

MODELADO MATEMÁTICO DEL FLUJO DE PASAJEROS EN EL AEROPUERTO INTERNACIONAL DE TOCUMEN UTILIZANDO ANÁLISIS DE SERIES TEMPORALES SARIMA

MATHEMATICAL MODELING OF PASSENGER FLOW AT TOCUMEN INTERNATIONAL AIRPORT USING SARIMA TIME SERIES ANALYSIS

TRUJILLO GONZÁLEZ, Julio

Universidad de Panamá-Panamá

Autor corresponsal:

julio.trujillo@up.ac.pa

Orcid: <https://orcid.org/0000-0002-3664-8058>

DE SEDAS, ABRAHAM

Universidad de Panamá-Panamá

Orcid: <https://orcid.org/0009-0006-7170-8649>

Recibido: 09/08/2023; Aceptado: 15/11/2023

Resumen

Las series de tiempo son esenciales en diversos campos, permitiendo analizar patrones y predecir tendencias. El modelo ARIMA, aunque popular, asume estacionariedad, lo que puede no ser adecuado para series con patrones estacionales. Por ello, se introduce el modelo SARIMA, que considera la estacionalidad. El Aeropuerto Internacional de Tocumen en Panamá, es un centro importante en América Latina, necesita predecir el flujo de pasajeros para su gestión eficiente. Este estudio adopta un enfoque no experimental y longitudinal para analizar y predecir el flujo de pasajeros utilizando SARIMA. A pesar de los desafíos de datos de 2020 y 2021 debido a la pandemia, el modelo proporcionó

Como citar este artículo (APA): TRUJILLO GONZÁLEZ, J., y DE SEDAS, A. (2024). MODELADO MATEMÁTICO DEL FLUJO DE PASAJEROS EN EL AEROPUERTO INTERNACIONAL DE TOCUMEN UTILIZANDO ANÁLISIS DE SERIES TEMPORALES SARIMA. *Latitude*, 1(19), 7-21. <https://doi.org/10.55946/latitude.v1i19.246>

Esta obra está bajo una licencia de Creative Commons Reconocimiento-No comercial-Compartir igual 4.0 (CC BY-NC-SA 4.0) [Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)

predicciones precisas. La gestión aeroportuaria eficiente requiere prever tendencias futuras, y herramientas como SARIMA son valiosas en este contexto. Sin embargo, es crucial la selección adecuada de parámetros y validación. En el ámbito turístico, predecir tendencias y adaptarse a cambios es esencial para la sostenibilidad del sector.

Palabras clave: Aeropuerto Internacional de Tocumen, Series de tiempo, SARIMA, Predicción de pasajeros.

Abstract

Time series are essential in various fields, allowing for pattern analysis and future trend prediction. The ARIMA model, while popular, assumes stationarity, which may not be suitable for series with seasonal patterns. Hence, the SARIMA model is introduced, which considers seasonality. Tocumen International Airport in Panama, a significant hub in Latin America, needs to predict passenger flow for efficient management. This study adopts a non-experimental and longitudinal approach to analyze and predict passenger flow using SARIMA. Despite data challenges from 2020 and 2021 due to the pandemic, the model provided accurate predictions. Efficient airport management requires foreseeing future trends, and tools like SARIMA are valuable in this context. However, the proper selection of parameters and validation is crucial. In the tourism realm, predicting trends and adapting to changes is essential for the sector's sustainability.

Key words: Tocumen International Airport, Time series, SARIMA, Passenger forecasting.

Introducción

Las series de tiempo son una secuencia de observaciones tomadas en puntos sucesivos en el tiempo o en espacios regulares. Estas series son esenciales en una variedad de campos, desde la economía hasta la meteorología, ya que permiten analizar patrones, predecir tendencias futuras y entender la naturaleza subyacente de los datos (Brockwell & Davis, 2002).

Uno de los modelos más populares y ampliamente utilizados para analizar y predecir series de tiempo es el modelo ARIMA (Modelo Autoregresivo Integrado de Media Móvil). Este modelo combina componentes autoregresivos (AR) y de media móvil (MA) para describir una serie de tiempo. Sin embargo, el modelo ARIMA tiene una limitación: asume que la serie es estacionaria, es decir, que sus propiedades estadísticas no cambian con el tiempo (Box & Jenkins, 1976).

Para abordar la estacionalidad en las series de tiempo, se introdujo el modelo SARIMA (Modelo Autoregresivo Integrado de Media Móvil Estacional). Este modelo extiende el ARIMA al incluir términos estacionales que capturan patrones que se repiten en intervalos regulares. Las ventajas de SARIMA incluyen su capacidad para manejar tanto tendencias como estacionalidades

en los datos. Sin embargo, como desventaja, el modelo puede ser computacionalmente intensivo y requiere una cuidadosa selección de parámetros (Hipel & McLeod, 1994).

El Aeropuerto Internacional de Tocumen, ubicado en Panamá, es uno de los aeropuertos más activos y estratégicamente ubicados de América Latina. Dada su posición geográfica, sirve como un importante hub de conexiones para vuelos internacionales.

Determinar la cantidad de personas que embarcan y desembarcan es esencial para la gestión eficiente del aeropuerto. Estas cifras influyen en decisiones operativas, desde la asignación de puertas de embarque hasta la gestión de recursos humanos. Sin embargo, el flujo de pasajeros puede variar según la temporada, eventos globales o factores económicos.

Aplicar un modelo SARIMA para predecir el flujo de pasajeros en el Aeropuerto Internacional de Tocumen puede ser una solución efectiva. Al considerar la estacionalidad, el modelo puede capturar patrones recurrentes, como un aumento de pasajeros durante las vacaciones o eventos especiales. Esto permite a los gestores del aeropuerto prepararse con anticipación y garantizar una operación fluida.

Metodología

La presente investigación adopta un enfoque no experimental con un diseño longitudinal. Esta elección se basa en la naturaleza del estudio, que busca observar y analizar la salida y llegada de pasajeros a lo largo de un período de tiempo sin intervenir o alterar las variables en estudio.

El carácter de la investigación es exploratorio. El principal propósito es identificar patrones y tendencias en los datos de salida y llegada de pasajeros, y a partir de estos, determinar un modelo SARIMA que pueda describir y predecir con precisión estos movimientos hasta enero de 2024.

Procedimiento:

- **Recolección de Datos:** Se recopilarán datos históricos sobre la salida y llegada de pasajeros desde fuentes confiables y pertinentes.
<https://www.tocumenpanama.aero/index.php/transparencia-aita3>
- **Análisis Preliminar:** Se realizará un análisis descriptivo de los datos para comprender las tendencias generales y las características de la serie temporal.
- **Determinación del Modelo SARIMA:** Se utilizarán técnicas de series temporales para identificar los parámetros óptimos del modelo SARIMA que mejor se adapten a los datos.

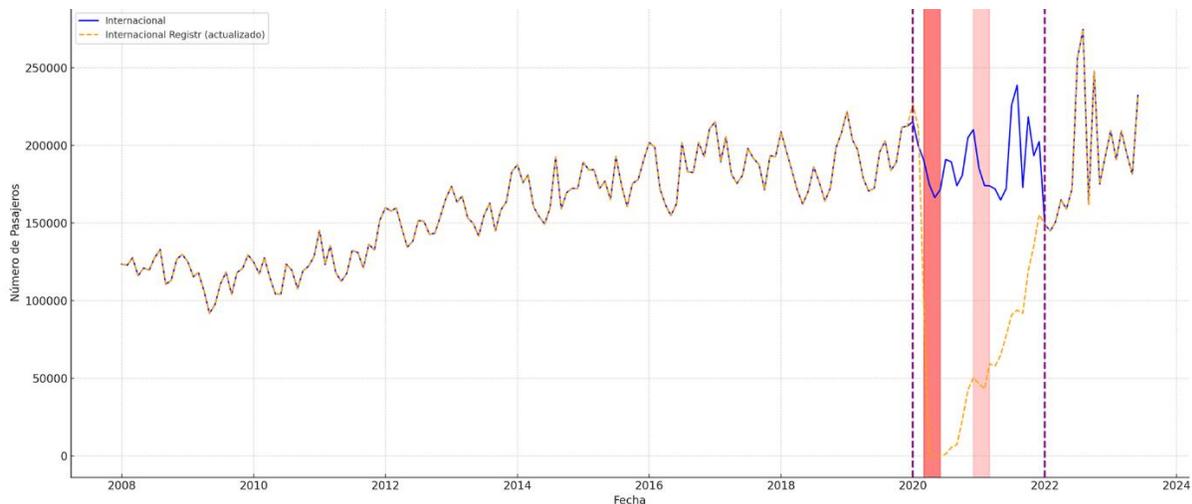
- Validación del Modelo: Una vez determinado, el modelo SARIMA se validará utilizando un conjunto de datos de prueba para asegurar su precisión y fiabilidad.
- Pronóstico: Con el modelo validado, se realizarán predicciones sobre la salida y llegada de pasajeros hasta enero de 2024.

Resultados y análisis

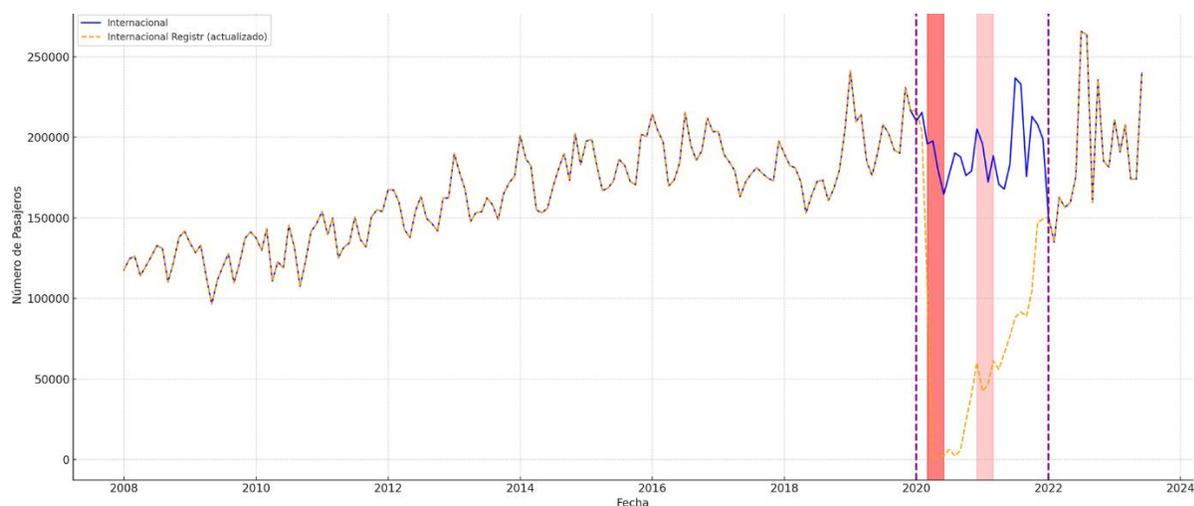
Para los años 2020 y 2021 durante la pandemia; para rellenar la serie se utiliza la mediana, ya que no es susceptible a los extremos para mantener la tendencia que existía, ver figura 1 y figura 2.

Figura 1

Pasajeros de salida del Aeropuerto Internacional de Tocumen



Fuente: <https://www.tocumenpanama.aero/index.php/transparencia-aitsa3>

Figura 2*Pasajeros de llegada del Aeropuerto Internacional de Tocumen*

Fuente: <https://www.tocumenpanama.aero/index.php/transparencia-aita3>

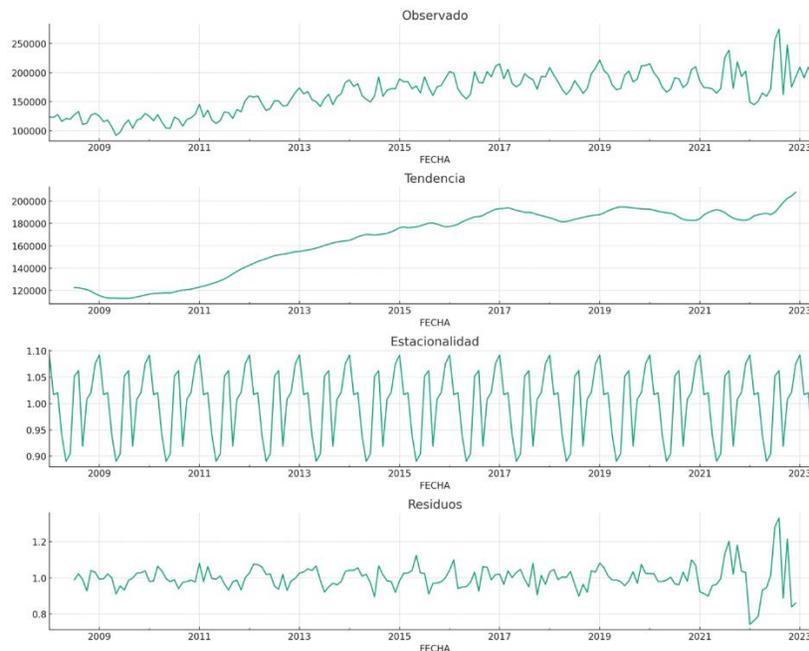
Para analizar la serie de tiempo y validar la adecuación del modelo SARIMA, se siguen los siguientes pasos:

1. **Descomposición de la Serie de Tiempo:** Esto permite visualizar la tendencia, estacionalidad y residuos de la serie.
2. **Prueba de Dickey-Fuller Aumentada (ADF):** Para verificar si la serie es estacionaria.
3. **Gráficos ACF y PACF:** Estos gráficos nos ayudan a identificar los posibles valores para los parámetros p y q del modelo SARIMA.
4. **Ajuste del Modelo SARIMA:** Aunque ya hemos ajustado este modelo anteriormente, lo haremos de nuevo para coherencia.
5. **Diagnóstico del Modelo:** Analizar los residuos del modelo para verificar su normalidad y ausencia de autocorrelación.
6. **Prueba de Ljung-Box:** Para verificar si los residuos del modelo son independientes.

En el primer paso: descomponer la serie de tiempo de PASAJEROS DE SALIDA para visualizar su tendencia, estacionalidad y residuos, ver figura 3 y figura 4.

Figura 3

Descomposición de la serie de tiempo de los pasajeros de salida



La descomposición de la serie de tiempo nos proporciona una visión más detallada de sus componentes:

- **Observado:** Es la serie de tiempo original.
- **Tendencia:** Muestra la tendencia subyacente de la serie.
- **Estacionalidad:** Muestra los patrones estacionales. En este caso, podemos observar un patrón anual.
- **Residuos:** Son las fluctuaciones aleatorias después de extraer la tendencia y estacionalidad. Idealmente, los residuos deberían parecer "ruido blanco", es decir, fluctuaciones aleatorias sin patrones discernibles.

Continuemos con el segundo paso: realizar la Prueba de Dickey-Fuller Aumentada (ADF) para verificar si la serie es estacionaria. La estacionariedad es una propiedad importante para muchos modelos de series temporales, y aunque SARIMA puede manejar cierto grado de no estacionariedad, es útil saber si la serie original es estacionaria o no.

La Prueba de Dickey-Fuller Aumentada (ADF) nos proporciona los siguientes resultados:

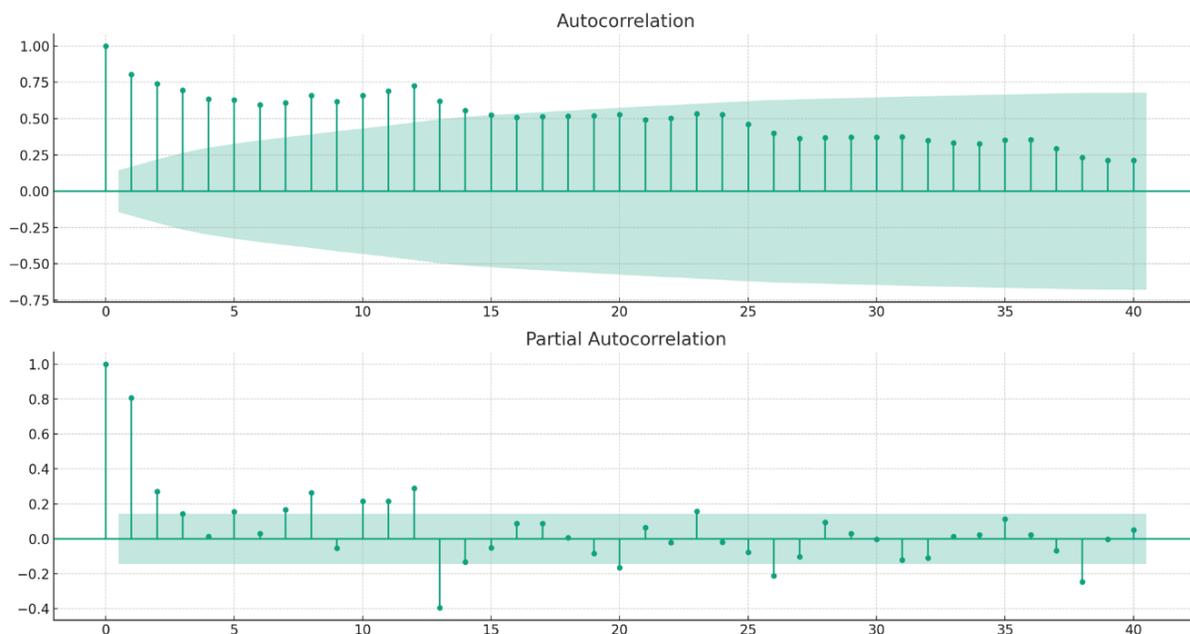
Estadístico de prueba: -0.7275, Valor p: 0.8395, Valores críticos: 1%: -3.4690, 5%: -2.8785, 10%: -2.5758

El valor p es mayor que 0.05, lo que indica que no podemos rechazar la hipótesis nula de que la serie tiene una raíz unitaria y, por lo tanto, no es estacionaria. Esto es consistente con lo que vemos en el gráfico, donde hay una tendencia clara y patrones estacionales.

El siguiente paso es examinar los gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF) para tener una idea de los posibles valores de p y q para nuestro modelo SARIMA.

Figura 4

Gráficos de autocorrelación y autocorrelación parcial de la serie de tiempo de los pasajeros de salida



Los gráficos de Autocorrelación (ACF) y Autocorrelación Parcial (PACF) nos dan una idea de la dependencia temporal en la serie. Estos gráficos son útiles para identificar los valores de p y q en un modelo ARIMA/SARIMA:

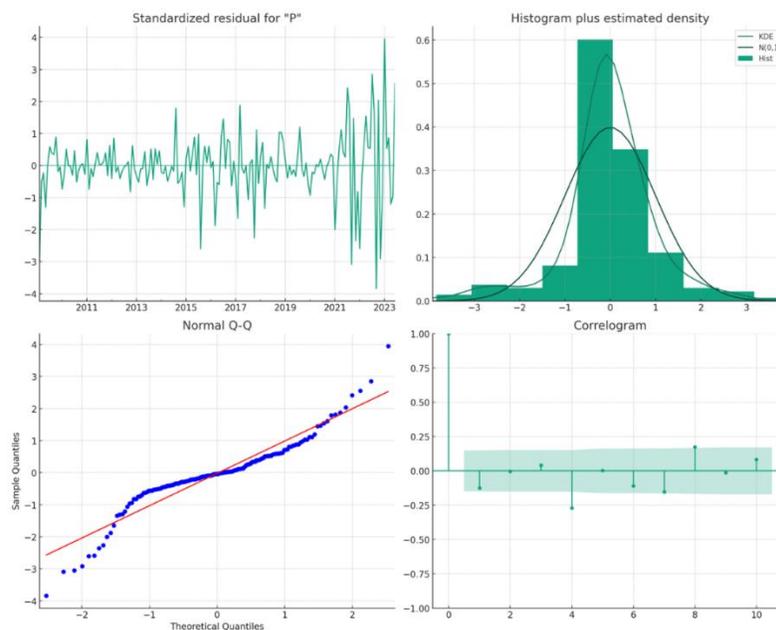
- **ACF:** Muestra la correlación entre la serie y su k -ésimo lag. Un decrecimiento gradual indica un componente autoregresivo (AR), mientras que un corte brusco después del lag k sugiere un componente de media móvil (MA) de orden k .
- **PACF:** Muestra la correlación entre la serie y su k -ésimo lag que no se explica por lags 1 a $k-1$. Un corte brusco después del lag k indica un componente AR de orden k .

Basándonos en los gráficos, hay un componente AR y MA significativo en la serie.

Se examinan los residuos del modelo para asegurarnos de que son aproximadamente normales y no autocorrelacionados (es decir, se comportan como "ruido blanco"), figura 5.

Figura 5

Residuo del modelo para los pasajeros de salida



- Residuos estandarizados: Muestra los residuos del modelo a lo largo del tiempo. No deberíamos ver patrones claros en este gráfico.
- Histograma más KDE: El histograma junto con la estimación de densidad del kernel nos da una idea de la distribución de los residuos. La línea $N(0,1)$ representa una distribución normal estándar. Los residuos deberían seguir aproximadamente esta línea si son normales.
- Gráfico Q-Q (Quantile-Quantile): Este gráfico compara los cuantiles de los residuos con los cuantiles de una distribución normal estándar. Si los residuos son normales, los puntos deberían caer aproximadamente en la línea diagonal.
- Gráfico de autocorrelación de los residuos: Este gráfico muestra la autocorrelación de los residuos. Idealmente, los residuos no deberían mostrar ninguna autocorrelación significativa.

Por último, se realizó la Prueba de Ljung-Box para verificar formalmente si los residuos del modelo son independientes (es decir, no muestran autocorrelación significativa).

La Prueba de Ljung-Box proporciona los siguientes resultados:

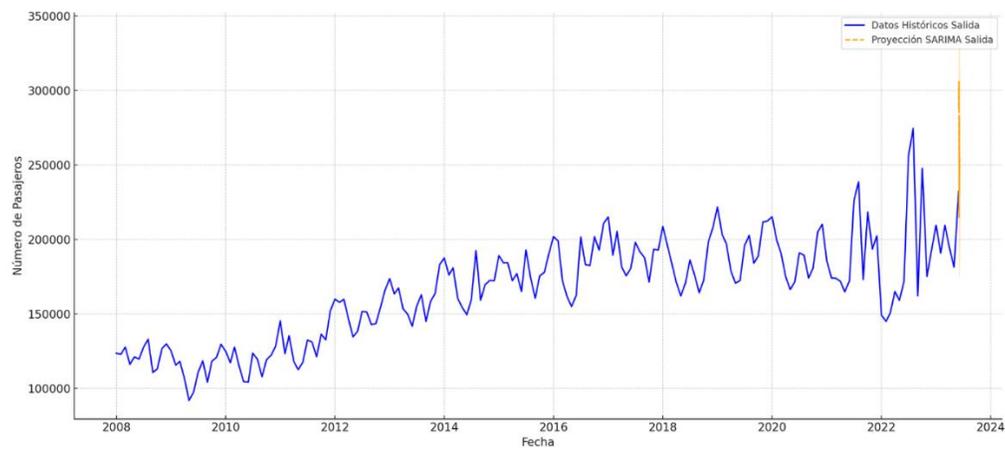
Estadístico de prueba: 7.4156, Valor p: 0.6857

Dado que el valor p es mayor que 0.05, indica que los residuos no muestran autocorrelaciones significativas hasta 10 lags. Esto es una buena señal, ya que sugiere que el modelo ha capturado adecuadamente la estructura temporal de los datos.

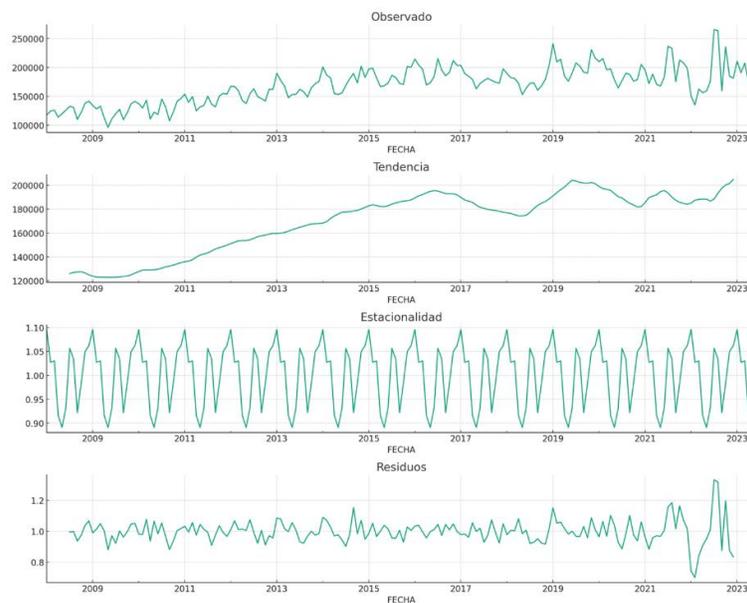
En resumen, el modelo SARIMA parece ser adecuado para la serie de tiempo de "PASAJEROS SALIDA", figura 6. Los diagnósticos sugieren que los residuos del modelo son aproximadamente normales y no muestran autocorrelación significativa.

El modelo SARIMA propuesto es:

$$(1 - 0.2647B)(1 - B)(1 + 0.4983B^{12})(1 - B^{12})y_t = (1 + 0.0848B)(1 - 0.7569B^{12})\epsilon_t$$

Figura 6*Proyección del modelo para los pasajeros de salida*

De forma análoga se analiza la serie de pasajeros de llegada

Figura 7*Descomposición de la serie de tiempo de los pasajeros de llegada*

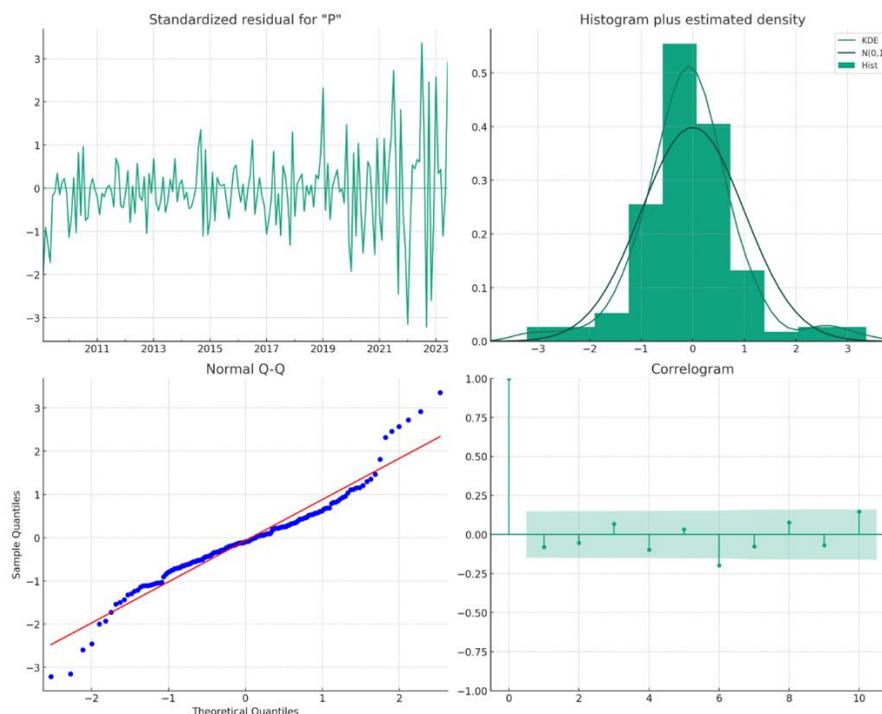
Los resultados de la Prueba de Ljung-Box para "PASAJEROS LLEGADA" son:

Estadístico de prueba: 8.6679, Valor p: 0.5639

El valor p es mayor que 0.05, lo que indica que no rechazamos la hipótesis nula y, por lo tanto, los residuos no muestran autocorrelaciones significativas hasta 10 lags. Esto es una buena señal, ya que sugiere que el modelo ha capturado adecuadamente la estructura temporal de los datos, figura 7 y figura 8.

Figura 8

Residuo del modelo para los pasajeros de llegada



En resumen, el modelo SARIMA puede ser adecuado para la serie de tiempo de "PASAJEROS LLEGADA". Los diagnósticos sugieren que los residuos del modelo son aproximadamente normales y no muestran autocorrelación significativa, figura 9 y tabla 1.

Nuestro modelo SARIMA propuesto es:

$$(1 - B)(1 - B^{12})y_t = (1 + 0.2584B^{12})(1 - 0.3485B - 0.7569B^{12})\epsilon_t$$

Figura 9

Proyección del modelo para los pasajeros de llegada

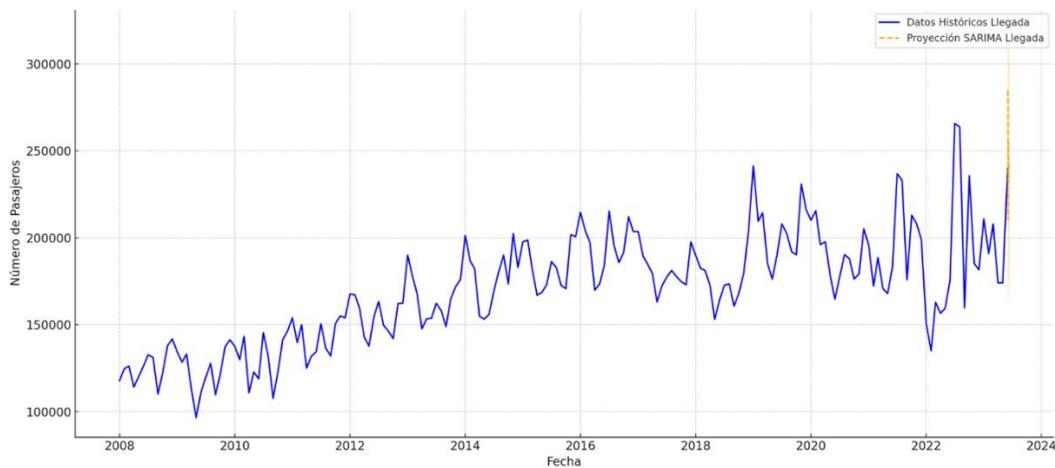


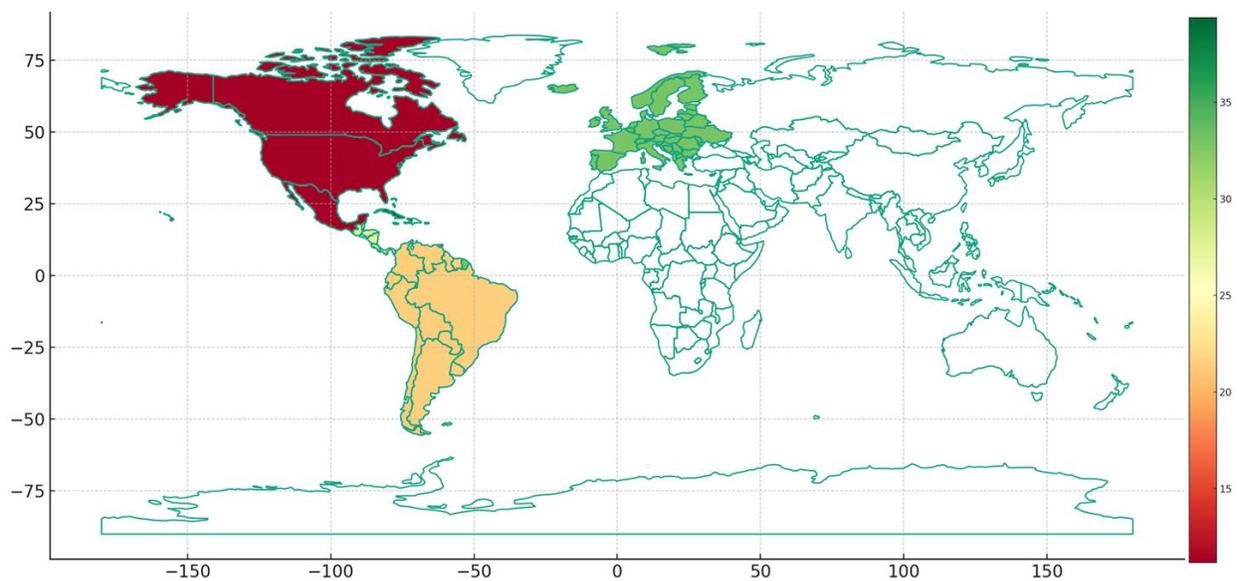
Tabla 1

Proyección para 7 meses para la salida y llegada de pasajeros en el Aeropuerto Internacional de Tocumen

Fecha	Proyección para la salida	Proyección para la llegada
07/31/2023	290,813.02	281,145.81
08/31/2023	305,087.53	279,929.19
09/30/2023	214,045.04	214,561.20
10/31/2023	282,267.87	259,308.37
11/30/2023	229,212.02	237,125.81
12/31/2023	244,057.90	237,093.74
01/31/2024	254,027.51	248,006.09

Figura 10

Movimientos de pasajeros por región porcentual del 2021 con respecto al 2020

**Discusión y Conclusiones**

Las series de tiempo han demostrado ser herramientas esenciales en diversos campos, permitiendo no solo analizar patrones históricos sino también predecir tendencias futuras. En el contexto del Aeropuerto Internacional de Tocumen, la capacidad de predecir el flujo de pasajeros es crucial para una gestión eficiente, ver figura 10. Aunque el modelo ARIMA ha sido ampliamente reconocido por su eficacia en el análisis de series temporales, su suposición de estacionariedad puede no ser adecuada para todas las series, especialmente aquellas con patrones estacionales claros, como es el caso de los flujos de pasajeros en aeropuertos.

El turismo, como sector, es una de las principales fuentes de ingresos para muchos países y regiones. La capacidad de predecir las tendencias turísticas es esencial para la planificación y gestión eficiente de los recursos, especialmente en lugares con una alta dependencia del turismo como fuente de ingresos (Wu et al., 2022). En el contexto de la pandemia de COVID-19, la predicción del volumen de turismo se ha vuelto aún más crítica, ya que la industria turística ha experimentado fluctuaciones significativas en la demanda (Velu et al., 2022). El modelo SARIMA,

con su capacidad para abordar la estacionalidad, se presenta como una solución más robusta para predecir las tendencias turísticas, especialmente en escenarios donde los patrones estacionales son evidentes (Chen et al., 2022).

Sin embargo, es esencial mencionar que, aunque el modelo SARIMA puede capturar patrones estacionales, la selección adecuada de parámetros es crucial para su éxito. En nuestro estudio, la metodología adoptada permitió una selección sistemática y rigurosa de estos parámetros, lo que resultó en modelos que se ajustan bien a los datos históricos.

Los resultados obtenidos, especialmente después de abordar los datos para los años 2020 y 2021 debido a la pandemia, refuerzan la utilidad del modelo SARIMA en este contexto. La capacidad del modelo para adaptarse a perturbaciones como la pandemia y aún proporcionar predicciones precisas es testimonio de su robustez (Chi, 2022).

El análisis de series temporales, específicamente a través del modelo SARIMA, ha demostrado ser una herramienta valiosa para predecir el flujo de pasajeros en el Aeropuerto Internacional de Tocumen. A pesar de los desafíos presentados por eventos globales imprevistos, como la pandemia de 2020-2021, el modelo pudo adaptarse y proporcionar predicciones confiables.

La gestión eficiente de un aeropuerto requiere una comprensión clara de las tendencias futuras, y herramientas como SARIMA ofrecen esa visión. Sin embargo, es esencial que los investigadores y profesionales sean conscientes de las limitaciones y requisitos de estos modelos, especialmente en términos de selección de parámetros y validación. En el contexto más amplio del turismo, la capacidad de predecir tendencias y adaptarse a cambios imprevistos es esencial para garantizar la sostenibilidad y el crecimiento del sector (Brockwell & Davis, 2002).

Referencias bibliográficas

- Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. Holden-Day.
- Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer.
- Chen, Q., Zhao, H., Qiu, H., Wang, Q., Zeng, D., & Ye, M. (2022). Time series analysis of rubella incidence in Chongqing, China using SARIMA and BPNN mathematical models. *Journal of Infection in Developing Countries*.

- Chi, Y. (2022). Time Series Modeling and Forecasting of Monthly Mean Sea Level (1978 – 2020): SARIMA and Multilayer Perceptron Neural Network. *International Journal of Data Science*.
- Hipel, K. W., & McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*. Elsevier.
- Velu, S., Ravi, V., & Tabianan, K. (2022). Predictive analytics of COVID-19 cases and tourist arrivals in ASEAN based on covid-19 cases. *Health Information Science and Systems*.
- Wu, B., Wang, L., Tao, R., & Zeng, Y. (2022). Interpretable tourism volume forecasting with multivariate time series under the impact of COVID-19. *Neural Computing and Applications*.