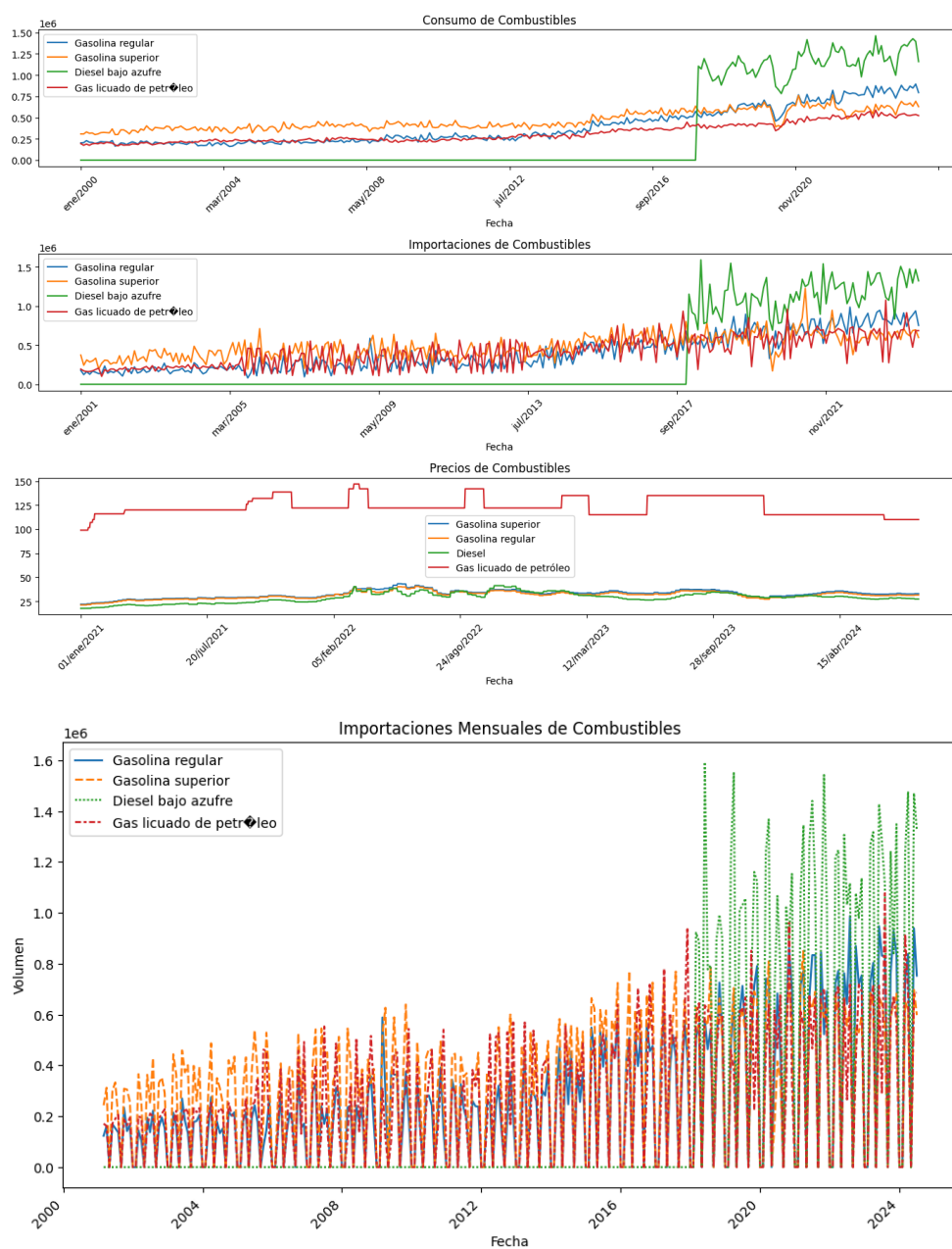
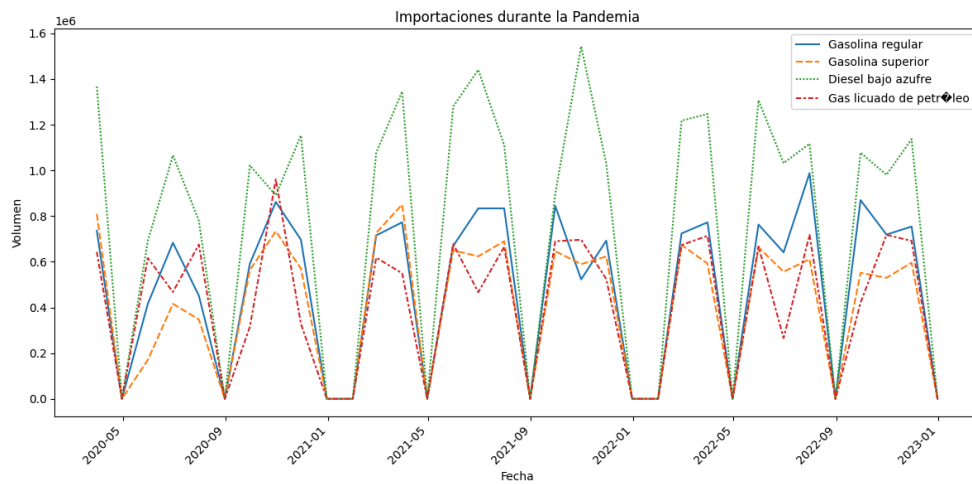


Data Science

Laboratorio 02

Análisis exploratorio





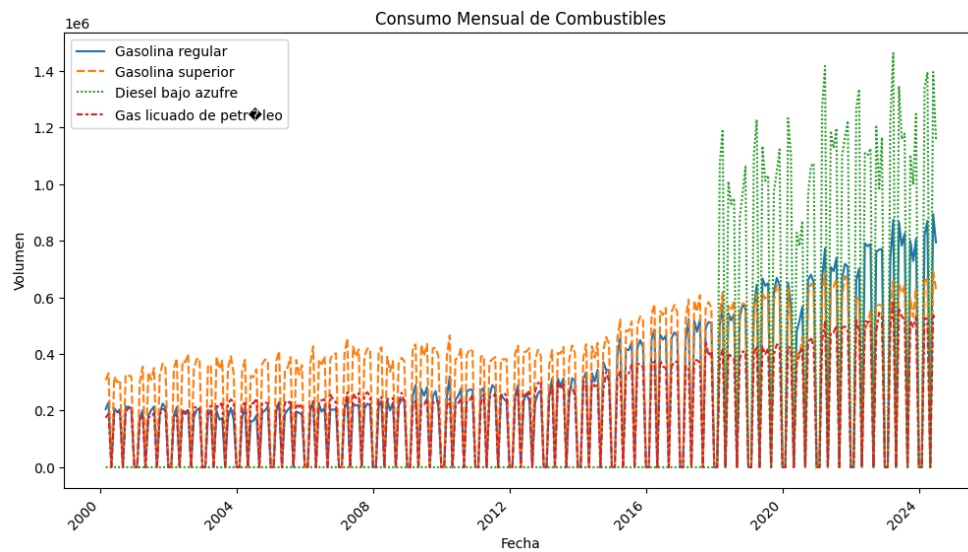
Análisis de las series de tiempo

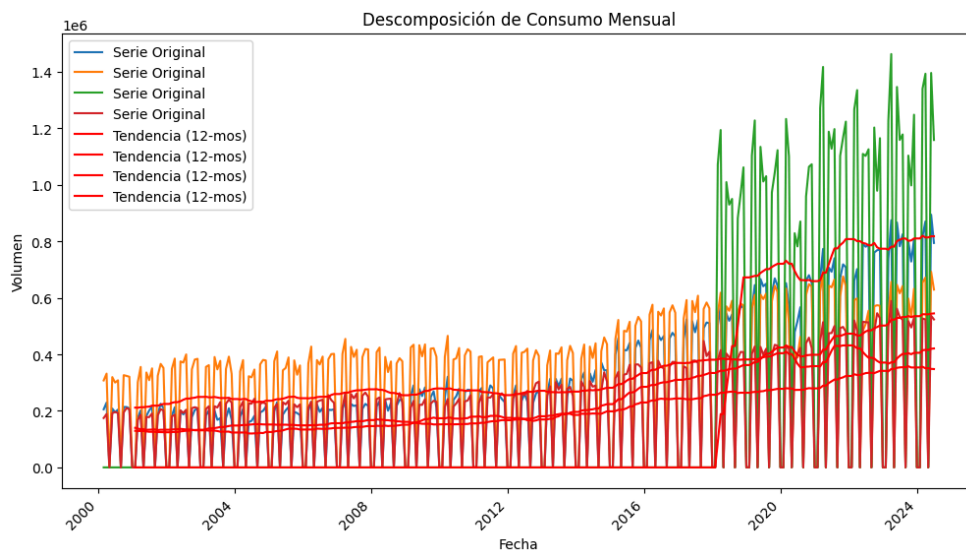
- Serie de consumo (Gasolina regular)

```

Inicio de los datos: 2000-02-01 00:00:00
Fin de los datos: 2024-06-01 00:00:00
Frecuencia de los datos: Fecha
31 days    49
30 days    48
28 days    18
29 days     7
Name: count, dtype: int64

```



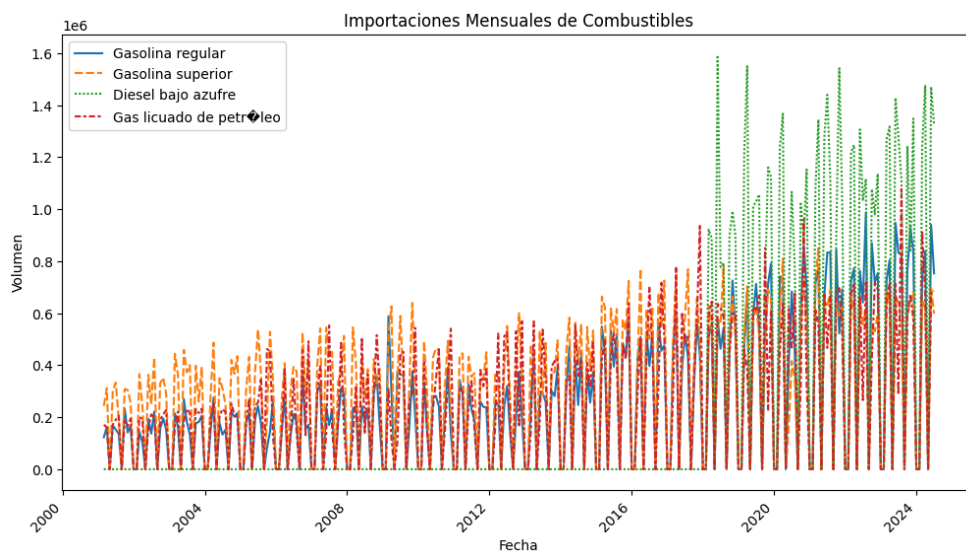


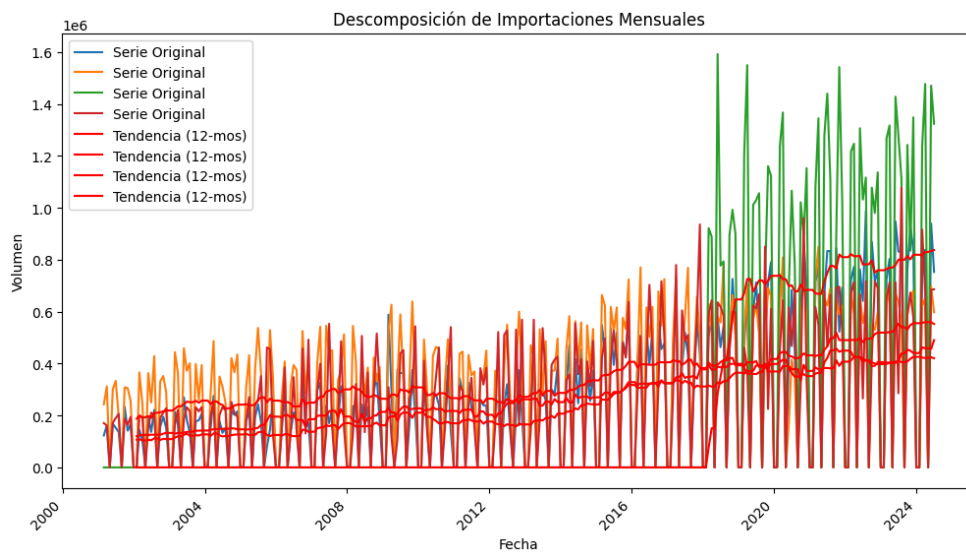
- Serie de importacion (Gasolina regular)

```

Inicio de los datos: 2001-02-01 00:00:00
Fin de los datos: 2024-06-01 00:00:00
Frecuencia de los datos: Fecha
31 days    47
30 days    46
28 days    18
29 days     6
Name: count, dtype: int64

```

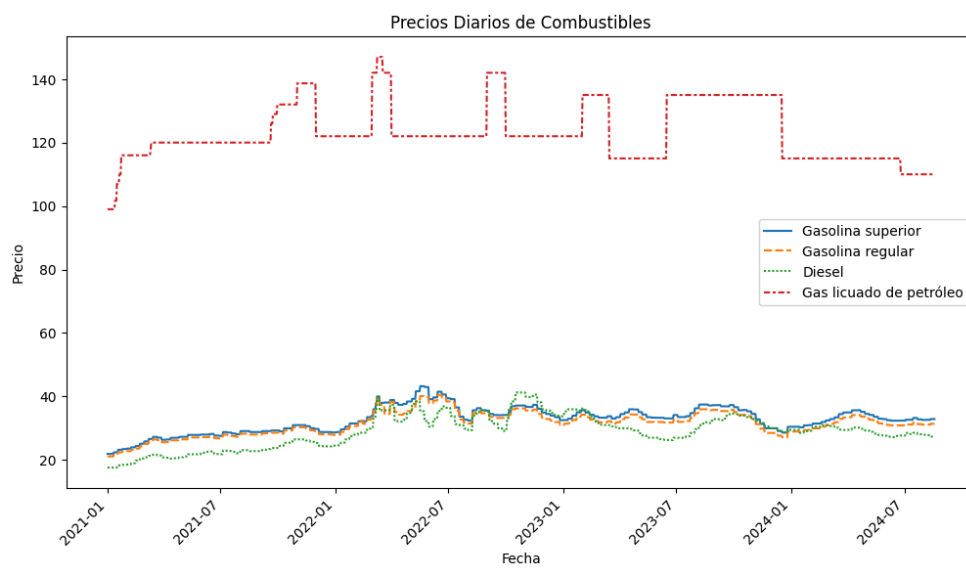




- Serie de precio (Gasolina regular)

```

Inicio de los datos: 2021-01-01 00:00:00
Fin de los datos: 2024-08-18 00:00:00
Frecuencia de los datos: Fecha
1 days    1325
Name: count, dtype: int64
  
```

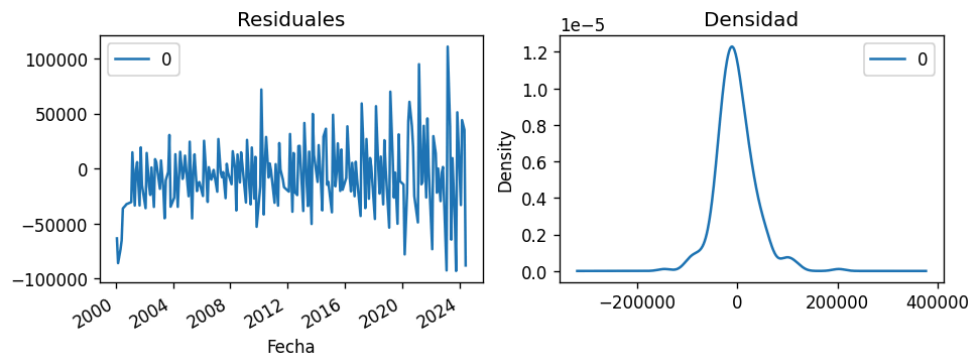


```

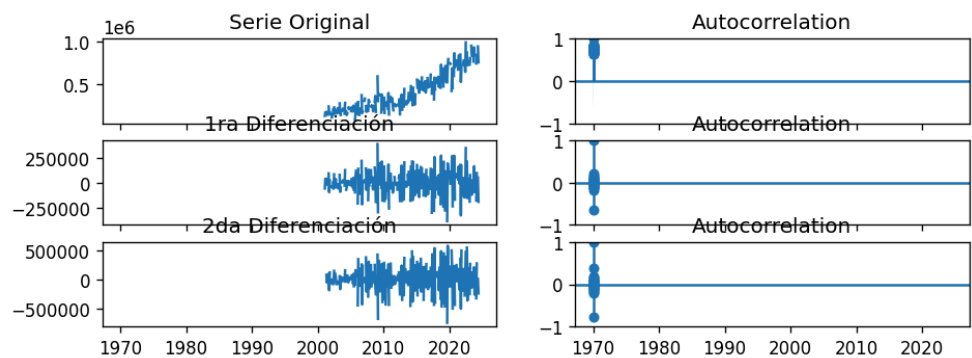
=====
SARIMAX Results
=====
Dep. Variable:      Gasolina regular      No. Observations:      294
Model:              ARIMA(1, 2, 1)         Log Likelihood         -3493.176
Date:              Sun, 18 Aug 2024       AIC                    6992.352
Time:              22:00:28               BIC                    7003.382
Sample:            0                      HQIC                   6996.770
                  - 294
Covariance Type:   opg

=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          -0.3458      0.053      -6.532      0.000      -0.450      -0.242
ma.L1          -0.9972      0.054     -18.363      0.000      -1.104      -0.891
sigma2         1.765e+09     1.1e-11     1.61e+20      0.000     1.77e+09     1.77e+09
=====
Ljung-Box (L1) (Q):              2.26      Jarque-Bera (JB):              38.64
Prob(Q):                        0.13      Prob(JB):                  0.00
Heteroskedasticity (H):          3.80      Skew:                      0.35
Prob(H) (two-sided):             0.00      Kurtosis:                  4.64
=====

```

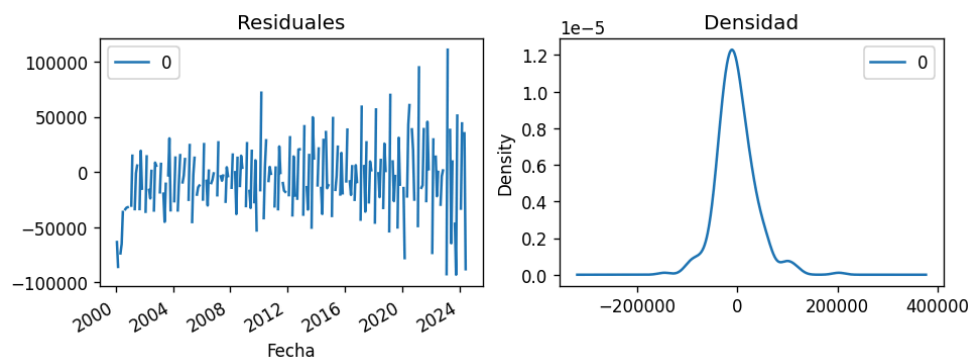


- Serie de importacion (Gasolina regular)

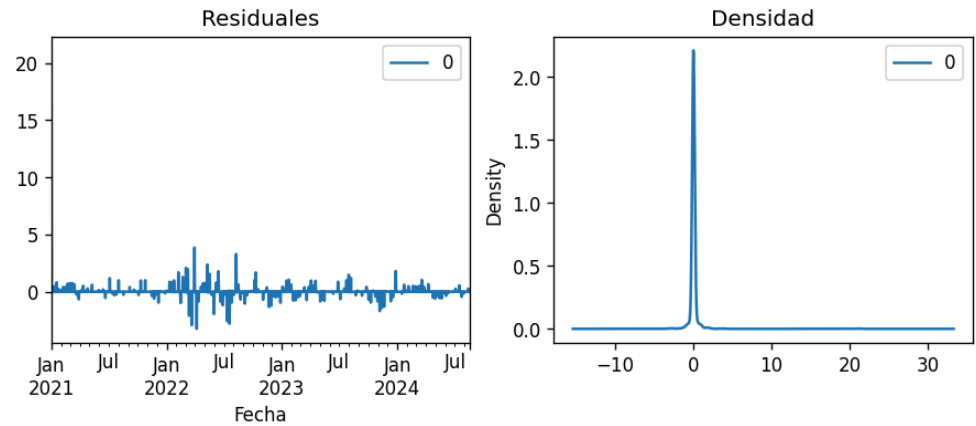
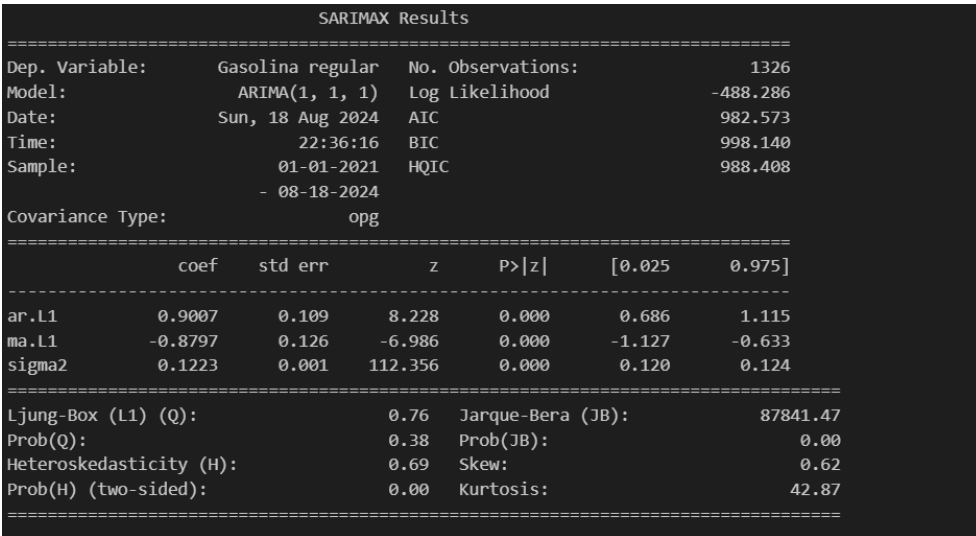
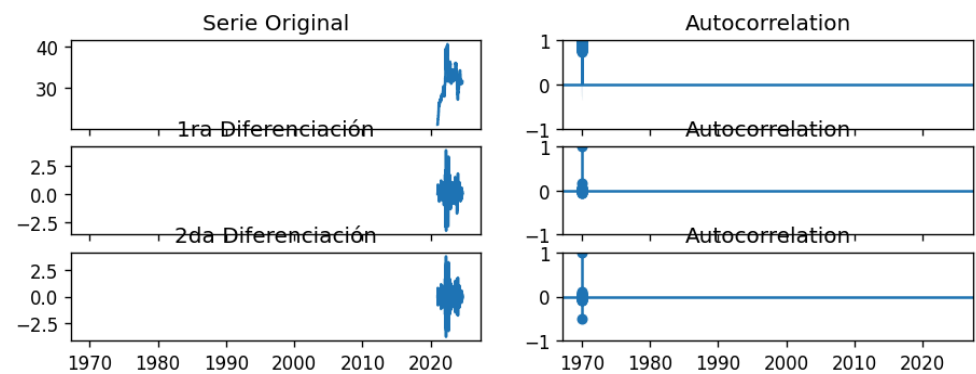


SARIMAX Results						
=====						
Dep. Variable:	Gasolina regular	No. Observations:	294			
Model:	ARIMA(1, 2, 1)	Log Likelihood	-3493.176			
Date:	Sun, 18 Aug 2024	AIC	6992.352			
Time:	21:45:09	BIC	7003.382			
Sample:	0	HQIC	6996.770			
	- 294					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

ar.L1	-0.3458	0.053	-6.532	0.000	-0.450	-0.242
ma.L1	-0.9972	0.054	-18.363	0.000	-1.104	-0.891
sigma2	1.765e+09	1.1e-11	1.61e+20	0.000	1.77e+09	1.77e+09
=====						
Ljung-Box (L1) (Q):	2.26	Jarque-Bera (JB):	38.64			
Prob(Q):	0.13	Prob(JB):	0.00			
Heteroskedasticity (H):	3.80	Skew:	0.35			
Prob(H) (two-sided):	0.00	Kurtosis:	4.64			
=====						

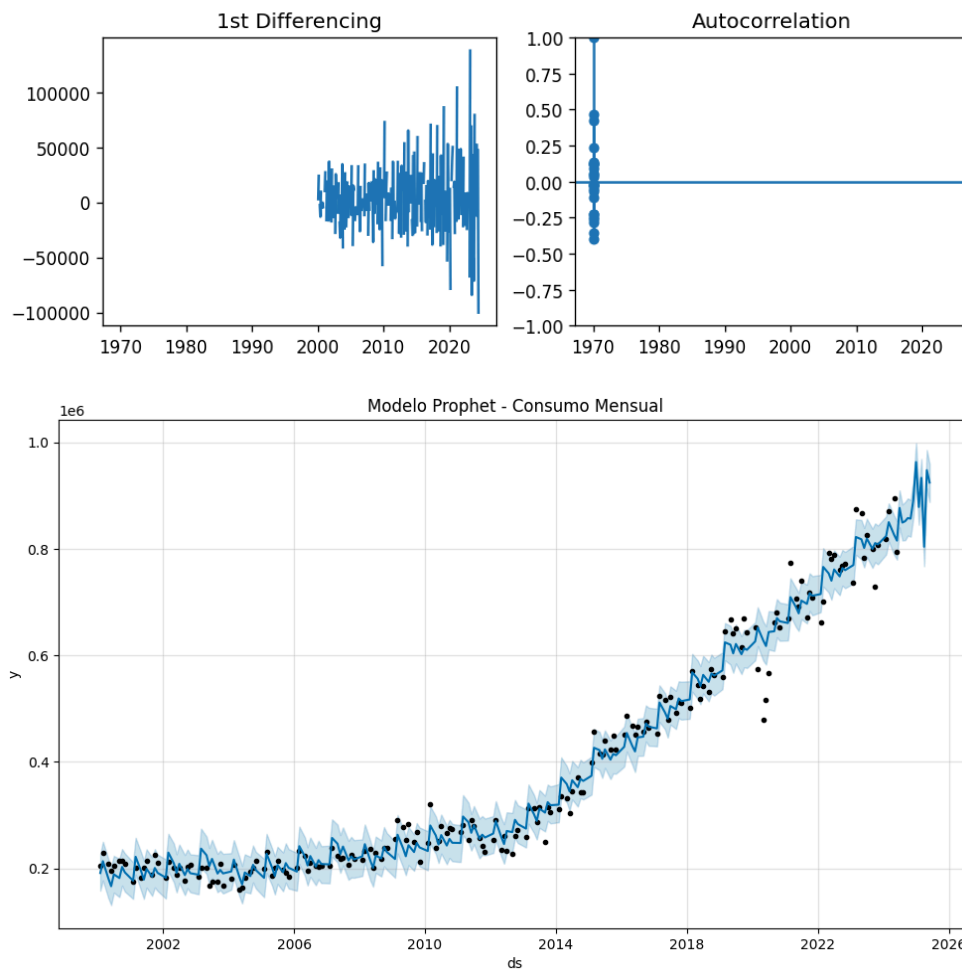


- Serie de precio (Gasolina regular)

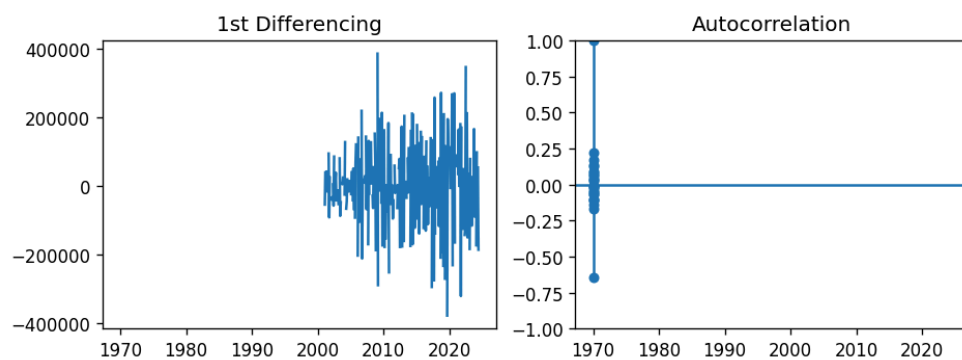


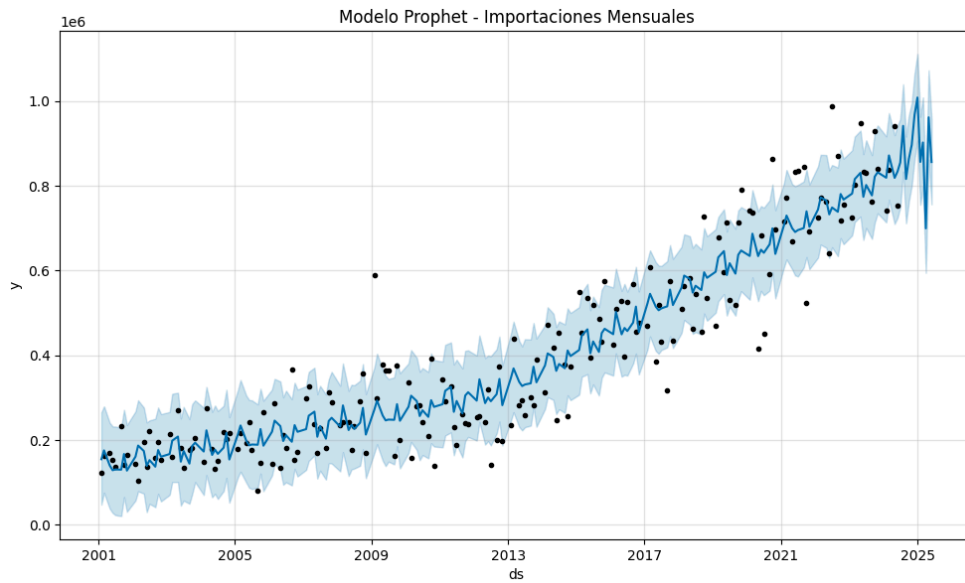
Generación de modelos

- Serie de consumo (Gasolina regular)

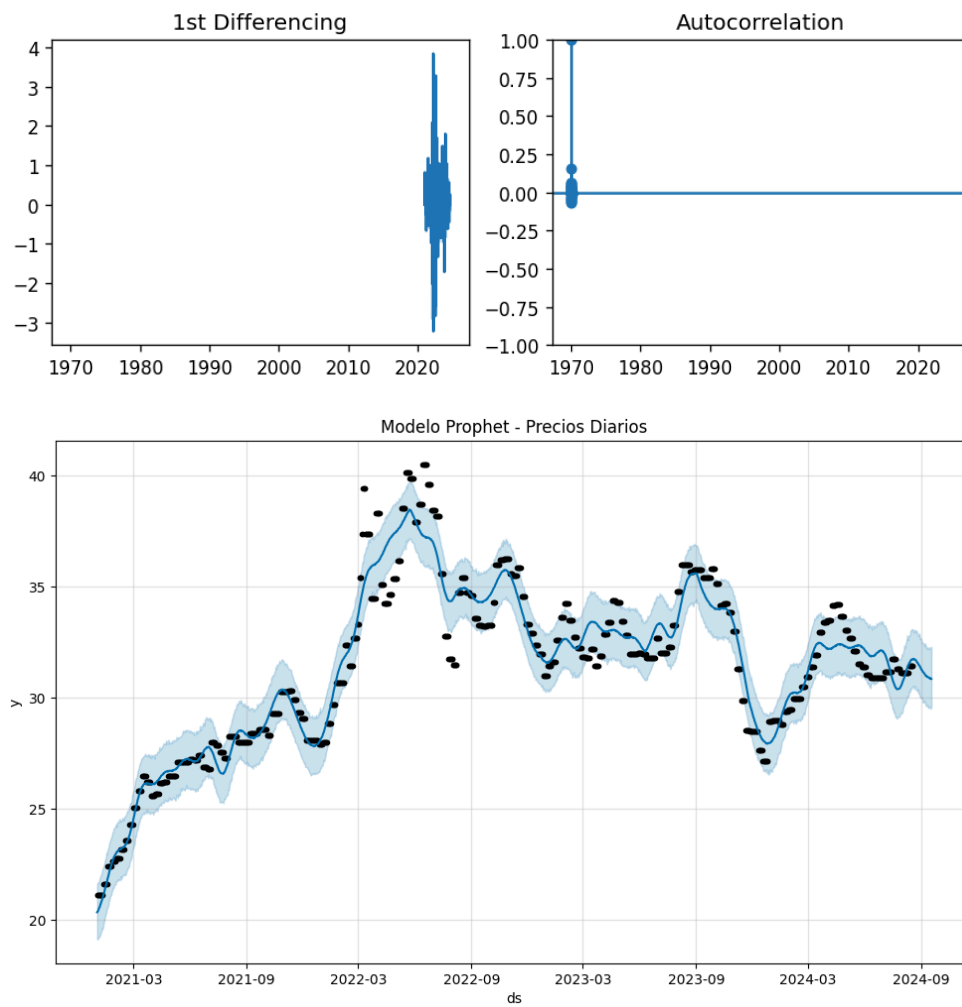


- Serie de importacion (Gasolina regular)





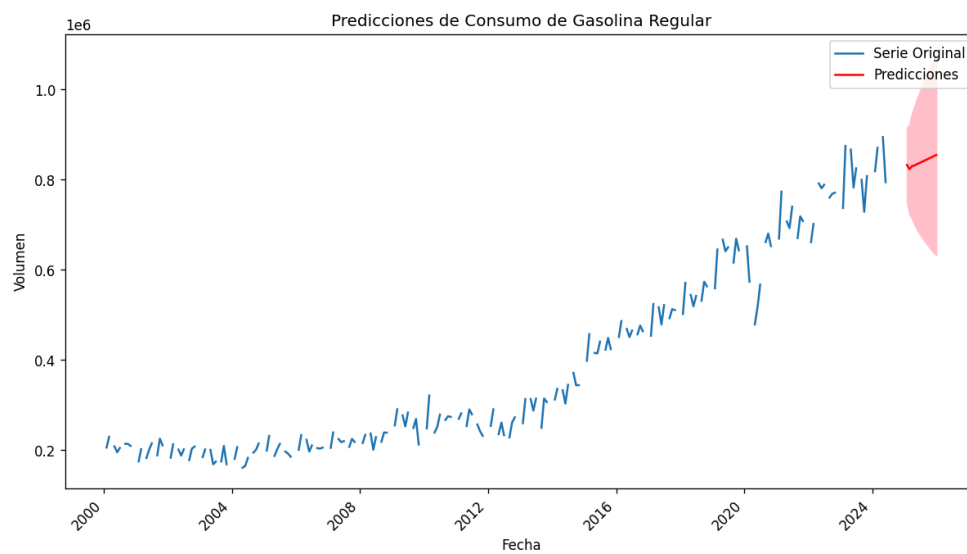
- Serie de precio (Gasolina regular)



Predicción con los modelos generados

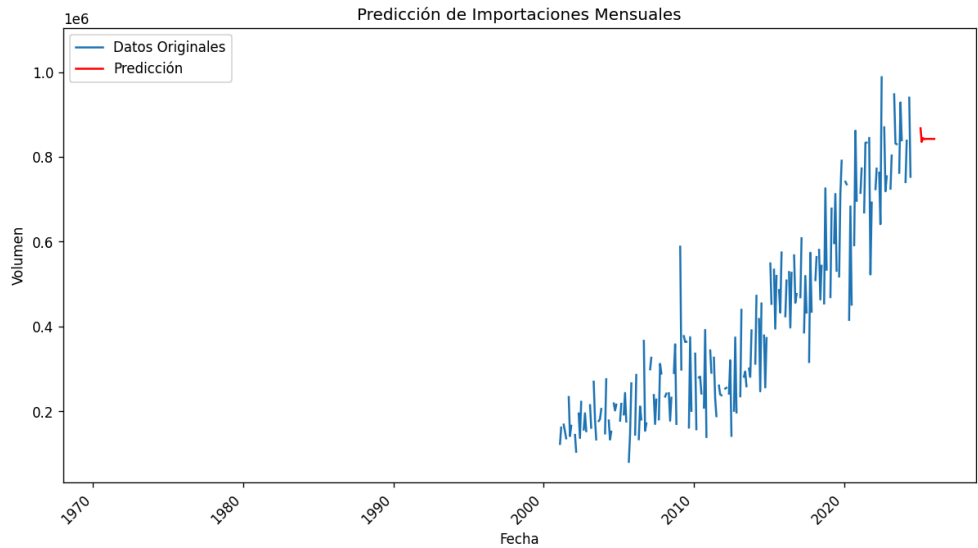
- Serie de consumo (Gasolina regular)

Gasolina regular	
Fecha	
2025-01-31	832853.935994
2025-02-28	823381.767871
2025-03-31	830479.618058
2025-04-30	831847.540316
2025-05-31	835196.877046
2025-06-30	837861.038786
2025-07-31	840762.134682
2025-08-31	843581.298524
2025-09-30	846428.794548
2025-10-31	849266.493277
2025-11-30	852107.579919
2025-12-31	854947.495018



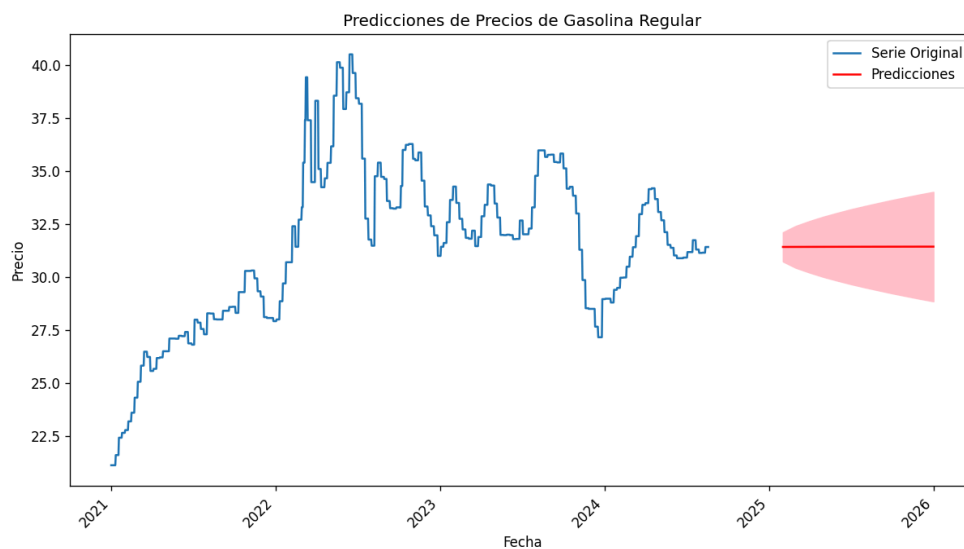
- Serie de importacion (Gasolina regular)

Gasolina regular	
Fecha	
2025-01-31	867631.969215
2025-02-28	835385.998035
2025-03-31	844467.105219
2025-04-30	841909.684721
2025-05-31	842629.905226
2025-06-30	842427.076799
2025-07-31	842484.197321
2025-08-31	842468.111045
2025-09-30	842472.641261
2025-10-31	842471.365462
2025-11-30	842471.724752
2025-12-31	842471.623569



- Serie de precio (Gasolina regular)

Gasolina regular	
Fecha	
2025-01-31	31.412279
2025-02-28	31.414333
2025-03-31	31.416182
2025-04-30	31.417847
2025-05-31	31.419347
2025-06-30	31.420698
2025-07-31	31.421915
2025-08-31	31.423011
2025-09-30	31.423998
2025-10-31	31.424887
2025-11-30	31.425688
2025-12-31	31.426409



Predicciones

Para este laboratorio pudimos utilizar ARIMA (AutoRegressive Integrated Moving Average) el cual es un modelo estadístico utilizado para analizar y predecir series temporales. Para entender mejor qué es ARIMA podemos ver que esta combina tres componentes principales:

1. **AR (AutoRegresivo):** El modelo usa las dependencias entre una observación y un número de observaciones anteriores (lags). Está representado por el parámetro p .
2. **I (Integrado):** Involucra la diferenciación de la serie de tiempo para hacerla estacionaria, es decir, eliminar tendencias o fluctuaciones que no son estables en el tiempo. El grado de diferenciación es representado por el parámetro d .

3. **MA (Media Móvil):** Captura la dependencia entre una observación y el residuo de un modelo de media móvil aplicado a observaciones anteriores. Es representado por el parámetro q .

En este caso se utilizó tres modelos distintos:

- Series de consumo con gasolina regular
- Series de importación con gasolina regular
- Series de precios con gasolina regular

Para estas tres series, se realizó el mismo procedimiento, el cual primero se buscó obtener el valor de la variable d (como se vio antes, es el orden de diferenciación), esto se realizó con ADF statics, que nos brinda el valor de p -value, con ello nos podemos guiar si es necesario diferenciar la serie y ver diagramas de correlación (se puede observar en las imágenes de arriba, como se observa la correlación) con ello podemos ver hasta qué punto la serie se vuelve estacionario, y con ello se puede obtener el valor de d . Luego se busca obtener el valor de p , esto por medio de Lags, (segunda gráfica de predicción), con ello se puede ver que tan significativo es el lag y ahí decidir el valor de p . Por último en esta parte, se busca obtener el valor de q , esto es posible de obtener por medio de la gráfica '1st Differencing' en donde gracias al área de confianza podemos obtener el valor necesario para q .

Luego es necesario imprimir e iniciar el modelo ARIMA, es ahí donde podemos observar nuestro resultado de si escogemos de manera adecuada nuestros valores de p, q, d esto lo hacemos observado el valor de 'ar.L1' y 'ma.L1', se espera y se busca que su valor sea menor que 0.05 en sus p -value. Luego se observan los residuos generados por el modelo (Gráfica Residuals y Density).

Por último se realiza la predicción que se desea, colocando cuanto tiempo es que se decide predecir, y que valor se desea predecir. Al final se genera una gráfica que contenga tanto la predicción como los valores anteriores (Gráfica de Predicciones)

Con los resultados obtenidos, es posible concluir que la gasolina como tal, tiene un valor muy volátil e inestable, esto debido a una combinación de factores económicos, geopolíticos y ambientales:

Oferta y Demanda Global: Los precios de la gasolina están fuertemente influenciados por la oferta y la demanda global de petróleo crudo. Cualquier fluctuación en la producción, ya sea por conflictos geopolíticos, decisiones de la OPEP, o cambios en la producción de países clave, puede afectar los precios.

Factores Geopolíticos: Conflictos en regiones productoras de petróleo, sanciones económicas, y decisiones políticas pueden reducir la oferta de petróleo en el mercado global, lo que generalmente lleva a aumentos en los precios de la gasolina.

Impacto de la Pandemia: La pandemia de COVID-19 tuvo un impacto significativo en la demanda de gasolina debido a las restricciones de movilidad, la caída en la actividad económica, y la reducción de los viajes internacionales. Esto llevó a una disminución abrupta en el consumo de gasolina y una acumulación de inventarios, causando una caída drástica en los precios. Sin embargo, con la recuperación económica y el levantamiento de restricciones, la demanda ha ido subiendo, lo que genera una nueva ola de inestabilidad en los precios.

En resumen, la inestabilidad en el valor, consumo e importación de la gasolina es el resultado de una interacción compleja entre factores económicos, geopolíticos y circunstancias excepcionales como la pandemia, que alteran tanto la oferta como la demanda a nivel global.

Repositorio

<https://github.com/Sebas021210/Lab2-DS>