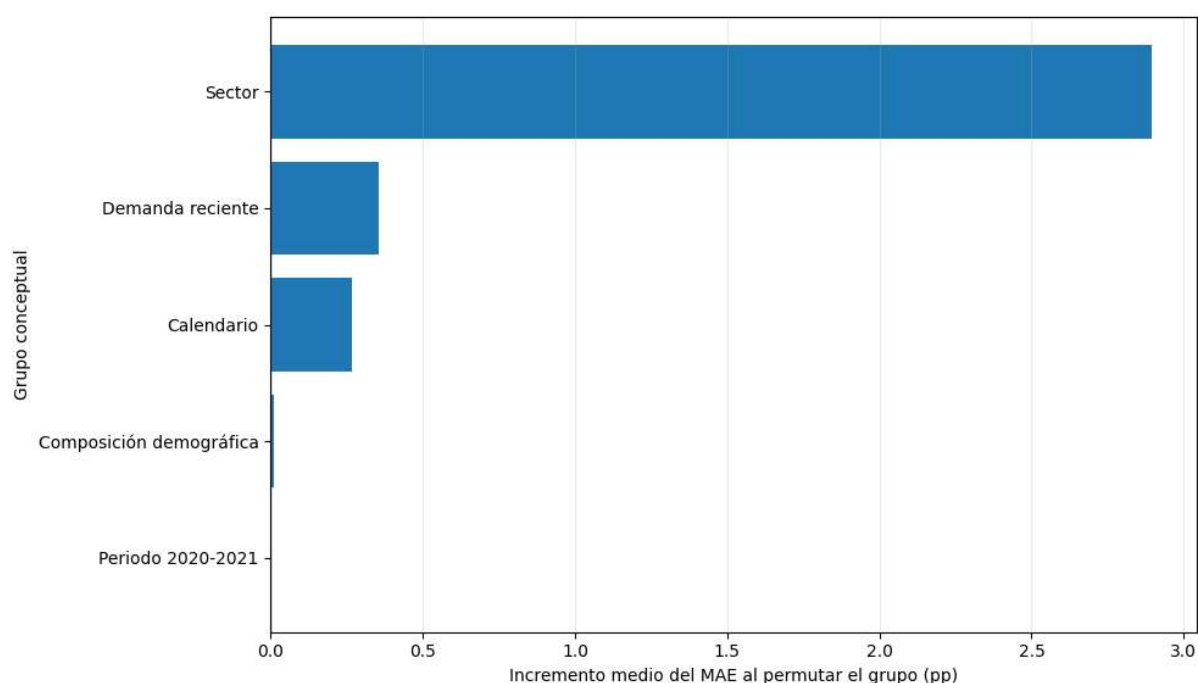
Nota. Elaboración propia. MAE representa el aumento promedio del error al permutar cada grupo de variables; pp = puntos porcentuales.

El sector del establecimiento presentó la mayor importancia agrupada, con un incremento promedio del MAE de aproximadamente 2,90 puntos porcentuales al alterar su información. La demanda reciente ocupó el segundo lugar, con 0,36 puntos, seguida por las variables de calendario, con 0,27 puntos. En contraste, la composición demográfica presentó una contribución prácticamente nula, de aproximadamente 0,01 puntos porcentuales.

Este resultado es especialmente importante para la interpretación del Modelo B: aunque determinadas variables demográficas aparecieron en posiciones elevadas cuando fueron evaluadas individualmente mediante SHAP, su contribución conjunta prácticamente desaparece cuando se considera la redundancia existente entre los predictores. Por ello, la evidencia más robusta se encuentra en los grupos de variables y no en las posiciones individuales.

Figura 13

Importancia agrupada de los factores asociados a la presión hospitalaria en el Modelo B



Nota. Elaboración propia. La importancia representa el aumento del MAE al permutar cada grupo de variables.

4.9.3. Análisis de robustez

El análisis de sensibilidad restringido al periodo posterior a 2020 permitió evaluar si la jerarquía observada se mantenía al excluir la etapa de mayor alteración de la dinámica hospitalaria. Los resultados conservaron el mismo orden general: sector, demanda reciente y calendario continuaron ocupando las primeras posiciones.

En este análisis, la importancia del sector aumentó hasta 4,8275 puntos porcentuales, mientras que la demanda reciente alcanzó 0,7570 y las variables de calendario 0,6196. La composición demográfica presentó un valor de $-0,0882$, lo que indica que su permutación no produjo un deterioro del modelo y, por tanto, no mostró una contribución predictiva robusta en este escenario.

Tabla 24

Importancia agrupada de los factores del Modelo B: análisis principal y sensibilidad post-2020

Grupo	Análisis principal	Sensibilidad post-2020
Sector	2,8965	4,8275
Demanda reciente	0,3550	0,7570
Calendario	0,2687	0,6196
Composición demográfica	0,0099	-0,0882

Nota. Elaboración propia. Los valores representan el cambio promedio del MAE tras permutar cada grupo de variables.

En consecuencia, dentro de las variables consideradas por el Modelo B, las diferencias entre sectores constituyen el factor asociado más consistente con el comportamiento de la presión hospitalaria, seguido por la dinámica reciente de la demanda y los patrones de calendario. La composición demográfica, en cambio, no presentó una contribución estable una vez consideradas conjuntamente las variables correlacionadas.

4.10. Contraste de hipótesis

Los resultados obtenidos permiten integrar los tres componentes principales del estudio: la predicción de la demanda hospitalaria, la caracterización de la presión hospitalaria relativa a la capacidad y la identificación de factores asociados a su comportamiento.

4.10.1. Respuesta a la pregunta de investigación

Los resultados muestran que los modelos analizados permiten predecir la demanda hospitalaria diaria con niveles de error diferenciados según el periodo evaluado. Durante la etapa de desarrollo 2020–2023, el modelo Seasonal Naive $t-7$ presentó el menor MAE promedio, con 50,07 ingresos diarios, evidenciando la relevancia del componente semanal de la demanda. Sin embargo, en el conjunto independiente de 2024, LSTM-28 obtuvo el mejor desempeño observado, con un MAE de 39,77, seguido por Prophet con 42,32 y Seasonal Naive con 54,64. Esto muestra que la superioridad de un modelo no fue constante entre periodos y que los cambios en la dinámica temporal de la demanda influyeron directamente sobre su desempeño.

Respecto a la ocupación hospitalaria, el análisis evidenció que los resultados agregados del DMQ pueden ocultar diferencias importantes entre sectores y establecimientos. Aunque la ocupación agregada se mantuvo por debajo del 85 % durante todo el periodo analizado, la reconstrucción diaria por sector permitió identificar 66 observaciones diarias sectoriales con presión igual o superior al 85 %, agrupadas en nueve episodios. La mayor concentración ocurrió en el sector público durante 2020, incluyendo un episodio de 41 días consecutivos de alta presión.

Finalmente, el análisis de factores asociados mostró que el sector del establecimiento constituye el componente más relevante y consistente para explicar las diferencias en la presión hospitalaria dentro del Modelo B. Le siguieron la demanda reciente y las variables de calendario, mientras que la composición demográfica presentó una contribución considerablemente menor. Esta jerarquía se mantuvo en el análisis de sensibilidad, obteniéndose el siguiente orden:

Sector > Demanda reciente > Calendario >> Composición demográfica

En conjunto, la respuesta a la pregunta de investigación indica que la demanda hospitalaria presenta una estructura temporal susceptible de ser modelada predictivamente, mientras que la presión hospitalaria no depende únicamente del volumen de ingresos, sino también de diferencias estructurales entre los sectores y de su comportamiento reciente.

4.10.2. Contraste de hipótesis

La hipótesis nula establecida señalaba que:

H0: Los modelos de inteligencia artificial y series temporales evaluados no presentan una reducción del error de predicción de la demanda hospitalaria diaria frente a un modelo de referencia estacional, considerando su desempeño mediante validación temporal fuera de muestra.

Mientras que la hipótesis alternativa planteaba que:

H1: Al menos uno de los modelos de inteligencia artificial o series temporales evaluados presenta una reducción del error de predicción de la demanda hospitalaria diaria frente a un modelo de referencia estacional, considerando su desempeño mediante validación temporal fuera de muestra.

Los resultados no muestran una superioridad uniforme de los modelos avanzados frente al modelo estacional. En la validación 2020–2023, Seasonal Naive t–7 presentó el menor MAE promedio y, por tanto, fue seleccionado durante la etapa de desarrollo. No obstante, en el conjunto independiente de 2024, LSTM-28 redujo el MAE de 54,64 a 39,77, equivalente a una disminución aproximada del 27,2 % respecto del modelo estacional.

Por tanto, existe evidencia empírica a favor de la hipótesis alternativa, debido a que al menos uno de los modelos evaluados logró reducir el error frente al modelo estacional en un periodo independiente fuera de muestra. Sin embargo, este resultado debe interpretarse considerando que dicha mejora no fue uniforme en todos los periodos de validación, especialmente ante la ruptura observada durante 2020.

CAPITULO 5:

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

El análisis histórico permitió identificar una dinámica diferenciada de la demanda hospitalaria durante el periodo estudiado. Entre 2017 y 2019 se observó un comportamiento relativamente estable, mientras que 2020 constituyó una ruptura importante en la serie, con una reducción considerable de los ingresos hospitalarios. Posteriormente, la demanda presentó una recuperación progresiva hasta alcanzar nuevamente niveles elevados durante 2022 y 2023. Este comportamiento confirma la necesidad de utilizar esquemas de evaluación temporal capaces de considerar cambios estructurales y no únicamente patrones históricos regulares.

En relación con la predicción de la demanda hospitalaria diaria, el modelo Seasonal Naive t-7 presentó el menor MAE promedio durante la validación temporal 2020–2023, con 50,07 ingresos diarios, superando globalmente a LSTM-28 y Prophet. Este resultado evidenció que la demanda presenta una estructura semanal importante y que un modelo relativamente sencillo puede constituir una referencia predictiva competitiva. Sin embargo, en el conjunto independiente de 2024, LSTM-28 obtuvo el menor error observado, con un MAE de 39,77, frente a 54,64 del modelo estacional. Esto demuestra que el desempeño relativo de los modelos puede modificarse entre periodos y que una mayor complejidad algorítmica no garantiza una superioridad constante.

Respecto a la hipótesis de investigación, los resultados proporcionan evidencia empírica favorable a la hipótesis alternativa, debido a que al menos uno de los modelos evaluados logró reducir el error de predicción frente al modelo estacional en un periodo independiente fuera de muestra. No obstante, esta mejora no fue uniforme durante todos los periodos analizados, por lo que no puede afirmarse una superioridad general de los modelos avanzados bajo cualquier condición temporal.

El análisis de capacidad permitió diferenciar claramente los conceptos de demanda y ocupación hospitalaria. La ocupación agregada del DMQ se mantuvo por debajo del 85 % durante todos los años, registrándose su valor más bajo en 2020, con 57,55 %, y una

recuperación posterior hasta 68,57 % en 2024. Sin embargo, el análisis por establecimiento y sector mostró que los resultados agregados pueden ocultar niveles considerablemente superiores de utilización de la capacidad.

La reconstrucción del proxy diario permitió profundizar en estas diferencias. Aunque el indicador agregado del DMQ no alcanzó el umbral del 85 %, el análisis sectorial identificó 66 observaciones diarias sectoriales iguales o superiores al 85 %, agrupadas en nueve episodios. La mayor concentración ocurrió en el sector público durante 2020, donde se registraron 59 días por encima del 85 % y 35 días por encima del 90 %. El episodio de mayor duración se extendió durante 41 días consecutivos, con una presión media de 91,06 % y un máximo de 96,00 %. Estos resultados evidencian que el análisis desagregado permite identificar situaciones de presión que no son visibles mediante indicadores globales.

El Modelo A demostró una alta capacidad para anticipar el comportamiento del proxy diario de presión hospitalaria. XGBoost alcanzó un MAE promedio de 1,5260 puntos porcentuales y un R^2

de 0,9760, reduciendo el error en 42,81 % respecto al baseline de persistencia del día anterior. Los análisis de interpretabilidad mostraron que esta capacidad predictiva depende principalmente de la información histórica reciente del propio indicador, especialmente de `proxy_lag1`, lo que confirma un marcado componente autorregresivo en el comportamiento diario de la presión hospitalaria.

Por su parte, el Modelo B no logró superar globalmente al baseline basado en la media sectorial, por lo que sus resultados deben considerarse exploratorios y no como evidencia de superioridad predictiva. Sin embargo, el análisis agrupado permitió identificar una jerarquía consistente de factores asociados: Sector > Demanda reciente > Calendario >> Composición demográfica. La importancia del sector alcanzó 2,8965 puntos porcentuales en el análisis principal y 4,8275 en la sensibilidad post-2020, manteniéndose como el componente más relevante en ambos escenarios.

En conjunto, la investigación evidencia que la planificación hospitalaria no debería sustentarse únicamente en la predicción del número de ingresos. La presión sobre la capacidad resulta de la interacción entre la demanda, la ocupación acumulada, la estructura

sectorial y la dinámica temporal reciente. Por ello, la combinación de modelos de predicción de demanda con indicadores de presión relativa a capacidad proporciona una visión más completa para el seguimiento de la utilización hospitalaria en el Distrito Metropolitano de Quito.

5.2.Recomendaciones

Se recomienda que futuras herramientas de monitoreo hospitalario incorporen simultáneamente indicadores de demanda y capacidad, evitando utilizar el número de ingresos hospitalarios como una medida directa de saturación u ocupación. La disponibilidad de información conjunta sobre ingresos, días-paciente, camas disponibles y presión relativa permitiría identificar con mayor precisión los periodos de mayor utilización de los servicios.

Para la predicción de la demanda hospitalaria se recomienda mantener modelos de referencia estacionales dentro de cualquier sistema de evaluación. El buen desempeño de Seasonal Naive t-7 demuestra que los modelos avanzados deben compararse siempre frente a alternativas sencillas y competitivas antes de justificar su utilización. Asimismo, modelos como LSTM deberían ser reevaluados periódicamente mediante ventanas temporales actualizadas, debido a que su desempeño puede mejorar cuando incorporan información correspondiente a cambios recientes de la serie.

En aplicaciones operacionales de corto plazo, el Modelo A basado en XGBoost podría considerarse como una herramienta complementaria para anticipar cambios en la presión hospitalaria, debido a la mejora observada frente al baseline de persistencia. No obstante, antes de su implementación institucional se recomienda realizar validaciones adicionales con información de periodos posteriores y, de ser posible, a nivel de establecimiento individual.

También se recomienda fortalecer la disponibilidad de información diaria de capacidad hospitalaria. En esta investigación, la presión diaria fue reconstruida a partir de las estancias hospitalarias y de una capacidad media anual derivada de los días-cama disponibles. Contar con registros diarios de camas habilitadas, camas ocupadas y modificaciones temporales de capacidad permitiría construir indicadores diarios de ocupación directamente observados y reducir la necesidad de utilizar aproximaciones analíticas.

Los resultados sectoriales sugieren que los sistemas de monitoreo deberían permitir análisis diferenciados por tipo de establecimiento. El sector público presentó los niveles más persistentes de presión hospitalaria, por lo que un indicador exclusivamente agregado para todo el DMQ podría ocultar situaciones relevantes dentro de determinados componentes de la red de salud.

Finalmente, los resultados del Modelo B deben utilizarse como evidencia de asociación y no de causalidad. Debido a la multicolinealidad identificada entre varios predictores y al desempeño limitado del modelo frente al baseline sectorial, futuras investigaciones podrían ampliar el análisis incorporando nuevas variables relacionadas con características de los establecimientos, especialidades, disponibilidad de recursos, complejidad asistencial y otros factores que permitan explicar con mayor profundidad las diferencias observadas en la presión hospitalaria.

Bibliografía

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160.
<https://doi.org/10.1109/ACCESS.2018.2870052>
- Avinash, G., Pachori, H., Sharma, A., & Mishra, S. (2025). Time series forecasting of bed occupancy in mental health facilities in India using machine learning. *Scientific Reports*, 15, Article 2686. <https://doi.org/10.1038/s41598-025-86418-9>
- Bagust, A., Place, M., & Posnett, J. W. (1999). Dynamics of bed use in accommodating emergency admissions: Stochastic simulation model. *BMJ*, 319(7203), 155–158.
<https://doi.org/10.1136/bmj.319.7203.155>
- Bravo Silva, C. M. (2024). Modelos de predicción de la demanda hospitalaria en pediatría utilizando modelos epidemiológicos y aprendizaje profundo [Tesis de magíster, Universidad de Chile]. Repositorio Académico de la Universidad de Chile.
<https://repositorio.uchile.cl/handle/2250/202949>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Calegari, R., Fogliatto, F. S., Lucini, F. R., Neyeloff, J., Kuchenbecker, R. S., & Schaan, B. D. (2016). Forecasting daily volume and acuity of patients in the emergency department. *Computational and Mathematical Methods in Medicine*, 2016, 3863268. DOI: 10.1155/2016/3863268.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. En *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery.
<https://doi.org/10.1145/2939672.2939785>
- Erkamp, N. S., van Dalen, D. H., & de Vries, E. (2021). Predicting emergency department visits in a large teaching hospital. *International Journal of Emergency Medicine*, 14, 34. DOI: 10.1186/s12245-021-00357-6.
- Eyles, E., Redaniel, M. T., Jones, T., Prat, M., & Keen, T. (2022). Can we accurately forecast non-elective bed occupancy and admissions in the NHS? A time-series MSARIMA analysis of longitudinal data from an NHS Trust. *BMJ Open*, 12(4), e056523.
<https://doi.org/10.1136/bmjopen-2021-056523>
- García, M., & López, R. (2022). Modelos predictivos en sistemas de salud latinoamericanos. *Revista de Salud Pública*, 24(3).
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
<https://www.deeplearningbook.org/>

- Hernández-Sampieri, R., & Mendoza Torres, C. P. (2018). Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta. McGraw-Hill Education.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. <https://otexts.com/fpp2/>
- Instituto Nacional de Estadística y Censos. (s. f.). Camas y egresos hospitalarios. Recuperado el 16 de agosto de 2026, de <https://www.ecuadorencifras.gob.ec/camas-y-egresos-hospitalarios/>
- Instituto Nacional de Estadística y Censos. (2025). Metodología de los registros estadísticos de camas y egresos hospitalarios 2024. https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/Camas_Egresos_Hospitalarios/2024/Metodologia_de_Camas_y_Egresos_Hospitalarios_2024.pdf
- Jilani, T., Housley, G., Figueredo, G., Tang, P.-S., Hatton, J., & Shaw, D. (2019). Short and long term predictions of hospital emergency department attendances. *International Journal of Medical Informatics*, 129, 167–174. DOI: 10.1016/j.ijmedinf.2019.05.011.
- Juang, W.-C., Huang, S.-J., Huang, F.-D., Cheng, P.-W., & Wann, S.-R. (2017). Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in Southern Taiwan. *BMJ Open*, 7(11), e018628. DOI: 10.1136/bmjopen-2017-018628.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. En *Advances in Neural Information Processing Systems* 30 (pp. 4765–4774). Curran Associates, Inc. <https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions>
- Mahmoudian, Y., Nemati, A., & Safaei, A. S. (2023). A forecasting approach for hospital bed capacity planning using machine learning and deep learning with application to public hospitals. *Healthcare Analytics*, 4, Article 100245. <https://doi.org/10.1016/j.health.2023.100245>
- Organización Mundial de la Salud. (2019). Clasificación estadística internacional de enfermedades y problemas relacionados con la salud: Décima revisión (CIE-10) (6.^a ed.). https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf
- Porto, B. M., & Fogliatto, F. S. (2024). Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning. *BMC Medical Informatics and Decision Making*, 24, 377. DOI: 10.1186/s12911-024-02788-6.

- Rodríguez, J., Pérez, M., & Gómez, L. (2021). Forecasting hospital demand using SARIMA models. *Health Systems Journal*, 10(2).
- Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B.-W., Aslam, W., & Choi, G. S. (2020). COVID-19 future forecasting using supervised machine learning models. *IEEE Access*, 8, 101489–101499. <https://doi.org/10.1109/ACCESS.2020.2997311>
- Schwaber, K., & Sutherland, J. (2020). La guía de Scrum: La guía definitiva de Scrum: Las reglas del juego. <https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-Spanish-European.pdf>
- Seo, H., Ahn, I., Gwon, H., Kang, H., Kim, Y., Choi, H., Kim, M., Han, J., Kee, G., Park, S., Ko, S., Jung, H., Kim, B., Oh, J., Jun, T. J., & Kim, Y.-H. (2024). Forecasting hospital room and ward occupancy using static and dynamic information concurrently: Retrospective single-center cohort study. *JMIR Medical Informatics*, 12, e53400. <https://doi.org/10.2196/53400>
- Shahid, F., Zameer, A., & Muneeb, M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals*, 140, Article 110212. <https://doi.org/10.1016/j.chaos.2020.110212>
- Silva, E., Pereira, M. F., Vieira, J. T., Ferreira-Coimbra, J., Henriques, M., & Rodrigues, N. F. (2023). Predicting hospital emergency department visits accurately: A systematic review. *The International Journal of Health Planning and Management*, 38(4), 904–917. DOI: 10.1002/hpm.3629.
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
- Tello, M., Reich, E. S., Puckey, J., Maff, R., Garcia-Arce, A., Bhattacharya, B. S., & Feijoo, F. (2022). Machine learning based forecast for the prediction of inpatient bed demand. *BMC Medical Informatics and Decision Making*, 22, Article 55. <https://doi.org/10.1186/s12911-022-01787-9>

Apéndices

Apéndice A. Código y repositorio del proyecto

Este apéndice presenta el acceso al repositorio que contiene el código desarrollado durante la investigación para el procesamiento y preparación de los datos, la implementación de los modelos predictivos, la generación de resultados y la construcción del visualizador. Su inclusión permite fortalecer la trazabilidad del proceso técnico y facilitar la revisión, comprensión y reproducibilidad de los procedimientos aplicados en el estudio. visualizador (Ayala López et al., 2026a)

https://github.com/jordanayala96/demanda_hospitalaria

Apéndice B. Diccionario de variables

Este apéndice presenta las principales variables originales y derivadas utilizadas en la preparación de los datos, el análisis de la demanda, el cálculo de la ocupación y la construcción de los modelos.

Tabla B1

Diccionario de variables utilizadas en la investigación

Variable	Fuente/Tipo	Descripción	Uso
fecha_ingreso	Derivada de egresos	Fecha de ingreso del paciente	Construcción de demanda diaria
fecha_egreso	Derivada de egresos	Fecha de egreso hospitalario	Reconstrucción de estancias
dia_estad	Original	Número de días de estancia hospitalaria	Duración de estancia
sector	Derivada/categorica	Sector del establecimiento	Análisis sectorial
camas_dnor	Original Camas	Dotación normal de camas	Capacidad hospitalaria
camas_disp	Original Camas	Camas disponibles	Capacidad hospitalaria
dia_estad / días-paciente	Original Camas	Total anual de días-paciente	Ocupación oficial
dias_cdis	Original Camas	Días-cama disponibles	Denominador de ocupación
ocupacion_agregada_pct	Derivada	Relación entre días-paciente y días-cama disponibles	Ocupación anual

ingresos_diarios	Derivada	Número de ingresos por fecha	Predicción de demanda
ocupacion_diaria_proxy_pct	Derivada	Proxy diario de presión relativa a capacidad	Análisis de presión
proxy_lag1	Derivada	Presión observada un día antes	Modelo A
proxy_lag7	Derivada	Presión observada siete días antes	Modelo A
ingresos_lag1	Derivada	Demanda hospitalaria rezagada un día	Modelos de factores
ingresos_media_7_prev	Derivada	Media de ingresos de los siete días previos	Modelo B
es_fin_semana	Derivada	Indicador de sábado o domingo	Componente calendario

Nota. Elaboración propia a partir de variables de los registros del INEC (2017–2024).

Apéndice C. Configuración de los modelos

Este apéndice presenta las configuraciones utilizadas en los modelos predictivos y en el análisis de factores asociados, con el propósito de documentar los parámetros definidos durante el desarrollo de la investigación y facilitar la comprensión y reproducción del proceso analítico aplicado.

Tabla C1

Configuración de los modelos implementados

Modelo	Parámetro	Configuración
LSTM	Lookback seleccionado	28 días
LSTM	Variable objetivo	Ingresos hospitalarios diarios
LSTM	Validación	Expansión temporal
Prophet	Variable objetivo	Ingresos hospitalarios diarios
Prophet	Validación	Expansión temporal
Random Forest	n_estimators	500
Random Forest	max_depth	12
Random Forest	min_samples_split	4
Random Forest	min_samples_leaf	3
Random Forest	max_features	1
Random Forest	bootstrap	True
XGBoost	n_estimators	500
XGBoost	learning_rate	0,03

XGBoost	max_depth	4
XGBoost	min_child_weight	5
XGBoost	subsample	0,85
XGBoost	colsample_bytree	0,85
XGBoost	reg_alpha	0
XGBoost	reg_lambda	1

Nota. Elaboración propia. Las configuraciones corresponden a los modelos implementados en la investigación.

Apéndice D. Diseño de validación temporal

Este apéndice describe los esquemas de entrenamiento, validación y prueba aplicados durante el desarrollo de los modelos, considerando la secuencia cronológica de los datos. Su finalidad es documentar la distribución temporal utilizada para el ajuste y evaluación de los modelos, garantizando que su desempeño sea analizado sobre periodos diferentes a los empleados durante el proceso de entrenamiento.

Tabla D1

Validación temporal de los modelos de demanda hospitalaria

Fold	Entrenamiento	Validación
1	2017–2019	2020
2	2017–2020	2021
3	2017–2021	2022
4	2017–2022	2023
Test independiente	2017–2023	2024

Nota. Elaboración propia. El año 2024 se reservó como conjunto de prueba independiente.

Tabla D2

Validación temporal de los modelos de presión hospitalaria

Fold	Entrenamiento	Validación
1	2017–2020	2021
2	2017–2021	2022
3	2017–2022	2023

Nota. Elaboración propia. Cada periodo de validación utilizó únicamente información cronológicamente anterior

Tabla D3*Análisis de sensibilidad temporal del Modelo B*

Escenario	Entrenamiento	Validación
S1	2021	2022
S2	2021–2022	2023

Nota. Elaboración propia. El análisis se restringió a periodos posteriores a 2020.