

wUniversidad del Valle de Guatemala
Facultad de Ingeniería
Departamento de Ciencias de la Computación



Laboratorio No. 2

Series de Tiempo

Daniela Villamar 19086
Diego Crespo 19541

Los datos de series temporales son datos que se registran en intervalos regulares o períodos de tiempo. Cualquier valor no estacionario que dependa del tiempo puede formar parte de una serie temporal. Como analista de datos, se puede usar datos de series temporales para descubrir tendencias subyacentes o causas de ciertos patrones a lo largo del tiempo.

Cualquier conjunto de datos de series temporales puede incluir uno o más de estos principales componentes:

- **Nivel:** Índice de nivel de los datos de series de tiempo se refiere a la media de la serie.
- **Tendencia:** una tendencia se refiere a un movimiento ascendente o descendente constante y a largo plazo en una serie. A diferencia de la variación estacional, una tendencia es inesperada y no identificable de inmediato. Una tendencia en la que podemos encontrar la causa se llama determinista, mientras que una tendencia que es inexplicable se llama estocástica.
- **Ciclo:** Un ciclo es un movimiento hacia arriba y hacia abajo que ocurre alrededor de una tendencia. A diferencia de la variación estacional, un ciclo no tiene un tiempo preciso e igual entre períodos de tiempo y, por lo tanto, no es predecible.
- **Estacionalidad:** A diferencia de una tendencia, la estacionalidad se refiere a las variaciones que ocurren con una frecuencia predecible y fija. Por ejemplo, las ventas de chumpas aumentan en el invierno porque la gente tiene frío y necesita ropa adecuada para el clima.
- **Irregularidad:** También conocida como ruido, la irregularidad es lo que queda cuando se elimina la estacionalidad o las tendencias del conjunto de datos. Las irregularidades son aleatorias e impredecibles. Un buen ejemplo de variaciones irregulares son el cambio en el precio de acciones.

Cuando los valores se recopilan en intervalos de tiempo iguales y consistentes, la serie de tiempo se denomina regular. Pero cuando las mediciones se recopilan en intervalos impredecibles e irregulares, la serie temporal se denomina irregular.

La estacionalidad, como sugiere su nombre, se refiere a las características estacionales de los datos de la serie temporal. Es el patrón predecible que se repite con cierta frecuencia dentro de un año, como semanal, mensual, trimestral, etc. Antes de incluir la estacionalidad en los modelos, necesitamos saber cómo se repiten los datos y con qué frecuencia. Detectar la estacionalidad puede ser sencillo si comprendemos muy bien el contexto de los datos. Por ejemplo, sabemos que la gasolina será más alta en las vacaciones y más baja en el invierno en un año. Para descubrir la estacionalidad de los datos, y si no estamos familiarizados con el contexto, la forma simple es graficar los datos y observar las señales periódicas junto con la serie temporal.

Después de detectar la estacionalidad, hay varias formas de incorporar la estacionalidad en el modelo para realizar mejor la previsión de series temporales. Así como hay muchos tipos, componentes y modelos en el análisis de series de tiempo, también hay una variedad de métodos/herramientas para estudiar datos:

- Modelos ARIMA de Box-Jenkins: estos modelos se utilizan para comprender una sola variable dependiente del tiempo, como la temperatura a lo largo del tiempo, y predecir futuros puntos de datos de variables. Estos modelos asumen que los datos son estacionarios. Los analistas tienen que eliminar tantas diferencias y estacionalidad en puntos de datos anteriores como puedan. Afortunadamente, el modelo ARIMA incluye términos para tener en cuenta operadores de diferencia estacional, promedios móviles y términos autorregresivos dentro del modelo.
 - El promedio móvil integrado autorregresivo o ARIMA es una parte muy importante de las estadísticas, la econometría y, en particular, el análisis de series temporales.
 - ARIMA es una técnica de pronóstico que nos da valores futuros completamente basados en su inercia.
 - Los modelos de media móvil integrada autorregresiva (ARIMA) incluyen un modelo estadístico claro para el componente asimétrico de una serie de tiempo que permite autocorrelaciones distintas de cero en el componente irregular.
 - Los modelos ARIMA se definen para series temporales estacionarias. Por lo tanto, si comienza con una serie de tiempo no estacionaria, primero se debe "diferenciar" la serie

de tiempo hasta que obtenga una serie de tiempo estacionaria.

- Modelos multivariados de Box-Jenkins: estos modelos se utilizan para analizar más de una variable dependiente del tiempo, como la temperatura y la humedad, a lo largo del tiempo.
- Método Holt-Winters: El método Holt-Winters es una técnica de suavizado exponencial. Está diseñado para predecir los resultados de los puntos de datos que incluyen la estacionalidad.

Prophet es un procedimiento para pronosticar datos de series temporales basado en un modelo aditivo en el que las tendencias no lineales se ajustan a la estacionalidad anual, semanal y diaria, además de los efectos de las vacaciones. Funciona mejor con series temporales que tienen fuertes efectos estacionales y varias temporadas de datos históricos. Prophet es resistente a los datos faltantes y los cambios en la tendencia, y por lo general maneja bien los valores atípicos. Prophet es un software de código abierto lanzado por el equipo central de ciencia de datos de Facebook.

Prophet es especialmente útil para conjuntos de datos que:

- ★ Contener un período de tiempo extendido (meses o años) de observaciones históricas detalladas (por hora, día o semana).
- ★ Tener múltiples estacionalidades fuertes.
- ★ Incluir eventos importantes, pero irregulares, previamente conocidos.
- ★ Tienen puntos de datos faltantes o grandes valores atípicos.
- ★ Tienen tendencias de crecimiento no lineales que se acercan a un límite.

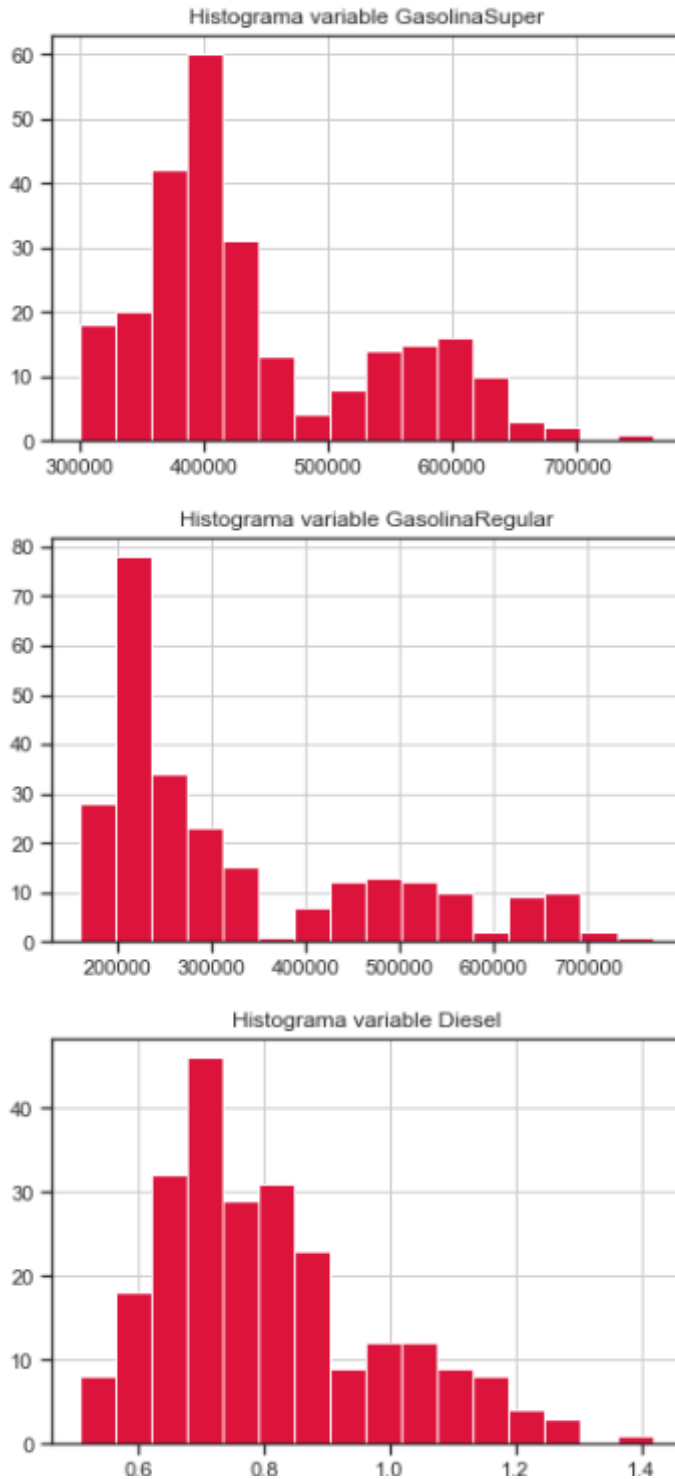
Prophet es un modelo de regresión aditivo con una tendencia de curva de crecimiento lineal o logística por partes. Incluye un componente estacional anual modelado utilizando series de Fourier y un componente estacional semanal modelado utilizando variables ficticias o más conocidas como dummy.

Las organizaciones utilizan el análisis de series temporales para comprender los patrones sistémicos y las causas subyacentes de las tendencias a lo largo del tiempo. Usando visualizaciones de datos, los usuarios comerciales pueden profundizar en por qué ocurren estas

tendencias. Al analizar los datos en intervalos constantes, las organizaciones pueden predecir la probabilidad de eventos futuros. Puede mostrar cambios como el comportamiento cíclico o la estacionalidad, lo que brinda una mejor comprensión de las variables de datos y ayuda a pronosticar mejor. El análisis de series temporales es uno de los aspectos más importantes del análisis de datos para cualquier organización grande, ya que ayuda a comprender la estacionalidad, las tendencias, los ciclos y la aleatoriedad en las ventas y la distribución y otros atributos. Estos factores ayudan a las empresas a tomar una decisión bien informada que es muy importante para los negocios. En este caso para la compra e importación de la gasolina es interesante ver de que manera los datos han ido aumentando como decreciendo. Vemos la diferencia entre la gasolina para los carros por ejemplo hasta de los aviones. Como todos lo saben, la pandemia influyo a que nosotros estuviéramos encerrados en casa y por ende los sistemas de transporte fueran parados por un tiempo y de igual manera aunque al pasar los primeros meses de pandemia eran pocos los transportes que podían funcionar. Gracias al análisis estadístico y trabajo de series de tiempo podemos notar que este si fue de gran impacto en Guatemala y no se diga global.

ANÁLISIS

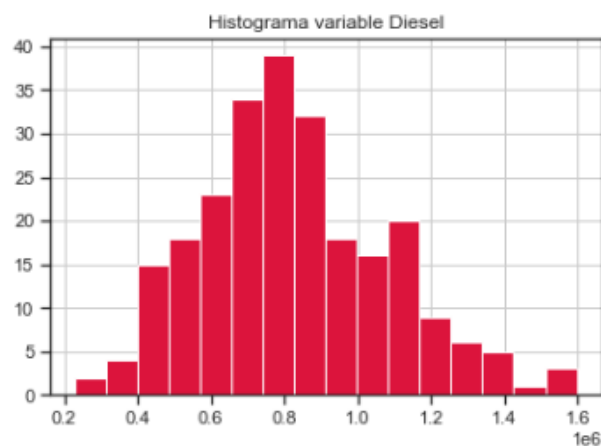
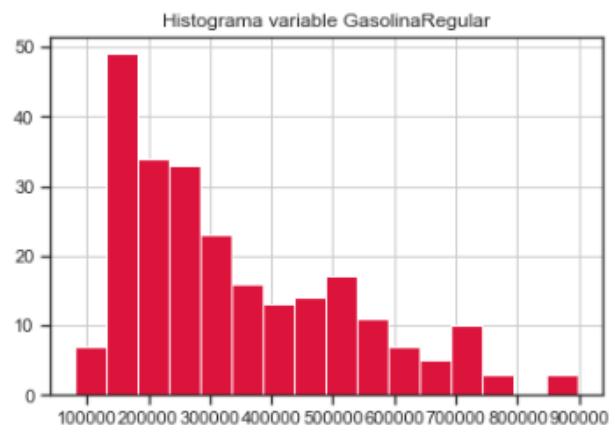
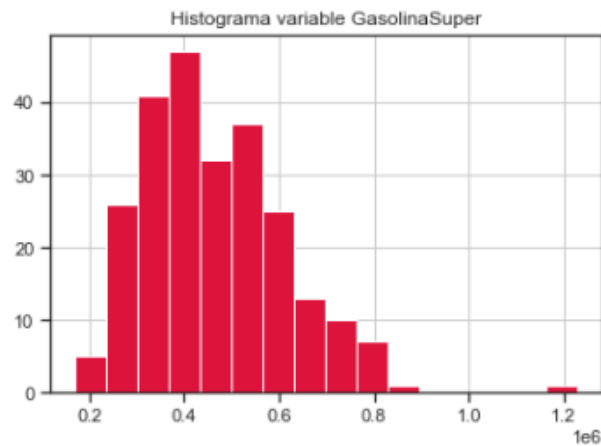
Histogramas para Consumo



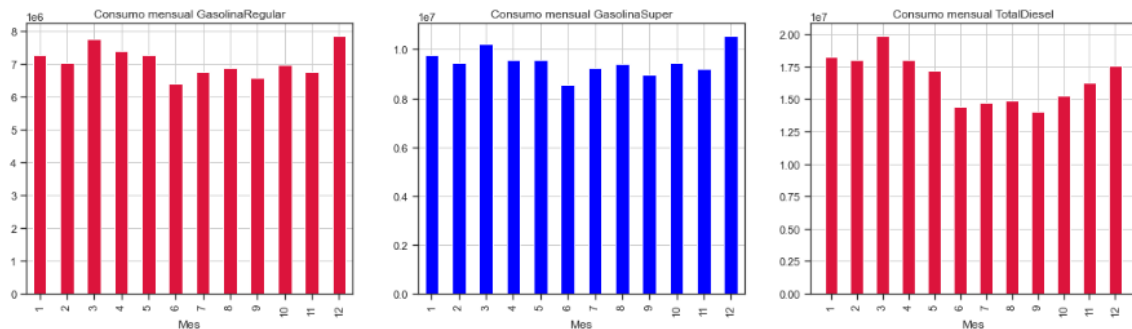
En los histogramas anteriores, es importante mencionar que ninguno parece tener una distribución normal. Para la **gasolina super**, puede observarse un sesgo hacia la

izquierda de los datos y una especie de aumento cerca del final. La **gasolina regular** tiene un pico pronunciado del lado izquierdo, así como una serie de datos atípicos más adelante, por lo que tampoco es normal. Para el **diesel**, se observa un comportamiento que se podría aproximar a normal, sin embargo, sería necesario manipular los datos.

Histogramas para Importaciones

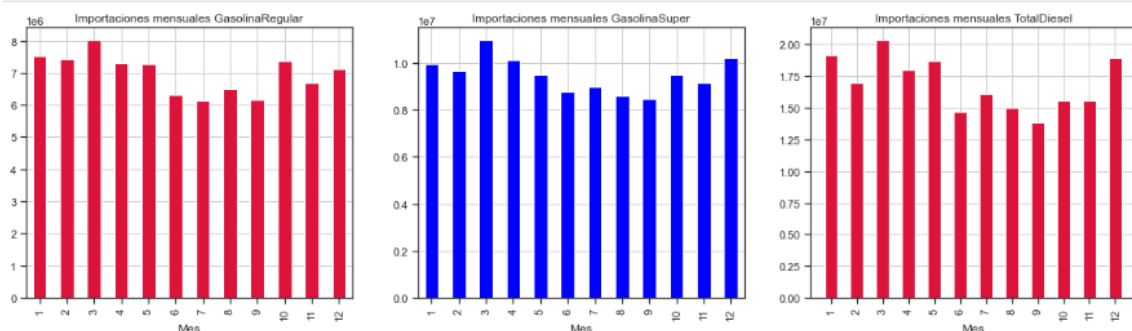


El comportamiento de estos datos es significativamente distinto al anterior. En primera mano, es posible observar que no existe una gran cantidad de datos atípicos, sino más bien se ve un comportamiento aproximadamente normal para los datos.



COMPARAMOS EL CONSUMO MENSUAL DE LOS DISTINTOS TIPOS DE GASOLINA

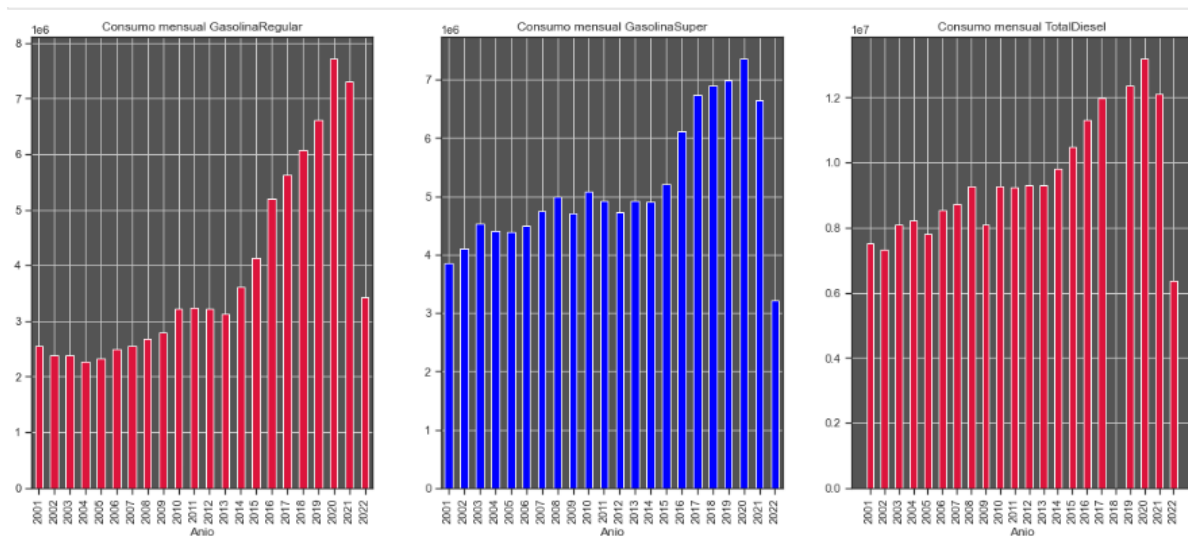
```
[25]: fig = plt.figure(figsize=(20,5))
ax = fig.add_subplot(131)
DATAIMPORTACION.groupby('Mes').GasolinaRegular.sum().plot.bar(color = "crimson")
ax.set_title("Importaciones mensuales GasolinaRegular")
plt.grid()
ax = fig.add_subplot(132)
DATAIMPORTACION.groupby('Mes').GasolinaSuper.sum().plot.bar(color = "blue")
ax.set_title("Importaciones mensuales GasolinaSuper")
plt.grid()
ax = fig.add_subplot(133)
DATAIMPORTACION.groupby('Mes').Diesel.sum().plot.bar(color = "crimson")
ax.set_title("Importaciones mensuales TotalDiesel")
plt.grid()
plt.axis('tight')
plt.show()
```



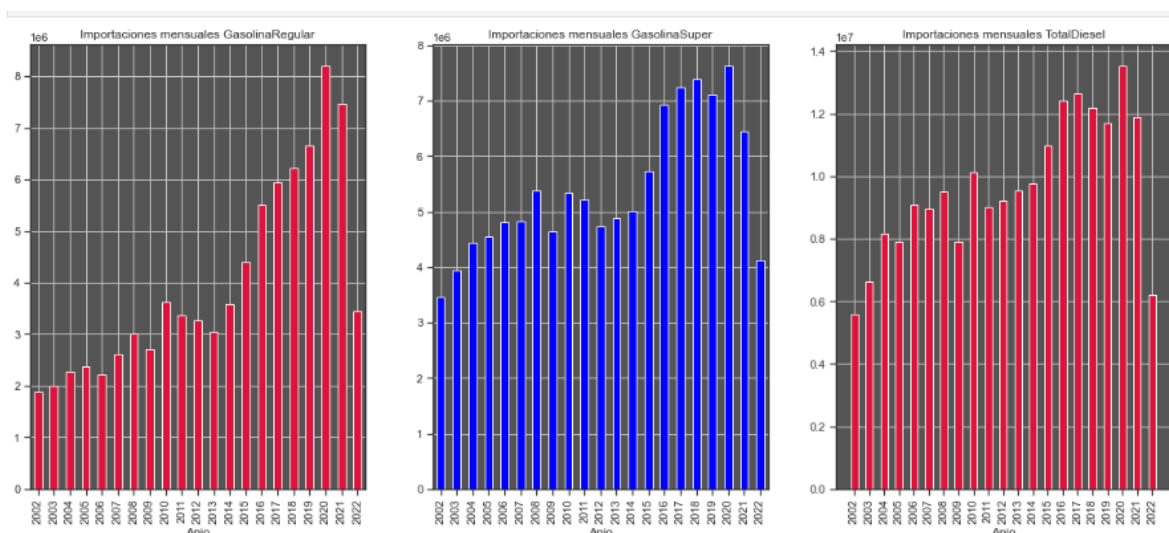
De forma inicial, para el **consumo mensual de gasolina regular** puede observarse que existe un consumo relativamente estable a lo largo de los meses. Puede verse un ligero aumento al final del primer trimestre del año, para luego reducirse significativamente en el primer semestre del año. Luego, vuelve a crecer, aunque no hay estabilidad en ello.

Para el **consumo mensual de gasolina super** inicialmente se observa un cambio en el factor del eje y. Si bien este se encuentra una decena por encima del anterior, al analizar los números, vemos que en realidad la cantidad está cercana, pues pocos meses alcanza el 1.

Ahora bien, el **consumo mensual de gasolina diesel** es significativamente más fluctuante que los dos anteriores, así como es más alto. Es lógico pensar que tenga una mayor popularidad, dado el transporte pesado, los servicios de transporte y la fácil adquisición económica. En la gráfica es posible observar una caída por el mes 7, que luego empieza a crecer, sin llegar a los valores del inicio de año.



El comportamiento en la gasolina de consumo tiene un comportamiento bastante definido, y es que ha aumentado de forma casi estable a lo largo de todos los años. Se observa además un pico pronunciado en el consumo del año, que sobresale ante los años anteriores. Sin embargo, vemos que el consumo para el año 2020 es significativamente menor en todos los casos, asemejándose sobremanera al comportamiento que se tuvo en 2012, 8 años antes. Este fenómeno sigue a 2022. Dicha fluctuación puede explicarse para el 2020 por la pandemia de la COVID19, así como la clausura masiva que nuestra nación tuvo como medida de prevención. Ahora bien, ese comportamiento se exporta hacia 2022, sí, sin embargo también cabe mencionar que los registros para el año actual no están completos, puesto que el año no ha terminado.

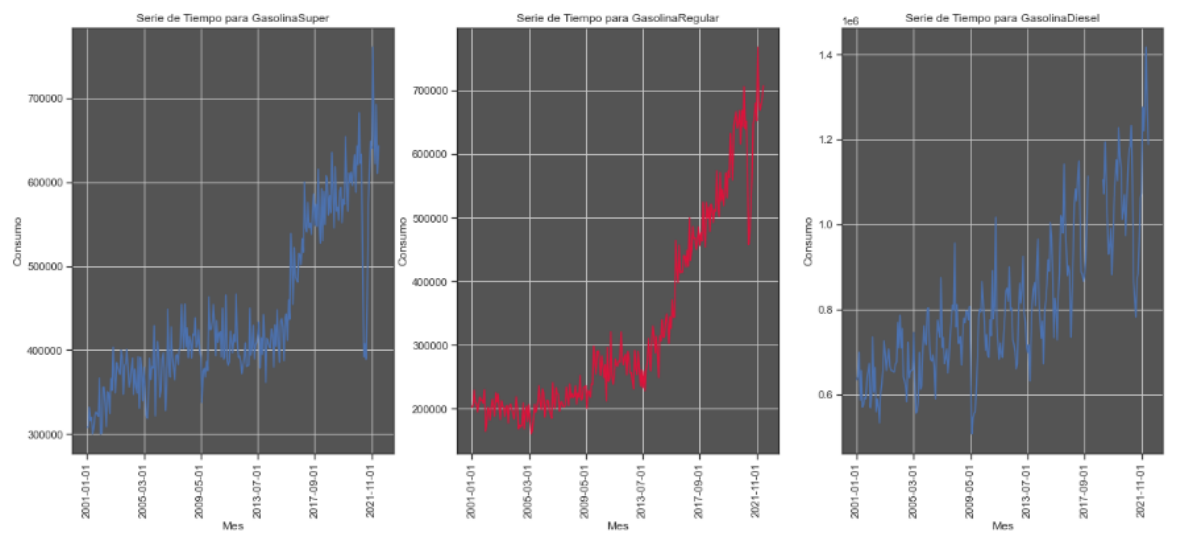


La importación anual de gasolina regular tiene una importación baja en los primeros años y luego podemos observar como esta va aumentando, formando una gráfica con un sesgo moderado hacia la izquierda.

En la grafica de importación anual de gasolina super podemos observar que inicia con una importación más alta que la gasolina regular, el cual va aumentando moderadamente formando un pequeño sesgo hacia la izquierda.

En la importación anual de Diesel podemos observar que va en aumento e incluso en cuatro años se mantuvo casi constante y luego incremento significativamente.

En estas tres gráficas podemos observar como el consumo disminuye en 2020 que es el año en el que todos nos vimos afectados por la pandemia y en 2021 es mucho mas bajo porque es el año en curso.



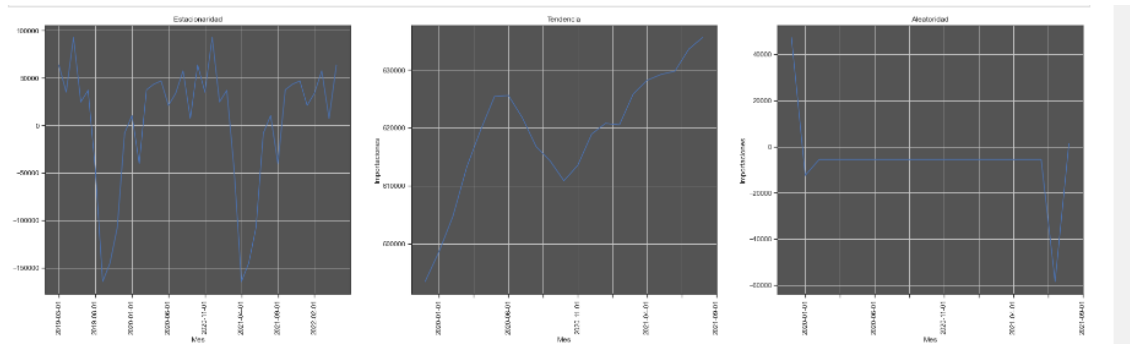
El consumo de gasolina presenta un consumo creciente a lo largo del tiempo, independientemente de la clase de gasolina que se esté analizando. Como es posible observar, la GasolinaSuper presenta un comportamiento ligeramente fluctuante en los primeros años, mostrando un aumento significativo en 2004. Luego, el crecimiento es estable, hasta 2014, en el que tiene un salto alto. Además, se observa una caída significativa alrededor de medio año del 2020, lo que es lógico al considerar los efectos de la pandemia y cómo alteró los hábitos de consumo.

Por su parte, el consumo de la GasolinaRegular muestra un crecimiento estable hasta 2014, punto en el cual da un salto hacia arriba y su consumo crece significativamente, en comparación a los datos históricos previos. Es posible observar una caída en el consumo cerca de agosto de 2020, aunque esta no es tan pronunciada como la anterior.

Por su parte, el consumo de Diesel es el más errático de todos. Si bien presenta una cota por arriba y por abajo, los saltos entre puntos son marcados y sumamente aleatorios. De forma similar a los casos anteriores, es posible observar una caída en el consumo cerca de 2020, explicable por la pandemia.

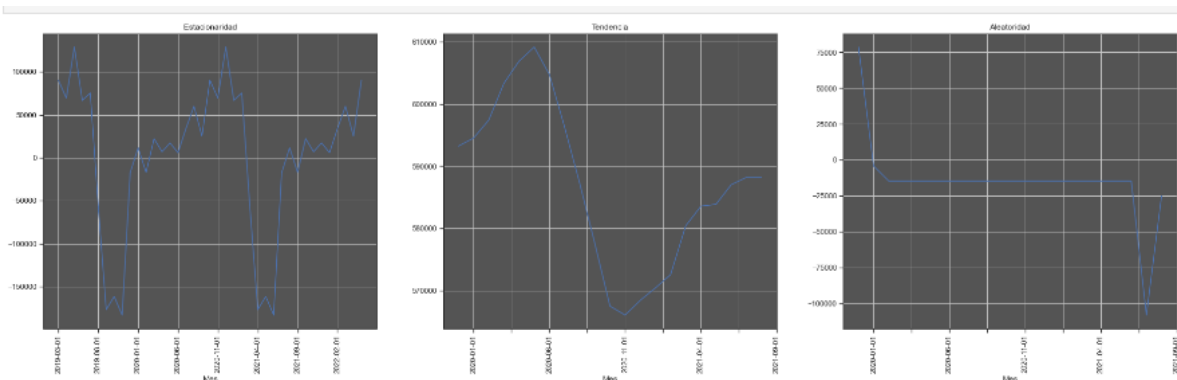
Descomposición para las series de tiempo

Consumo: Gasolina Regular



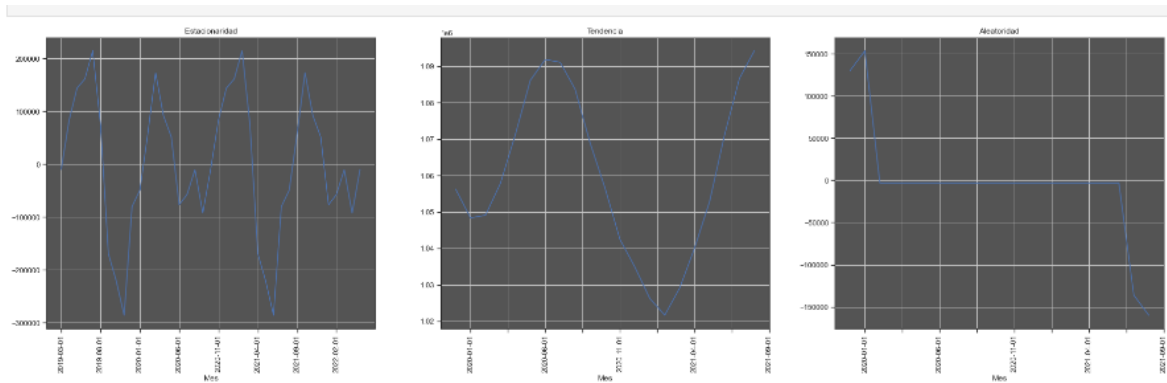
Al observar las gráficas, puede verse que presentan claramente un periodo en el que el consumo mantiene un mismo comportamiento, y este se replica a lo largo de los años. Por otra parte, es importante mencionar el crecimiento en tendencia a lo largo de los años, con pequeñas fluctuaciones entre 2002 y 2004, 2004 y 2007 y a finales de 2011. Luego, a partir de 2013, puede verse que la tendencia crece indiscretamente. Además, el comportamiento de aleatoriedad parece ser cíclico.

Consumo: Gasolina Super



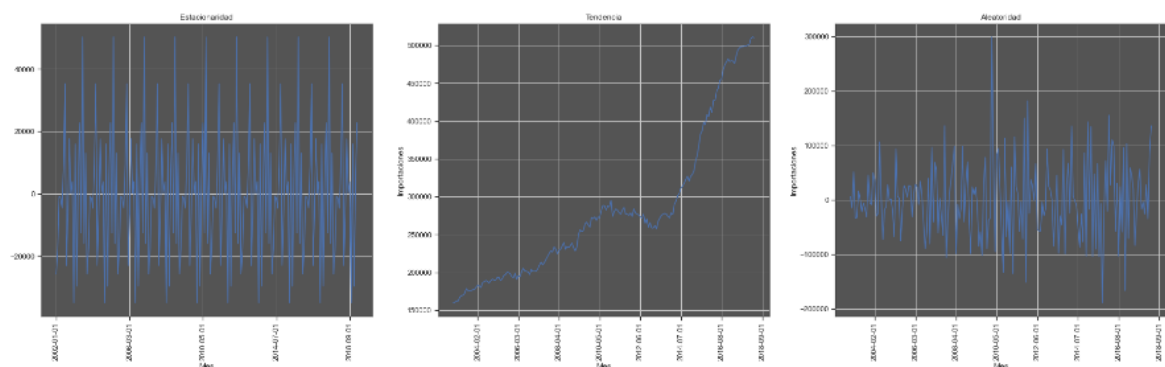
Al observar las gráficas, el comportamiento de la Gasolina Regular es tremendamente similar al de la Gasolina Super. Puede verse que presentan claramente un periodo en el que el consumo mantiene un mismo comportamiento, y este se replica a lo largo de los años. Por otra parte, es importante mencionar el crecimiento en tendencia a lo largo de los años, con pequeñas fluctuaciones entre 2002 y 2004, 2004 y 2007 y a finales de 2011. Luego, a partir de 2013, puede verse que la tendencia crece indiscretamente. Además, el comportamiento de aleatoriedad parece ser cíclico.

Consumo: Gasolina Diesel



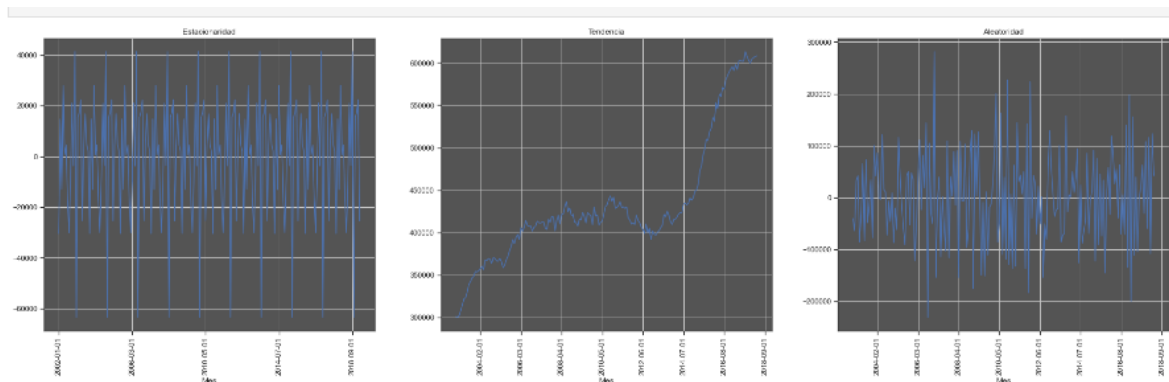
A diferencia de su contraparte, el consumo de diesel es un tanto más errático. Puede verse que es estacionario en cuanto a temporalidad, así como se observa un crecimiento a lo largo de la tendencia. Ahora bien, su elemento aleatorio parece ir en aumento hasta 2010, para luego empezar a disminuir.

Importación: Gasolina Regular



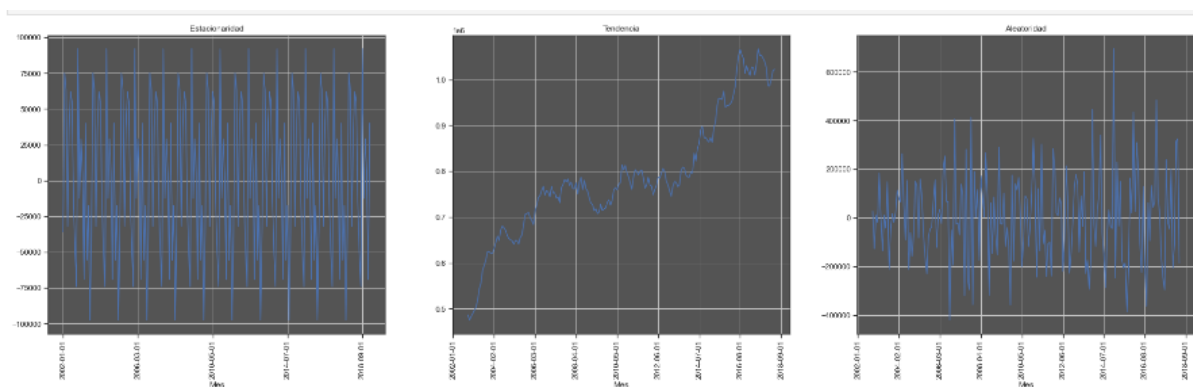
Al observar las gráficas, puede verse que presentan claramente un periodo en el que el consumo mantiene un mismo comportamiento, y este se replica a lo largo de los años. Por otra parte, es importante mencionar el crecimiento en tendencia a lo largo de los años, con pequeñas fluctuaciones en intervalos particulares. Luego, a partir de 2013, puede verse que la tendencia crece indiscretamente. Además, el comportamiento de aleatoriedad parece ser cíclico, con perturbaciones eminentes en el medio.

Importación: Gasolina Super

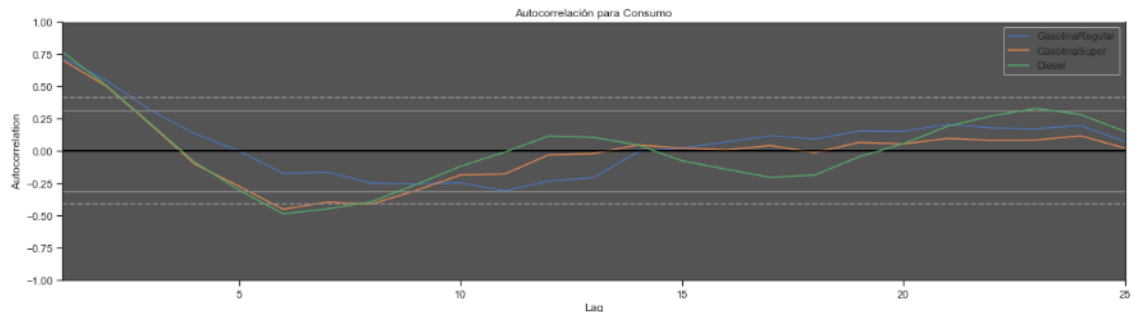


El comportamiento observado para la importación de gasolina súper es similar a su comportamiento en consumo. Al observar las gráficas, puede verse que presentan claramente un periodo en el que el consumo mantiene un mismo comportamiento, y este se replica a lo largo de los años. Por otra parte, es importante mencionar el crecimiento en tendencia a lo largo de los años, con pequeñas fluctuaciones en intervalos particulares. Luego, a partir de 2013, puede verse que la tendencia crece indiscretamente. Además, el comportamiento de aleatoriedad parece ser cíclico, con perturbaciones eminentes en el medio.

Importación: Gasolina Diesel



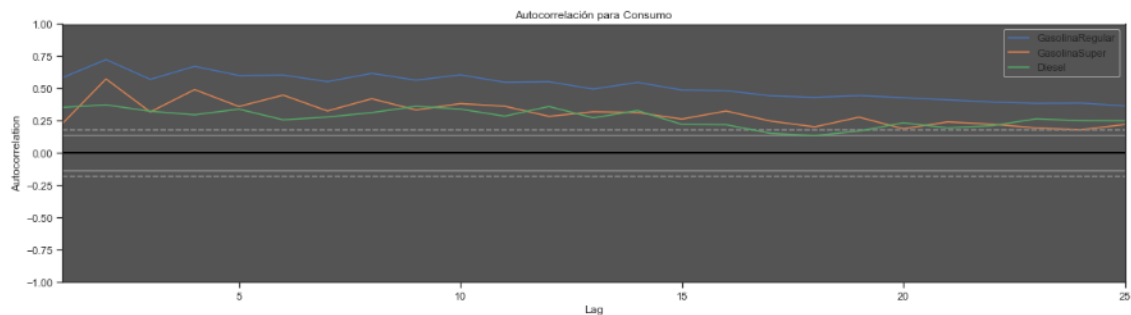
Graficas de Autocorrelación



IMPORTACION

```
n [60]: fig = plt.figure(figsize=(20,5))
pd.plotting.autocorrelation_plot(DATAIMPORTACION["GasolinaRegular"], label = "GasolinaRegular").set_xlim([1, 25])
pd.plotting.autocorrelation_plot(DATAIMPORTACION["GasolinaSuper"], label = "GasolinaSuper").set_xlim([1, 25])
pd.plotting.autocorrelation_plot(DATAIMPORTACION["Diesel"], label = "Diesel").set_xlim([1, 25])
plt.grid()
plt.title("Autocorrelación para Consumo")
plt.xlabel("Lag")
```

ut[60]: Text(0.5, 0, 'Lag')



Al observar las gráficas, es posible inferir que existe una mayor correlación entre los datos para el consumo, que el de importación. Sin embargo, es importante mencionar que en ambos casos, la GasolinaRegular presenta un comportamiento mayor y separado de los otros dos elementos.

Pruebas de Dickey - Fuller

En las pruebas estadísticas suponemos lo siguiente:

- H_0 : tiene una raíz unitaria. No es estacionaria en μ y tiene dependencia del tiempo.
- H_1 : no tiene una raíz unitaria. Es estacionaria en μ y no tiene una dependencia del tiempo.

Con el siguiente criterio:

- Si el valor - $p > 0.05$ no se rechaza la hipótesis nula.
- Si el valor - $p \leq 0.05$ se rechaza la hipótesis nula.

Consumo: Gasolina Regular

```
[62]: Xregularcon = DATACONSUMO["GasolinaRegular"].values
      ResultRegularcon = adfuller(Xregularcon)

      print('Estadístico: %f' % ResultRegularcon[0])
      print('Valor - p: %f' % ResultRegularcon[1])
      print('Valores críticos:')
      for key, value in ResultRegularcon[4].items():
          print('\t%s: %.3f' % (key, value))

Estadístico: -2.111769
Valor - p: 0.239809
Valores críticos:
    1%: -3.616
    5%: -2.941
   10%: -2.609
```

Consumo: Gasolina Super

```
[64]: Xsupercon = DATACONSUMO["GasolinaSuper"].values
      ResultSupercon = adfuller(Xsupercon)

      print('Estadístico: %f' % ResultSupercon[0])
      print('Valor - p: %f' % ResultSupercon[1])
      print('Valores críticos:')
      for key, value in ResultSupercon[4].items():
          print('\t%s: %.3f' % (key, value))

Estadístico: -2.610500
Valor - p: 0.090846
Valores críticos:
    1%: -3.689
    5%: -2.972
   10%: -2.625
```

Consumo: Gasolina Diesel

```
[66]: Xdieselcon = DATACONSUMO["Diesel"].values
      ResultDieselcon = adfuller(Xdieselcon)

      print('Estadístico: %f' % ResultDieselcon[0])
      print('Valor - p: %f' % ResultDieselcon[1])
      print('Valores críticos:')
      for key, value in ResultDieselcon[4].items():
          print('\t%s: %.3f' % (key, value))

Estadístico: -3.617503
Valor - p: 0.005435
Valores críticos:
    1%: -3.633
    5%: -2.949
   10%: -2.613
```

Para el consumo de gasolina Regular y Super Dado que el valor de $p > 0.05$, entonces no es posible rechazar la hipótesis nula, por lo que se tiene una raíz unitaria y los datos no es estacionaria en μ y tiene dependencia del tiempo. A comparación de la gasolina disel $p < 0.05$ por lo que se rechaza la hipótesis nula.

Importaciones: Gasolina Regular

```

:
Xregularimp = DATAIMPORTACION["GasolinaRegular"].va
ResultRegularimp = adfuller(Xregularimp)

print('Estadístico: %f' % ResultRegularimp[0])
print('Valor - p: %f' % ResultRegularimp[1])
print('Valores críticos:')
for key, value in ResultRegularimp[4].items():
    print('\t%s: %.3f' % (key, value))

Estadístico: 0.883629
Valor - p: 0.992873
Valores críticos:
    1%: -3.465
    5%: -2.877
    10%: -2.575

```

Importaciones: Gasolina Super

```

:
Xsuperimp = DATAIMPORTACION["GasolinaSuper"].values
ResultSuperimp = adfuller(Xsuperimp)

print('Estadístico: %f' % ResultSuperimp[0])
print('Valor - p: %f' % ResultSuperimp[1])
print('Valores críticos:')
for key, value in ResultSuperimp[4].items():
    print('\t%s: %.3f' % (key, value))

Estadístico: -1.487248
Valor - p: 0.539810
Valores críticos:
    1%: -3.463
    5%: -2.876
    10%: -2.575

```

Importaciones: Gasolina Diesel

```

:
Xdieselimp = DATAIMPORTACION["Diesel"].values
ResultDieselimp = adfuller(Xdieselimp)

print('Estadístico: %f' % ResultDieselimp[0])
print('Valor - p: %f' % ResultDieselimp[1])
print('Valores críticos:')
for key, value in ResultDieselimp[4].items():
    print('\t%s: %.3f' % (key, value))

Estadístico: -1.554660
Valor - p: 0.506320
Valores críticos:
    1%: -3.465
    5%: -2.877
    10%: -2.575

```

Para todas las importaciones de gasolina, Dado que el valor de $p > 0.05$, entonces no es posible rechazar la hipótesis nula, por lo que se tiene una raíz unitaria y los datos no es estacionaria en μ y tiene dependencia del tiempo.

Aplicamos una Diferencia para ver que ocurre - Primera diferencia.

Consumo

```
Estadístico: -6.917874
Valor - p: 0.000000
Valores críticos:
1%: -3.621
5%: -2.944
10%: -2.610
```

```
C:\Users\diego\AppData\Local\Temp\ipykernel_15480\3:
A value is trying to be set on a copy of a slice fr
Try using .loc[row_indexer,col_indexer] = value ins
```

```
See the caveats in the documentation: https://panda
DATACONSUMO["GasolinaRegular"] = np.log(DATACONSUMO
```

Consumo: Gasolina Super

```
5]: DATACONSUMO["GasolinaSuper"] = np.log(DATACONSUMO["GasolinaSuper"])
diffsupercon = DATACONSUMO["GasolinaSuper"].diff()

ResultSupercon = adfuller(diffsupercon)

print('Estadístico: %f' % ResultSupercon[0])
print('Valor - p: %f' % ResultSupercon[1])
print('Valores críticos:')
for key, value in ResultSupercon[4].items():
    print('\t%s: %.3f' % (key, value))
```

```
Estadístico: -2.780347
Valor - p: 0.061136
Valores críticos:
1%: -3.670
5%: -2.964
10%: -2.621
```

```
C:\Users\diego\AppData\Local\Temp\ipykernel_15480\2:
A value is trying to be set on a copy of a slice fr
Try using .loc[row_indexer,col_indexer] = value ins
```

```
See the caveats in the documentation: https://panda
DATACONSUMO["GasolinaSuper"] = np.log(DATACONSUMO
```

Consumo: Gasolina Diesel

```
5]: DATACONSUMO["Diesel"] = np.log(DATACONSUMO["Diesel"])
diffdieselcon = DATACONSUMO["Diesel"].diff().dropna()

ResultDieselcon = adfuller(diffdieselcon)

print('Estadístico: %f' % ResultDieselcon[0])
print('Valor - p: %f' % ResultDieselcon[1])
print('Valores críticos:')
for key, value in ResultDieselcon[4].items():
    print('\t%s: %.3f' % (key, value))
```

```
Estadístico: -5.232013
Valor - p: 0.000008
Valores críticos:
1%: -3.621
5%: -2.944
10%: -2.610
```

Importancia

```
Estadístico: -6.768123
Valor - p: 0.000000
Valores críticos:
    1%: -3.465
    5%: -2.877
    10%: -2.575
```

```
C:\Users\diego\AppData\Local\Temp\ipykernel_15480
A value is trying to be set on a copy of a slice
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/10min/05min.html#copy-on-write
DATAIMPORTACION["GasolinaRegular"] = np.log(DAT
```

Importaciones: Gasolina Super

```
]
DATAIMPORTACION["GasolinaSuper"] = np.log(DATAIMPORTACION["GasolinaSuper"].diff())
diffsuperimp = DATAIMPORTACION["GasolinaSuper"].diff()

ResultSuperimp = adfuller(diffsuperimp)

print('Estadístico: %f' % ResultSuperimp[0])
print('Valor - p: %f' % ResultSuperimp[1])
print('Valores críticos:')
for key, value in ResultSuperimp[4].items():
    print('\t%s: %.3f' % (key, value))
```

```
Estadístico: -7.795360
Valor - p: 0.000000
Valores críticos:
    1%: -3.464
    5%: -2.876
    10%: -2.575
```

```
C:\Users\diego\AppData\Local\Temp\ipykernel_15480
A value is trying to be set on a copy of a slice
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/10min/05min.html#copy-on-write
DATAIMPORTACION["GasolinaSuper"] = np.log(DATAIMPORTACION["GasolinaSuper"].diff())
```

Importaciones: Gasolina Diesel

```
]
DATAIMPORTACION["Diesel"] = np.log(DATAIMPORTACION["Diesel"].diff())
diffdieselimp = DATAIMPORTACION["Diesel"].diff()

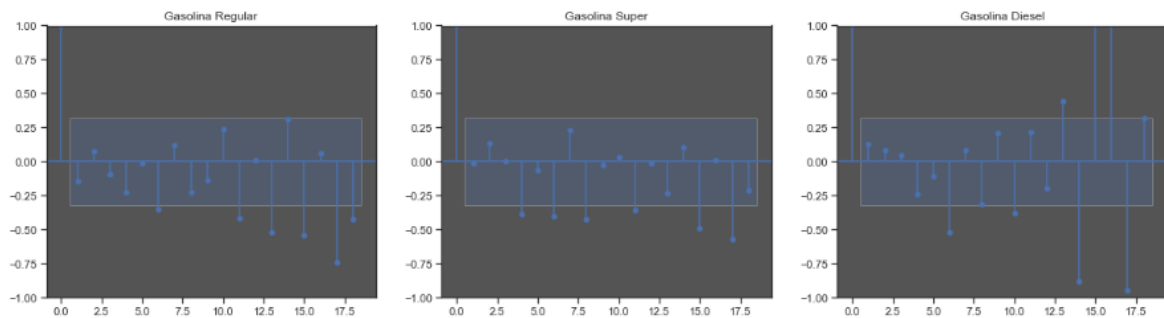
ResultDieselimp = adfuller(diffdieselimp)

print('Estadístico: %f' % ResultDieselimp[0])
print('Valor - p: %f' % ResultDieselimp[1])
print('Valores críticos:')
for key, value in ResultDieselimp[4].items():
    print('\t%s: %.3f' % (key, value))
```

```
Estadístico: -8.166520
Valor - p: 0.000000
Valores críticos:
    1%: -3.465
    5%: -2.877
    10%: -2.575
```

Dado que el valor de $p \leq 0.05$, entonces es posible rechazar la hipótesis nula, por lo que no se tiene una raíz unitaria y los datos presentan ser estacionaria en μ y no tiene dependencia del tiempo.

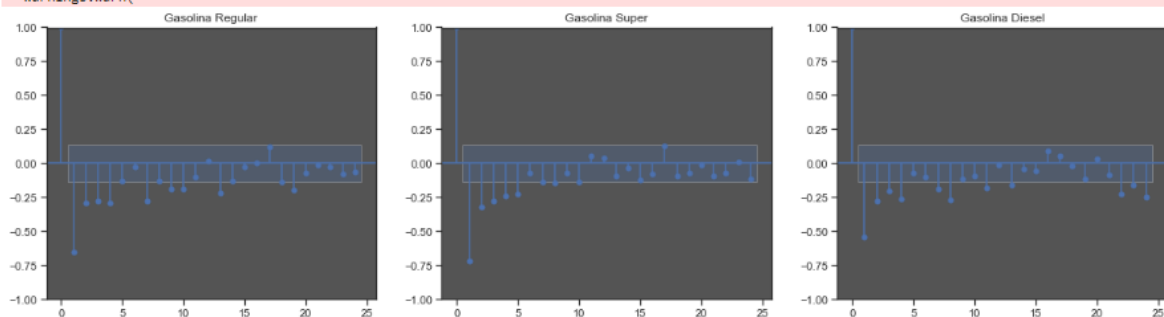
Para encontrar p necesitamos ver los gráficos de Autocorrelación Parcial



Importaciones: Autocorrelación Parcial

```
]: fig, axes = plt.subplots(1, 3, figsize=(20, 5))
plot_pacf(diffregularimp, title = "Gasolina Regular",ax=axes[0]) #p: 2
plot_pacf(diffsuperimp, title = "Gasolina Super",ax=axes[1]) #p: 2
plot_pacf(diffdieselimp, title = "Gasolina Diesel",ax=axes[2]) #p: 2
plt.show()
```

D:\Anaconda\lib\site-packages\statsmodels\graphics\tsaplots.py:348: FutureWarning: The default method 'yw' can produce PACF values outside of the [-1,1] interval. After 0.13, the default will change to unadjusted Yule-Walker ('ywm'). You can use this method now by setting method='ywm'.
warnings.warn(
D:\Anaconda\lib\site-packages\statsmodels\graphics\tsaplots.py:348: FutureWarning: The default method 'yw' can produce PACF values outside of the [-1,1] interval. After 0.13, the default will change to unadjusted Yule-Walker ('ywm'). You can use this method now by setting method='ywm'.
warnings.warn(
D:\Anaconda\lib\site-packages\statsmodels\graphics\tsaplots.py:348: FutureWarning: The default method 'yw' can produce PACF values outside of the [-1,1] interval. After 0.13, the default will change to unadjusted Yule-Walker ('ywm'). You can use this method now by setting method='ywm'.
warnings.warn(

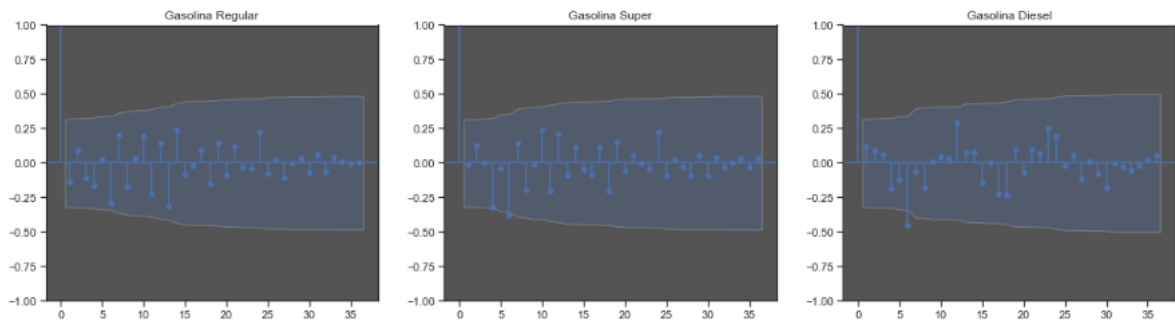


Dada la correlación presentada y la forma de presentarse los datos, puede decirse que no existe una correlación parcial entre los datos, para ningún caso.

Para encontrar q necesitamos ver los gráficos de Autocorrelación

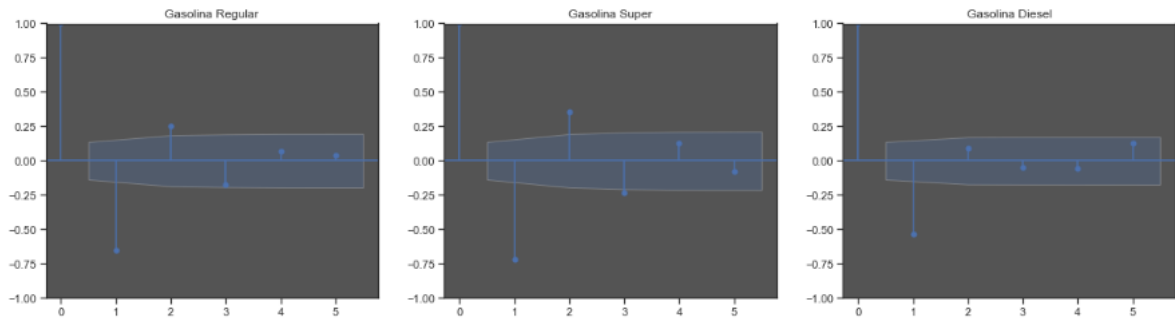
Consumo: Autocorrelación

```
1: fig, axes = plt.subplots(1, 3, figsize=(20, 5))
plot_acf(diffregularcon, title = "Gasolina Regular", ax=axes[0], lags = 36) #q: 0
plot_acf(difffsupercon, title = "Gasolina Super", ax=axes[1], lags = 36) #q: 2
plot_acf(difffdieselcon, title = "Gasolina Diesel", ax=axes[2], lags = 36) #p: 2
plt.show()
```



Importaciones: Autocorrelación

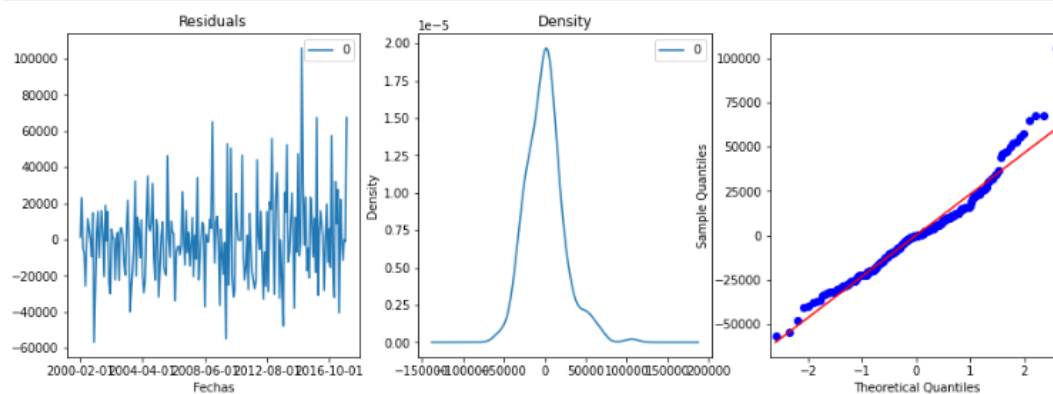
```
1: fig, axes = plt.subplots(1, 3, figsize=(20, 5))
plot_acf(diffregularimp, title = "Gasolina Regular", ax=axes[0], lags = 5) #q: 2
plot_acf(difffsuperimp, title = "Gasolina Super", ax=axes[1], lags = 5) #q: 3
plot_acf(difffdieselimp, title = "Gasolina Diesel", ax=axes[2], lags = 5) #q: 1
plt.show()
```



Dada la correlación presentada y la forma de presentarse los datos, puede decirse que no existe una correlación parcial entre los datos, para ningún caso.

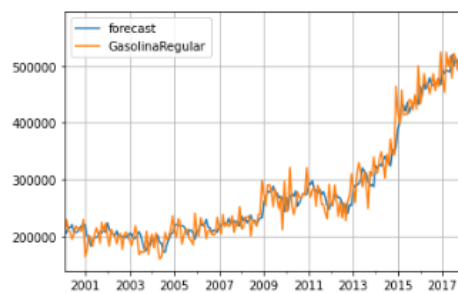
Modelos ARIMA

Modelo para el consumo de Gasolina Regular

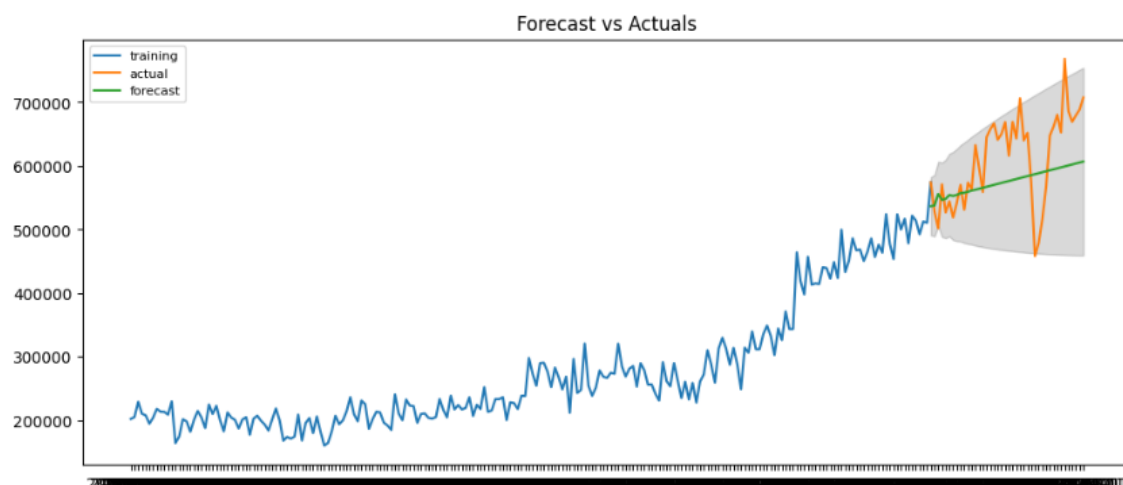


A partir de las gráficas, es posible observar un comportamiento normal para los residuales, así como para la densidad, lo que habla favorablemente del modelo.

```
model_fit_reg_con.plot_predict(dynamic=False)
plt.grid()
plt.show()
```

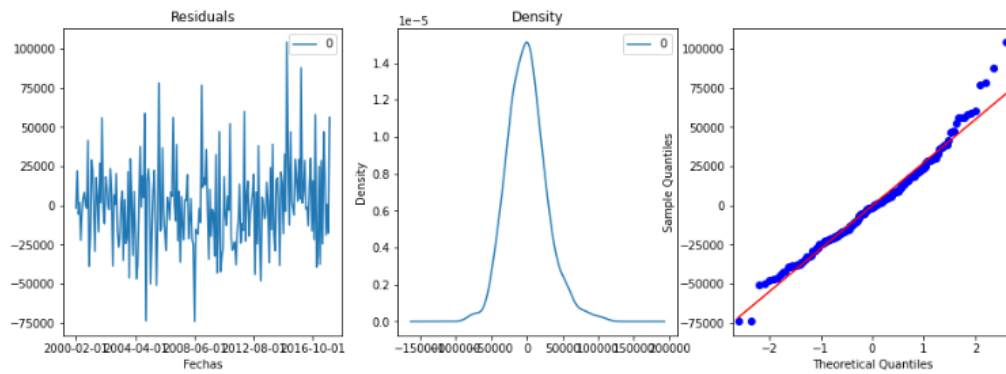


Comparando los resultados predichos, con los resultados reales, puede verse que el modelo realizado es apropiado y tiene una alta certeza.



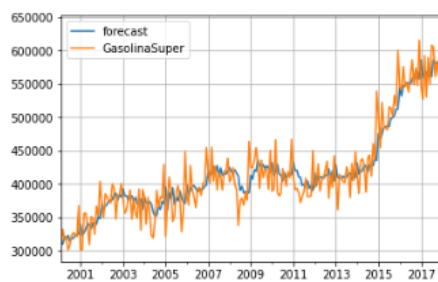
Como se observa en la gráfica, la data de entrenamiento presenta un comportamiento similar a la data real y, los procesos actuales están distintos a las predicciones, dado que las últimas son lineales.

Modelo para el consumo de Gasolina Super

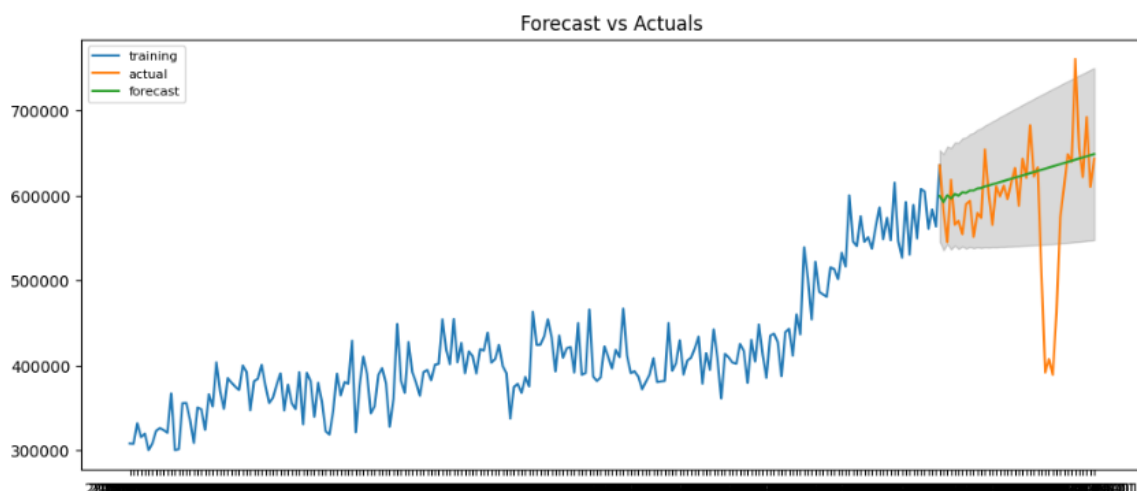


A partir de las gráficas, es posible observar un comportamiento normal para los residuales, así como para la densidad, lo que habla favorablemente del modelo.

```
]: model_fit_super_con.plot_predict(dynamic=False)
plt.grid()
plt.show()
```

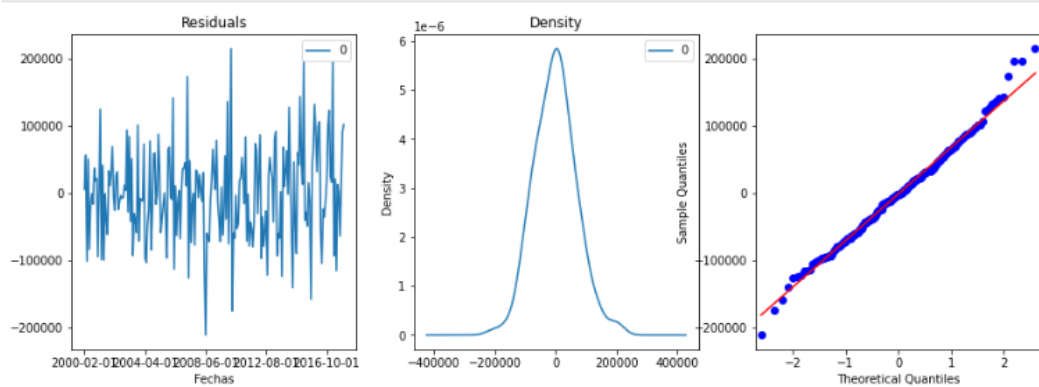


Comparando los resultados predichos, con los resultados reales, puede verse que el modelo realizado es apropiado y tiene una alta certeza.



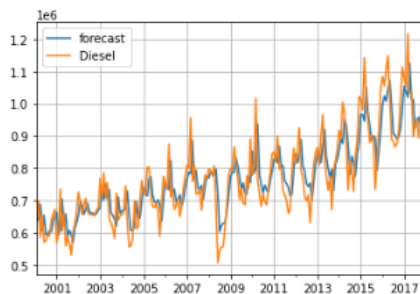
Como se observa en la gráfica, la data de entrenamiento presenta un comportamiento similar a la data real y, los procesos actuales están distintos a las predicciones, dado que las últimas son lineales.

Modelo para el consumo de Gasolina Diesel

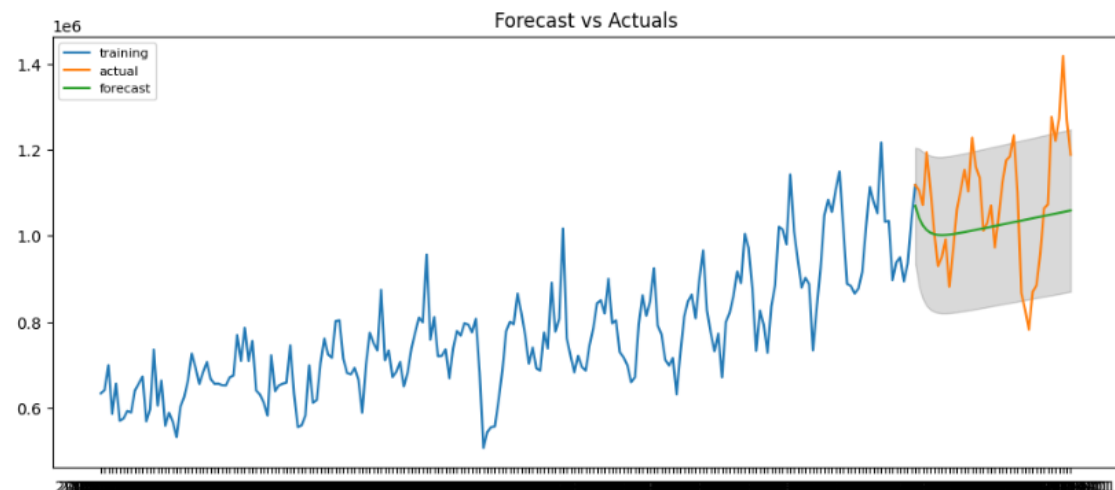


A partir de las gráficas, es posible observar un comportamiento normal para los residuales, así como para la densidad, lo que habla favorablemente del modelo.

```
model_fit_diesel_con.plot_predict(dynamic=False)
plt.grid()
plt.show()
```

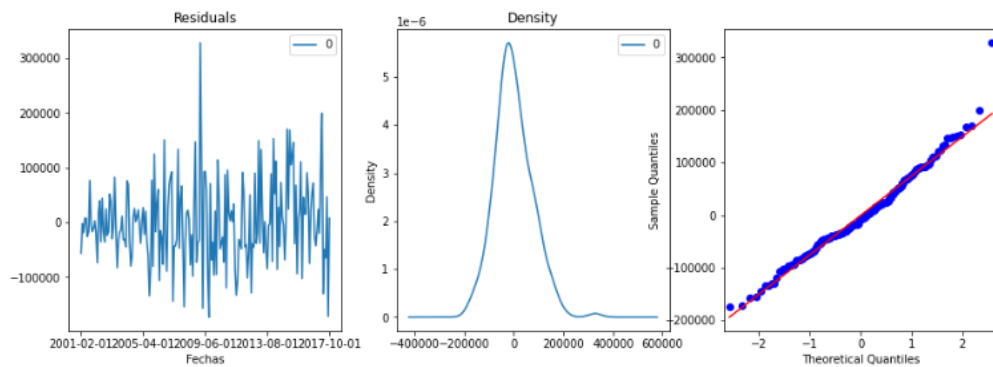


Comparando los resultados predichos, con los resultados reales, puede verse que el modelo realizado es apropiado y tiene una alta certeza.



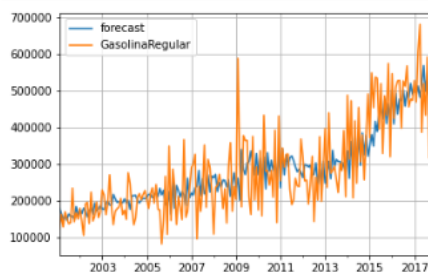
Como se observa en la gráfica, la data de entrenamiento presenta un comportamiento similar a la data real y, los procesos actuales están distintos a las predicciones, dado que las últimas son lineales.

Modelo para la importación de Gasolina Regular

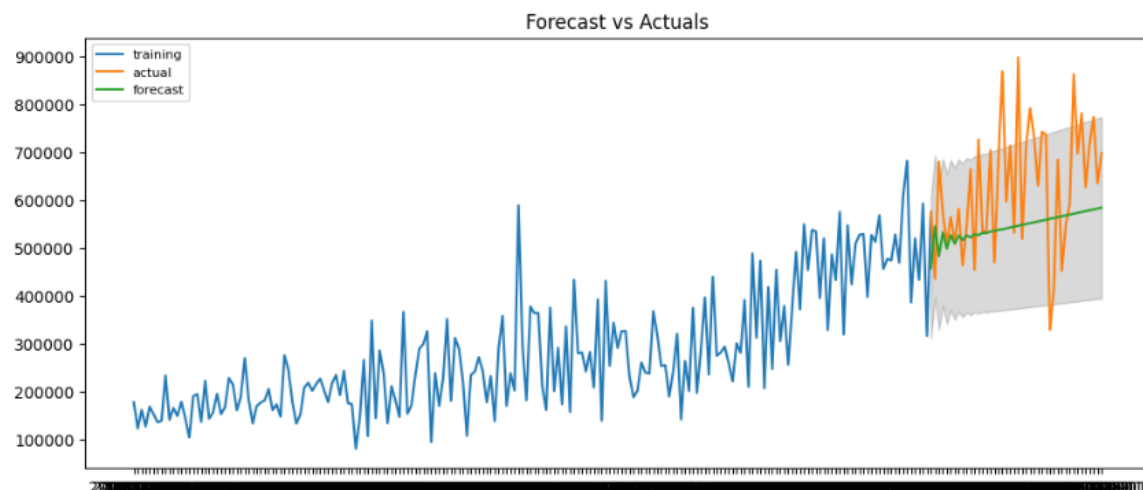


A partir de las gráficas, es posible observar un comportamiento normal para los residuos, así como para la densidad, lo que habla favorablemente del modelo.

```
]: model_fit_reg_imp.plot_predict(dynamic=False)
plt.grid()
plt.show()
```

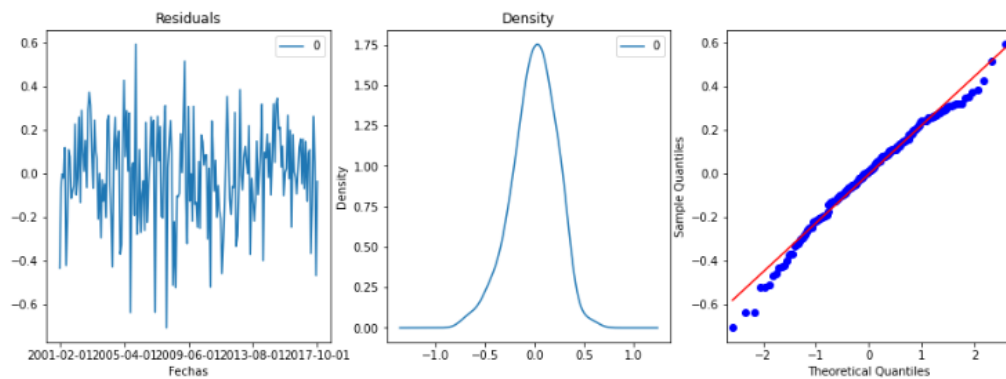


Comparando los resultados predichos, con los resultados reales, puede verse que el modelo realizado es apropiado y tiene una alta certeza.



Como se observa en la gráfica, la data de entrenamiento presenta un comportamiento similar a la data real y, los procesos actuales están distintos a las predicciones, dado que las últimas son lineales.

Modelo para la importación de Gasolina Super

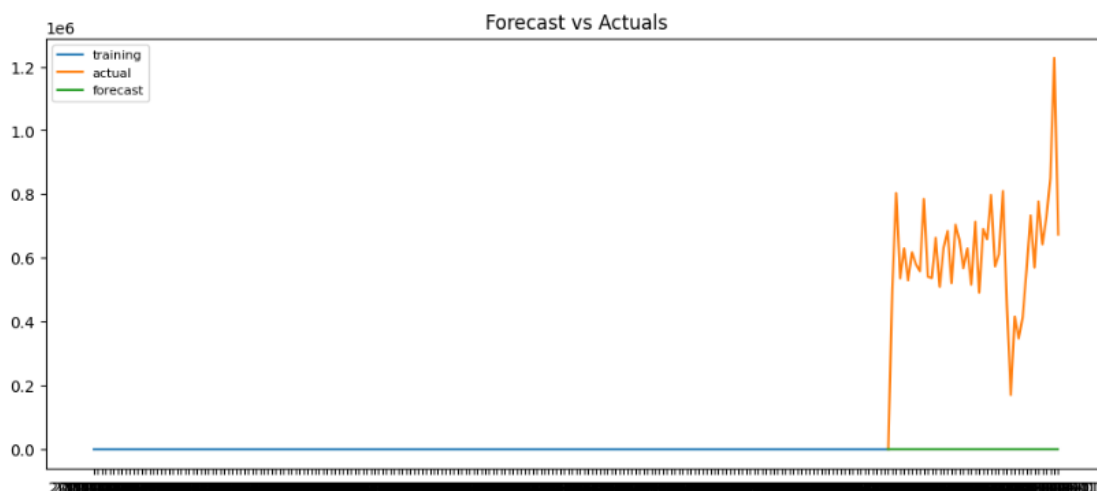


A partir de las gráficas, es posible observar un comportamiento normal para los residuales, así como para la densidad, lo que habla favorablemente del modelo.

```
model_fit_sup_imp.plot_predict(dynamic=False)
plt.grid()
plt.show()
```



Comparando los resultados predichos, con los resultados reales, puede verse que el modelo realizado es apropiado y tiene una alta certeza.



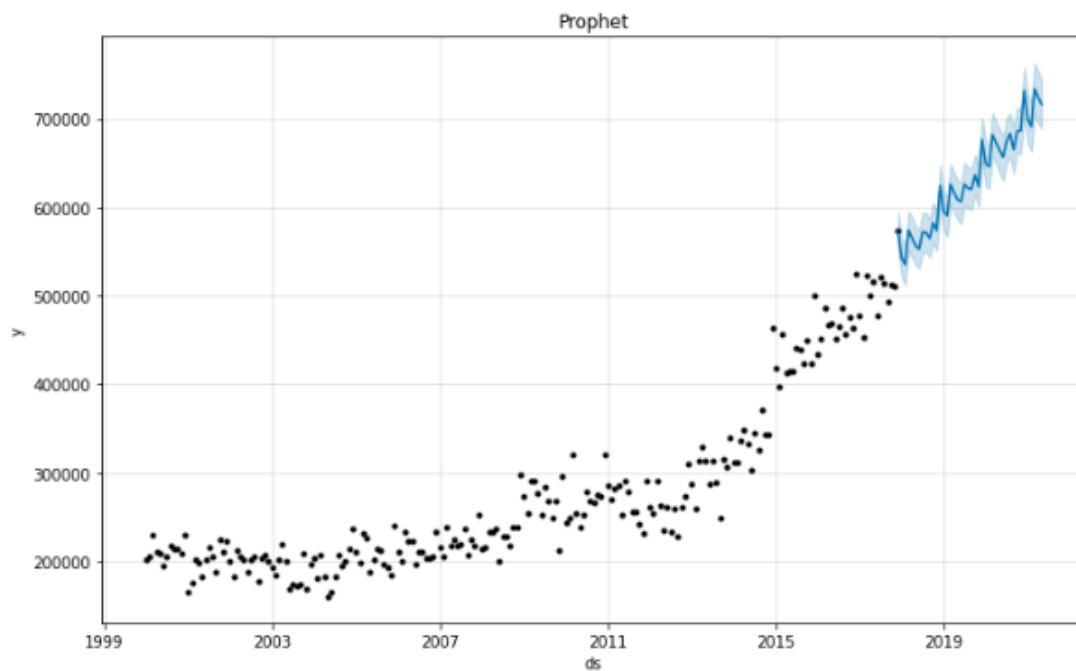
Como se observa en la gráfica, la data de entrenamiento presenta un comportamiento similar a la data real y, los procesos actuales están distintos a las predicciones, dado que las últimas son lineales.

Algoritmo Prophet

De forma inicial, dados los requisitos para la función de Prophet, creamos una función que transforme los datos a la forma deseada, para luego crear un dataframe con estos.

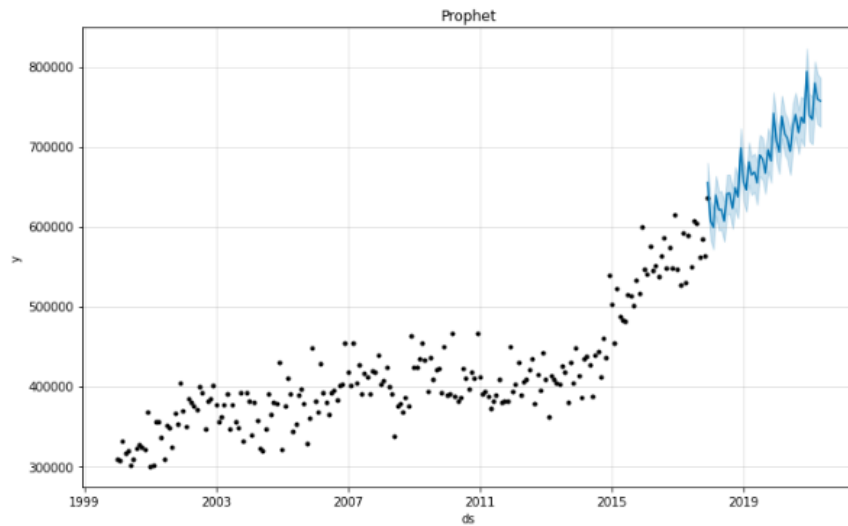
Consumo Gasolina Regular

	ds	yhat	yhat_lower	yhat_upper
0	2017-12-01	572495.820658	550604.156444	595514.073444
1	2018-01-01	542115.645864	520481.560092	564437.435163
2	2018-02-01	535319.405229	512561.877400	557789.704471
3	2018-03-01	574220.253760	551336.401739	595698.339780
4	2018-04-01	565051.060437	542530.626101	589117.671134



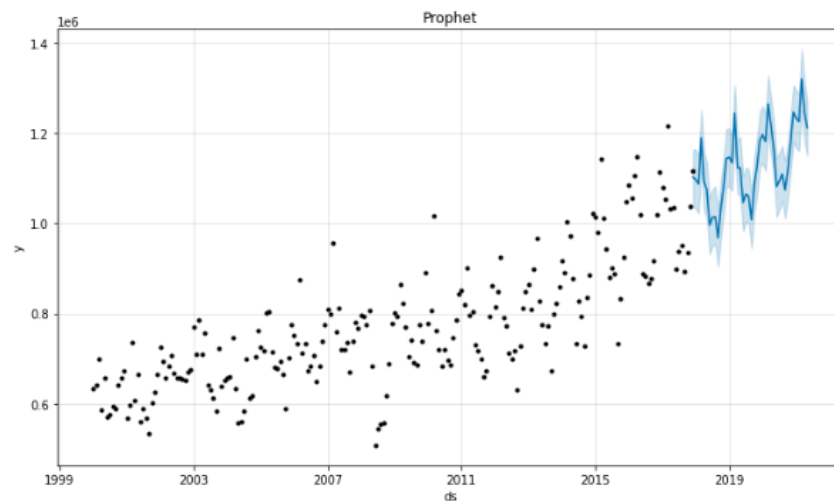
Consumo Gasolina Super

	ds	yhat	yhat_lower	yhat_upper
0	2017-12-01	654754.045289	629879.011291	679728.393883
1	2018-01-01	606494.282133	582258.972860	631591.845664
2	2018-02-01	598922.988772	571470.903697	621980.986132
3	2018-03-01	638965.370074	613567.408820	663420.913686
4	2018-04-01	620687.802400	595046.472286	643647.273609

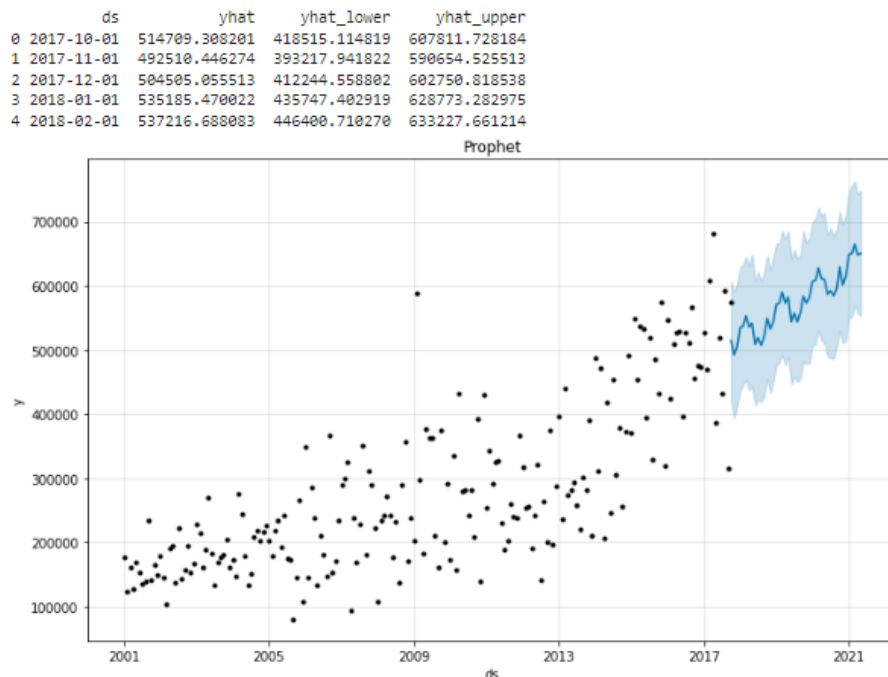


Consumo Gasolina Diesel

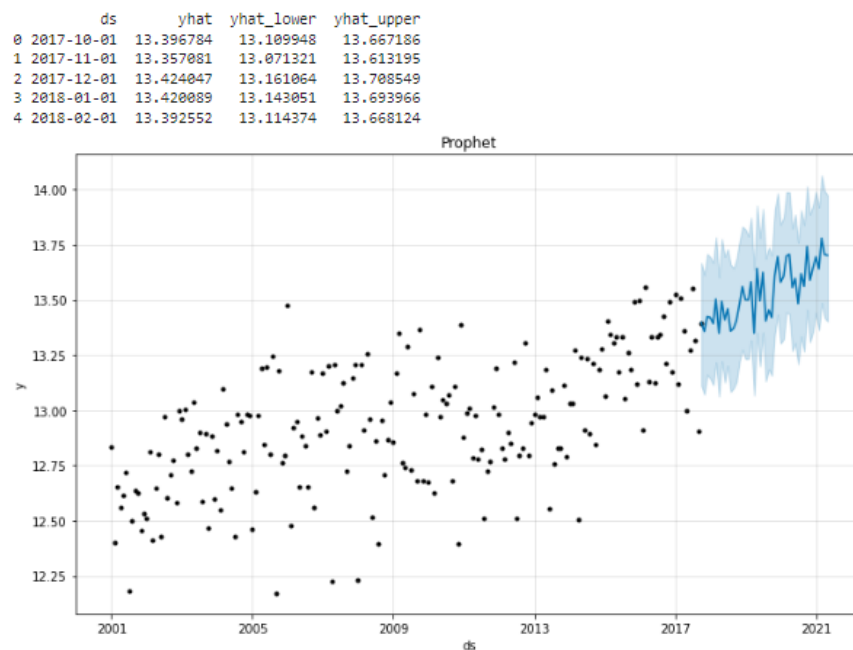
	ds	yhat	yhat_lower	yhat_upper
0	2017-12-01	1.102381e+06	1.037687e+06	1.163847e+06
1	2018-01-01	1.097183e+06	1.032564e+06	1.163456e+06
2	2018-02-01	1.087337e+06	1.022304e+06	1.150923e+06
3	2018-03-01	1.189663e+06	1.126957e+06	1.252046e+06
4	2018-04-01	1.093089e+06	1.027392e+06	1.158950e+06



Importaciones Gasolina Regular

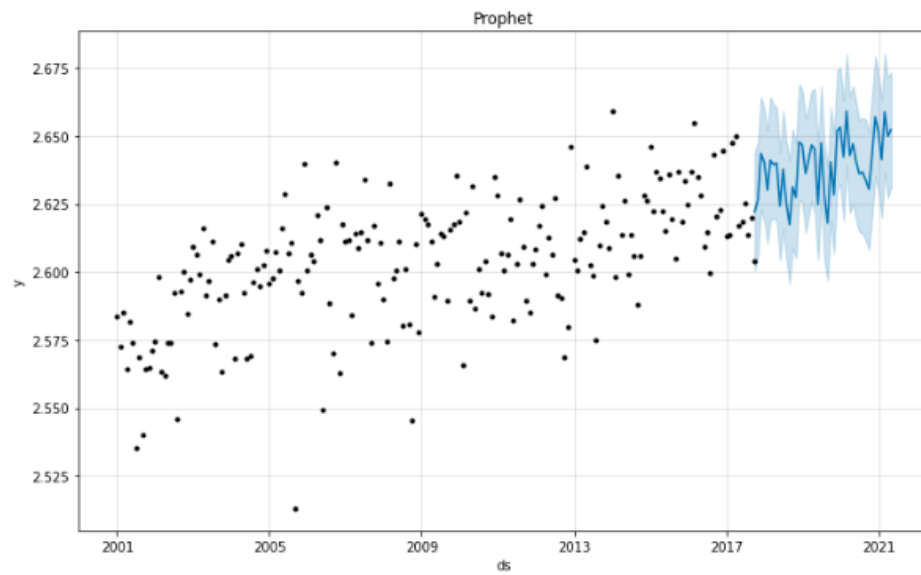


Importaciones Gasolina Super



Importaciones Gasolina Diesel

	ds	yhat	yhat_lower	yhat_upper
0	2017-10-01	2.622130	2.600530	2.643802
1	2017-11-01	2.626431	2.606052	2.647802
2	2017-12-01	2.643622	2.623212	2.664301
3	2018-01-01	2.640390	2.618698	2.659923
4	2018-02-01	2.630064	2.610539	2.651833



Referencias

Amat, J. (2021). Skforecast: forecasting series temporales con Python y Scikit-learn. 8/5/2022, de Ciencia de datos Sitio web: <https://www.cienciadedatos.net/documentos/py27-forecasting-series-temporales-python-scikitlearn.html>

Arzamendia, O. (2021). Predicción con Series de Tiempo - Una guía inicial. 8/5/2022, de Medium Sitio web: <https://medium.com/datos-y-ciencia/predicci%C3%B3n-con-series-de-tiempo-una-gu%C3%ADa-inicial-2bd62d55675a>

Barbier, M. (2021). SERIES DE TIEMPO (CADENA DE SUMINISTRO). 8/5/2022, de Lokad Sitio web: <https://www.lokad.com/es/series-de-tiempo-en-cadena-de-suministro>