



Universidad Tecnológica de Panamá
Facultad de Ingeniería de Sistemas Computacionales
Maestría en Analítica de Datos



Modelos Predictivos

Profesor:

Juan Marcos Castillo, PhD

Integrantes:

Aquino Rodríguez

Proyecto final

Predicción de la Demanda de pasajeros en USA Utilizando Series Temporales.

2025

Introducción

El transporte aéreo es una de las industrias más importantes a nivel mundial, ya que conecta países, mueve la economía, promueve el turismo, y facilita la movilidad de personas y bienes. En los Estados Unidos, este sector es crucial para la economía del país y facilitando la movilidad de las personas a nivel nacional como internacional. Sin embargo, como cualquier otra industria, la misma está sujeta a fluctuaciones y variaciones a lo largo del tiempo debido a diferentes factores, como crisis económicas, estacionalidades, cambios en las políticas de viaje o incluso eventos inesperados.

Ante este escenario, los modelos predictivos basados en series temporales o de regresión se convierten en una herramienta clave para anticipar la demanda futura y apoyar la toma de decisiones, proporcionando información valiosa para predecir el comportamiento o demanda de pasajero que se viajarán en el futuro en vuelos nacionales como internacionales, lo que puede ayudar a las aerolíneas y a las autoridades a tomar decisiones más certeras.

Justificación

El análisis predictivo del tráfico aéreo en Estados Unidos es crucial para optimizar los recursos y la eficiencia operativa de las aerolíneas con la finalidad de que puedan responder eficientemente a la demanda de pasajeros. En un entorno dinámico y altamente competitivo como lo es la aviación, contar con herramientas que permitan predecir la demanda de pasajeros puede marcar una diferencia significativa en la programación de vuelos, la asignación de aviones y la gestión de la carga y la capacidad de los vuelos. Este estudio permitirá anticipar o proyectar la demanda de pasajeros a futuro, lo que puede resultar en una mejor utilización y disponibilidad de vuelos y asignación de recursos.

La aplicación de modelos de series temporales a datos históricos del tráfico aéreo permite detectar patrones y estacionalidades que contribuyen al desarrollo de herramientas de apoyo a la decisión, reforzando la competitividad del sector y fortaleciendo la capacidad de adaptación ante escenarios inciertos.

NUEVA YORK — Las aerolíneas esperan un fuerte impulso de la demanda de viajes en 2025, aunque podrían verse limitadas por problemas de capacidad. United Airlines, Delta Air Lines y varias otras aerolíneas con base en Estados Unidos han dado a los inversores pronósticos alentadores para el año. Wall Street espera que las principales aumenten sus ganancias en 2025.

Según lo anunciado por diversos medios de comunicación se espera que este 2025 haya una gran demanda vuelos debido a la demanda de pasajero que esperan realizar viajes es por ellos que predecir esta demanda basadas en datos históricos es fundamental para responder de manera eficiente la alta demanda que se espera.

Antecedentes

Se han realizados numerosos estudios sobre el análisis y la predicción del tráfico aéreo utilizando datos históricos. Un análisis similar en el contexto de las aerolíneas podría ser encontrado en estudios previos que utilizan modelos de series temporales como ARIMA y Prophet. Por ejemplo:

Los estudios previos han demostrado que los modelos de predicción de series temporales, cuando se ajustan correctamente, pueden generar pronósticos precisos sobre el volumen de pasajeros, lo que hace que estos enfoques sean una excelente opción para este análisis.

Definición del problema

El principal problema que se aborda en esta investigación es la predicción de la demanda futura de pasajeros en vuelos nacionales e internacionales en Estados Unidos. Aunque se disponen de datos históricos, es difícil anticipar el volumen exacto de pasajeros debido a diversas influencias como:

- **Estacionalidad:** Picos de tráfico durante festividades, vacaciones o eventos especiales.
- **Factores externos:** Crisis económicas, pandemias y fluctuaciones políticas o sociales que pueden afectar la movilidad de personas.
- **Innovaciones tecnológicas:** Avances en la aviación que permiten una mayor oferta de vuelos.

Como mencione anteriormente diversos factores pueden afectar la predicción de la demanda de pasajeros en los vuelos de USA, pero para este estudio no considerare factores como pandemias u otros.

El objetivo es utilizar modelos de series temporales para prever el número de pasajeros para los próximos meses, lo que permitirá a las aerolíneas y autoridades tomar decisiones informadas sobre la asignación de recursos y la programación de vuelos de forma que se pueda responder de forma eficaz a la demanda proyectada.

Análisis predictivo

A. Determinación de la base de datos.

He escogido el dataset "*U.S. Airline Traffic Data (2003-2023)*" de Kaggle el cual contiene los datos mensuales sobre el tráfico aéreo en los Estados Unidos desde enero de 2003 hasta 2023, el mismo es útil para realizar análisis detallados del tráfico aéreo, así como para la predicción de la demanda de pasajeros y la planificación de vuelos futuros en función de las tendencias históricas.

Se revisa hace una revisión exploratoria del dataset elegido "*U.S. Airline Traffic Data (2003-2023)*"

El dataset contiene 249 registros.

Las variables son:

- ❖ **Year (Año):** El año del registro.
- ❖ **Month (Mes):** El mes del registro.
- ❖ **Dom_Pax (Pasajeros Domésticos):** Número de pasajeros en vuelos nacionales
- ❖ **Int_Pax (Pasajeros Internacionales):** Número de pasajeros en vuelos internacionales.
- ❖ **Pax (Pasajeros Totales):** Número total de pasajeros (nacionales + internacionales).
- ❖ **Dom_Flt (Vuelos Nacionales):** Número de vuelos nacionales.
- ❖ **Int_Flt (Vuelos Internacionales):** Número de vuelos internacionales.
- ❖ **Flt (Vuelos Totales):** Número total de vuelos (nacionales + internacionales).
- ❖ **Dom_RPM (Millas de Pasajeros Ingresadas Nacionales):** Millas recorridas por pasajeros en vuelos nacionales.
- ❖ **Int_RPM (Millas de Pasajeros Ingresadas Internacionales):** Millas recorridas por pasajeros en vuelos internacionales.
- ❖ **RPM (Millas de Pasajeros Ingresadas Totales):** Millas totales recorridas por pasajeros.
- ❖ **Dom_ASM (Millas de Asientos Disponibles Nacionales):** Millas de asientos disponibles en vuelos domésticos.
- ❖ **Int_ASM (Millas de Asientos Disponibles Internacionales):** Millas de asientos disponibles en vuelos internacionales.
- ❖ **ASM (Millas de Asientos Disponibles Totales):** Millas totales de asientos disponibles.
- ❖ **Dom_LF (Factor de Carga Nacional):** Porcentaje de asientos ocupados en vuelos domésticos.
- ❖ **Int_LF (Factor de Carga Internacional):** Porcentaje de asientos ocupados en vuelos internacionales.

❖ **LF (Factor de Carga Total):** Porcentaje total de asientos ocupados.

Se realiza un análisis exploratorio de los datos del dataset

EDA

Variables y tipos de datos:

Generado mediante R Markdown

```
## 'data.frame': 249 obs. of 17 variables:
## $ Year : int 2003 2003 2003 2003 2003 2003 2003 2003 2003 2003 2003 ...
## $ Month : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Dom_Pax: num 43032450 41166780 49992700 47033260 49152352 ...
## $ Int_Pax: num 4905830 4245366 5008613 4345444 4610834 ...
## $ Pax : num 47938280 45412146 55001313 51378704 53763186 ...
## $ Dom_Flt: num 785160 690351 797194 766260 789397 ...
## $ Int_Flt: num 57667 51259 58926 55005 55265 ...
## $ Flt : num 842827 741610 856120 821265 844662 ...
## $ Dom_RPM: num 36211422 34148439 41774564 39465980 41001934 ...
## $ Int_RPM: num 12885980 10715468 12567068 10370592 11575026 ...
## $ RPM : num 49097402 44863907 54341633 49836572 52576960 ...
## $ Dom_ASM: num 56191300 50088434 57592901 54639679 55349897 ...
## $ Int_ASM: num 17968572 15587880 17753174 15528761 15629821 ...
## $ ASM : num 74159872 65676314 75346075 70168440 70979718 ...
## $ Dom_LF : num 64.4 68.2 72.5 72.2 74.1 ...
## $ Int_LF : num 71.7 68.7 70.8 66.8 74.1 ...
## $ LF : num 66.2 68.3 72.1 71 74.1 ...
```

Resumen estadístico de los datos:

Generado mediante R Markdown

##	Year	Month	Dom_Pax	Int_Pax
##	Min. :2003	Min. : 1.000	Min. : 2877290	Min. : 136609
##	1st Qu.:2008	1st Qu.: 3.000	1st Qu.:50982170	1st Qu.: 6395022
##	Median :2013	Median : 6.000	Median :56200104	Median : 7419187
##	Mean :2013	Mean : 6.446	Mean :55209710	Mean : 7392209
##	3rd Qu.:2018	3rd Qu.: 9.000	3rd Qu.:60892131	3rd Qu.: 8567847
##	Max. :2023	Max. :12.000	Max. :75378157	Max. :12432615
##	Pax	Dom_Flt	Int_Flt	Flt
##	Min. : 3013899	Min. :217262	Min. : 4996	Min. :222280
##	1st Qu.:57664576	1st Qu.:662000	1st Qu.:61615	1st Qu.:727898
##	Median :63899130	Median :709933	Median :66557	Median :779011
##	Mean :62601919	Mean :706751	Mean :64736	Mean :771487
##	3rd Qu.:69447429	3rd Qu.:781804	3rd Qu.:71924	3rd Qu.:848650
##	Max. :87810772	Max. :890938	Max. :82681	Max. :964102
##	Dom_RPM	Int_RPM	RPM	Dom_ASM
##	Min. : 2551127	Min. : 356762	Min. : 2907889	Min. :19489079
##	1st Qu.:44580290	1st Qu.:16602844	1st Qu.: 61825679	1st Qu.:56555517
##	Median :49857937	Median :19689006	Median : 69005186	Median :60889086
##	Mean :49578034	Mean :19181027	Mean : 68759061	Mean :61318217
##	3rd Qu.:55682567	3rd Qu.:22667597	3rd Qu.: 78039874	3rd Qu.:66268767
##	Max. :72267904	Max. :31376000	Max. :103643904	Max. :81997399
##	Int_ASM	ASM	Dom_LF	Int_LF
##	Min. : 1541419	Min. : 21030499	Min. :13.09	Min. :23.15
##	1st Qu.:21731193	1st Qu.: 79197800	1st Qu.:77.72	1st Qu.:76.43
##	Median :24645715	Median : 85051909	Median :82.12	Median :79.91
##	Mean :24139025	Mean : 85457242	Mean :79.94	Mean :77.30
##	3rd Qu.:27705349	3rd Qu.: 92551868	3rd Qu.:85.24	3rd Qu.:82.93
##	Max. :35326191	Max. :117312202	Max. :89.96	Max. :89.44
##	LF			
##	Min. :13.83			
##	1st Qu.:76.94			
##	Median :81.58			
##	Mean :79.29			
##	3rd Qu.:84.07			
##	Max. :89.14			

B. Pre-Procesamiento y Limpieza

Preprocesamiento de datos

- Se quitan los separadores de miles de los datos numéricos.
- Se realiza validación de los campos con datos nulos, no se encontró registros nulos.

Se identifican si hay valores nulos

```
## [1] "Year tiene 0 nulos"
## [1] "Month tiene 0 nulos"
## [1] "Dom_Pax tiene 0 nulos"
## [1] "Int_Pax tiene 0 nulos"
## [1] "Pax tiene 0 nulos"
## [1] "Dom_Flt tiene 0 nulos"
## [1] "Int_Flt tiene 0 nulos"
## [1] "Flt tiene 0 nulos"
## [1] "Dom_RPM tiene 0 nulos"
## [1] "Int_RPM tiene 0 nulos"
## [1] "RPM tiene 0 nulos"
## [1] "Dom_ASM tiene 0 nulos"
## [1] "Int_ASM tiene 0 nulos"
## [1] "ASM tiene 0 nulos"
## [1] "Dom_LF tiene 0 nulos"
## [1] "Int_LF tiene 0 nulos"
## [1] "LF tiene 0 nulos"
```

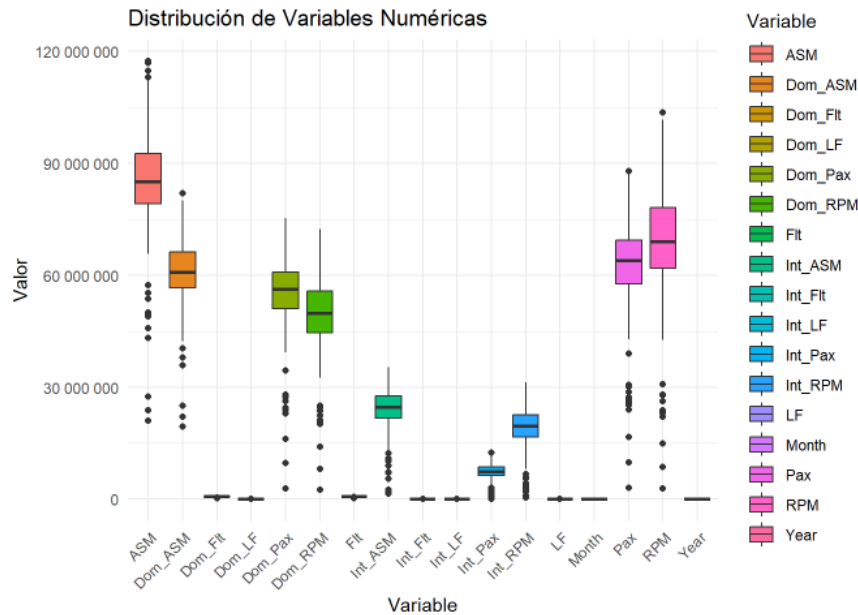
- No se encontró registros duplicados en los datos de las columnas de Año y Mes.

Se valida si hay registros duplicados

```
## [1] "Cantidad de duplicados (año y mes): 0"
```

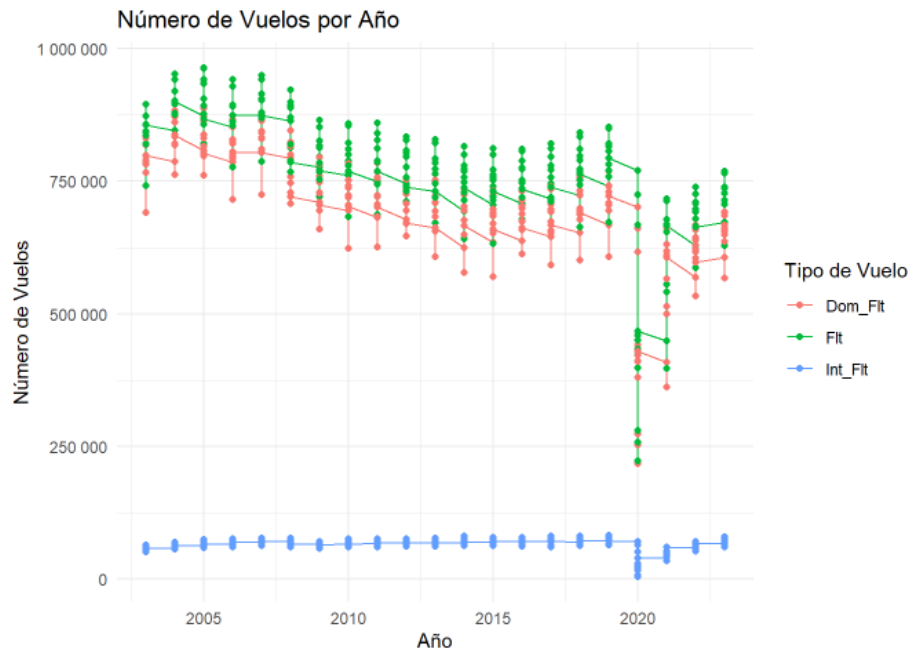
C. Análisis descriptivo

- ✓ Se realizó un análisis inicial de las tendencias, estacionalidades y patrones dentro de los datos, especialmente observando el comportamiento de las columnas clave como Pax (pasajeros), RPM (Revenue Passenger Miles) y LF (Load Factor).



Grafica 1: Distribución de variables numéricas.

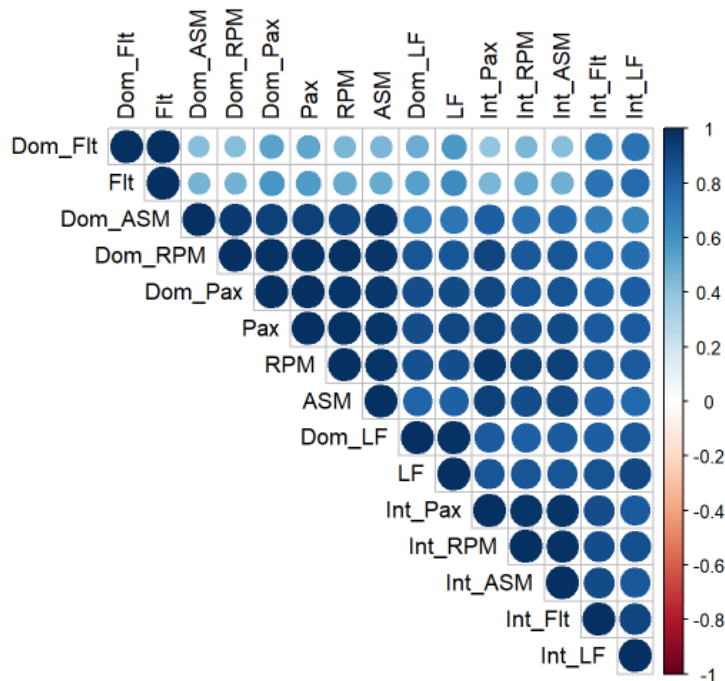
Esta gráfica de distribución de las variables numéricas del dataset en estudio, muestra cómo están distribuidos los datos y los valores y en el mismo se puede apreciar que hay valores (outliers) por debajo y encima del rango esperado



Gráfica 2: Número de vuelos por año

Grafica 2: Muestra el comportamiento de los vuelos nacionales e internacionales y el total de ellos en USA desde el 2003-2023, se muestra una baja entre 2020-2021, el mismo se debe a la pandemia, pero luego con el pasar de los meses se va recuperando de ese impacto. El número de vuelos internacionales se mantuvo relativamente bajo y estable en comparación con los vuelos domésticos y el total de vuelos durante la mayor parte del período observado.

Matriz de Correlación de la variables



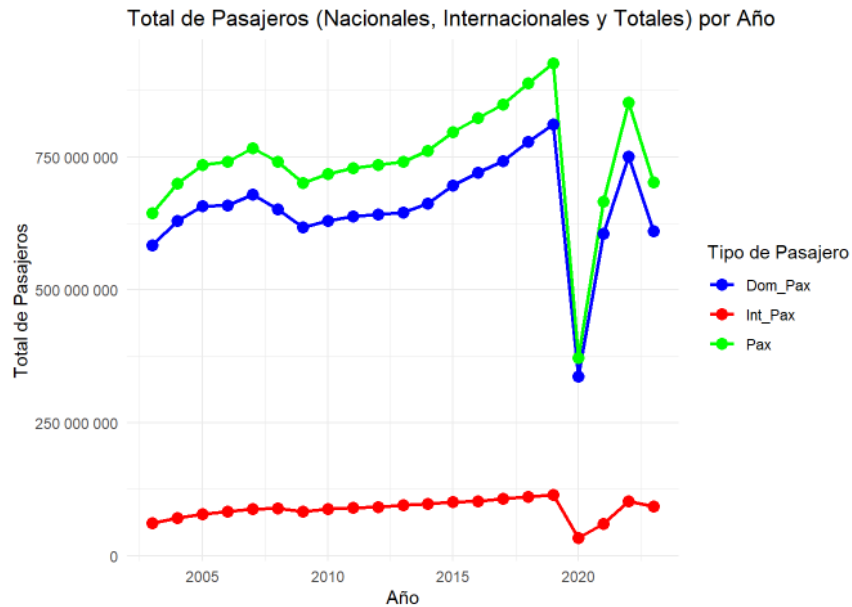
Grafica 3: Matriz de correlación de todas las variables numéricas

Grafica 3: Muestra Fuertes Correlaciones Positivas entre las siguientes variables.

- **Dom_Flt y Flt:** Existe una correlación positiva muy fuerte entre el número de vuelos domésticos y el número total de vuelos. Esto es esperable ya que los vuelos domésticos probablemente constituyen una parte importante del total.
- **Variables Domésticas entre sí (Dom_ASM, Dom_RPM, Dom_Pax):** Las variables relacionadas con la capacidad ofertada en vuelos domésticos (ASM), los pasajeros-kilómetro recorridos (RPM) y el número de pasajeros (Pax) muestran fuertes correlaciones positivas entre sí. Esto indica que cuando aumenta la capacidad, también tienden a aumentar los pasajeros transportados y los kilómetros recorridos.
- **Variables Internacionales entre sí (Int_ASM, Int_RPM, Int_Pax):** De manera similar, las variables de capacidad, pasajeros-kilómetro y pasajeros

para vuelos internacionales también presentan fuertes correlaciones positivas entre ellas.

- **RPM y ASM (general):** Existe una correlación positiva fuerte entre los pasajeros-kilómetro recorridos (RPM) y la capacidad ofertada (ASM) a nivel general.



Grafica 4: Movimientos de pasajeros (Nacionales, Internacionales y totales)

Grafica 4: podemos sacar obtener las siguientes conclusiones:

Pasajeros Domésticos (Línea Azul):

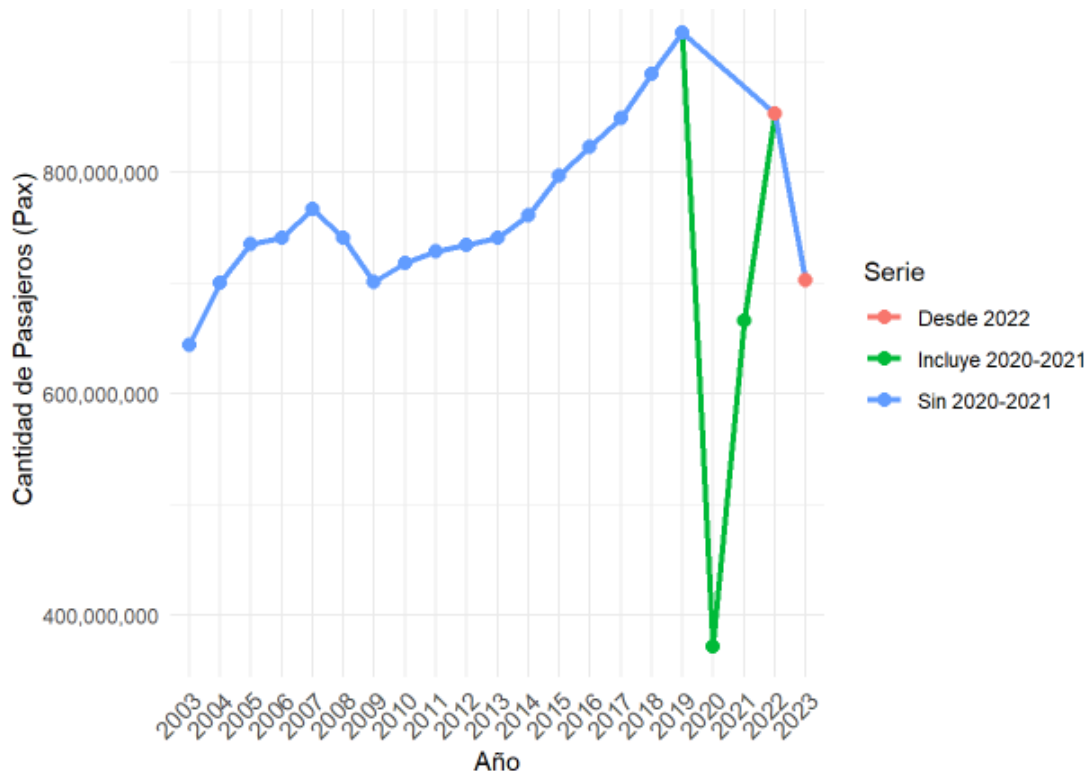
- Desde aproximadamente 2003 hasta 2008, se observa una tendencia al aumento en el número de pasajeros domésticos.
- Entre 2008 y 2011, hay una ligera fluctuación con una pequeña disminución y posterior recuperación.
- Desde 2011 hasta 2019, se aprecia una tendencia general al crecimiento, aunque con algunas variaciones anuales.
- En 2020, hay una caída drástica y muy pronunciada en el número de pasajeros domésticos, mientras que para el 2021 se refleja una recuperación con tendencia al aumento.

Pasajeros Internacionales (Línea Roja):

- A lo largo del período analizado hasta 2019, el número de pasajeros internacionales muestra una tendencia general al aumento gradual y constante, aunque a niveles significativamente más bajos que los pasajeros domésticos.
- En 2020, al igual que con los pasajeros domésticos, se produce una caída abrupta y muy marcada en el número de pasajeros internacionales. La disminución parece proporcionalmente mayor que la de los pasajeros domésticos.
- En 2021, se observa un ligero repunte, pero el número de pasajeros internacionales se mantiene muy por debajo de los niveles previos a 2020.
- En 2022, continúa la recuperación, pero aún lejos de los valores anteriores.

El número de pasajeros en vuelos domésticos es significativamente mayor que el número de pasajeros en vuelos internacionales durante todo el período analizado. Antes del año 2020, tanto el número de pasajeros domésticos como internacionales mostraban una tendencia general al crecimiento, lo que sugiere una expansión del sector de la aviación.

El año 2020 marca un punto de inflexión dramático, con una reducción masiva en el número de pasajeros en todas las categorías. Este colapso está muy probablemente relacionado con la pandemia global de COVID-19 y las restricciones de viaje implementadas a nivel mundial. A partir de 2021, se observa una tendencia clara a la recuperación tanto en los vuelos domésticos como internacionales. La recuperación en el número de pasajeros domésticos parece ser más rápida y pronunciada que la de los pasajeros internacionales.



Grafica 5: Cantidad de pasajeros por año

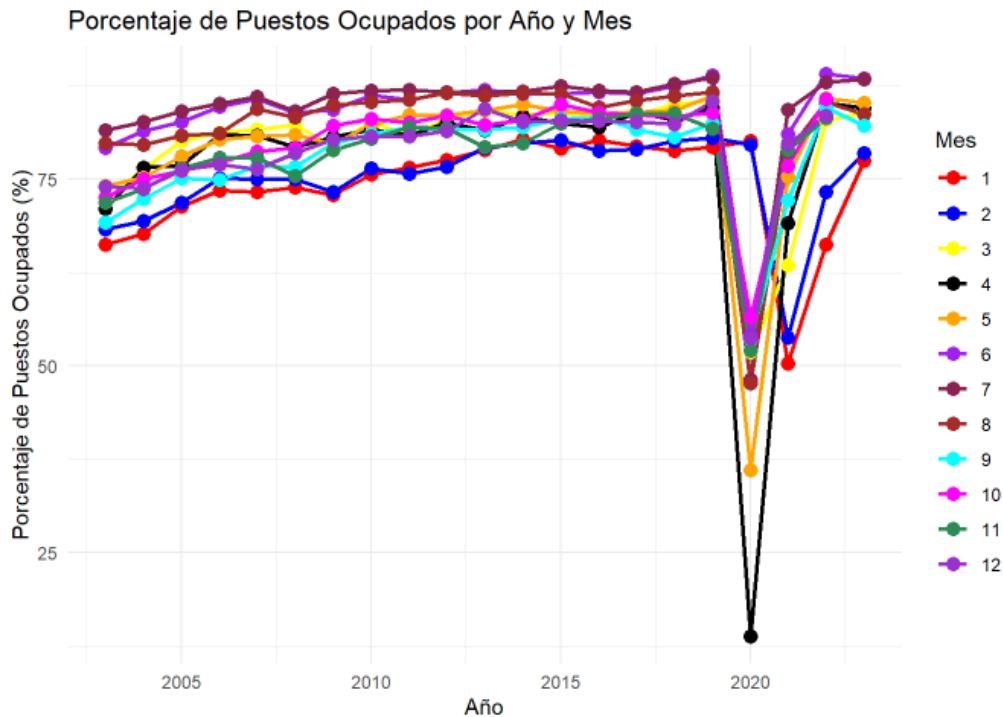
La grafica anterior muestra diferentes tendencias en base a algunos análisis a comparar.

En base a la línea azul más la verde que es el histórico de cantidad de pasajeros en USA en vuelos nacionales e internacionales.

- Desde 2003 hasta 2008, se observa una tendencia general al aumento en la cantidad de pasajeros, con algunas fluctuaciones anuales.
- Entre 2008 y 2010, hay una ligera disminución y posterior recuperación.
- Desde 2010 hasta 2019, se aprecia una tendencia clara y sostenida al **crecimiento significativo** en la cantidad de pasajeros, alcanzando un pico alrededor de 2019.
- En **2020**, se observa una **caída drástica y masiva** en la cantidad de pasajeros, llegando a un mínimo. Este descenso abrupto es casi con certeza el resultado de la pandemia global de COVID-19 y las restricciones de viaje asociadas.
- Desde el 2022 se muestra una recuperación.

En base a la línea azul (que no considera los datos (2020-2021)).

- Esta línea sigue la misma trayectoria ante mencionada (la línea azul más la verde)
- Se excluyen los años 2020-2021 de forma que se pueda interpretar o ver la gráfica desde otra perspectiva.



Grafica 6: Porcentajes de Puestos Ocupados.

Tendencia General Pre-Pandemia (2003-2019):

- En general, el porcentaje de puestos ocupados se mantiene relativamente alto y estable, oscilando principalmente entre el 75% y el 90% durante este período.
- Se pueden observar ligeras variaciones estacionales dentro de cada año, con algunos meses mostrando consistentemente porcentajes ligeramente más altos o más bajos que otros. Sin embargo, estas diferencias no son extremadamente pronunciadas.
- No se aprecia una tendencia clara y uniforme de aumento o disminución significativa del porcentaje de puestos ocupados a lo largo de estos años. Parece haber una cierta estabilidad en la ocupación.
- Algunos meses, como julio y diciembre, tienden a mostrar porcentajes de ocupación ligeramente más altos en algunos años, lo que podría estar relacionado con períodos vacacionales o festivos que impulsan la actividad.

- En el año **2020**, se observa un colapso dramático y repentino en el porcentaje de puestos ocupados para todos los meses. Los porcentajes caen a niveles muy bajos, algunos incluso por debajo del 25%, debido a las restricciones de movilidad producto de la pandemia
- En el año **2022**, la tendencia de recuperación continúa. Los porcentajes de ocupación para la mayoría de los meses se acercan o incluso superan los niveles observados antes de la pandemia en algunos casos. La estacionalidad parece volver a patrones más similares a los del período pre-pandémico.

Mediante el análisis y exploración del dataset, mismo que contiene muchos datos históricos, vemos que muestra una baja o un comportamiento no esperado en el 2020 debido probablemente a las restricciones de movilidad producto de la pandemia. Pero luego a partir del 2021-2022 vemos que nuevamente empieza a incrementar la cantidad de pasajeros.

D. Selección de variables

En base al análisis descriptivo las variables seleccionadas y que son de e interés para este estudio serán las siguientes:

- ✓ Year: Los años de información históricas que contiene el dataset con respecto a la movilidad de pasajeros.
- ✓ Month: Los meses de cada año de la información histórica del dataset.
- ✓ Pax: contiene la cantidad de pasajeros totales (Nacionales e Internacionales)

Con este estudio lo deseo es predecir la cantidad de pasajeros desde 2023¹⁰ al 2025¹² y ver el comportamiento del mismo, todo esto mediante el uso de modelos predictivos que permitan predecir lo más real posible estas tendencias.

E. Selección de modelos

Partiendo de las variables de estudios definidas y considerando lo visto en el análisis descriptivo referente al comportamiento de la gráfica de cantidad de pasajeros se procede a realizar validación para obtener el mejor modelo predictivo posible.

Modelo de predicción evaluados para todo el conjunto de datos históricos.

- ✓ Regresión Lineal
- ✓ Random Forest
- ✓ ARIMA
- ✓

Resultados:

	R2	MAE	MSE	RMSE
Regresión Lineal	-0.323685	1.721966e+07	6.620755e+14	2.573083e+07
Random Forest	-0.645189	1.873365e+07	8.228841e+14	2.868596e+07
ARIMA	-1.048078	2.439221e+07	1.024400e+15	3.200624e+07

El resultado nos arroja un R2 negativos para todos los modelos elegidos, lo que nos dice que la predicción del modelo no se ajusta a lo esperado.

Se realiza una nueva evaluación de los mismos modelos, pero se excluyen del análisis de los datos de los años 2020-2021, debido a que estos años fueron afectados por la pandemia.

Resultado:

	R2	MAE	MSE	RMSE
Regresión Lineal	-0.396642	7.542465e+06	8.318995e+13	9.120852e+06
Random Forest	0.468830	4.971609e+06	3.163877e+13	5.624835e+06
ARIMA	0.670698	3.067302e+06	1.961464e+13	4.428841e+06

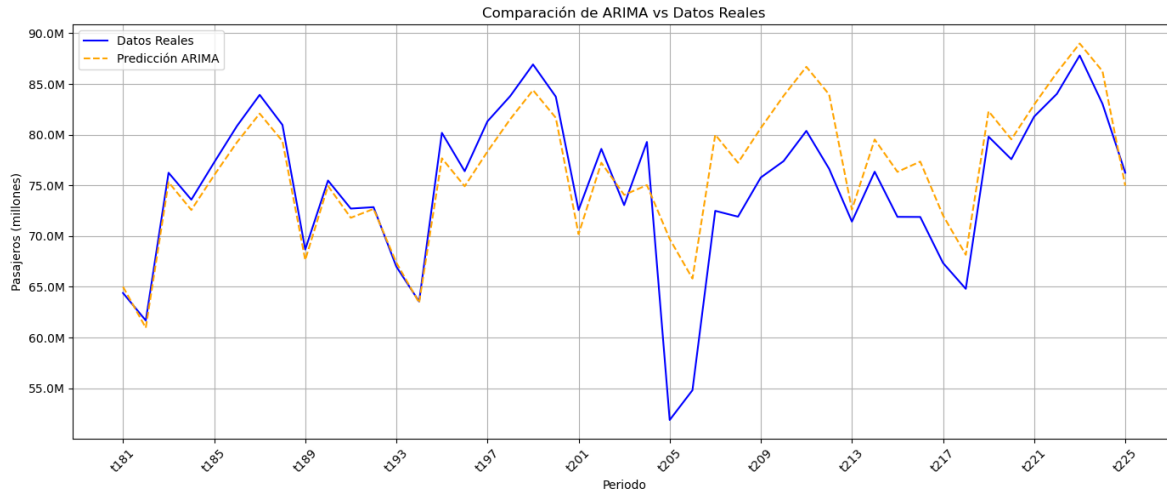
Los resultados indican que dos de los modelos tienen un R2 positivo (Random Forest y ARIMA) y de ellos ARIMA es el que mejor se ajusta.

ARIMA es el mejor modelo en este caso, por varias razones:

- ✓ Mayor R²: 0.67 → Significa que explica el 67% de la variabilidad de los datos. Muy decente.
- ✓ Menor MAE: Tiene el error absoluto promedio más bajo (≈ 3 millones de pasajeros).
- ✓ Menor MSE y RMSE: Comete menos errores, incluso penalizando los errores grandes.

Modelo ARIMA

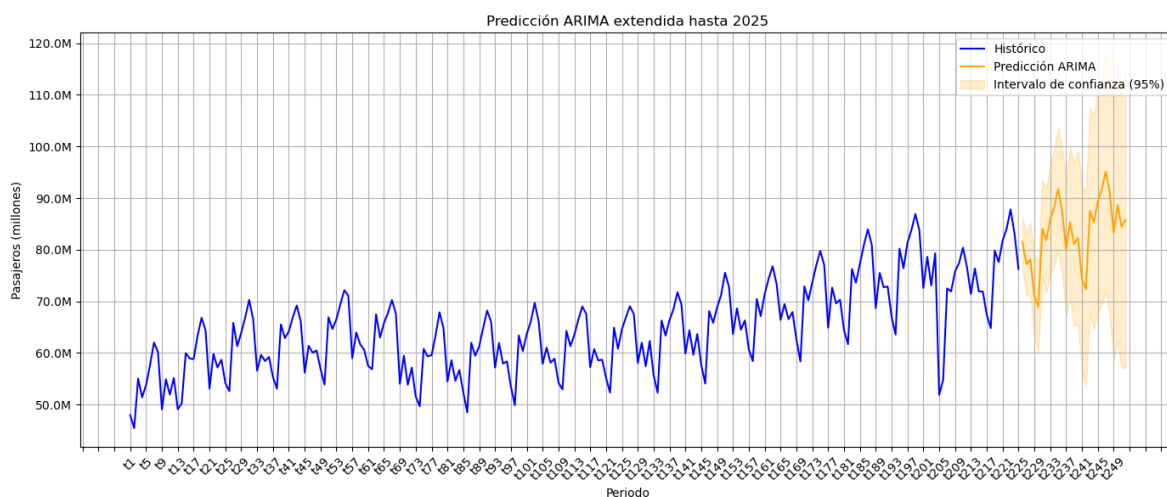
Realizamos el modelado, podemos ver el comportamiento de los datos reales vs la predicción del modelo, debido a que se excluyó los años 2020-2021, se ajustan los datos de año-mes por indicadores de tiempo



En la siguiente grafica podemos hacer la predicción en base al modelo de ARIMA ya entrenado para el periodo 202310-202512 y obviando el 2020-2021

- ✓ **Picos y Valles:** Se pueden observar varios picos y valles tanto en los datos reales como en las predicciones.
- ✓ **Retrasos:** En algunos casos, la predicción del modelo ARIMA parece seguir los movimientos de los datos reales con un ligero retraso.

En resumen, la gráfica proporciona una visión visual de la capacidad del modelo ARIMA para predecir los datos reales de pasajeros a lo largo del tiempo. Si bien el modelo parece capturar algunas de las tendencias generales, existen períodos de diferencia entre las predicciones y los valores reales.



Se observa una tendencia general ascendente en el número de pasajeros a lo largo del tiempo histórico. Esto sugiere un crecimiento constante en la variable analizada.

- ✓ La predicción del modelo ARIMA parece continuar la tendencia ascendente observada en los datos históricos.
- ✓ El modelo también proyecta la continuación del patrón estacional, con picos y valles que se repiten anualmente.
- ✓ El modelo ARIMA ha logrado capturar la tendencia general y la estacionalidad de los datos históricos. Esto sugiere que el modelo es razonablemente adecuado para predecir el comportamiento futuro de la variable analizada.
- ✓ La predicción indica un crecimiento continuo en el número de pasajeros, manteniendo el patrón estacional observado en el pasado.

Igual procedemos a imprimir los valores predichos para cada fecha en base al modelo ARIMA desde 202310 - 202512

Date	Periodo	Prediccion_Pax_M	Lower_CI_M	Upper_CI_M	
225	2023-10-01	t226	81.5M	76.7M	86.2M
226	2023-11-01	t227	77.2M	71.2M	83.2M
227	2023-12-01	t228	78.1M	71.0M	85.1M
228	2024-01-01	t229	71.1M	63.1M	79.0M
229	2024-02-01	t230	68.9M	60.2M	77.6M
230	2024-03-01	t231	84.1M	74.7M	93.5M
231	2024-04-01	t232	81.8M	71.7M	91.9M
232	2024-05-01	t233	85.9M	75.2M	96.6M
233	2024-06-01	t234	88.2M	76.9M	99.5M
234	2024-07-01	t235	91.8M	79.9M	103.6M
235	2024-08-01	t236	87.5M	75.1M	99.9M
236	2024-09-01	t237	80.0M	67.1M	92.9M
237	2024-10-01	t238	85.3M	70.7M	99.9M
238	2024-11-01	t239	81.1M	65.3M	96.9M
239	2024-12-01	t240	82.3M	65.4M	99.2M
240	2025-01-01	t241	74.3M	56.4M	92.3M
241	2025-02-01	t242	72.3M	53.4M	91.3M
242	2025-03-01	t243	87.6M	67.7M	107.5M
243	2025-04-01	t244	85.3M	64.5M	106.1M
244	2025-05-01	t245	89.3M	67.7M	111.0M
245	2025-06-01	t246	91.7M	69.2M	114.2M
246	2025-07-01	t247	95.1M	71.9M	118.4M
247	2025-08-01	t248	91.1M	67.1M	115.1M
248	2025-09-01	t249	83.3M	58.5M	108.0M
249	2025-10-01	t250	88.6M	62.4M	114.8M
250	2025-11-01	t251	84.4M	57.0M	111.9M
251	2025-12-01	t252	85.8M	57.2M	114.4M

Modelo holtwinters

Procedo a realizar una evaluación del modelo holtwinters para validar el comportamiento del mismo.

Evaluación del modelo excluyendo los años 2020-2021

Resultado:

MAE: 3,719,941

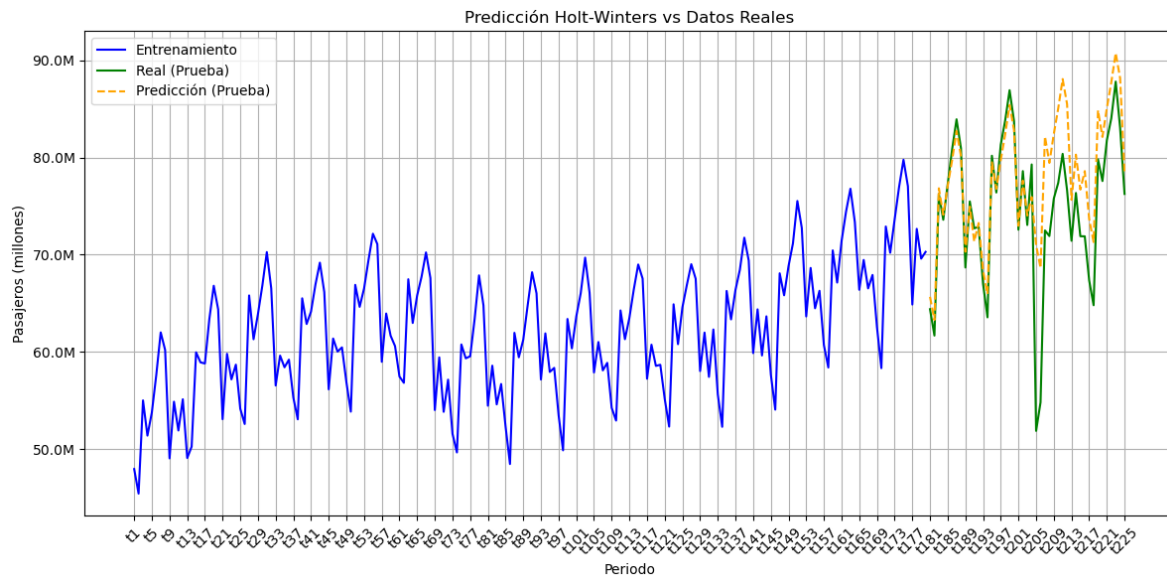
RMSE: 5,342,086

MAPE: 1.52%

- **MAE (Error Absoluto Medio):** En promedio, el modelo se equivoca por ~3.7 millones de pasajeros. Esto ya es mucho más aceptable, considerando que probablemente estás trabajando con volúmenes mensuales altos (por ejemplo, 1M a 2M pasajeros por mes).
- **RMSE (Raíz del Error Cuadrático Medio):** El RMSE de ~5.3 millones sigue siendo razonable. Es un poco más alto que el MAE, lo cual es normal y te dice que no hay errores extremos muy graves.
- **MAPE (Error Porcentual Absoluto Medio):** Un MAPE de 1.52% es *excelente*, Esto significa que, en promedio, el modelo solo se equivoca en 1.5% respecto al valor real. Este tipo de precisión es ideal y habla de un modelo muy bien ajustado.

Este modelo Holt-Winters que generó esas métricas es *altamente confiable* y probablemente esté haciendo predicciones muy cercanas a la realidad.

Se realiza entrenamiento del modelo holtwinters



Análisis de los Datos Reales de Prueba (Línea Verde):

- **Continuación de Patrones:** Los datos reales del período de prueba parecen continuar con la tendencia ascendente y el patrón estacional observados en el período de entrenamiento. Esto sugiere que los patrones identificados en los datos históricos se mantienen en el nuevo período.

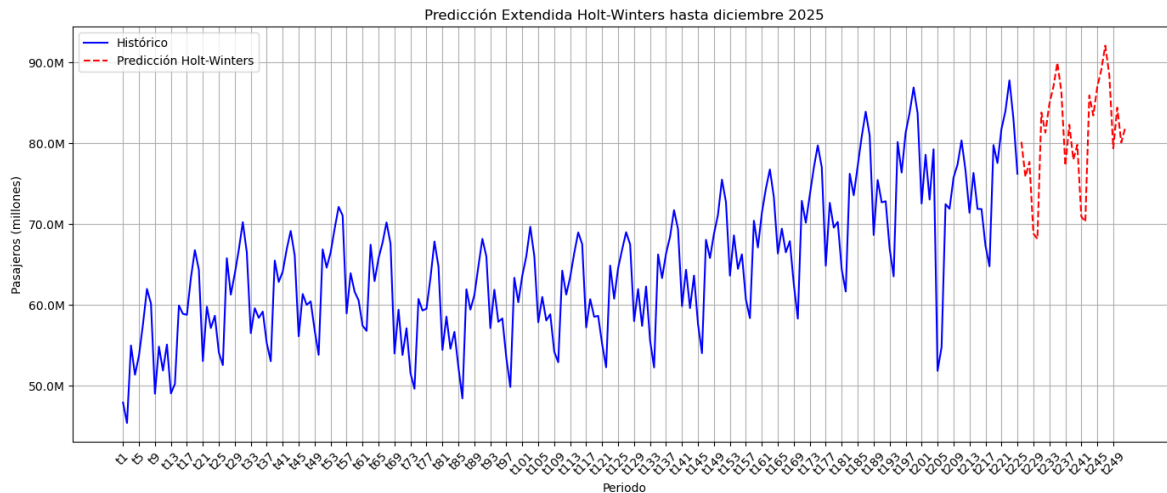
Análisis de la Predicción Holt-Winters (Línea Punteada Naranja):

- **Captura de la Tendencia:** La predicción del modelo Holt-Winters parece haber capturado la tendencia ascendente general de los datos de prueba. La línea naranja sigue una trayectoria similar a la línea verde en términos de dirección general.
- **Captura de la Estacionalidad:** El modelo también intenta replicar el patrón estacional observado en los datos reales de prueba, con picos y valles que se alinean aproximadamente con los de la línea verde.

Comparación entre la Predicción y los Datos Reales de Prueba:

- **Ajuste General:** En general, el modelo Holt-Winters parece haber logrado un ajuste razonable a los datos reales de prueba, capturando la dirección de la tendencia y la naturaleza cíclica de la estacionalidad.

En base al modelo entrenado se realiza un nuevo entrenamiento y se realiza la predicción de 202310 – 202512



Comparación entre la Predicción y los Datos Históricos Recientes:

- **Ajuste al Final del Histórico:** Es relevante observar cómo la predicción inicial se alinea con los últimos puntos de datos históricos antes de extenderse hacia el futuro. Un buen ajuste en esta etapa sugiere que el modelo ha capturado los patrones recientes de la serie.
- **Proyección Futura:** La línea punteada roja muestra la trayectoria esperada según el modelo Holt-Winters, incluyendo la continuación de la tendencia y las fluctuaciones estacionales

la gráfica presenta una predicción del número de pasajeros hasta diciembre de 2025 utilizando el método Holt-Winters, basada en la extrapolación de la tendencia y la estacionalidad observadas en los datos históricos. La predicción sugiere una continuación del crecimiento con las fluctuaciones estacionales trimestrales características. Sin embargo, la precisión real de esta predicción en el futuro dependerá de la estabilidad de los patrones históricos.

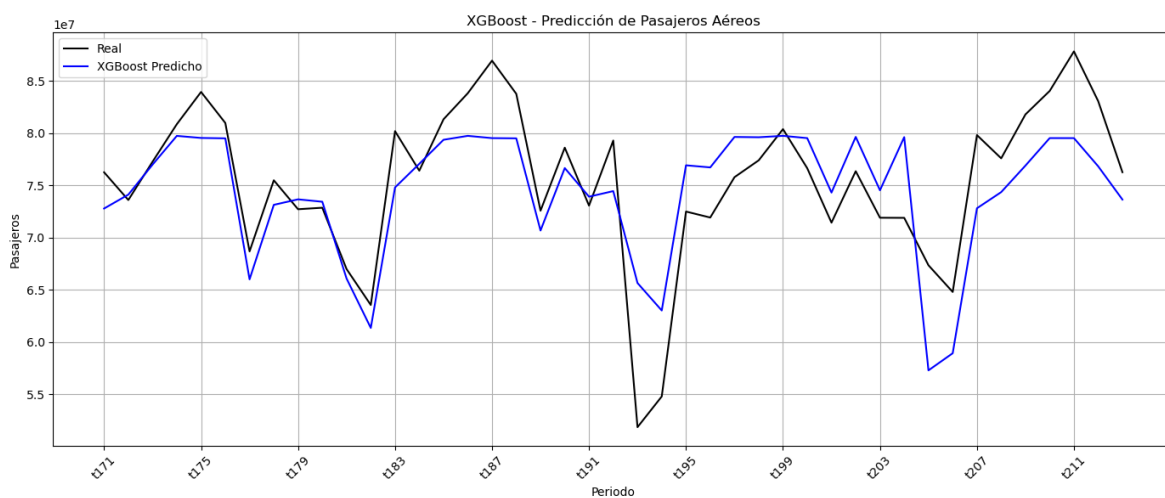
Igual procedemos a imprimir los valores predichos para cada fecha en base al modelo HOLT-WINTER

	Año	Periodo	Pronóstico_Pax_Millones
0	2023	t226	80.2M
1	2023	t227	75.9M
2	2023	t228	77.7M
3	2024	t229	68.8M
4	2024	t230	68.2M
5	2024	t231	83.8M
6	2024-04	t232	81.3M
7	2024-05	t233	84.9M
8	2024-06	t234	87.0M
9	2024-07	t235	90.0M
10	2024-08	t236	86.4M
11	2024-09	t237	77.3M
12	2024-10	t238	82.3M

13	2024-11	t239	78.0M
14	2024-12	t240	79.8M
15	2025-01	t241	71.0M
16	2025-02	t242	70.3M
17	2025-03	t243	85.9M
18	2025-04	t244	83.5M
19	2025-05	t245	87.0M
20	2025-06	t246	89.1M
21	2025-07	t247	92.1M
22	2025-08	t248	88.5M
23	2025-09	t249	79.4M
24	2025-10	t250	84.4M
25	2025-11	t251	80.1M
26	2025-12	t252	81.9M

Modelo XGBOOST

RMSE: 4,767,127
R²: 0.5898



Análisis de los Datos Reales (Línea Negra):

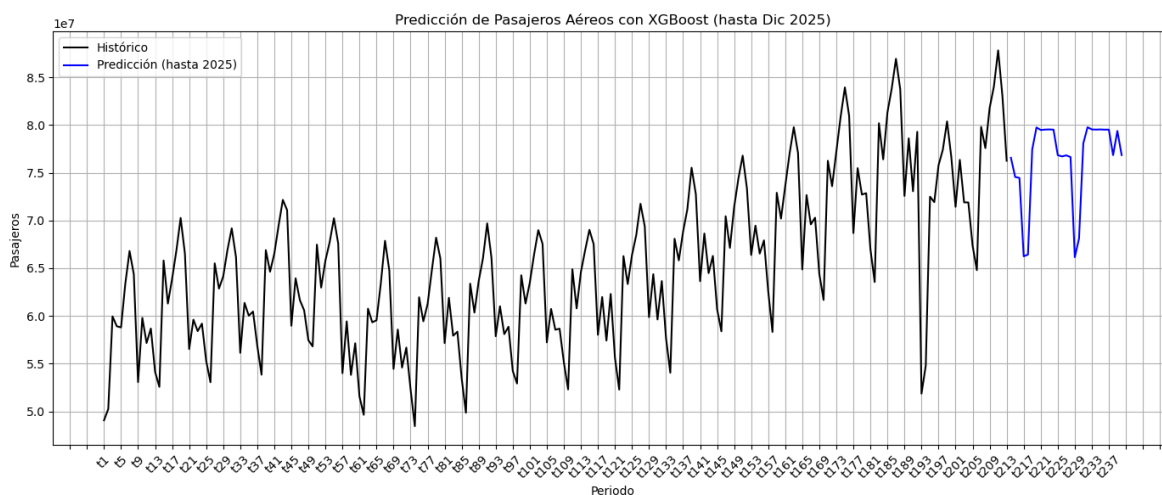
- **Fluctuaciones:** La serie temporal real muestra fluctuaciones significativas a lo largo del tiempo, con picos y valles irregulares. Esto sugiere que el número de pasajeros aéreos está influenciado por diversos factores que varían con el tiempo.
- **Tendencia Aparente:** Si bien hay mucha variabilidad, no se observa una tendencia lineal clara ascendente o descendente pronunciada en todo el período mostrado. Podría haber tendencias locales o patrones

cíclicos de mayor duración que no son inmediatamente evidentes debido a la volatilidad.

Comparación entre la Predicción y los Datos Reales:

- **Ajuste Variable:** El ajuste del modelo XGBoost a los datos reales no es constante en el tiempo. Hay períodos de buen ajuste donde las líneas están cerca una de la otra, y otros períodos donde la diferencia es mayor.

En resumen, la gráfica muestra que el modelo XGBoost ha logrado capturar algunas de las dinámicas de la serie temporal de pasajeros aéreos, pero también presenta errores de predicción en ciertos períodos. La variabilidad de los datos reales representa un desafío para cualquier modelo de predicción



Tendencia en la Predicción: En el horizonte de predicción (hasta diciembre de 2025), la línea azul muestra una tendencia que parece estabilizarse o incluso disminuir ligeramente después de un período inicial. Esto podría indicar que el modelo XGBoost prevé una desaceleración del crecimiento o una estabilización en el número de pasajeros para ese período.

La predicción sugiere una posible estabilización o ligera disminución en el crecimiento, manteniendo un patrón estacional anual. Sin embargo, la fiabilidad de esta predicción depende de la capacidad del modelo para capturar los patrones subyacentes y de la estabilidad de estos patrones en el futuro

Conclusiones:

En este proyecto se realizó el análisis de la demanda de pasajeros con respecto a al total de pasajeros en vuelos a nivel local e internacional de USA utilizando modelos predictivos que pudiesen predecir la demanda de pasajeros para 2025.

Se realizó la limpieza de datos, análisis descriptivos, selección de las variables de estudio, comportamiento de la data a través de graficos.

Se validaron tres modelos predictivos:

- ARIMA
- HoltWinters
- XGBOOST

De los 3 modelos predictivos estudiados, podemos indicar que el modelo ARIMA es el que más se ajusta al comportamiento históricos de los datos y por ende a la predicción de los mismos para el año 2025, el mismo presenta un R2 de 0.67

Recomendaciones

Aunque el modelo ARIMA es que más se ajusta, se recomienda evaluar otros modelos más y poder también evaluar las métricas de los datos del 2025 arrojados en el modelo predictivo para validar que tanto se ajustan a los reales.

Bibliografías

<https://uni.edu.gt/noticias/modelos-predictivos/>

<https://keyrus.com/sp/es/insights/las-11-tecnicas-mas-utilizadas-en-el-modelado-de-analisis-predictivos>

<https://nuclio.school/blog/modelos-predictivos-que-son-y-usos/>

<https://www.ibm.com/es-es/think/topics/arima-model>

<https://www.kaggle.com/datasets/yyxian/u-s-airline-traffic-data>

Anexos

En el siguiente enlace podrás ver los anexos.

https://github.com/Akaaino/Modelo_Predictivo.git

Base de datos

- ✓ air_traffic
- ✓ air_traffic_final

Limpieza y análisis descriptivo

- ✓ Validacion_Analisis_BD.Rmd
- ✓ Validacion_Analisis_BD.html

Modelos predictivos

- ✓ Modelo_Predictivo_holtwinters.ipynb
- ✓ Modelo_Predictivo_ARIMA.ipynb
- ✓ Modelo_Predictivo_xgboost.ipynb

Documentación Final

- ✓ Modelos-Predictivos-Proyecto-Final.pdf
- ✓ Proyecto_Final_Presentacion