MODELADO PREDICTIVO DEL FLUJO DE PASAJEROS EN EL AEROPUERTO EL DORADO USANDO SARIMA, ARIMA Y LSTM (1996–2024)

ERIKA MILDREEY BELTRAN PAEZ

UNIVERSIDAD JORGE TADEO LOZANO

FACULTAD DE CIENCIAS NATURALES E INGENIERIA

PROGRAMA DE MAESTRIA DE INGENIERIA Y ANALITICA DE DATOS

BOGOTA

2025

TRABAJO DE GRADO PARA OPTAR POR EL TITULO DE MÁSTER EN INGENIERIA Y ANALITICA DE DATOS.

IXENT GALPIN

UNIVERSIDAD JORGE TADEO LOZANO

FACULTAD DE CIENCIAS NATURALES E INGENIERIA

PROGRAMA DE MAESTRIA DE INGENIERIA Y ANALITICA DE DATOS

BOGOTA

2025

Tabla de Contenido

Contenido de figuras

Contenido de Tablas

Resumen

El Aeropuerto Internacional El Dorado, en Bogotá, se destaca por su eficiencia operativa en los últimos años y se posiciona como un hub clave para el tránsito de pasajeros en Latinoamérica. Esta investigación se centra en la implementación de modelos predictivos para estimar el flujo mensual de pasajeros, utilizando datos entre 1996 y 2024. Los modelos ARIMA, SARIMA y LSTM se emplean para realizar las predicciones.

La validación se lleva a cabo mediante validación cruzada, adaptada específicamente para series temporales. La evaluación del desempeño de los modelos se realiza utilizando las métricas de error RMSE (Root Mean Square Error) y MAE (Mean Absolute Error), seleccionadas por su capacidad para cuantificar con precisión la desviación entre los valores predichos y los observados.

Los datos utilizados provienen de la Unidad Administrativa Especial de Aeronáutica Civil de Colombia, e incluyen tanto el flujo de pasajeros nacionales como internacionales. El proceso analítico sigue la metodología CRISP-DM, que abarca las etapas de recolección, limpieza, normalización, suavización y modelado de los datos, seguido de la evaluación mediante las métricas mencionadas.

Los resultados indican que el modelo SARIMA presenta el menor error en comparación con ARIMA y LSTM, especialmente en flujos de carácter estacional. SARIMA muestra valores más bajos y estables de RMSE y MAE, lo que lo convierte en el modelo más adecuado para este tipo de predicción.

Finalmente, se propone el despliegue de las predicciones a través de un dashboard interactivo o un chatbot, lo que facilitaría el acceso a la información en tiempo real y apoyaría la toma de decisiones estratégicas en la gestión operativa del aeropuerto, mejorando tanto la experiencia del usuario como la planificación de recursos.

Introducción

El Aeropuerto Internacional El Dorado ha sido reconocido como el aeropuerto mejor conectado de América Latina y ocupa el puesto número 20 en el ranking global de conectividad aérea, según el informe OAG Megahubs 2024. Este informe es un estudio anual elaborado por OAG (Official Airline Guide), una de las principales fuentes globales de inteligencia en aviación. Entre las variables críticas para la gestión operativa de un aeropuerto se encuentra el flujo de pasajeros, cuya correcta estimación permite una planificación más eficiente de recursos, procesos y servicios.

Además, El Dorado ha sido distinguido por sexta vez, y por tercer año consecutivo, como el Mejor Aeropuerto de Sudamérica por Skytrax, una organización internacional que evalúa la experiencia de los viajeros en diversos aeropuertos, considerando factores como la atención del personal, los procesos de seguridad y los servicios de check-in, entre otros.

La predicción del flujo de pasajeros es crucial para una planificación efectiva, especialmente en aeropuertos con alta demanda y operaciones complejas. Esta tarea requiere técnicas analíticas robustas que permitan capturar la estacionalidad, las tendencias y los patrones ocultos en los datos históricos. En este estudio, se propone la aplicación de modelos avanzados de series temporales, como ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA) y LSTM (Long Short-Term Memory), conocidos por su capacidad para modelar tanto patrones lineales como no lineales en secuencias temporales.

El análisis se fundamenta en una base de datos pública proporcionada por la Unidad Administrativa Especial de Aeronáutica Civil de Colombia, la cual contiene registros desde 1996 hasta 2024. Estos datos, presentados en archivos de Excel, presentan variaciones en su estructura, como nombres de columnas inconsistentes, campos faltantes o adicionales según el año. Esto requiere un proceso riguroso de limpieza y transformación para lograr un conjunto de datos homogéneo y fiable. A partir de esta depuración, se segmentan los vuelos en categorías nacionales e internacionales, se eliminan valores atípicos y se aplican técnicas de suavización estadística para facilitar la identificación de tendencias.

El desarrollo de este estudio sigue la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), que consta de seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue (Shearer, 2000). Para evaluar el rendimiento de los modelos, se implementa la técnica de validación cruzada específica para series temporales, utilizando el método TimeSeriesSplit de la biblioteca scikit-learn, que permite preservar el orden cronológico al dividir los datos en subconjuntos de entrenamiento y prueba.

Los modelos SARIMA, ARIMA y LSTM se evalúan en distintos periodos utilizando métricas como el RMSE (Root Mean Squared Error) y el MAE (Mean Absolute Error), que ofrecen una medición objetiva del error promedio de predicción en comparación con los valores reales.

Finalmente, los modelos se despliegan de manera interactiva, priorizando la visualización y la capacidad de generar herramientas aplicables a contextos reales. Esto facilita la toma de decisiones gerenciales y optimiza el uso de recursos en la operación aeroportuaria.

Este trabajo tiene como propósito proporcionar un análisis sólido del comportamiento del flujo de pasajeros mediante modelos predictivos, ofreciendo una solución práctica y eficiente para mejorar la gestión del tiempo y los recursos en el Aeropuerto El Dorado. Se presentan herramientas, métodos y técnicas de gran valor para uno de los activos más importantes en la gestión operativa: el conocimiento anticipado del comportamiento de la demanda.

Estado del arte

En 2012, Bergmeir y Benítez analizan la aplicabilidad de métodos de validación cruzada en series temporales, un aspecto clave en la evaluación de modelos predictivos. Aunque este trabajo no se centra en datos aeroportuarios, establece bases fundamentales sobre la medición de desempeño en modelos de series temporales. (Bergmeir & Benítez, 2012)

En 2014, Laik et al. desarrollan un modelo de predicción del flujo de pasajeros en aeropuertos asiáticos utilizando algoritmos de árboles de decisión. Los datos reales permiten evaluar el rendimiento del modelo, que obtiene un error cuadrático medio (RMSE) relativo entre el 3% y el 12%. Además, se integra una simulación para optimizar la asignación de recursos en el proceso de check-in, abordando así problemas prácticos en la operación aeroportuaria.(Laik et al., 2014)

En 2016, Li et al. aplicaron el modelo SARIMA para predecir el volumen diario de pasajeros en el aeropuerto de Kunming Changshui (China). Los resultados mostraron una alta precisión, con errores absolutos medios (MAE) entre el 1 % y el 3 %, lo que permitió sugerir ajustes operativos más eficientes.

En 2019, Orsini et al. proponen una comparación entre tres tipos de redes neuronales para predecir el comportamiento de los pasajeros utilizando trazas WiFi anónimas en el aeropuerto de Bolonia. Las redes evaluadas incluyen modelos feedforward (FNN), LSTM y combinaciones de ambas. Los mejores resultados se obtienen con LSTM, especialmente para predicciones a corto plazo (20 minutos), lo que demuestra el potencial de los datos en tiempo real para la gestión dinámica del aeropuerto. (Orsini et al., 2019).

En 2021, Guo et al. desarrollan un modelo híbrido que combina árboles de regresión con simulaciones para pronosticar el flujo de pasajeros en tránsito (transferencias). Utilizando datos en tiempo real, el modelo logra predecir la distribución de tiempos de conexión y anticipar los volúmenes en procesos como inmigración, lo que tiene un impacto significativo en la experiencia del pasajero y la asignación de recursos.

Más recientemente, en 2022, Guimarães et al. abordan la predicción de la pérdida de conexiones aéreas, un problema crítico para la rentabilidad de las aerolíneas. Para ello, emplean técnicas avanzadas de aprendizaje automático adaptadas a datos heterogéneos, ruidosos y desequilibrados. Aplican codificación probabilística, técnicas de balanceo, modelos gaussianos y boosting, alcanzando un AUC superior a 0.93. El estudio concluye que los tiempos de conexión programados son el principal predictor de la pérdida de enlaces. (Guimarães et al., 2022)

En el contexto del análisis de series temporales, también es relevante mencionar la prueba de Dickey-Fuller, que permite determinar si una serie es estacionaria o posee una raíz unitaria. Esta prueba estadística es clave para la selección del modelo apropiado, ya que muchos modelos requieren estacionariedad para ofrecer buenos resultados. (Harris, 1992)

A diferencia de los trabajos previos, el presente estudio propone un enfoque integrador que combina técnicas tradicionales y modernas para la predicción del flujo de pasajeros en el Aeropuerto Internacional El Dorado. En concreto, se comparan modelos clásicos como ARIMA y SARIMA con redes neuronales LSTM. Esta combinación permite capturar tanto

patrones lineales como no lineales, aprovechando las ventajas tanto de la estadística tradicional como del aprendizaje profundo. Además, se utiliza un conjunto de datos extenso y actualizado (1996–2024) proveniente del aeropuerto más importante de Latinoamérica, lo que aporta mayor relevancia y aplicabilidad a los resultados.

Tabla 1. Comparación de estudios previos en predicción de flujo de pasajeros

Año	**Autor(es)**	**Técnica principal**	**Tipo de datos**	**Métrica de desempeño**
2014	Laik et al.	Árboles de decisión	Datos históricos de aeropuertos asiáticos	RMSE: 3–12 %
2016	Li et al.	SARIMA	Datos diarios del aeropuerto de Kunming	MAE: 1–3 %
2019	Orsini et al.	FNN, LSTM	Datos WiFi en tiempo real	Mejor desempeño: LSTM
2021	Guo et al.	Árbol + simulación	Datos en tiempo real de pasajeros en tránsito	No especificada
2022	Guimarães et al.	ML con *boosting*, GaussianNB	Datos operacionales y categóricos	AUC > 0.93

Entendimiento del negocio

El Aeropuerto Internacional El Dorado, ubicado en Bogotá, Colombia, es el principal hub aéreo del país y uno de los más importantes de América Latina. Su rol es estratégico tanto a nivel nacional como regional, ya que conecta a Colombia con más de 70 destinos internacionales y más de 30 nacionales. Según datos de la Aeronáutica Civil de Colombia, en 2023 El Dorado movilizó más de 38 millones de pasajeros, cifra que representa cerca del 60 % del tráfico aéreo total del país, reafirmando su posición como nodo clave en la conectividad aérea de Sudamérica. (Prensa Latina, 2025; Aviacionline, 2025)

Desde su remodelación y ampliación en la década de 2010, el aeropuerto ha experimentado un crecimiento sostenido en capacidad y volumen de pasajeros, con una tasa promedio de crecimiento del tráfico de alrededor del 4–5 % anual en la última década, exceptuando el descenso registrado durante la pandemia de COVID-19 (El País, 2023). El crecimiento del flujo de pasajeros plantea desafíos significativos para la operación diaria, incluyendo congestión en procesos como el check-in, migración, seguridad y entrega de equipaje. Por ello, la capacidad de predecir con precisión el flujo de pasajeros se convierte en una herramienta clave para mejorar la eficiencia operativa, optimizar la asignación de recursos y reducir tiempos de espera, generando así una mejor experiencia para el viajero.

El Dorado ha sido galardonado en seis ocasiones como el mejor aeropuerto de Sudamérica por los *World Airport Awards* de Skytrax (2015, 2016, 2017, 2021, 2023 y 2024). Skytrax basa sus rankings en encuestas de satisfacción aplicadas a más de 100 países, evaluando aspectos como limpieza, confort, procesos de embarque, señalización, amabilidad del personal y eficiencia de los servicios. En este sentido, El Dorado ha destacado frente a competidores regionales como los aeropuertos de Lima, Santiago y São Paulo.

En febrero de 2025, El Dorado recibió el premio platino a la excelencia operativa otorgado por *Cirium*, una firma global de análisis del sector aeronáutico. (Cirium, 2025). Este premio destaca el uso de tecnologías emergentes como más de 50 estaciones de auto check-in, inteligencia artificial en puntos de atención al cliente, más de 40 estaciones de entrega automática de equipaje y un sistema biométrico para procesos migratorios (*Biomig*), que ha contribuido a reducir los tiempos de tránsito. Según lo público la página oficial del Aeropuerto El Dorado. (*El Aeropuerto El Dorado Continúa Recibiendo Premios Por Esta Razón | Infraestructura | Economía | Portafolio*, n.d.)

Desde el punto de vista económico, El Dorado genera diariamente una gran demanda de pasajeros y contribuye de forma significativa al PIB de Bogotá y del país, tanto por el turismo como por el comercio internacional. Su eficiencia y competitividad impactan directamente en la atracción de inversión extranjera, la promoción del turismo, y la posición estratégica de Colombia en el comercio regional.

La predicción del flujo de pasajeros en este contexto se convierte en una herramienta indispensable para anticipar picos de demanda, preparar recursos humanos y logísticos, y minimizar cuellos de botella en procesos críticos. Modelos como ARIMA, SARIMA y LSTM permiten capturar las características de las series temporales del tráfico de pasajeros — como estacionalidad, tendencia y patrones no lineales— y ofrecen un marco técnico para implementar sistemas de gestión proactivos.

Esta investigación, al aplicar y comparar estos modelos sobre datos históricos y actualizados de El Dorado, contribuye no solo al conocimiento técnico del área, sino también al fortalecimiento de la competitividad del aeropuerto y, por extensión, de la infraestructura logística del país.

Entendimiento de los datos

El conjunto de datos utilizado en este análisis se basa en la información de pasajeros del Aeropuerto El Dorado de Bogotá, desde 1996 hasta principios de 2024. Esta información fue obtenida de una fuente pública de la Unidad Administrativa Especial Aeronáutica Civil de Colombia, específicamente de su base de datos origen–destino, dentro de las estadísticas de las actividades aeronáuticas.

Los datos están disponibles en varios archivos en formato Excel, organizados por columnas. Los archivos presentan una estructura heterogénea: algunos están organizados por año y otros por mes, lo que requiere un proceso previo de estandarización antes de su análisis.

Cada archivo contiene información de pasajeros organizada por origen y destino, con datos específicos sobre el tráfico de pasajeros, carga y correo en meses y años específicos para las rutas establecidas por las aerolíneas autorizadas en cada periodo. Además, cada archivo proporciona información sobre las cantidades totales de pasajeros y permite identificar patrones emergentes en diferentes periodos de tiempo, ya sea por ruta, ciudad o país.

Nombre	**Descripción**
Origen	Es la sigla IATA del aeropuerto donde se realiza el proceso de embarque de pasajeros
Destino	Es la sigla IATA del aeropuerto donde se realiza el proceso de desembarque de pasajeros
Pasajeros	Unidad numérica que corresponde al pasajero que ha realizado el proceso de embarque y desembarque, incluyendo los siguientes casos: Ofertas y promociones, programas de fidelidad, pasajeros que viajan con descuentos por empresas, pasajeros que son identificados como funcionarios gubernamentales, militares, marinos, estudiantes etc. Por ultimo los bebes
Tipo de vuelo	R: Operación regular, son los servicios aéreos públicos que son anunciados y tiene horario fijo A: Vuelos adicionales, son aquellos vuelos que son realizados debido a exceso de trafico C: vuelos chárteres, Son vuelos autorizados para atender situaciones especiales de demanda en el aeropuerto. T: Taxi aéreo, operación realizada por las empresas de taxi aéreo
Trafico	N tráfico domestico I tráfico internacional E tráfico entre dos aeropuertos fuera de Colombia
Fecha	Mes y año
Ciudad de origen	Nombre de la ciudad de origen donde embarcan los pasajeros.
Ciudad de	Nombre de la ciudad de destino donde desembarcaron los

destino	pasajeros.
Sigla de la empresa	Sigla OACI con la cual es identificada la empresa o aerolínea en las autoridades aeronáuticas.
Nombre Empresa	Nombre comercial de la empresa

Tabla 2 Las variables y su descripción que corresponden a las columnas en los archivos Excel recolectados en la base de datos de la Aeronáutica Civil de Colombia.(*Bases de Datos*, n.d.) La grafica está dividida en 4 líneas, la línea amarilla y roja muestra el flujo de salidas nacionales e internacionales y las líneas azul y verde muestran el flujo de llegadas nacionales e internacionales.

La exploración de los datos se analizó con los gráficos de estadística descriptiva de la distribución de pasajeros durante los 27 años.

Las gráficas desde la figura 1 a la figura 6, ilustran mejor el comportamiento de los datos con su analítica descriptiva básica.

La información recolectada se agrupo por año luego de un análisis y filtrado para poder visualizar y entender los datos actuales y poder encontrar algunos patrones o tendencias, como se muestra en la figura 1 y figura 2.

Entre las anomalías y valores atípicos se encontraron los correspondientes años a la pandemia mundial COVID 19 donde se muestra una disminución significativa al flujo de pasajeros en algunos años y por motivos a nivel mundial hasta el momento claros y justificables se realiza la omisión de ese conjunto de datos que comprende los años 2019-2020 y 2021

En la gráfica a continuación se realizó la lectura de los datos donde el eje x se observan la fecha y el eje y la cantidad de pasajeros en 4 series de tiempo del flujo correspondiente para las salidas llegadas nacionales e internacionales del aeropuerto El Dorado.

Figura 1

La grafica muestra la evolución de las llegadas y salidas naciones e internacionales en el aeropuerto El Dorado de Bogotá entre los años 1996 al 2023.

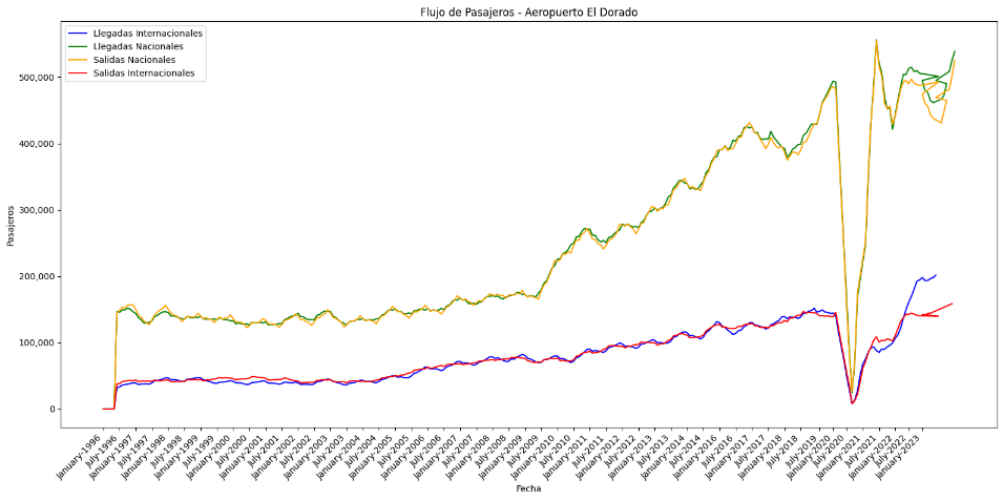
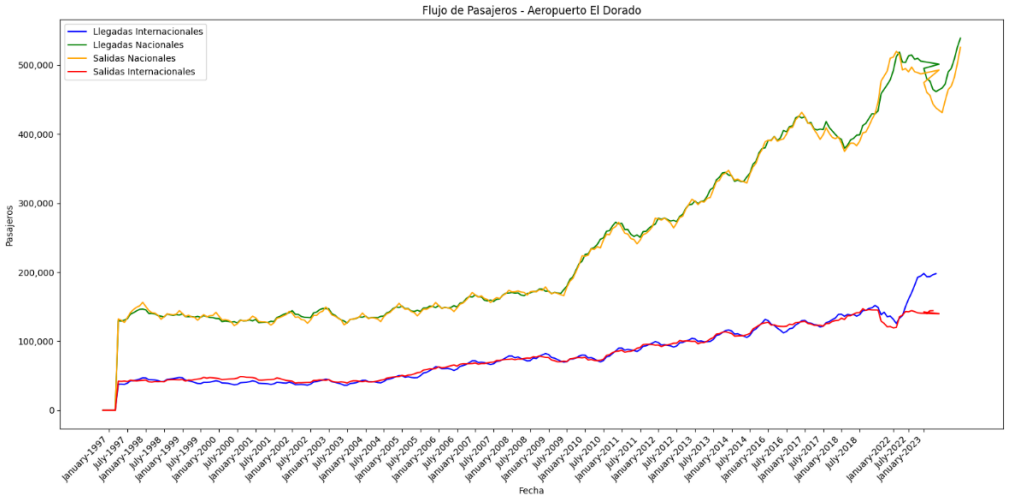


Figura 2



La serie de tiempo evidencia una tendencia creciente en el número de pasajeros desde 1996 hasta 2019, interrumpida temporalmente durante el periodo de pandemia. Se realizo la omisión de los datos que correspondían a la fecha de la pandemia mundial COVID 19 por que se afectó el flujo de pasajeros de forma significativa, a diferencia de la figura 1, esta se muestra con los datos más estables sin fluctuaciones.

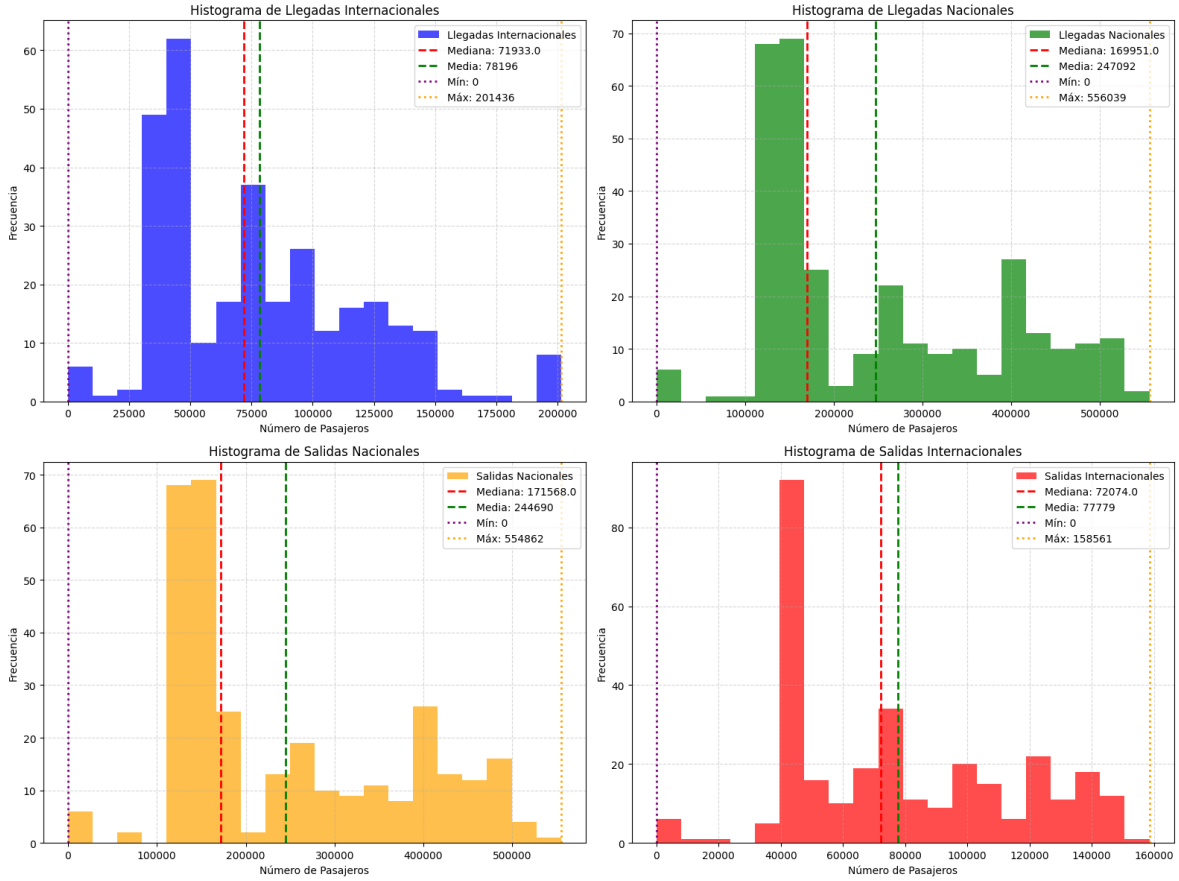


Figura 3

El histograma de las 4 series de tiempo del flujo de pasajeros de las entradas y salidas nacionales e internacionales del aeropuerto El Dorado muestra los datos de los años 1996 al 2023 con estadísticas descriptivas como la media, mediana, max y min.

La distribución de las 4 graficas de la figura 3, corresponde a las series de tiempo analizadas representado en una barra que muestra la frecuencia con ciertos intervalos relacionada la cantidad de pasajeros para los meses con mayor demanda y en ese orden se muestra la media indicando el valor promedio para cada serie de tiempo con el fin de visualizar la tendencia central de los datos analizados. La mediana es generada para poder tener un orden al entender la distribución de los datos si llegamos a tener asimetrías y con lo que efectivamente nos encontramos en las 4 series de tiempo. Los valores máximos y mínimos permiten identificar la amplitud de las fluctuaciones en el flujo de pasajeros, útiles para detectar picos estacionales o caídas abruptas.

En las gráficas de la figura 3, las llegadas y salidas nacionales poseen una similitud de los registros de pasajeros agrupados por cantidad con una tendencia en común con un rango de 0 a 300000 de pasajeros con unas frecuencias de 0 a 70 en intervalos con picos altos en algunos casos pronunciados. Esos cambios pueden ser debidos a las temporadas altas y bajas o eventos especiales. Las llegadas y salidas internacionales indican una ligera simetría con la mayoría de flujo en rangos bajos y pocos picos elevados. También se observa que el flujo de pasajeros de las salidas internacionales es menor al de las llegadas internacionales.

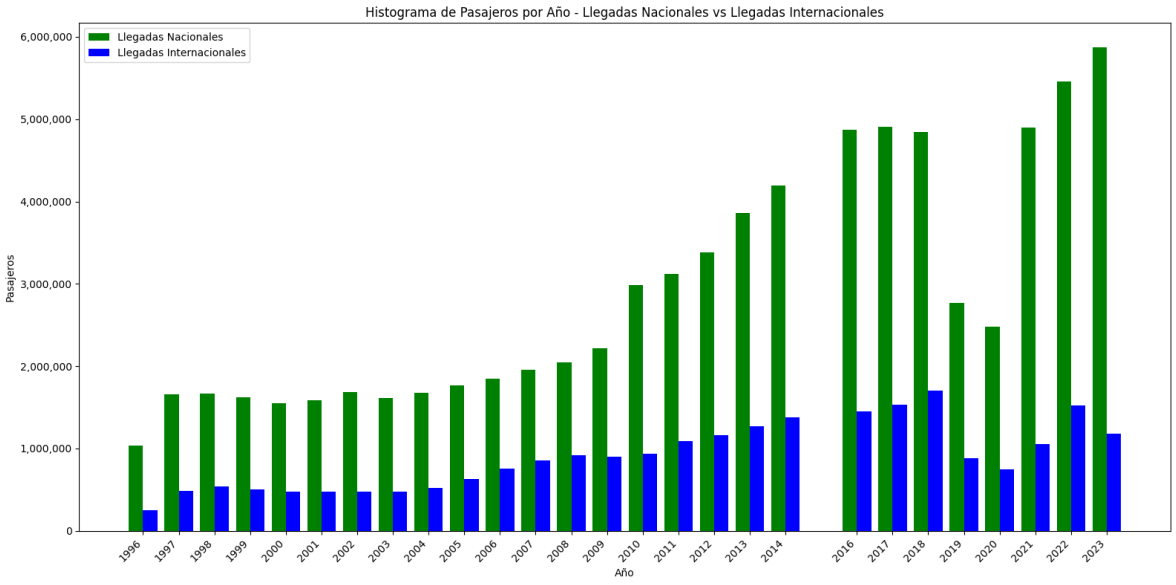


Figura 4

El histograma muestra la cantidad de pasajeros por año con la totalidad de los datos presentes y actualmente recolectados, con unas variaciones en el año 2015 donde no se pudo obtener información compatible y en los años de la pandemia donde los valores atípicos tienen una razón.

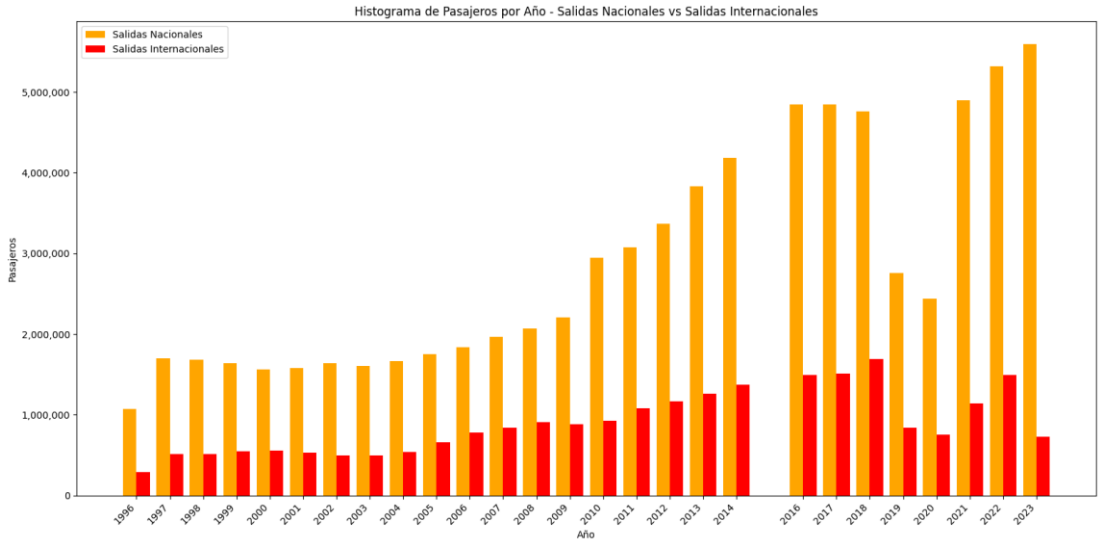
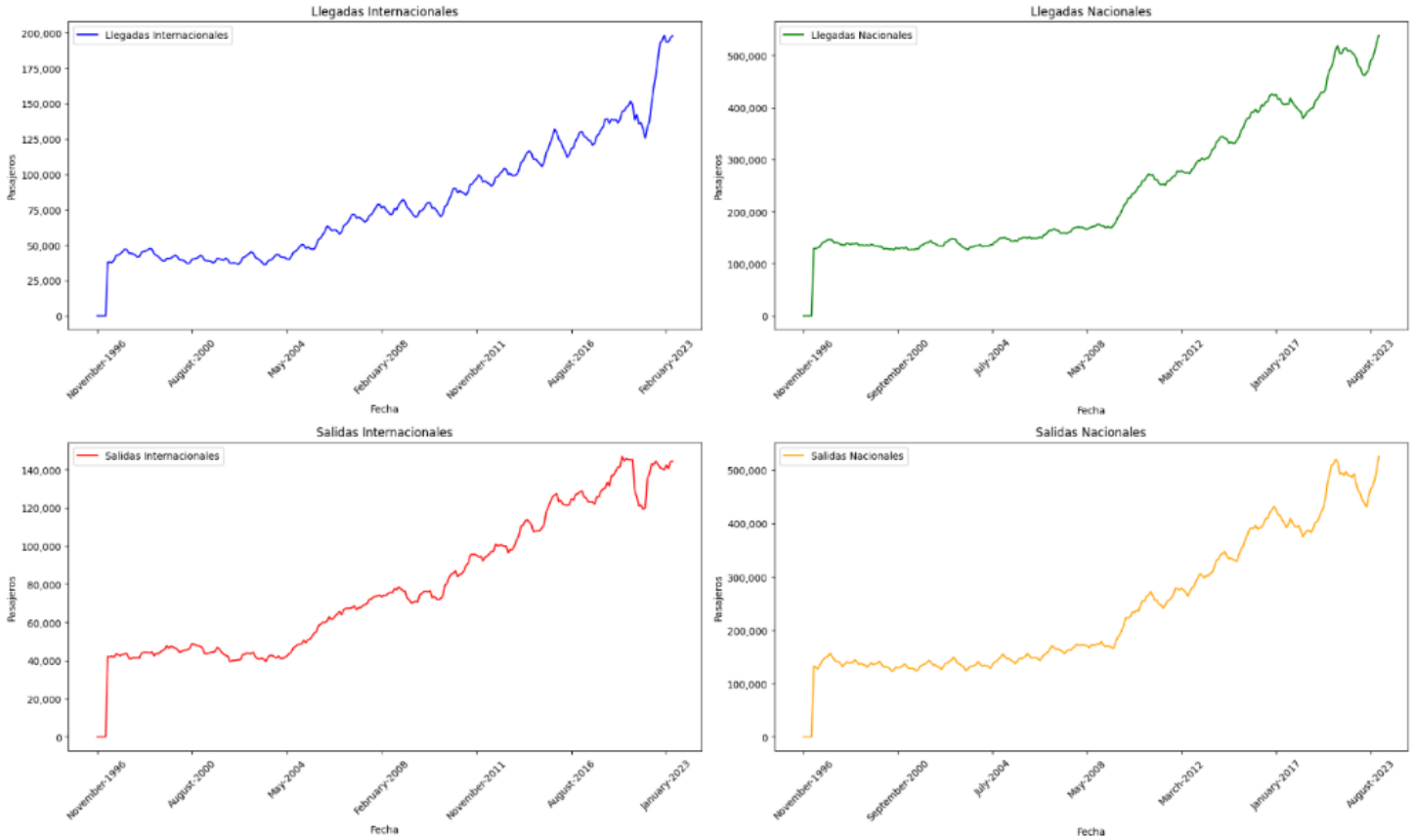


Figura 5

El histograma muestra la cantidad de pasajeros por año con la totalidad de los datos presentes y actualmente recolectados, con unas variaciones en el año 2015 donde no se pudo obtener información compatible y en los años de la pandemia donde los valores atípicos tienen una razón

Figura 6



Las 4 graficas muestran las 4 líneas de tiempo con cada punto y se observa cómo se comporta en el tiempo quitando los puntos y valores atípicos mencionados en las grafica 4 y 5. Obteniendo los datos claro para el proceso de análisis de los datos.

El objetivo principal del análisis es mejorar la planificación operativa del aeropuerto mediante la predicción precisa del flujo de pasajeros, optimizando recursos como personal, mostradores y bandas de equipaje.

Por otro lado, y el enfoque más importante el cual se centra el análisis es la implementación de varios modelos de predicción a este conjunto de datos para comparar su rendimiento y eficiencia en cada uno, teniendo finalmente un argumento basado en estadística y evidencia.

Preparación de los datos

La fase de preparación de los datos, enmarcada dentro de la metodología CRISP-DM, se compone de las siguientes subetapas: selección de datos, limpieza de datos, construcción de variables, integración de fuentes y formateo de datos. Esta etapa fue crucial para asegurar la calidad y homogeneidad del conjunto de datos antes de aplicar los modelos predictivos.

Selección de los datos: Se seleccionaron tres conjuntos de datos históricos sobre tráfico aéreo en Colombia, proporcionados por la Unidad Administrativa Especial de Aeronáutica Civil. Estos archivos, en formato Excel, cubren los periodos 1996–2013, 2014–2019 y 2020–2024. La información incluye variables como año, mes, ciudad de origen, ciudad de destino y número de pasajeros.

Los archivos fueron cargados usando la función pd.read_excel() de la librería Pandas, y posteriormente combinados en un único DataFrame mediante pd.concat().

```
df1 = pd.read_excel('1996-2013.xlsx')
df2 = pd.read_excel('2014-2019.xlsx')
df3 = pd.read_excel('2020-2024.xlsx')
df = pd.concat([df1, df2, df3], ignore_index=True)
```

Limpieza de datos: Debido a inconsistencias en nombres de columnas entre archivos, se estandarizaron los nombres usando

```
df.columns = df.columns.str.upper()
```

Posteriormente, se identificaron y eliminaron valores atípicos mediante la función personalizada remove_outliers(), basada en el rango intercuartil (IQR). Esta técnica permite eliminar registros extremos que podrían distorsionar el análisis y los modelos:

```
def remove_outliers(data, column):
    Q1 = data[column].quantile(0.25)
    Q3 = data[column].quantile(0.75)
    IQR = Q3 - Q1
    return data[(data[column] >= Q1 - 1.5 * IQR) & (data[column] <= Q3 + 1.5 * IQR)]
```

Construcción de variables: Para mejorar el análisis temporal, se creó una variable de fecha combinando año y mes, y se aplicó un promedio móvil de seis meses para suavizar las fluctuaciones mensuales y detectar tendencias a largo plazo:

```
df['PASAJEROS_SUAVIZADOS'] = df['PASAJEROS'].rolling(window=6).mean()
```

También se formatearon los valores para facilitar la interpretación en gráficas usando:

```
def format_passengers(x, pos):
    return f'{x:,.0f}'
```

Integración de datos: Los datos se integraron en función de la categoría del vuelo (nacional o internacional), utilizando filtros por ciudades de origen y destino. Se crearon listas de aeropuertos internacionales (international_airports) y colombianos (colombia_airports), y se segmentó el DataFrame en función de estas condiciones:

```
df_internacionales = df[df['ORIGEN'].isin(international_airports) | df['DESTINO'].isin(internatic
df_nacionales = df[df['ORIGEN'].isin(colombia_airports) & df['DESTINO'].isin(colombia_airports)]
```

Formateo de datos: Para facilitar los análisis temporales y modelado, los datos fueron agregados por año y mes usando groupby() y sum():

```
df_grouped = df.groupby(['AÑO', 'MES']).sum().reset_index()
```

Visualización de los datos: Se usó la librería Matplotlib para visualizar las series de tiempo de llegadas y salidas (nacionales e internacionales) con ejes X (tiempo) e Y (número de pasajeros).

Evaluación y validación de modelos: Los modelos predictivos SARIMA, ARIMA y LSTM fueron implementados utilizando las funciones SARIMAX(), ARIMA() (de statsmodels) y Keras para redes neuronales. La arquitectura del modelo LSTM incluyó capas Sequential, LSTM y Dense:

```
model = Sequential()
model.add(LSTM(50, return_sequences=True, input_shape=(X_train.shape[1], 1)))
model.add(LSTM(50))
model.add(Dense(1))
```

Validación cruzada temporal: Se utilizó TimeSeriesSplit de scikit-learn, una técnica adecuada para series temporales ya que respeta la secuencia cronológica de los datos, a diferencia de la validación cruzada aleatoria, que rompería la dependencia temporal:

```
from sklearn.model_selection import TimeSeriesSplit


tscv = TimeSeriesSplit(n_splits=5)
```

Las métricas de evaluación fueron el RMSE (Root Mean Squared Error) y MAE (Mean Absolute Error), calculadas con mean_squared_error() y mean_absolute_error():

```
from sklearn.metrics import mean_squared_error, mean_absolute_error
```

La elección de esta técnica se fundamenta en su capacidad para simular escenarios reales en predicciones futuras, reduciendo el riesgo de sobreajuste (Bergmeir & Benítez, 2012)

Modelado de los datos

El objetivo de esta etapa fue seleccionar, configurar y entrenar modelos adecuados para predecir el flujo de pasajeros en el Aeropuerto El Dorado, con base en datos históricos mensuales entre 1996 y 2024. Se implementaron tres enfoques principales para series temporales: ARIMA, SARIMA y LSTM, seleccionados por su capacidad para modelar relaciones lineales, estacionales y no lineales, respectivamente.

Selección de modelos consistió en los siguientes descritos a continuación:

ARIMA

El modelo ARIMA (AutoRegressive Integrated Moving Average) es apropiado para series temporales estacionarias o que pueden hacerse estacionarias mediante diferenciación. Este modelo combina tres componentes: el autorregresivo (p), el integrado (d) y la media móvil (q). En este estudio se configuró con parámetros (5, 1, 0), definidos tras un análisis de autocorrelación (ACF) y autocorrelación parcial (PACF), y mediante la función auto_arima() de la librería **pmdarima**, que selecciona el modelo óptimo minimizando los criterios AIC y . (R. J. Hyndman & Khandakar, 2008)

SARIMA

Para capturar la estacionalidad evidente en los datos —por ejemplo, las variaciones durante vacaciones o temporada alta—, se utilizó el modelo SARIMA, una extensión del modelo ARIMA con términos estacionales. Se aplicó la configuración (1, 1, 1) × (1, 1, 1, 12), considerando la estacionalidad anual. Este tipo de modelo ha demostrado ser útil en entornos como aeropuertos, donde los patrones de tráfico de pasajeros son cíclicos . (R. J. Hyndman & Khandakar, 2008)

LSTM

El modelo LSTM (Long Short-Term Memory) pertenece a la familia de redes neuronales recurrentes (RNN), y está diseñado específicamente para trabajar con datos secuenciales como las series temporales. En este proyecto se implementó usando la librería **Keras**, con una arquitectura que incluyó dos capas LSTM y una capa densa de salida. Se entrenó con secuencias de 12 meses para capturar relaciones de largo plazo en los datos. (R. J. Hyndman & Khandakar, 2008)

Division del conjunto de datos: Los datos fueron divididos en un 80 % para entrenamiento y un 20 % para validación, manteniendo la secuencia cronológica. Se implementó la técnica **Time Series Split** con k=5, respetando el orden temporal y evitando el uso de validación cruzada aleatoria, que no es adecuada para este tipo de datos. Esta técnica es recomendada para evitar fuga de información y se alinea con lo propuesto por Bergmeir y Benítez (2012) sobre validación cruzada en series de tiempo. (Bergmeir & Benítez, 2012)

La evaluación y desempeño consistió en utilizar dos métricas para evaluar el rendimiento de los modelos:

- **RMSE (Root Mean Squared Error)**: penaliza los errores grandes y proporciona una estimación robusta del ajuste.
- **MAE (Mean Absolute Error)**: ofrece una interpretación directa del error promedio.

Estas métricas se calcularon utilizando las funciones mean_squared_error() y mean_absolute_error() de la librería scikit-learn. Este enfoque metodológico está en línea con estudios como los de .(Donate et al., 2013). (Neunhoeer & Sternberg, n.d.) y quienes destacan la importancia de evaluar modelos con múltiples métricas para mejorar la robustez del análisis.

La combinación de ARIMA y SARIMA permitió contrastar el rendimiento de modelos lineales con y sin estacionalidad. Por su parte, el modelo LSTM ofreció una alternativa no lineal capaz de adaptarse a patrones más complejos e inestables, como los ocasionados por la pandemia o eventos extraordinarios. Esta variedad de enfoques permitió explorar diferentes dinámicas presentes en los datos, optimizando la capacidad de predicción en distintos escenarios temporales.

Evaluación

Prueba de Estacionariedad (ADF)

la prueba de Dickey-Fuller aumentada (ADF) es una técnica de evaluación estadística, su uso se ubica metodológicamente dentro de la etapa de preparación de los datos en el ciclo CRISP-DM, ya que su función principal es verificar el supuesto de estacionariedad en las series temporales antes de aplicar modelos. (D. G. Dickey, 2011)

Para cada una de las cuatro series temporales analizadas —llegadas internacionales, llegadas nacionales, salidas internacionales y salidas nacionales— se aplicó esta prueba. En todos los casos, el valor p fue superior al umbral comúnmente utilizado de 0.05, por lo que no se pudo rechazar la hipótesis nula, lo cual indica que las series presentan una raíz unitaria y no son estacionarias. Esto justifica la necesidad de diferenciación para aplicar modelos como ARIMA y SARIMA, que requieren estacionariedad en la serie.

Resultados de la Prueba ADF

Llegadas internacionales:

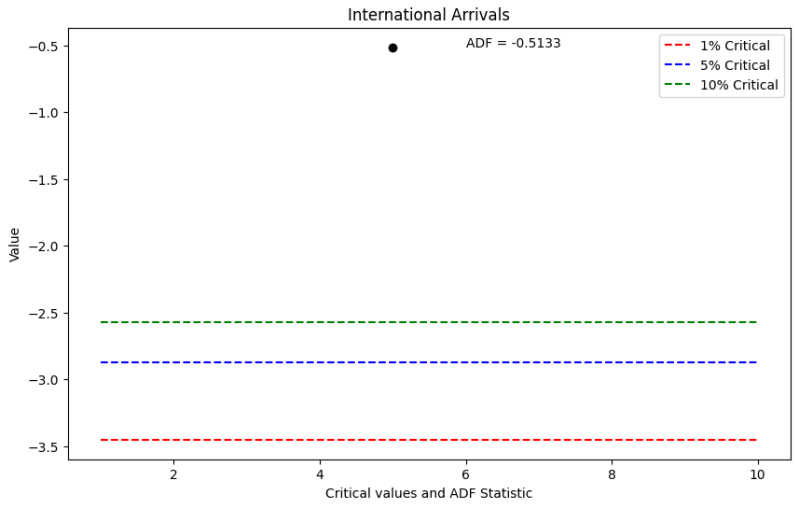
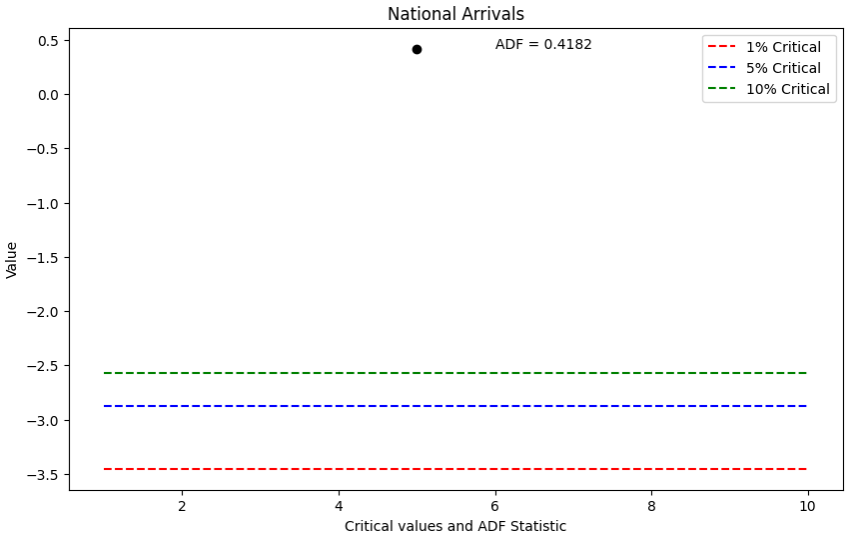


Figura 7

- **Estadístico ADF**: -0.5133
- **Valor p**: 0.8894
- **Conclusión**: no se rechaza la hipótesis nula, no es estacionaria.



Llegadas nacionales:

Figura 8

- **Estadístico ADF**: 0.4182
- **Valor p**: 0.9822
- **Conclusión**: no se rechaza la hipótesis nula, no es estacionaria.

Salidas internacionales:

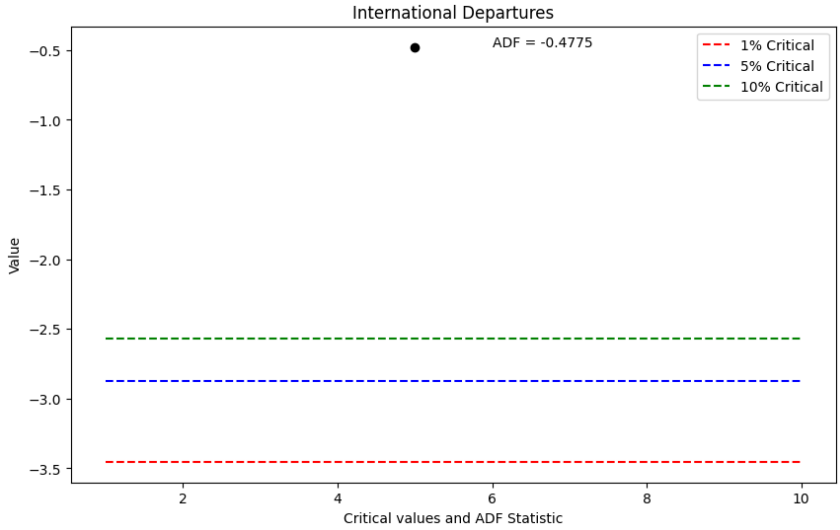


Figura 9

- **Estadístico ADF**: -0.4775
- **Valor p**: 0.8963
- **Conclusión**: no se rechaza la hipótesis nula, no es estacionaria.

Salidas nacionales:

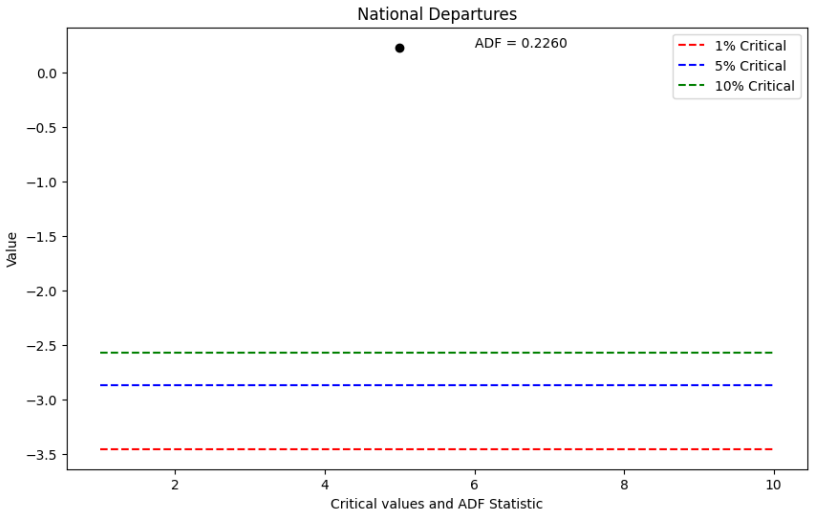


Figura 10

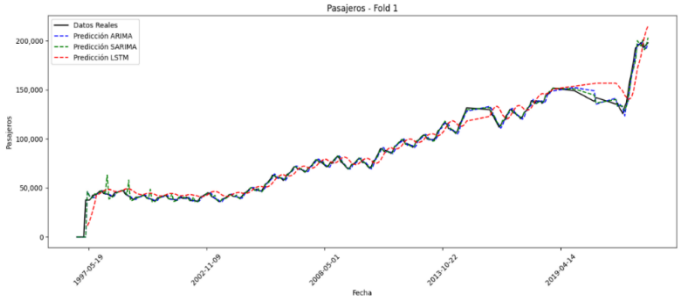
- **Estadístico ADF**: 0.2260
- **Valor p**: 0.9737
- **Conclusión**: no se rechaza la hipótesis nula, no es estacionaria

Para todas las series de tiempo los valores fueron mayores a 0.05 por lo tanto no se puede rechazar la hipótesis nula y que se posee una raíz unitaria como dice el teorema para inferir que es estacionaria.

Con la prueba de la prueba de dickey Fuller y como se puede observar en las gráficas 7,8,9 y 10 las líneas rojas azules y verdes representan los valores críticos y niveles de 1%, 5% y 10% y el punto negro ADF como se observa en todas las gráficas está por encima de los valores quiere decir que la serie no es estacionaria. (D. A. Dickey & Fuller, 1979)

Los siguientes graficas de predicción vs. realidad para cada modelo (SARIMA, ARIMA, LSTM), se muestra cada modelo en cuatro líneas de tiempo como lo son llegadas salidas nacionales e internacionales

Llegadas internacionales



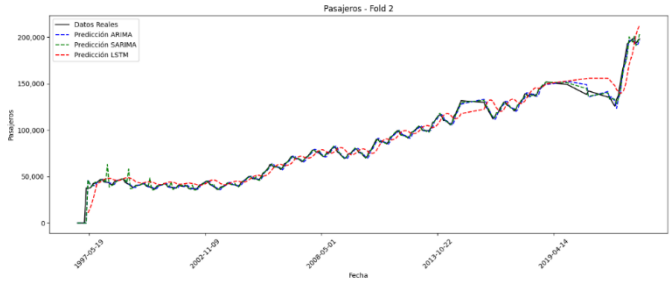


Figura 11

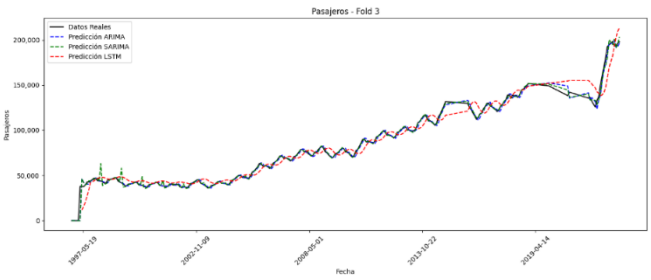


Figura 12

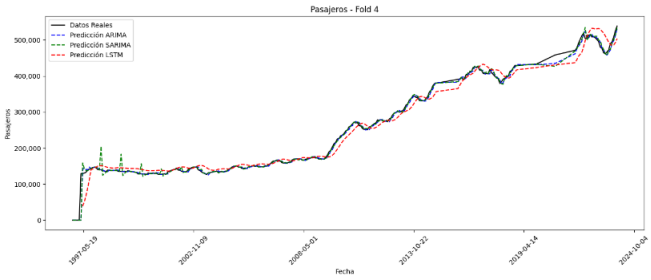


Figura 13

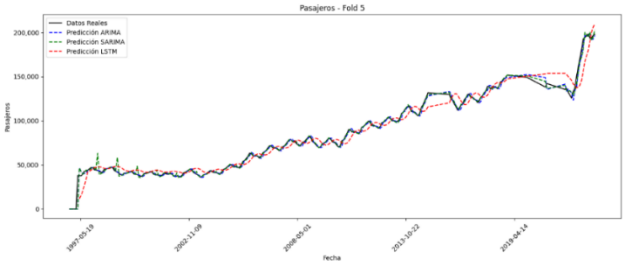


Figura 14

Figura 15

La figura 7 a la figura 11, muestra la implementación de ARIMA como el modelo más adecuado para las llegadas internacionales con valores RMSE y MAE bajos. SARIMA tiene un desempeño similar a ARIMA. El modelo LSTM muestra que las redes neuronales para estos casos son la mejor opción esto se comprueba con los siguientes datos.

ARIMA: muestra estabilidad en los 5 folds con los mismos valores de RMSE y MAE con valores de error bajos lo que indica un buen ajuste de precisión en los datos.

SARIMA: Los resultados que muestra son bajos en su RMSE y mas bajo en MAE, indicando que la estacionalidad no mejora los resultados.

LSTM: Los valores de RMSE y MAE son altos a diferencia de SARIMA y ARIMA indicando que la precisión para las llegadas internacionales no tiene un buen rendimiento.

Tabla 1 .ARIMA

Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
Fold 1	3445.43	1945.03	49	49	45	45
Fold 2	3445.43	1945.03	49	49	45	45
Fold 3	3445.43	1945.03	49	49	45	45
Fold 4	3445.43	1945.03	49	49	45	45
Fold 5	3445.43	1945.03	49	49	45	45

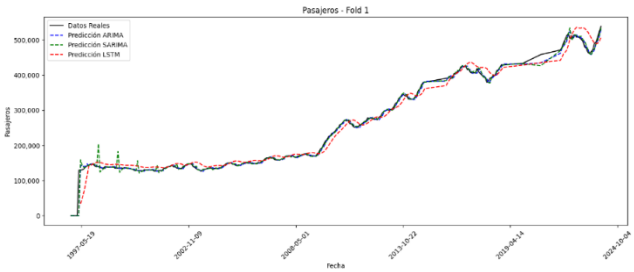
Tabla 2 .SARIMA

Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
Fold 1	3439.82	1428.91	94	94	45	45
Fold 2	3439.82	1428.91	94	94	45	45
Fold 3	3439.82	1428.91	94	94	45	45
Fold 4	3439.82	1428.91	94	94	45	45
Fold 5	3439.82	1428.91	94	94	45	45

Tabla 3. LSTM

Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
Fold 1	7809.65	5689.12	94	94	45	45
Fold 2	7709.97	5570.32	94	94	45	45
Fold 3	7726.59	5527.15	94	94	45	45
Fold 4	7726.59	5527.15	94	94	45	45
Fold 5	7726.59	5527.15	94	94	45	45

Llegadas nacionales



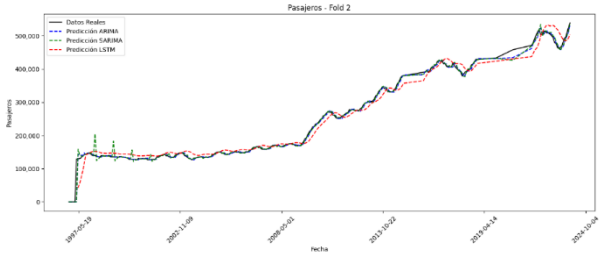


Figura 16

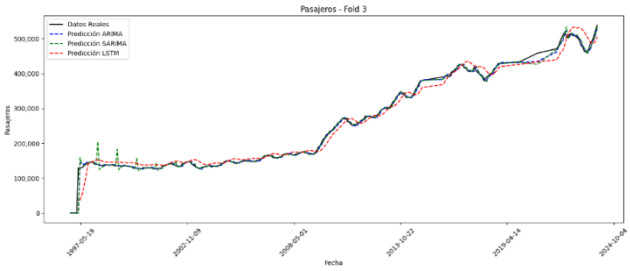


Figura 17

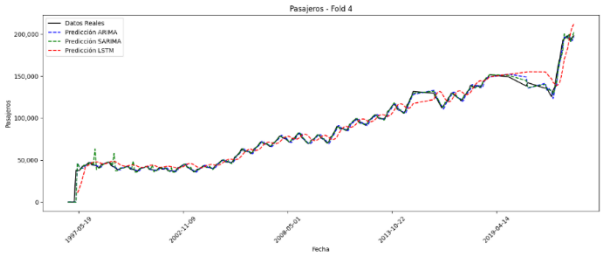


Figura 18

Figura 19

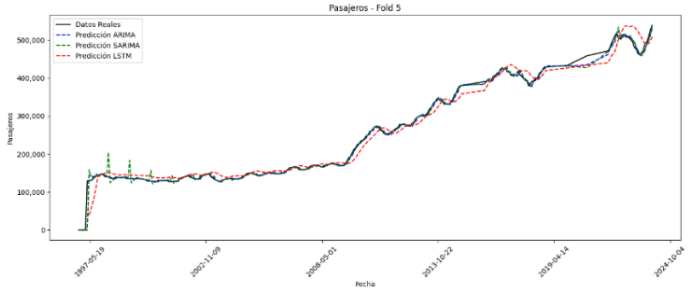


Figura 20

La figura 12 a la figura 16 muestran en sus graficas los puntos y los resultados de los modelos predictivos aplicados donde el modelo ARIMA es el modelo más adecuado para las llegadas nacionales por sus valores bajos mostrando mejor desempeño que los modelos de SARIMA y LSTM.
ARIMA: los resultados obtenidos para este modelo son consistentes en los 5 folds con los valores de RMSE y MAE constantes, indicando que ARIMA tienen mejor desempeño en la predicción de este conjunto de datos.

SARIMA: Este modelo presenta valores RMSE y MAE más altos que el modelo ARIMA lo que indica que la estacionalidad que incorpora este modelo en sus parámetros no mejora el proceso de precisión de los resultados.

LSTM: Los resultados para el modelo fueron mucho más altos que los otros modelos evaluados considerandosen como los valores más altos con el desempeño no eficiente para este conjunto de datos.

Tabla 4 .ARIMA

Modelo	Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
ARIMA	1	9186.02	3842.94	51	51	46	46
ARIMA	2	9186.02	3842.94	97	97	46	46
ARIMA	3	9186.02	3842.94	143	143	46	46
ARIMA	4	9186.02	3842.94	189	189	46	46
ARIMA	5	9186.02	3842.94	235	235	46	46

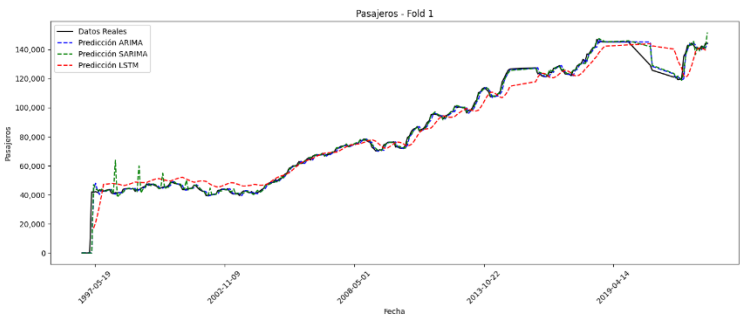
Tabla 5 .SARIMA

Modelo	Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
SARIMA	1	10501.46	3774.42	51	51	46	46
SARIMA	2	10501.46	3774.42	97	97	46	46
SARIMA	3	10501.46	3774.42	143	143	46	46
SARIMA	4	10501.46	3774.42	189	189	46	46
SARIMA	5	10501.46	3774.42	235	235	46	46

Tabla 6 .LSTM

Modelo	Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
LSTM	1	17593.58	12315.10	51	51	46	46
LSTM	2	17383.71	12784.30	97	97	46	46
LSTM	3	17458.19	12437.08	143	143	46	46
LSTM	4	17588.93	12559.53	189	189	46	46
LSTM	5	17589.72	12412.18	235	235	46	46

Salidas internacionales



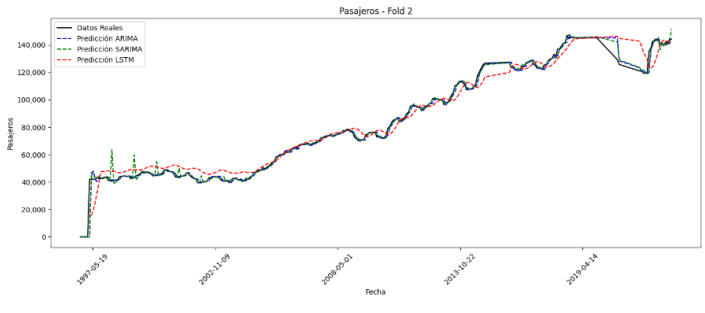


Figura 21

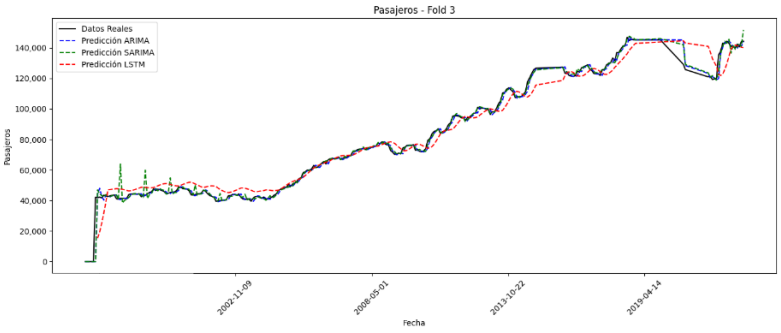


Figura 22

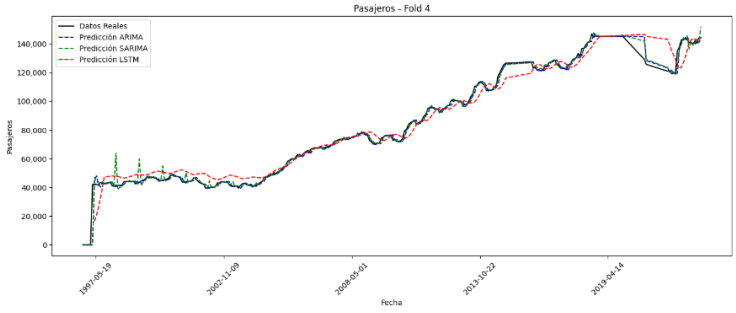


Figura 23

Figura 24

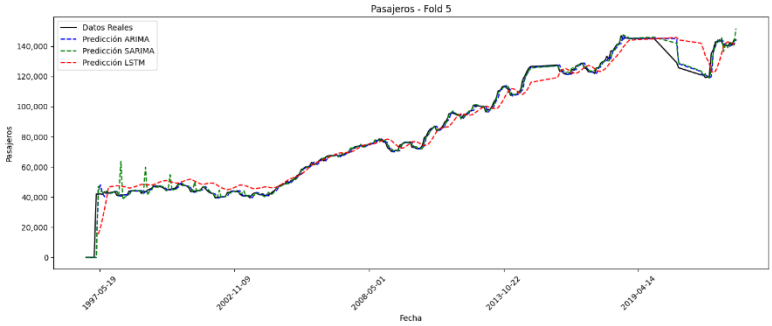


Figura 25

Las figuras 16 a la figura 21 muestran los resultados y puntos para los modelos aplicados al conjunto de datos en este caso para las salidas internacionales el modelo de predicción ARIMA fue el modelo más apropiado por sus valores de error más bajos.

ARIMA: En los 5 folds los resultados fueron constantes lo que proporciona una estabilidad en el modelo siendo eficiente.

SARIMA: presenta en los resultados un rendimiento similar a ARIMA, pero con los valores un poco mas altos, por lo tanto, para este conjunto de datos la estacionalidad no mejora la predicción.

LSTM: Los resultados de este modelo fueron los resultados mas altos obtenidos a comparación de ARIMA y SARIMA por lo que para este conjunto de datos este modelo fue el menos eficiente.

Tabla 7 .ARIMA

Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
1	3286.12	1388.83	49	49	45	45
2	3286.12	1388.83	94	94	45	45
3	3286.12	1388.83	139	139	45	45
4	3286.12	1388.83	184	184	45	45
5	3286.12	1388.83	229	229	45	45

Tabla 8 .SARIMA

Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
1	3535.72	1360.44	49	49	45	45
2	3535.72	1360.44	94	94	45	45
3	3535.72	1360.44	139	139	45	45
4	3535.72	1360.44	184	184	45	45
5	3535.72	1360.44	229	229	45	45

Tabla 9 .LSTM

Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
1	6161.02	4594.02	49	49	45	45
2	6031.93	4449.82	94	94	45	45
3	6041.16	4415.89	139	139	45	45
4	6009.69	4406.37	184	184	45	45
5	5918.03	4273.41	229	229	45	45

Salidas nacionales

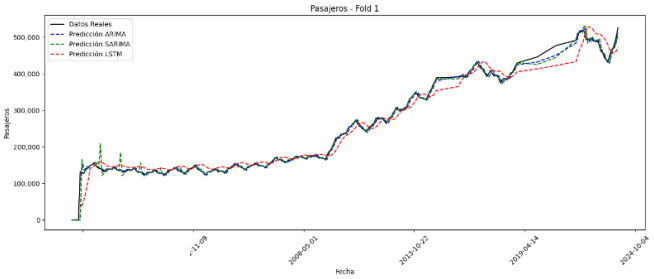


Figura 26

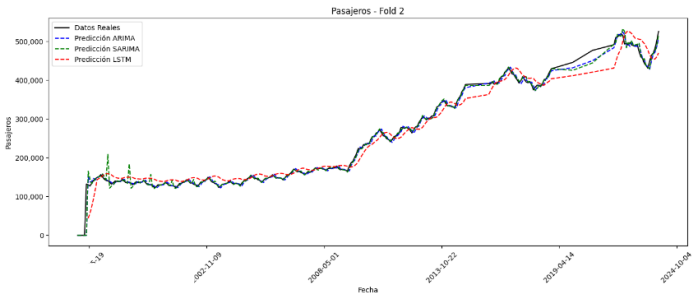


Figura 27

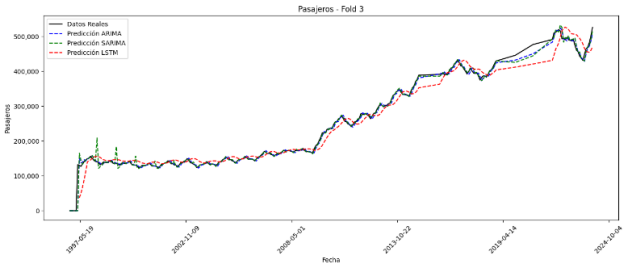


Figura 28

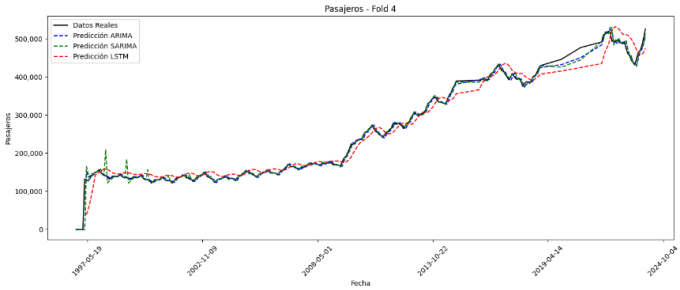


Figura 29

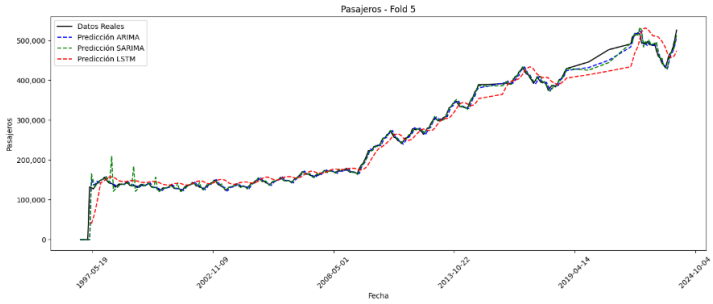


Figura 30

Las figuras 22 a la figura 26 muestran el comportamiento y los resultados de la implementación de los modelos de predicción ARIMA, SARIMA y LSTM para el conjunto de datos de las salidas nacionales.

ARIMA: Los valores de RMSE y MAE en los 5 folds fueron constantes y estables lo que quiere decir una buena precisión en la implementación de este modelo en el conjunto de datos evaluado.

SARIMA: Los resultados de este modelo y la incorporación de la estacionalidad en el modelo no garantiza mejores predicciones, sus resultados fueron altos y el modelo ARIMA supera la estabilidad de los datos.

LSTM: los valores de RMSE y MAE fueron altos lo cual para este conjunto de datos la implementación de este modelo no es la más adecuada.

Tabla 10. ARIMA

Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
Fold 1	9969.67	5006.26	51	51	46	46
Fold 2	9969.67	5006.26	97	97	46	46
Fold 3	9969.67	5006.26	143	143	46	46
Fold 4	9969.67	5006.26	189	189	46	46
Fold 5	9969.67	5006.26	235	235	46	46

Tabla 11 .SARIMA

Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
Fold 1	11057.34	4143.02	51	51	46	46
Fold 2	11057.34	4143.02	97	97	46	46
Fold 3	11057.34	4143.02	143	143	46	46
Fold 4	11057.34	4143.02	189	189	46	46
Fold 5	11057.34	4143.02	235	235	46	46

Tabla 12 .LSTM

Fold	RMSE	MAE	Meses de Entrenamiento	Puntos de Entrenamiento	Meses de Validación	Puntos de Validación
Fold 1	19562.92	14313.34	51	51	46	46
Fold 2	19574.37	14747.38	97	97	46	46
Fold 3	19546.85	14010.63	143	143	46	46
Fold 4	19655.65	14395.59	189	189	46	46
Fold 5	19595.99	14366.64	235	235	46	46

En las figuras 7 a 27, que muestran los resultados de las salidas nacionales e internacionales, la comparación de los modelos revela que SARIMA y ARIMA son más efectivos para predecir el flujo de pasajeros en todas las rutas analizadas. Ambos modelos generan valores de RMSE y MAE más estables, con menos variaciones en comparación con otros enfoques.

En términos de RAE, el modelo SARIMA presenta mejores resultados que ARIMA. La inclusión de la estacionalidad en el modelo SARIMA, adaptada a este conjunto de datos, contribuye significativamente a la mejora de la precisión en las predicciones.

Por otro lado, el modelo LSTM, aunque es adecuado para el análisis de series temporales, no es la mejor opción para este conjunto de datos. Los errores generados por LSTM son elevados, y no se observa una mejora significativa al aumentar la cantidad de datos de entrenamiento, lo que limita su efectividad en este caso.

Además, los modelos de predicción se aplican a una ruta específica, Bogotá-Nueva York, con el objetivo de observar el comportamiento del flujo de pasajeros en un grupo de datos más reducido y específico.

.

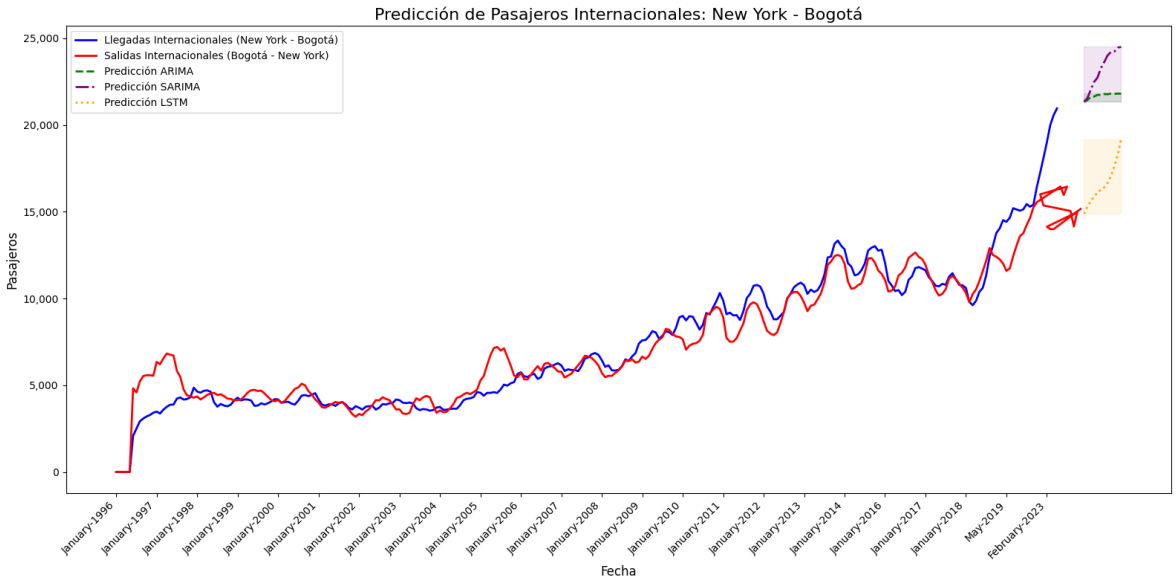


Figura 31 *muestra el comportamiento de los modelos ARIMA, SARIMA y LSTM en el tiempo*

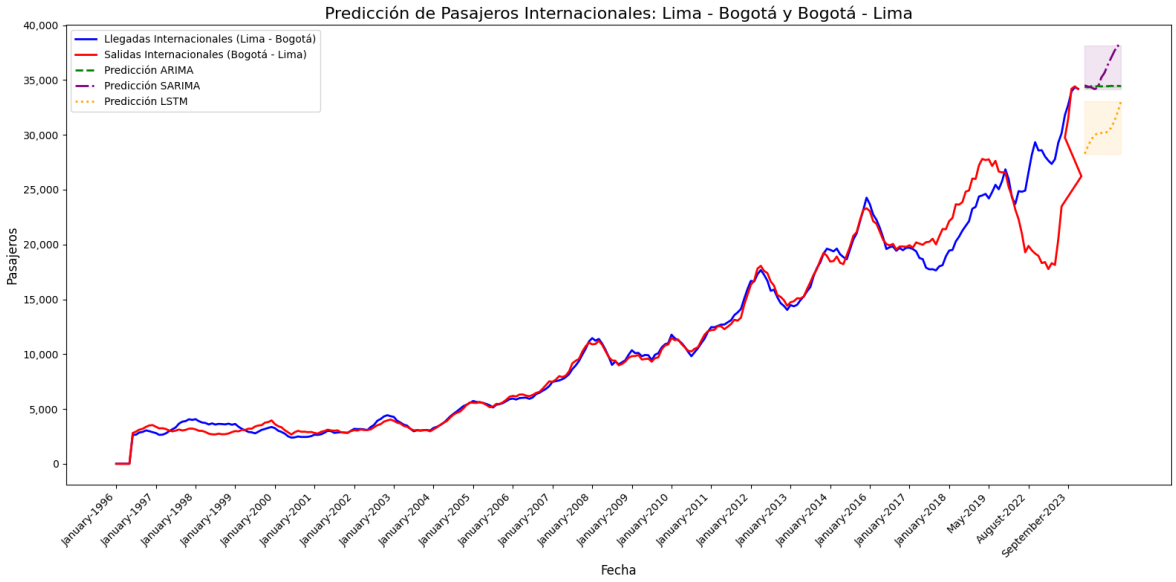


Figura 32 *flujo de pasajeros Bogotá-Lima con los modelos de predicción SARIMA, ARIMA, LSTM.*

Discusiones y observaciones

Para la evaluación del rendimiento de los modelos predictivos ARIMA, SARIMA y LSTM en la estimación del flujo de pasajeros del Aeropuerto El Dorado (Bogotá) entre los años 1996 y 2024, se diseñó un marco metodológico que incluyó técnicas de validación cruzada adaptadas a series temporales. Este enfoque buscó garantizar una comparación justa entre modelos al mantener la dependencia temporal de los datos durante la validación, como se sugiere en trabajos como Hyndman & Athanasopoulos (2021)

Se implementa una estrategia metodológica estructurada que integra análisis exploratorio, preprocesamiento de datos, modelado, validación y visualización de resultados. Aunque inicialmente se hace referencia a una metodología denominada “XLM”, este término no corresponde a una técnica reconocida en el ámbito del modelado de series temporales. En su lugar, el enfoque adoptado en este proyecto puede entenderse como una extensión del flujo tradicional de aprendizaje automático, adaptado específicamente al análisis de datos temporales con fines comparativos.

Los resultados obtenidos se integran en un tablero interactivo (dashboard) desarrollado en Microsoft Power BI (ver Figura 29), el cual permite a los usuarios explorar de forma visual las predicciones y tendencias generadas por los modelos. Esta herramienta de visualización facilita una toma de decisiones más ágil y fundamentada para los responsables de la gestión operativa del Aeropuerto El Dorado.

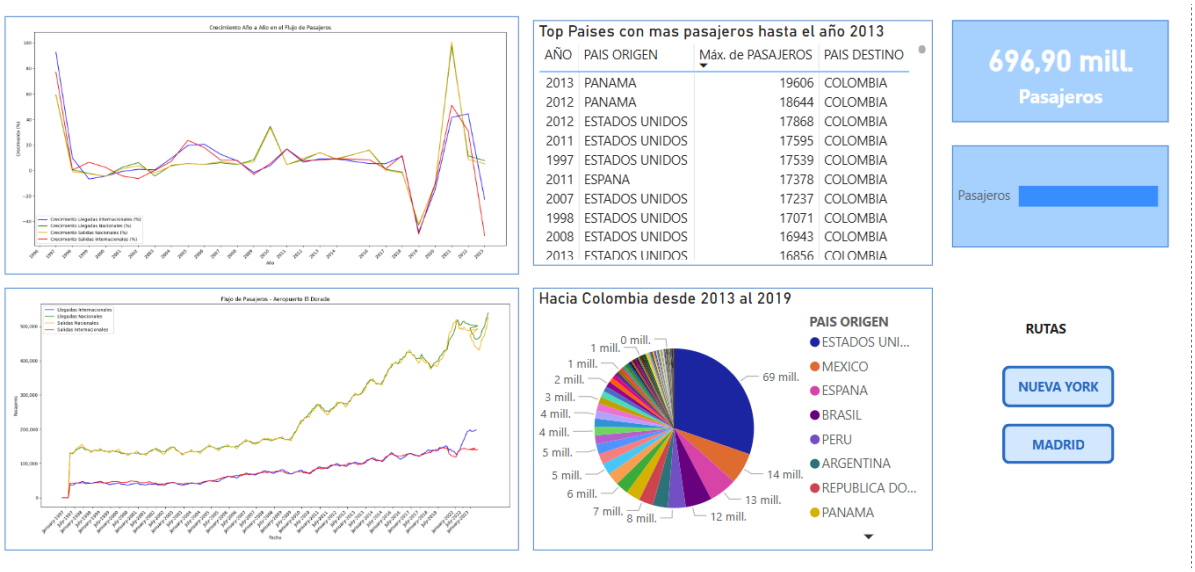


Figura 33

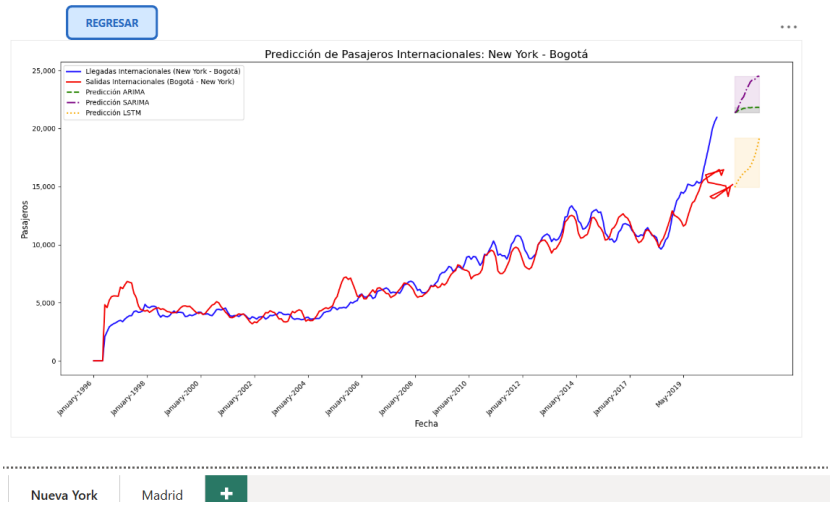


Figura 34

Finalmente se plantea como implementación futura la integración de un chatbot basado en lenguaje natural que ofrezca acceso automatizado a estas predicciones. Esto permitiría consultas rápidas y accesibles para diferentes tipos de usuarios interesados en el comportamiento proyectado del tráfico aéreo.

Conclusiones

Los resultados obtenidos en este estudio demuestran que el modelo SARIMA ofrece un mejor desempeño en la predicción del flujo de pasajeros en el Aeropuerto El Dorado en comparación con los modelos ARIMA y LSTM. Esto se debe a que SARIMA incorpora componentes estacionales que capturan de manera más precisa los patrones recurrentes observados en los datos históricos, lo cual es especialmente relevante en series temporales con comportamientos cíclicos como el tráfico aéreo anual.

Por otro lado, el modelo LSTM, a pesar de su capacidad para modelar secuencias complejas y dependencias a largo plazo, no alcanza el mismo nivel de precisión en este caso. Esto puede atribuirse a varios factores, entre ellos, el tamaño del conjunto de datos disponible, la sensibilidad del modelo a la configuración de hiperparámetros y la necesidad de una mayor cantidad de datos para aprovechar plenamente su potencial. Además, la naturaleza estacional y relativamente estructurada de los datos favorece enfoques estadísticos clásicos como SARIMA.

Estos resultados pueden servir como base para el desarrollo de sistemas interactivos de apoyo a la toma de decisiones en la gestión operativa del aeropuerto, al integrar modelos precisos de predicción en herramientas visuales como dashboards o asistentes inteligentes. La implementación de estas soluciones puede facilitar la planificación de recursos, la asignación de personal y la optimización de operaciones aeroportuarias de manera más eficiente.

Finalmente, se recomienda extender esta metodología a otros aeropuertos o sistemas de transporte con dinámicas similares, así como explorar modelos híbridos que combinen enfoques estadísticos y de aprendizaje profundo. Esto permitiría mejorar la robustez y generalización de los modelos predictivos, contribuyendo al diseño de herramientas analíticas más completas y adaptativas para la gestión de infraestructuras de transporte en diferentes contextos.

Referencias Bibliográficas

1. Instituto Distrital de Turismo. (2024). *El Dorado, el aeropuerto mejor conectado de América Latina y destaca en el top 20 global*. https://www.idt.gov.co/es/el-dorado-el-aeropuerto-mejor-conectado-de-america-latina-y-destaca-en-el-top-20-global
2. Shearer, C. (2000). *The CRISP-DM model: The new blueprint for data mining*. Journal of Data Warehousing, 5(4), 13–22
3. Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (n.d.). *The "K" in K-fold Cross Validation*. Retrieved March 22, 2025, from http://www.i6doc.com/en/livre/?GCOI=28001100967420.
4. *Bases de Datos*. (n.d.). Retrieved March 22, 2025, from https://www.aerocivil.gov.co/atencion/estadisticas-de-las-actividades-aeronauticas/bases-de-datos
5. Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–213. https://doi.org/10.1016/J.INS.2011.12.028
6. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427. https://doi.org/10.2307/2286348
7. Dickey, D. G. (2011). Dickey-Fuller Tests. *International Encyclopedia of Statistical Science*, 385–388. https://doi.org/10.1007/978-3-642-04898-2_210
8. Donate, J. P., Cortez, P., Sánchez, G. G., & De Miguel, A. S. (2013). Time series forecasting using a weighted cross-validation evolutionary artificial neural network ensemble. *Neurocomputing*, *109*, 27–32. https://doi.org/10.1016/J.NEUCOM.2012.02.053
9. *El Aeropuerto El Dorado continúa recibiendo premios por esta razón | Infraestructura | Economía | Portafolio*. (n.d.). Retrieved March 22, 2025, from https://www.portafolio.co/economia/infraestructura/el-aeropuerto-el-dorado-continua-recibiendo-premios-por-esta-razon-623809
10. *Forecasting: principles and practice - Rob J Hyndman, George Athanasopoulos - Google Libros*. (n.d.). Retrieved March 22, 2025, from https://books.google.com.co/books?hl=es&lr=&id=_bBhDwAAQBAJ&oi=fnd&pg=PA7&dq=10.+Hyndman,+R.+J.,+%26+Athanasopoulos,+G.+(2018).+Forecasting:+Principles+and+Practice+(2nd+ed.).+OTexts.&ots=TjiXAkTPLL&sig=QXHDvy8FJ9GvKKjFvnFwXXV3yxQ#v=onepage&q&f=false
11. Guimarães, M., Soares, C., & Ventura, R. (2022). Decision Support Models for Predicting and Explaining Airport Passenger Connectivity From Data. *IEEE Transactions on Intelligent Transportation Systems*, *23*(9), 16005–16015. https://doi.org/10.1109/TITS.2022.3147155
12. Guo, X., Grushka-Cockayne, Y., & De Reyck, B. (2018). *Forecasting Airport Transfer Passenger Flow Using Real-Time Data and Machine Learning*.

13. Harris, R. I. D. (1992). Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test. *Economics Letters*, *38*(4), 381–386. https://doi.org/10.1016/0165-1765(92)90022-Q
14. Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, *27*(3), 1–22. https://doi.org/10.18637/JSS.V027.I03
15. Hyndman, R., Koehler, A., Ord, K., & Snyder, R. (2008). *Forecasting with Exponential Smoothing*. https://doi.org/10.1007/978-3-540-71918-2
16. Laik, M. N., Choy, M., & Sen, P. (2014). Predicting airline passenger load: A case study. *Proceedings - 16th IEEE Conference on Business Informatics, CBI 2014*, *1*, 33–38. https://doi.org/10.1109/CBI.2014.39
17. Li, Z., Bi, J., & Li, Z. (2017). Passenger Flow Forecasting Research for Airport Terminal Based on SARIMA Time Series Model. *IOP Conference Series: Earth and Environmental Science*, *100*(1). https://doi.org/10.1088/1755-1315/100/1/012146
18. Neunhoeer, M., & Sternberg, S. (n.d.). *How Cross-Validation Can Go Wrong and What to Do About it*. https://doi.org/10.7910/DVN/Y9KMJW
19. Orsini, F., Gastaldi, M., Mantecchini, L., & Rossi, R. (2019). *Neural networks trained with WiFi traces to predict airport passenger behavior*. 1–7. https://doi.org/10.1109/MTITS.2019.8883365
20. Wilson, G. T. (2016). Time Series Analysis: Forecasting and Control, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1. *Journal of Time Series Analysis*, *37*(5), 709–711. https://doi.org/10.1111/JTSA.12194
21. *World Airport Awards | SKYTRAX*. (n.d.). Retrieved March 22, 2025, from https://www.worldairportawards.com/es/
22. Aviacionline. (2025, febrero 4). *El Dorado de Bogotá supera al AICM de México y se convierte en el aeropuerto con más pasajeros de América Latina*. Recuperado de https://www.aviacionline.com/pendiente-bogota-el-dorado-crece-y-desbanca-al-aicm-de-ciudad-de-mexico-como-el-aeropuerto-con-mas-pasajeros-de-america-latina
23. Cirium. (2025, enero 15). *2024 Winner – The Platinum Award For Operational Excellence*. Recuperado de https://www.cirium.com/thoughtcloud/2024-airport-winner-platinum-award-for-operational-excellence/
24. El Dorado Continúa Recibiendo Premios Por Esta Razón | Infraestructura | Economía | Portafolio. (n.d.). *Portafolio*. Recuperado de https://www.portafolio.co/economia/infraestructura/primer-premio-platino-a-un-aeropuerto-mundial-de-cirium-es-para-el-dorado-de-bogota-620851
25. El País. (2023, septiembre 13). *El aeropuerto El Dorado se queda pequeño y exaspera al sector aéreo*. Recuperado de https://elpais.com/america-colombia/2023-09-13/el-aeropuerto-el-dorado-se-queda-pequeno-y-exaspera-al-sector-aereo.html
26. Prensa Latina. (2025, febrero 5). *Aeropuerto de Bogotá fue el de mayor tráfico en 2024 en Latinoamérica*. Recuperado de https://www.prensa-

latina.cu/2025/02/05/aeropuerto-de-bogota-fue-el-de-mayor-trafico-en-2024-en-latinoamerica/