

Pronóstico de demanda

Por: Leilan [REDACTED], Maria Jose [REDACTED] y Maryam [REDACTED]

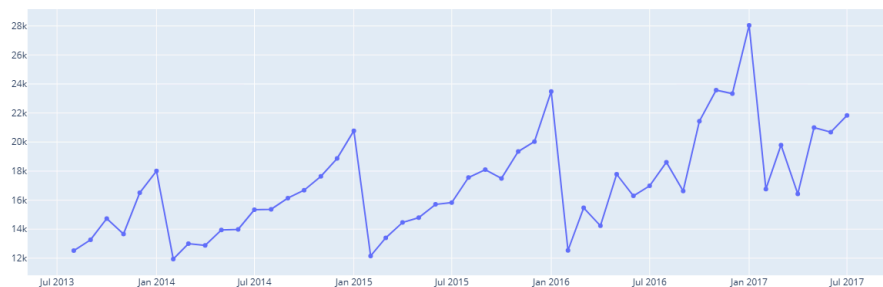
1. Introducción

El objetivo principal de este análisis es hacer un pronóstico de ventas para los próximos 12 meses. Los datos disponibles para este análisis incluyen ventas históricas de un producto a lo largo de cinco años. Esta data contiene 48 filas y está compuesta por 3 columnas: mes, año y demanda. El pronóstico de ventas se realizará por medio de dos modelos diferentes, justificando la selección de los mismos. El análisis se dividió en tres partes: análisis de los datos, construcción y evaluación de los modelos, y los pronósticos.

2. Limpieza y Análisis:

El proceso comenzó con el análisis de los datos, los cuales no presentaron valores nulos ni outliers que no pudieran ser explicados. Al estar trabajando con una serie de tiempo, era esencial que la data estuviera ordenada cronológicamente. Para ello, se creó una columna nueva llamada *mes_número* ya que la columna *mes* contenía los nombres de los meses en formato de texto. En *mes_número* se colocó un número del mes correspondiente con los meses en la columna *mes*. Luego se añadió otra columna llamada *fecha* donde combinamos la columna *año*, *mes_número* y el día que fue establecido como 1 para tener una fecha completa. Esta columna fue transformada a un índice y nos permitió organizar los datos por tiempo.

Una vez asegurada la consistencia y usabilidad de los datos, realizamos una primera visualización de la serie temporal con el fin de identificar las características principales que nos permitan reconocer los componentes clave a tener en cuenta para la elección del modelo de pronóstico a utilizar.



El gráfico representa la evolución de una demanda que muestra tener fluctuaciones regulares con un patrón que se repite anualmente donde se muestra una caída drástica de la demanda al principio de año y que empieza a aumentar progresivamente a partir del mes de febrero hasta alcanzar el máximo a finales de cada año. Vemos además que el nivel de la demanda va aumentando cada año que pasa, demostrando que hay una tendencia positiva. Así, observamos que se trata de una serie temporal que exhibe una tendencia y una clara estacionalidad.

3. Construcción del modelo

Para garantizar la selección de un modelo adecuado, consideramos aquellos modelos que incorporan los componentes de estacionalidad y tendencia que hemos observado en los datos. Por ello, el siguiente apartado explora la construcción de modelos de pronóstico utilizando, por un lado, un modelo de alisado exponencial triple y por otro, un modelo ARIMA.

A. Modelo Alisado Triple:

El método **Holt-Winters** es una extensión del método de alisado exponencial que incluye el efecto tanto del nivel como de la tendencia y de la estacionalidad. La naturaleza de éste último efecto, la estacionalidad, da lugar a dos posibles variaciones del modelo: el **aditivo** y el **multiplicativo**.

En el contexto de nuestros datos, el comportamiento de las oscilaciones en la demanda parecen mostrar un ligero aumento en la amplitud de la variación estacional a medida que la serie temporal avanza, lo cual sugeriría un modelo multiplicativo. Sin embargo, dado que se trata de un aumento sutil, procederemos con la construcción de ambas variaciones del modelo para evaluar con más exactitud cuál es el más adecuado.

```

=====
ExponentialSmoothing Model Results
=====
Dep. Variable:          demanda      No. Observations:      48
Model:                  ExponentialSmoothing      SSE      63116813.237
Optimized:              True          AIC      788.286
Trend:                  Additive      BIC      738.225
Seasonal:               Additive      AICC     731.872
Seasonal Periods:      12            Date:                Tue, 25 Mar 2025
Box-Cox:                False         Time:                15:13:03
Box-Cox Coeff.:        None
=====

```

	coeff	code	optimized
smoothing_level	0.3339483	alpha	True
smoothing_trend	0.0053982	beta	True
smoothing_seasonal	0.0106606	gamma	True
initial_level	14131.400	l.0	True
initial_trend	207.82325	b.0	True
initial_seasons.0	-71.660426	s.0	True
initial_seasons.1	-405.64685	s.1	True
initial_seasons.2	1032.2191	s.2	True
initial_seasons.3	2413.6000	s.3	True
initial_seasons.4	2891.2330	s.4	True
initial_seasons.5	5978.7318	s.5	True
initial_seasons.6	-3828.8698	s.6	True
initial_seasons.7	-2169.3404	s.7	True
initial_seasons.8	-2567.2110	s.8	True
initial_seasons.9	-1015.1217	s.9	True
initial_seasons.10	-1433.0064	s.10	True
initial_seasons.11	-924.28023	s.11	True

```

=====
/usr/local/lib/python3.11/dist-packages/statsmodels/tsa/base/tsa_model.py:473: Valu

```

Imagen 3.1 Alisado Triple - Aditivo

Los coeficientes de suavización son moderados en cuanto al nivel, mostrando que el modelo **multiplicativo** le da un poco más de peso a los valores recientes. El parámetro de suavizado para la estacionalidad es extremadamente bajo lo que indica estabilidad.

```

=====
ExponentialSmoothing Model Results
=====
Dep. Variable:          demanda      No. Observations:      48
Model:                  ExponentialSmoothing      SSE      49484984.581
Optimized:              True          AIC      696.607
Trend:                  Additive      BIC      726.546
Seasonal:               Multiplicative      AICC     720.193
Seasonal Periods:      12            Date:                Tue, 25 Mar 2025
Box-Cox:                False         Time:                15:13:03
Box-Cox Coeff.:        None
=====

```

	coeff	code	optimized
smoothing_level	0.3585714	alpha	True
smoothing_trend	0.0001	beta	True
smoothing_seasonal	0.0001	gamma	True
initial_level	14231.942	l.0	True
initial_trend	169.69015	b.0	True
initial_seasons.0	1.0024992	s.0	True
initial_seasons.1	0.9866651	s.1	True
initial_seasons.2	1.0608760	s.2	True
initial_seasons.3	1.1442169	s.3	True
initial_seasons.4	1.1672271	s.4	True
initial_seasons.5	1.3384662	s.5	True
initial_seasons.6	0.7687815	s.6	True
initial_seasons.7	0.8674786	s.7	True
initial_seasons.8	0.8552692	s.8	True
initial_seasons.9	0.9398444	s.9	True
initial_seasons.10	0.9186305	s.10	True
initial_seasons.11	0.9500452	s.11	True

```

=====
/usr/local/lib/python3.11/dist-packages/statsmodels/tsa/holtwinters/model.py:918: C

```

Imagen 3.2 Modelo Alisado Triple - Multiplicativo

Una comparación de los criterios de información nos permite ver que los valores del modelo **Holt-Winters** multiplicativo son menores en todas las instancias, lo cual refuerza la idea de que este es el modelo más adecuado ya que demostró un mejor ajuste y que, efectivamente, la amplitud de las fluctuaciones estacionales (la magnitud de los patrones estacionales) cambia con el nivel de la serie.

Pronóstico

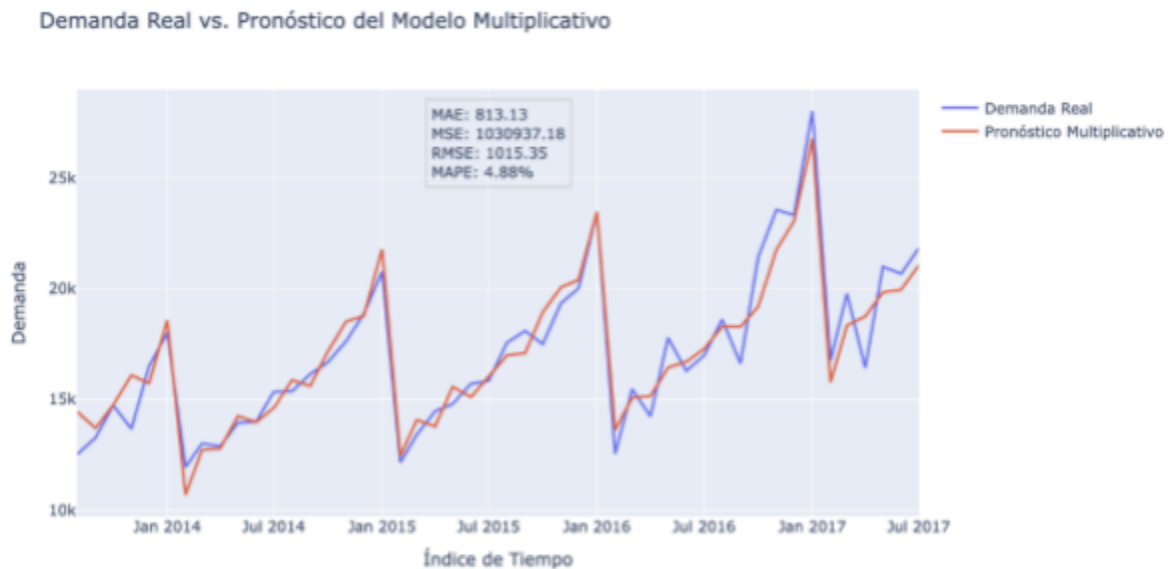


Imagen 3.3 Demanda Real vs Pronóstico del Modelo Multiplicativo

Podemos observar que el modelo multiplicativo captura de manera efectiva los patrones de la demanda, con una línea de pronóstico que sigue de cerca la línea (azul) de demanda real. También podemos observar que la demanda real ha fluctuado en el pasado, lo que nos permite identificar tendencias y patrones estacionales.



Imagen 3.4 Comparación de Alisado Triple Holt - Aditivo vs. Multiplicativo

Como podemos ver en el gráfico, ambos modelos capturan la estacionalidad de los datos, pero con ligeras variaciones en la magnitud de las fluctuaciones. En cuanto a la tendencia, ambos modelos proyectan una continuación de la tendencia ascendente en sus pronósticos.

Las diferencias entre los pronósticos aditivo y multiplicativo son más notables hacia el final del período de pronóstico. Esto nos sugiere que la elección entre un modelo aditivo y multiplicativo puede tener un impacto significativo en las predicciones a largo plazo. Esto significa que a corto plazo, ambos modelos pueden dar resultados similares, pero a largo plazo, las predicciones pueden variar significativamente.



Imagen 3.5 Comparación de Predicciones (Holt-Winters)

Las métricas de error indican que el modelo Holt-Winters tiene un margen de error significativo en la predicción de la demanda. Un MAPE de 15.36% nos sugiere que las predicciones se desvían en un 15.36% de los valores reales.

En la primera sección de este pronóstico de demanda, observamos que el uso del Alisado Exponencial de Holt-Winters nos permite modelar los datos de demanda, identificando la existencia de tendencias y estacionalidad en la serie de tiempo. Investigamos dos variantes: el modelo **Aditivo** que asume que el elemento estacional se añade a la tendencia, mientras que el modelo **Multiplicativo** que asume que el elemento estacional se incrementa con la tendencia. Nuestro análisis demostró que el modelo multiplicativo tiene una Suma de Errores al Cuadrado (SSE), AIC, BIC y AICC más bajos, lo que indica un mejor ajuste a los datos en comparación con el modelo aditivo. Los parámetros de suavizado (α , β y γ) en ambos modelos son relativamente bajos, lo que nos sugiere que los modelos se adaptan lentamente a los cambios recientes en el nivel, la tendencia y la estacionalidad. En resumen, podemos observar que el modelo multiplicativo parece ajustarse mejor durante los períodos de mayor demanda que el aditivo.

B. Modelo ARIMA:

Preparación del Modelo

El siguiente modelo que se construyó fue un **ARIMA**. Este modelo requiere que los datos sean estacionarios, dado que asume que la media, la varianza y la covarianza son constantes a lo largo del tiempo. Para tener una serie estacionaria era necesario eliminar la tendencia y estacionalidad de nuestra serie ya que, de no hacerlo, podría generar predicciones incorrectas. Para comprobar que nuestra serie no era estacionaria se utilizó la prueba de Dickey-Fuller. Los resultados confirmaron que nuestra serie no era estacionaria, significando que teníamos que aplicar la técnica de diferenciación para convertir nuestra serie a una estacionaria. Comenzamos con la primera diferenciación, la cual eliminó la tendencia.

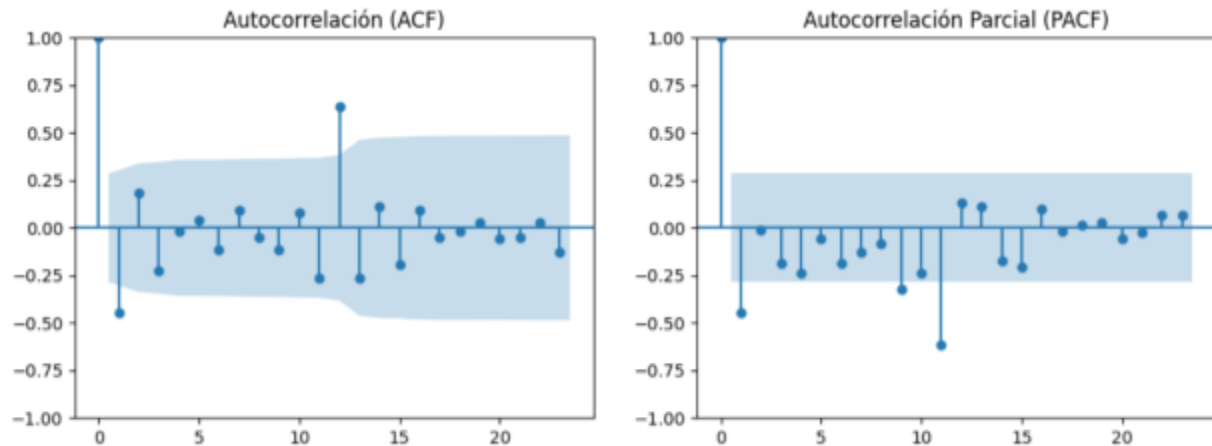


Imagen 3.4 Gráficos de ACF y PACF luego de la primera diferenciación

Después de la primera diferenciación, aplicamos nuevamente la prueba de Dickey-Fuller, la cual nos muestra que la serie había sido transformada en estacionaria. Sin embargo, realizamos una segunda diferenciación para eliminar los patrones estacionales que sabíamos que estaban presentes en la serie. En esta segunda diferenciación especificamos el patrón estacional de cada 12 meses.

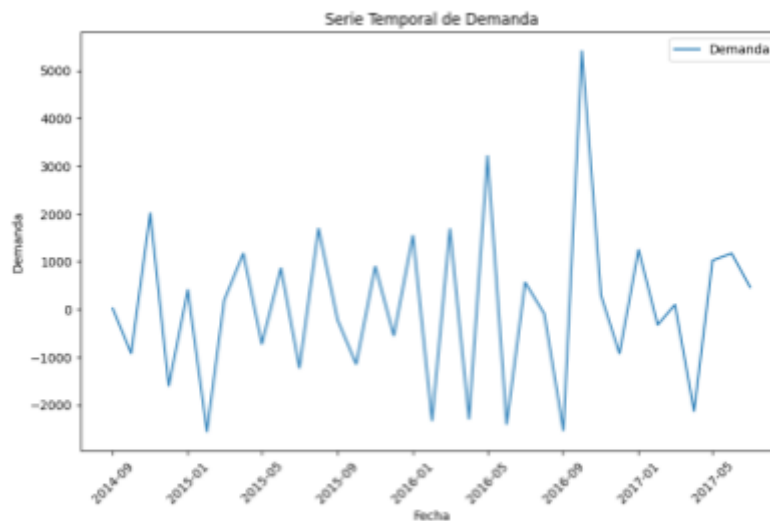


Imagen 3.5 Serie Temporal de Demanda después de segunda diferenciación

En el gráfico se observa que ya no hay una tendencia presente y hay una reducción de la estacionalidad. Mientras que no ha desaparecido completamente, hacer otra diferenciación no eliminaría la estacionalidad restante y empeoraría el modelo al añadir más ruido a la serie.

Los gráficos ACF y PACF resultantes después de hacer la doble diferenciación muestran una caída súbita en el primer elemento del lag en lugar de descender progresivamente, lo cual también es prueba de que la serie temporal ha alcanzado una estacionariedad suficiente.

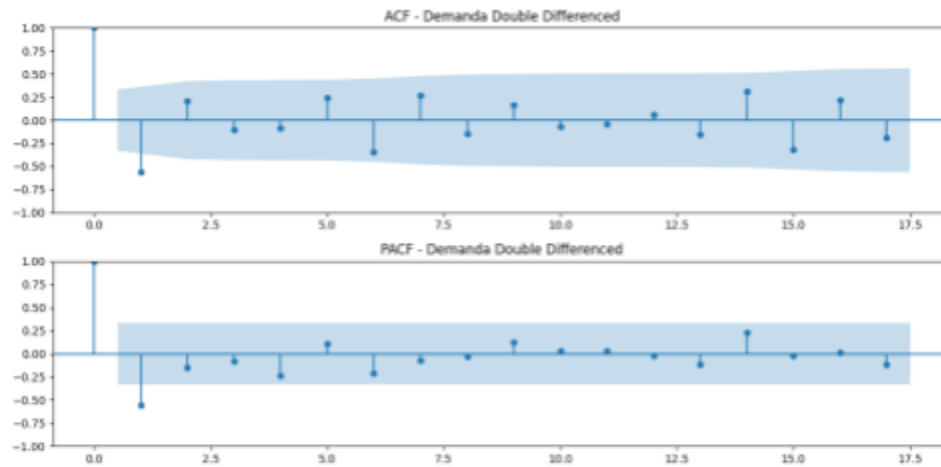


Imagen 3.6 Gráficos de ACF y PACF después de la segunda diferenciación

Los gráficos de ACF y PACF de ambas diferenciaciones nos ayudan a obtener los valores para nuestro modelo de orden (p, d, q) y el orden estacional (P, D, Q, S) . Empezamos con la selección del orden (p, d, q) . Para determinar el valor de p , es necesario observar el gráfico PACF de la primera diferenciación, en él se cuenta cuántas piruletas están fuera del intervalo de confianza antes de volverse insignificantes y entrar dentro de la capa azul en el gráfico. Después de la primera piruleta las siguientes 10 piruletas están dentro del intervalo de confianza, por lo tanto asignamos $p = 1$. Para encontrar el valor de q , miramos el gráfico de ACF de la primera diferenciación, y de la misma forma que se determina el valor de p , buscamos cuántas piruletas están fuera del intervalo de confianza antes de volverse insignificantes, por tal razón establecemos $q = 1$. El valor de d es determinado por la cantidad de diferenciaciones necesarias para convertir la serie en estacionaria. En nuestro caso, con solo la primera diferenciación la serie se convierte en estacionaria, por lo que asignamos $d = 1$. Nuestro orden en el modelo ARIMA será **(1,1,1)**.

Con los gráficos ACF y PACF de la segunda diferenciación, podemos determinar los valores del orden estacional (P, D, Q, S) . Para encontrar los valores de P y Q , necesitamos detectar si existe alguna piruleta significativa en los gráficos PACF y ACF, respectivamente, durante el patrón estacional, en este caso es 12. Al observar que no hay ninguna piruleta fuera del intervalo de confianza en el gráfico cuando el eje X está en 12, concluimos que $P = 0$ y $Q =$

0. Al necesitar hacer una diferenciación para eliminar la estacionalidad, esto nos lleva a que $D = 1$. Por último, $S = 12$, ya que S representa el patrón estacional de la serie. El primer modelo de ARIMA que probamos es **ARIMA(1,1,1)(0,1,0,12)**.

Creación del modelo

Al crear el modelo, observamos que los coeficientes no son significativos, pero puede funcionar como base. Con eso en mente, decidimos usar el *auto_arima* para obtener su sugerencia. El *auto_arima* nos sugiere dos modelos diferentes: ARIMA(1,0,0)(1,1,0)[12] y ARIMA(1,0,0)(0,1,0)[12]. Probamos ambos modelos, y entre los dos, el modelo ARIMA(1,0,0)(0,1,0)[12] presenta un AIC y un BIC más bajo, lo que indica que tiene un mejor ajuste. En términos de coeficientes, ese mismo modelo tiene menos coeficientes pero todos son significativos. Después de analizar los resultados llegamos a la conclusión que el mejor modelo entre los dos obtenido por el *auto_arima* es: **ARIMA(1,0,0)(0,1,0)[12]**.

SARIMAX Results						
Dep. Variable:	demanda		No. Observations:	48		
Model:	ARIMA(1, 0, 0)x(0, 1, 0, 12)		Log Likelihood	-322.810		
Date:	Sun, 23 Mar 2025		AIC	649.621		
Time:	12:50:04		BIC	652.788		
Sample:	00-01-2013		HQIC	650.726		
	- 07-01-2017					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.5438	0.052	10.513	0.000	0.442	0.645
sigma2	3.368e+06	9.08e+05	3.709	0.000	1.59e+06	5.15e+06
Ljung-Box (L1) (Q):	2.37	Jarque-Bera (JB):	3.96			
Prob(Q):	0.12	Prob(JB):	0.14			
Heteroskedasticity (H):	1.75	Skew:	-0.60			
Prob(H) (two-sided):	0.35	Kurtosis:	4.11			
Warnings:						
[1] Covariance matrix calculated using the outer product of gradients (complex-step).						

Imagen 3.7 Modelo ARIMA (1,0,0)(0,1,0,12)

Comparamos ese modelo con el modelo seleccionado al principio de ARIMA(1,1,1)(0,1,0,12). En esta comparación, se observa que el AIC y BIC del modelo al principio tiene un mejor AIC y BIC con un diferencia en el AIC de 31.56 y en el BIC de 30.061. Un AIC y BIC más bajo nos indica un mejor ajuste, mejor calidad de modelo y una capacidad de ajustarse a los datos sin sobre ajustarse. Sin embargo, ese mismo modelo tiene coeficientes no significativos que hace el modelo no tan fuerte ya que los componentes AR y MA no muestra tener un impacto real.

```

=====
SARIMAX Results
=====
Dep. Variable:          demanda    No. Observations:          48
Model:                ARIMA(1, 1, 1)x(0, 1, [ ], 12)    Log Likelihood          -306.030
Date:                  Sun, 23 Mar 2025    AIC                    618.061
Time:                  16:35:40    BIC                    622.727
Sample:                00-01-2013    HQIC                   619.672
                        - 07-01-2017
Covariance Type:      opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
ar.L1         -0.4261      0.424      -1.005      0.315      -1.257      0.405
ma.L1          0.0633      0.480       0.132      0.895      -0.878      1.005
sigma2         2.091e+06    4.55e+05     4.594      0.000      1.2e+06    2.98e+06
=====
Ljung-Box (L1) (Q):                1.69    Jarque-Bera (JB):                1.85
Prob(Q):                           0.19    Prob(JB):                     0.40
Heteroskedasticity (H):             1.96    Skew:                          0.49
Prob(H) (two-sided):                0.26    Kurtosis:                     3.55
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

```

Imagen 3.8 Modelo ARIMA(1,1,1)(0,1,0,12)

Encontrar un modelo con coeficientes significativos nos asegura que un modelo está capturando correctamente los patrones de la serie y puede hacer predicciones fiables. Dado que los modelos anteriores tienen sus desventajas (uno con coeficientes no significativos y el otro un AIC y un BIC más alto) decidimos probar con un nuevo modelo: **ARIMA(1,1,0)(0,1,0,12)**. Este modelo combina elementos de los dos anteriores (ARIMA(1,1,1)(0,1,0,12) y ARIMA(1,0,0)(0,1,0,12)). Para este nuevo modelo, se mantiene $d=1$ ya que sabemos que la serie no es estacionaria si no se hace una diferenciación y el modelo anterior con $d=0$ tenía un AIC y BIC más alto. En este nuevo modelo, $q=0$ pues en el modelo donde $q=1$, el componente MA era un coeficiente no significativo. Al no aportar al modelo ni mejorar su ajuste, el término de media móvil se excluye.

```

=====
SARIMAX Results
=====
Dep. Variable:          demanda    No. Observations:          48
Model:                ARIMA(1, 1, 0)x(0, 1, 0, 12)    Log Likelihood          -306.046
Date:                  Sun, 23 Mar 2025    AIC                    616.092
Time:                  12:49:34    BIC                    619.203
Sample:                00-01-2013    HQIC                   617.166
                        - 07-01-2017
Covariance Type:      opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
ar.L1         -0.3762      0.082     -4.566      0.000      -0.538      -0.215
sigma2         2.062e+06    4.3e+05     4.797      0.000      1.22e+06    2.9e+06
=====
Ljung-Box (L1) (Q):                1.54    Jarque-Bera (JB):                1.69
Prob(Q):                           0.21    Prob(JB):                     0.43
Heteroskedasticity (H):             1.89    Skew:                          0.47
Prob(H) (two-sided):                0.29    Kurtosis:                     3.51
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

```

Imagen 3.9 Modelo ARIMA(1,1,0)(0,1,0,12)

Al analizar este modelo, observamos que tiene los valores más bajos de AIC y BIC, lo que sugiere que es el que mejor se ajusta a los datos. Además, sus coeficientes son significativos lo que confirma que no era necesario incluir la media móvil en el modelo. Este modelo logra explicar la serie a través de componente autorregresivo (AR) y captura los patrones de la serie con la menor cantidad posible de parámetros, lo que ayuda a reducir el riesgo de *overfitting*. Comparado con los modelos anteriores, el valor de σ^2 es el más bajo, lo cual es positivo, ya que σ^2 mide la varianza del error en el modelo y un valor más bajo significa menos ruido en las predicciones.

Los resultados de este modelo y de la prueba de Ljung-Box muestran que los residuos no tienen autocorrelación significativa, son ruido blanco, siguen una distribución normal, tienen una varianza constante y los errores son aleatorios. Esto nos asegura que nuestro modelo tendrá predicciones estables y que no queda ninguna parte de los datos sin explicar. Este modelo es el más robusto y preciso de todos.

Validación del Modelo ARIMA

Después de tener nuestro modelo, es necesario validar que esté funcionando correctamente. Estaremos comprobando con las tres versiones hechas de este modelo para validar que el **ARIMA(1,1,0)(0,1,0,12)** es el mejor de todos. La data se dividió en entrenamiento y prueba, en donde la prueba son los últimos 12 datos de la serie original. Los modelos están prediciendo los últimos 12 meses de la data original para observar cual es el más cercano al dato original.



Imagen 3.10 Comparando los modelos de ARIMA

En el gráfico se observa que mientras todos los modelos hacen una predicción bastante similar, el último modelo **ARIMA(1,1,0)(0,1,0,12)** tiene los picos más altos y se acerca más a los datos reales. También, se hizo el RMSE y MSE de los tres modelos y se observa que el último modelo tiene los valores más bajos con un RMSE de 2878.064 y un MSE de 8283253.59.

```
Modelo 1 (ARIMA(1,1,1)(0,1,0,12)):
MSE: 9980120.628260635, RMSE: 3159.132891832921, MAE: 3008.0302598814887

Modelo 2 (ARIMA(1,0,0)(0,1,0,12)):
MSE: 13476996.475039959, RMSE: 3671.1028962751725, MAE: 3447.0395765418302

Modelo 3 (ARIMA(1,0,0)(0,1,0,12)):
MSE: 8283253.5943874465, RMSE: 2878.064209566466, MAE: 2687.4229358854827
```

Imagen 3.11 MSE, RMSE y MAE de los modelos

Con estos resultados podemos concluir que el mejor modelo es **ARIMA(1,1,0)(0,1,0,12)** y con ese modelo se hará el pronóstico de los próximos 12 meses.

Pronóstico

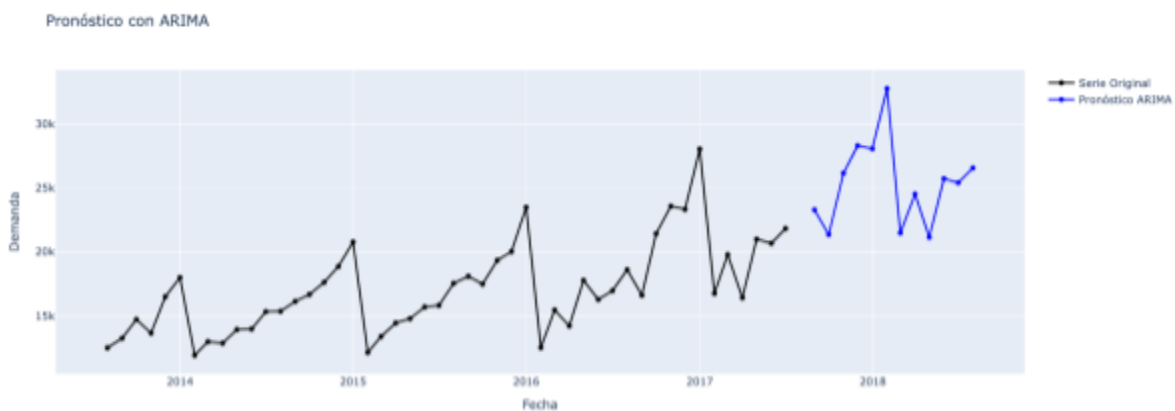


Imagen 3.12 Pronóstico de demanda - Modelo ARIMA

El pronóstico de este modelo muestra de la misma manera que la serie original, un crecimiento a través de los años. El mes de enero sigue teniendo la demanda más alta de todos y en el pronóstico demuestra tener una demanda de 32.76553k y una bajada de 11k el próximo mes con una demanda de 21.49004k.

4. Comparación del Modelo Alisado Triple y Modelo ARIMA

Habiendo desarrollado dos modelos distintos de pronóstico, procederemos con una comparación del rendimiento de cada uno de ellos con la finalidad de visualizar cómo se comportan y qué tan bien se adaptan a los movimientos reales de la demanda.

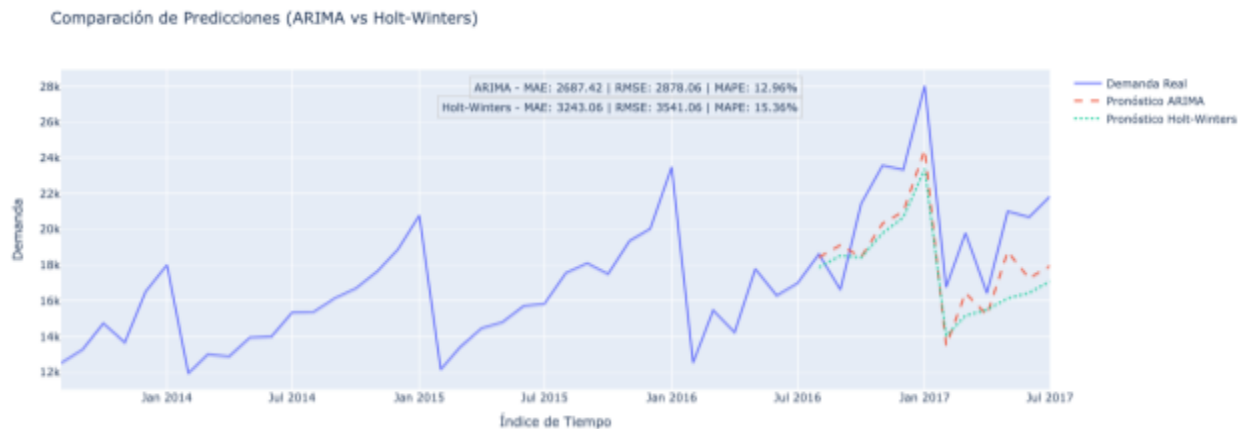


Imagen 4.1 Comparación entre modelo Holt y ARIMA

Para ello, se dividen las observaciones de nuestros datos en un conjunto de entrenamiento y uno de prueba, tomando los doce últimos datos como el conjunto de referencia para evaluar los pronósticos. Como se puede ver en la imagen 4.1, el desempeño de ambos modelos es relativamente similar, con la diferencia de que el ARIMA demuestra tener mejor capacidad para adaptarse a las variaciones mientras que el alisado exponencial es más estable.

Las métricas de evaluación MAE, RMSE y MAPE respaldan la superioridad del modelo ARIMA, permitiéndonos orientar nuestra preferencia hacia este modelo.

5. Conclusión

Al comparar ambos modelos, se concluyó que el modelo $ARIMA(1,1,0)(0,1,0,12)$ proporcionó los pronósticos más precisos para los próximos 12 meses. Este modelo captura eficazmente la tendencia y la estacionalidad de la demanda, ofreciendo predicciones confiables para la planificación de ventas y la toma de decisiones estratégicas. De cumplirse el pronóstico, la demanda tendrá un crecimiento proyectado para el futuro de 14.04%. Este incremento se alinea con la estacionalidad identificada en el análisis de datos y capturada por el modelo $ARIMA(1,1,0)(0,1,0,12)$. En resumen, este análisis proporciona una base sólida para la planificación de ventas y la toma de decisiones estratégicas.