Trabajo Práctico

Análisis de Series de Tiempo

CEIA – FIUBA

Alumno: Federico Paolino

Profesores: Carlos Romano

Magdalena Bouza

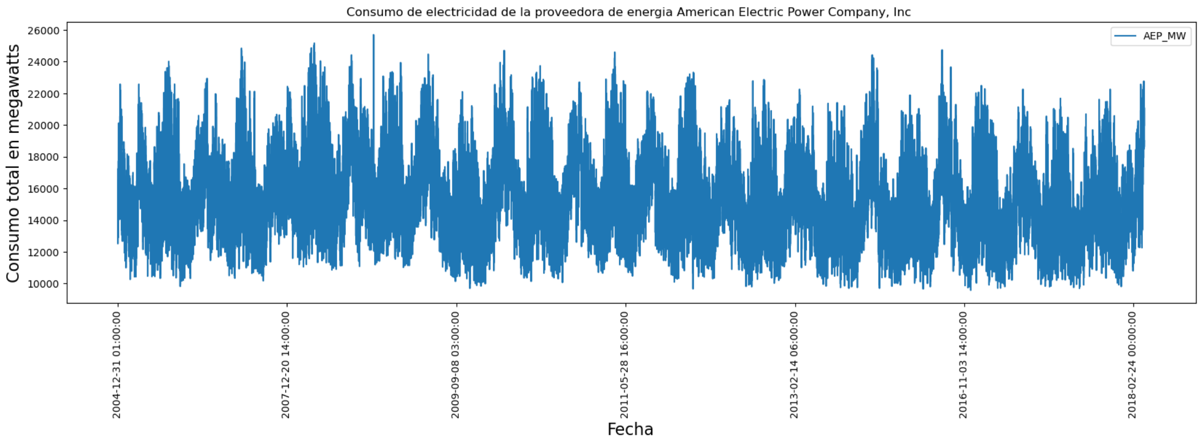
1- Graficar una serie a partir de un dataset relevante. Describir observaciones

El dataset fue tomado de este link, el dataset muestra los datos de consumo eléctrico de muchas proveedoras en la PJM Interconnection LLC desde el año 2004 al 2018.

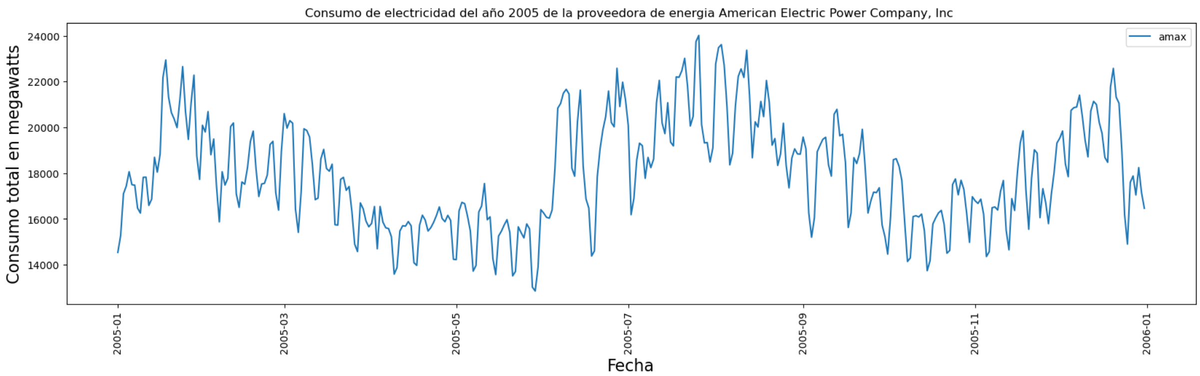
PJM Interconnection LLC es una organización regional que entrega suministro eléctrico a Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, y Columbia.

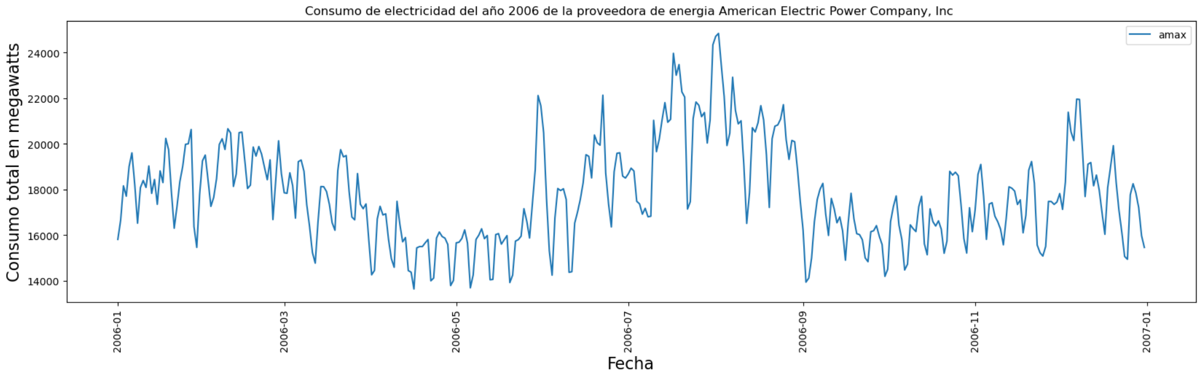
En este caso se redujo el dataset para que sea manejable a solo una empresa, la American Electric Power Company. Además, se tomó solo el máximo de consumo de cada día para que los datos fueran más manejables y además los gráficos fueran más claros.

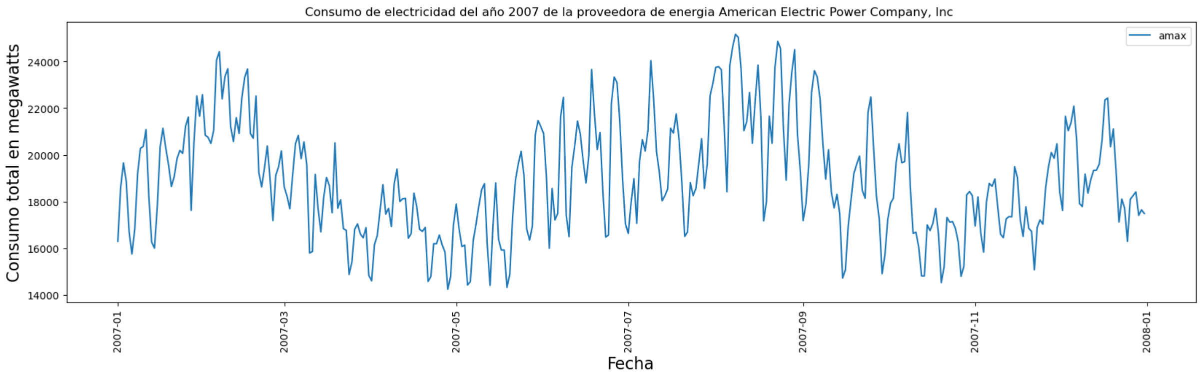
Dataset entero:



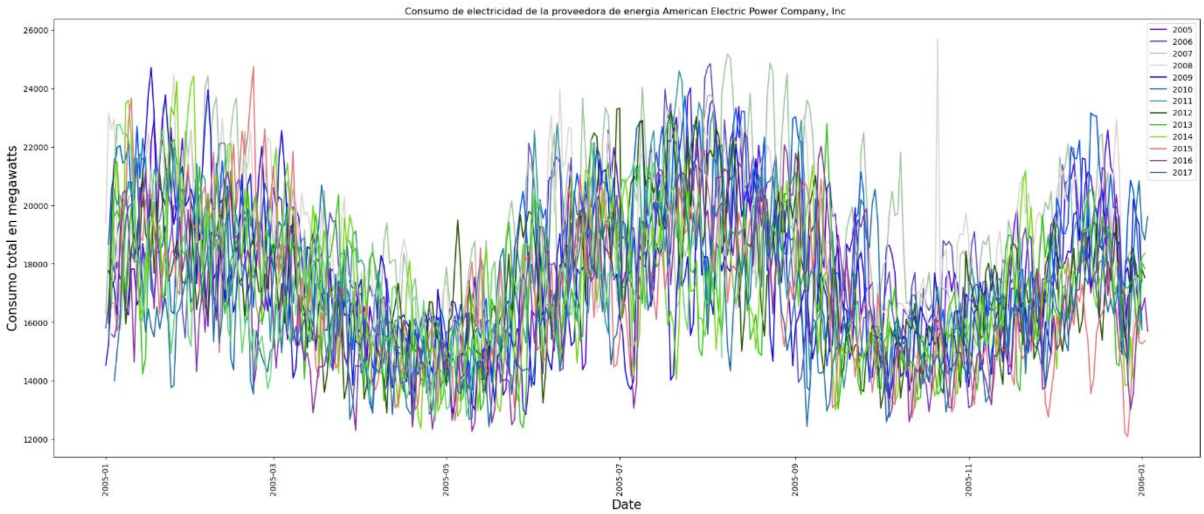
Dataset por año 2005-2007:



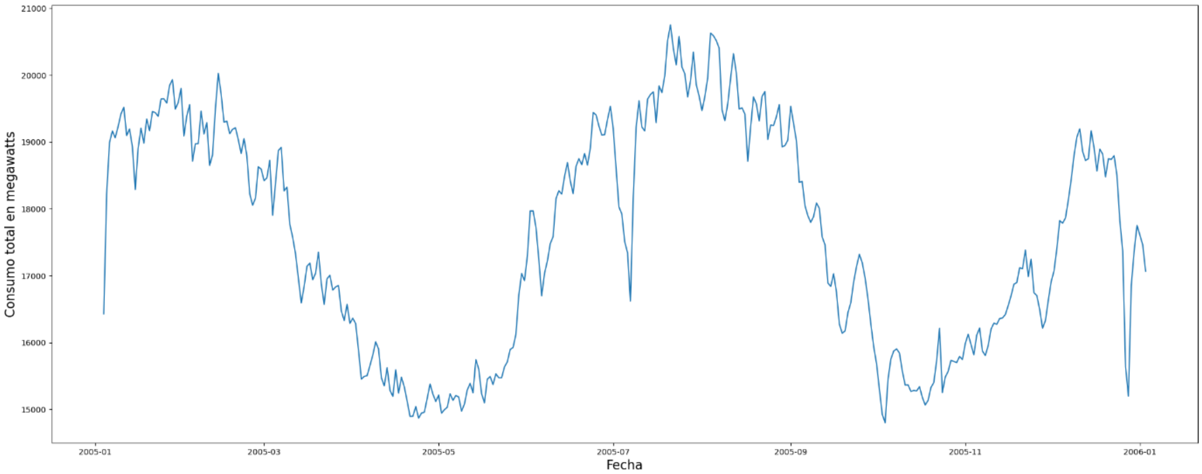




Consumo máximo por día dependiendo de la fecha:



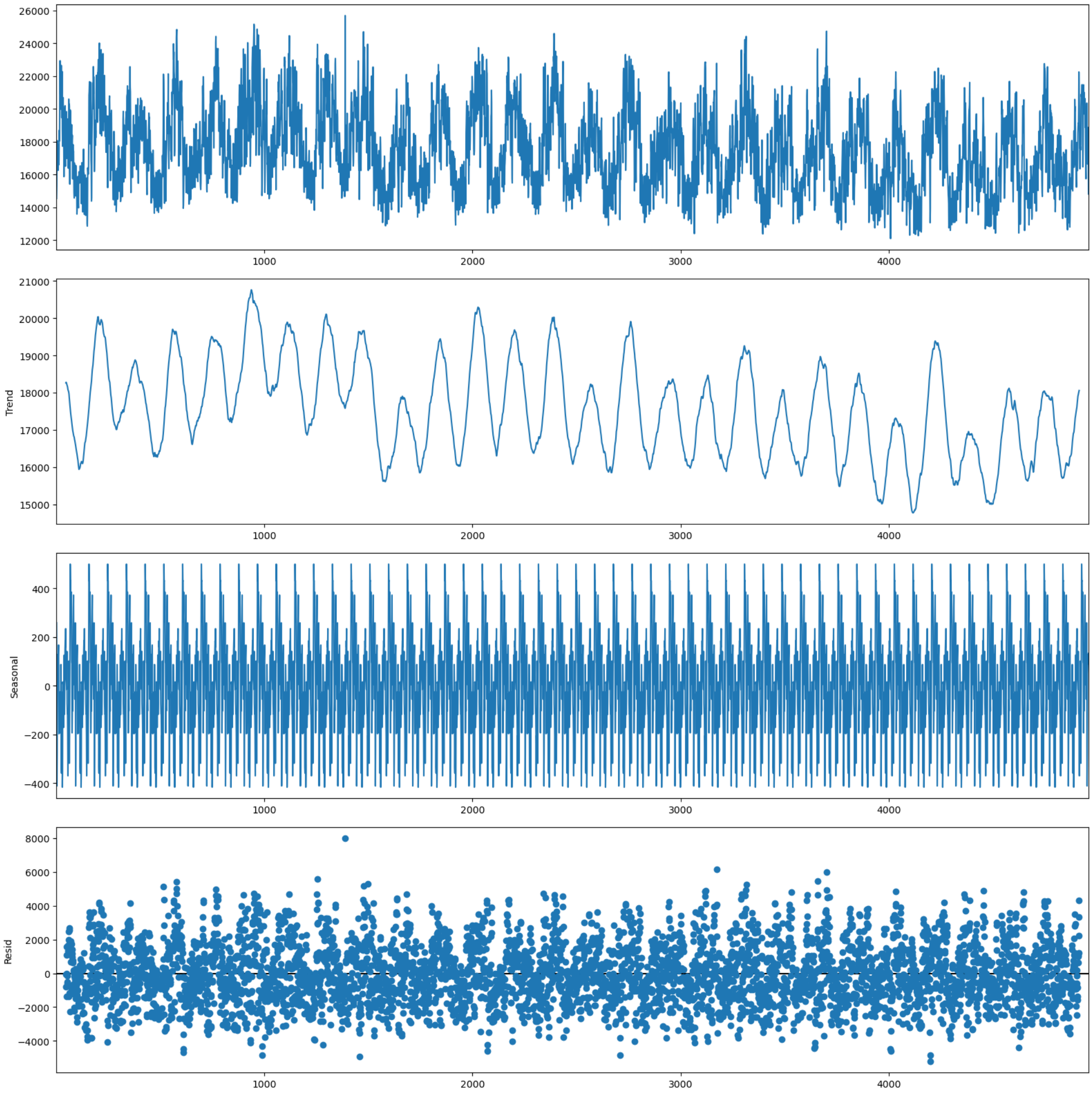
Promedio del consumo máximo por día dependiendo de la fecha:



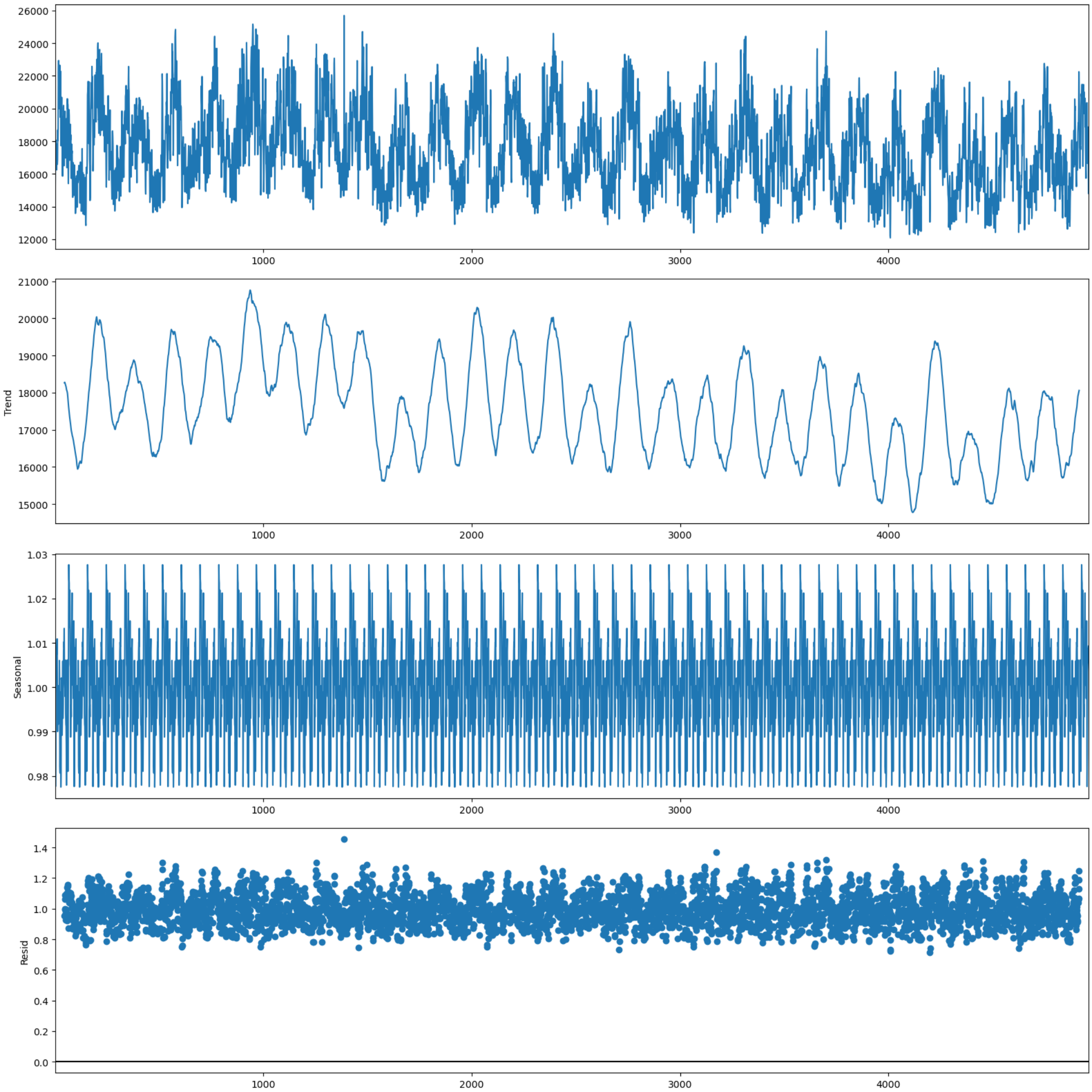
Viendo esta serie de tiempo podemos ver una estacionalidad al largo del año, que, si bien tiene cambios, en general el consumo sube bastante durante tanto los meses más calurosos del verano y los meses más fríos del invierno.

2- Descomponer una serie de tiempo usando el modelo aditivo de cuatro componentes.

a. Aditivo

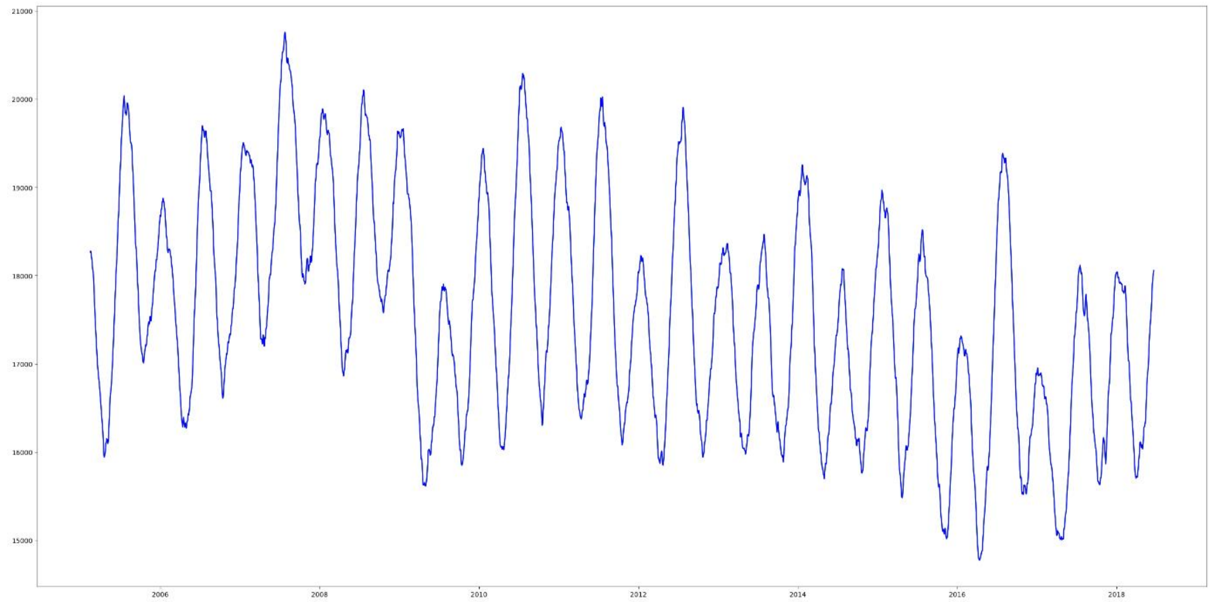


b. Multiplicativo

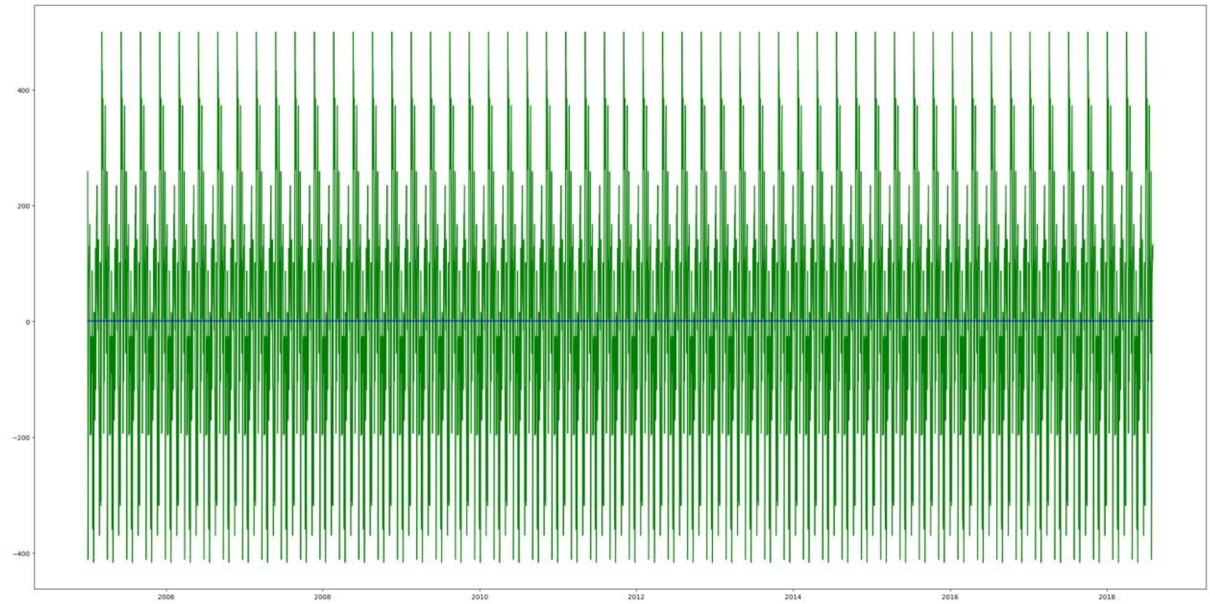


3- Extraer la tendencia y ajustar un modelo determinístico. Explicar su relación con el contexto. Obtener conclusiones acerca de la validez del modelo.

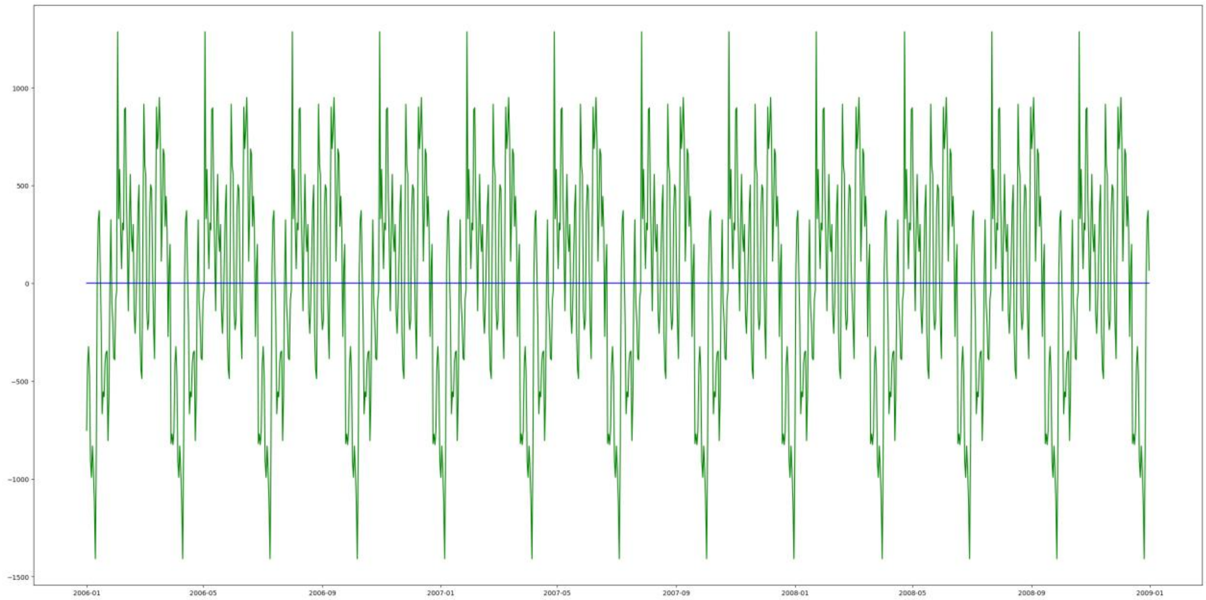
Modelo aditivo de tren(rojo) vs modelo multiplicativo de tren(azul)



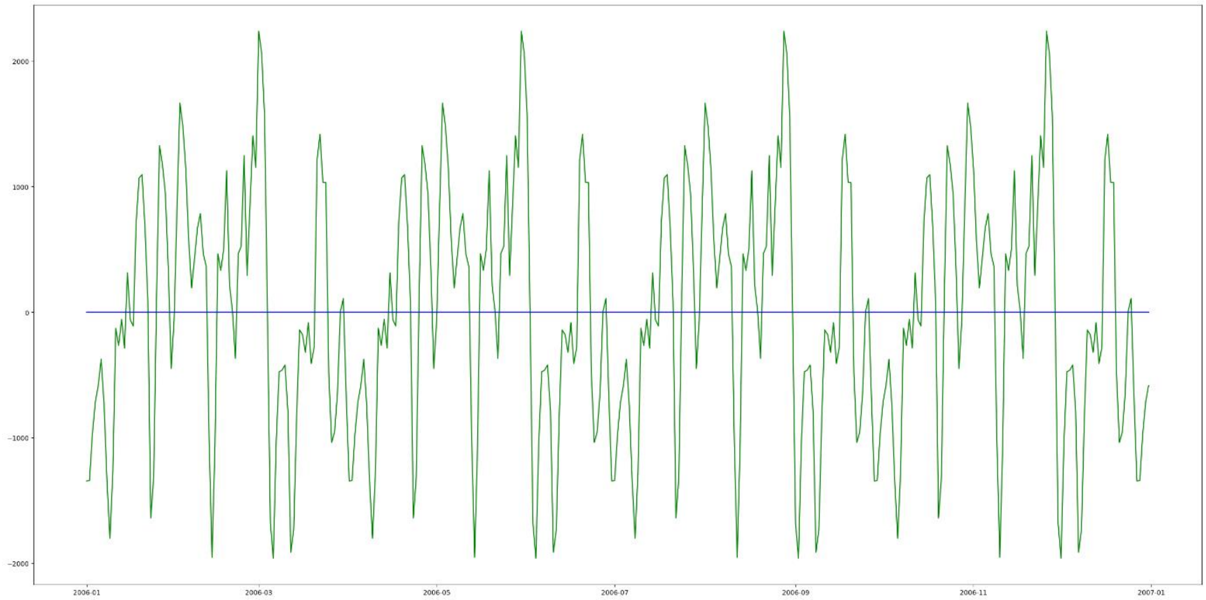
Modelo aditivo de estacionalidad(verde) vs modelo multiplicativo de estacionalidad(azul)



Modelo aditivo de estacionalidad(verde) vs modelo multiplicativo de estacionalidad(azul) año 2006 al 2008

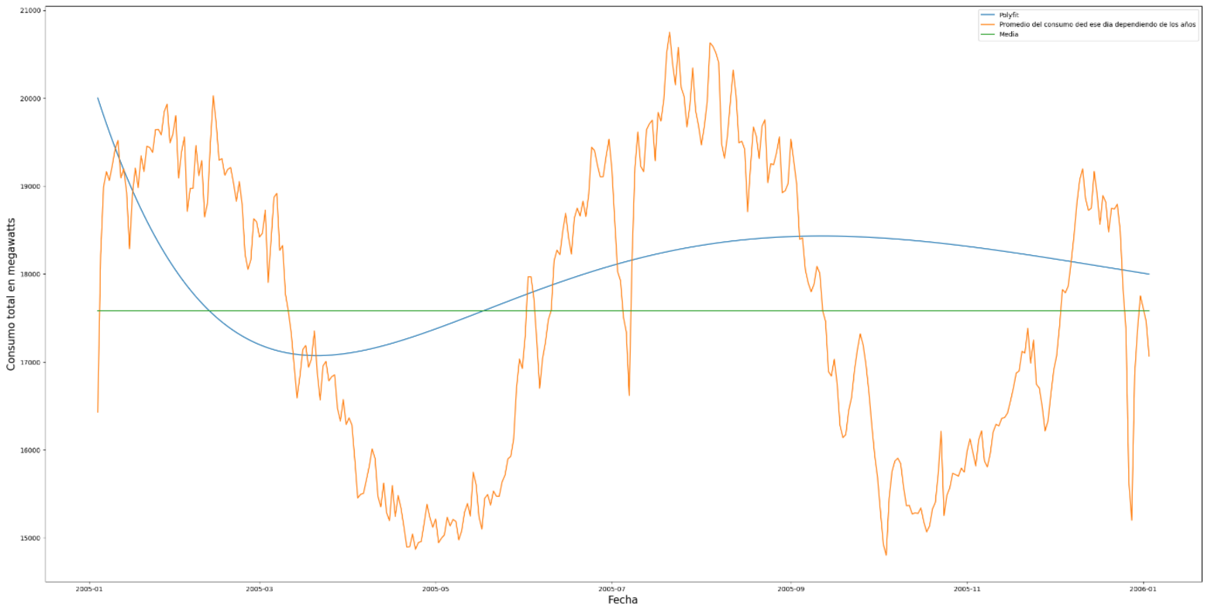


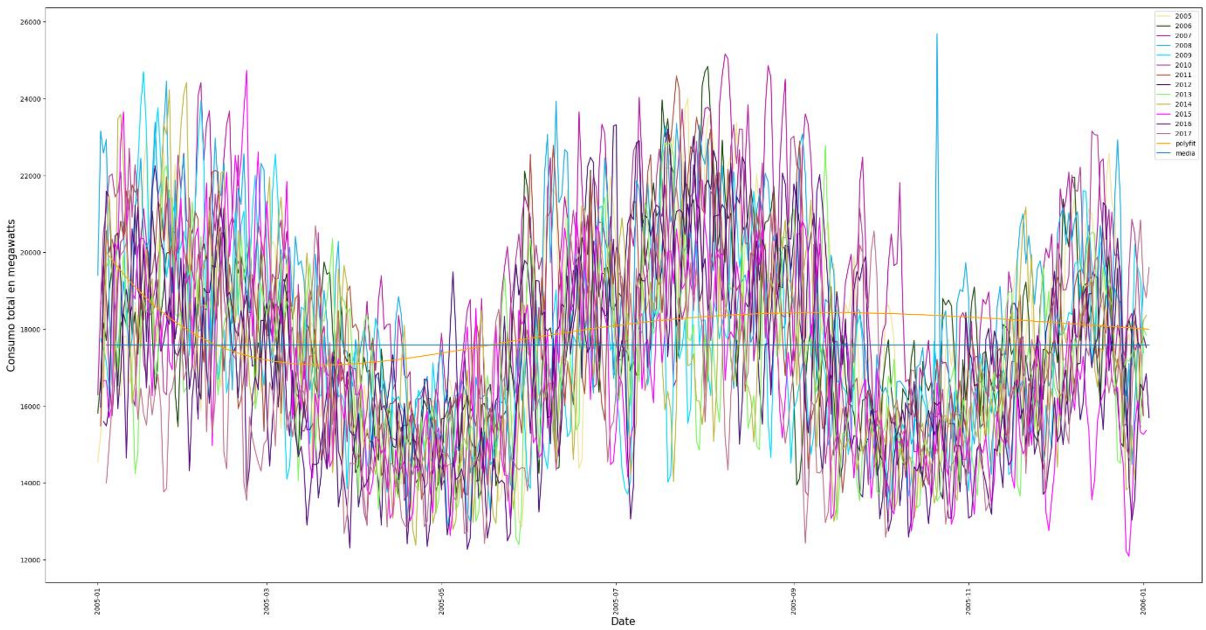
Modelo aditivo de estacionalidad(verde) vs modelo multiplicativo de estacionalidad(azul) año 2006



Viendo la serie descompuesta podemos ver una tendencia anual, como así también una tendencia trimestral que empieza a crecer desde el inicia exacto de cada trimestre del año. Podemos ver una clara estacionalidad viendo el modelo multiplicativo.

Para ajustar el modelo determinístico usamos un polinomio de grados 30, usando la función polyfit.





Aquí podemos ver que, si bien el polinomio intenta seguir la tendencia usando la media como guía, no es insuficiente a la hora de replicar la serie de tiempo.

4- Evaluar si la serie original es estacionaria

Para evaluar la serie original y ver si es estacionaria necesitamos que en el test de Dickey-Fuller rechacemos h0 y en KPSS no podamos rechazar h0.

Dickey-Fuller:

```
(-7.731802721663455,
 1.1202560657678411e-11,
 29,
 4930,
 {'1%': -3.43167712132583,
  '5%': -2.862126442286194,
  '10%': -2.5670821642549444},
 82998.99968660779)
```

Basado en el nivel de significancia de 0.05 y el p valor del test de Dickey-Fuller de 9.7e-12, podemos rechazar la hipótesis nula. Por lo tanto, según este test, la serie es estacionaria.

KPSS:

```
(1.3904345518598666,
 0.01,
 39,
 {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739})
```

Basado en el nivel de significancia de 0.05 y el p valor del test KPSS de 0.01, podemos rechazar la hipótesis nula. Por lo tanto, según este test, la serie NO es estacionaria.

Por lo tanto, en este caso la serie es estacionaria diferencial y, por lo tanto, debemos aplicar diferenciación para poder hacerla estacionaria.

Una vez hecho esto, volvemos a correr ambos tests:

Dickey-Fuller:

```
(-28.952018565926164,
 0.0,
 32,
 4924,
 {'1%': -3.4316787392959416,
  '5%': -2.862127157092089,
  '10%': -2.5670825447748093},
 83694.83312865635)
```

Basado en el nivel de significancia de 0.05 y el p valor del test de Dickey-Fuller de 3.8e-18, podemos rechazar la hipótesis nula. Por lo tanto, según este test, la serie es estacionaria.

KPSS:

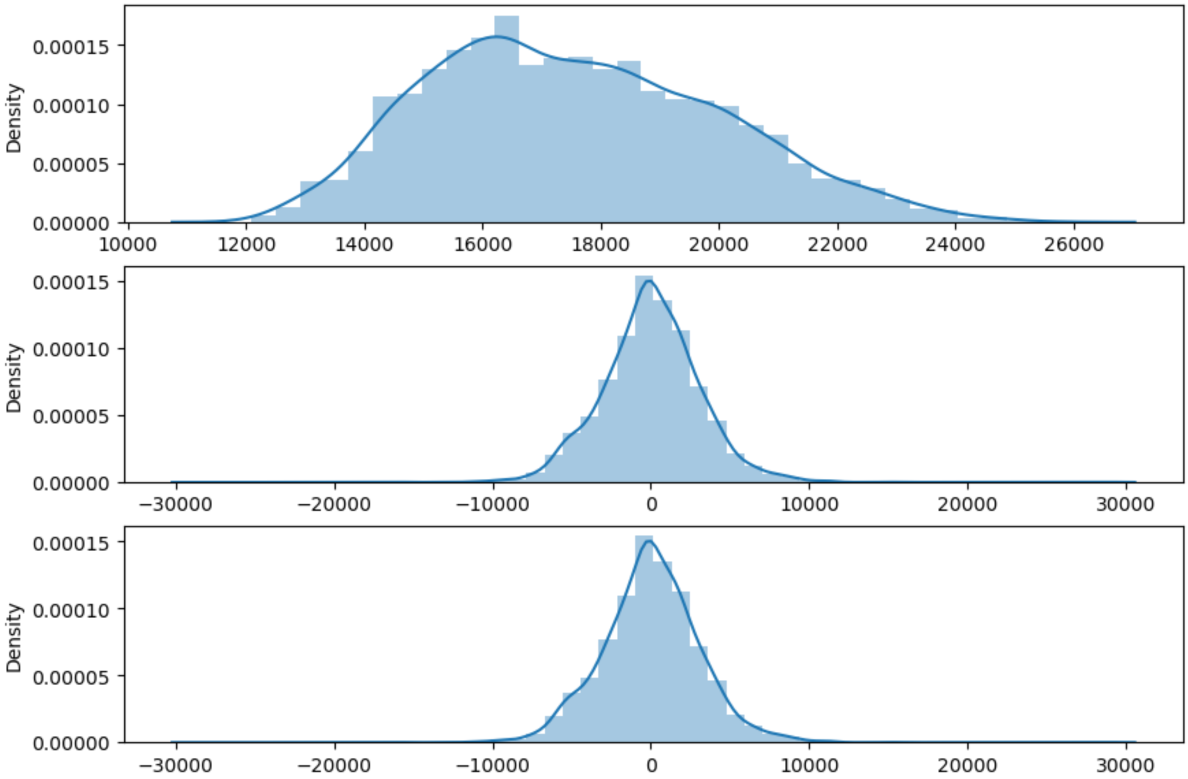
```
(0.028143556291208478,
 0.1,
 80,
 {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739})
```

Basado en el nivel de significancia de 0.05 y el p valor del test KPSS de 0.1, no podemos rechazar la hipótesis nula. Por lo tanto, según este test, la serie es estacionaria.

5- Preprocesamiento:

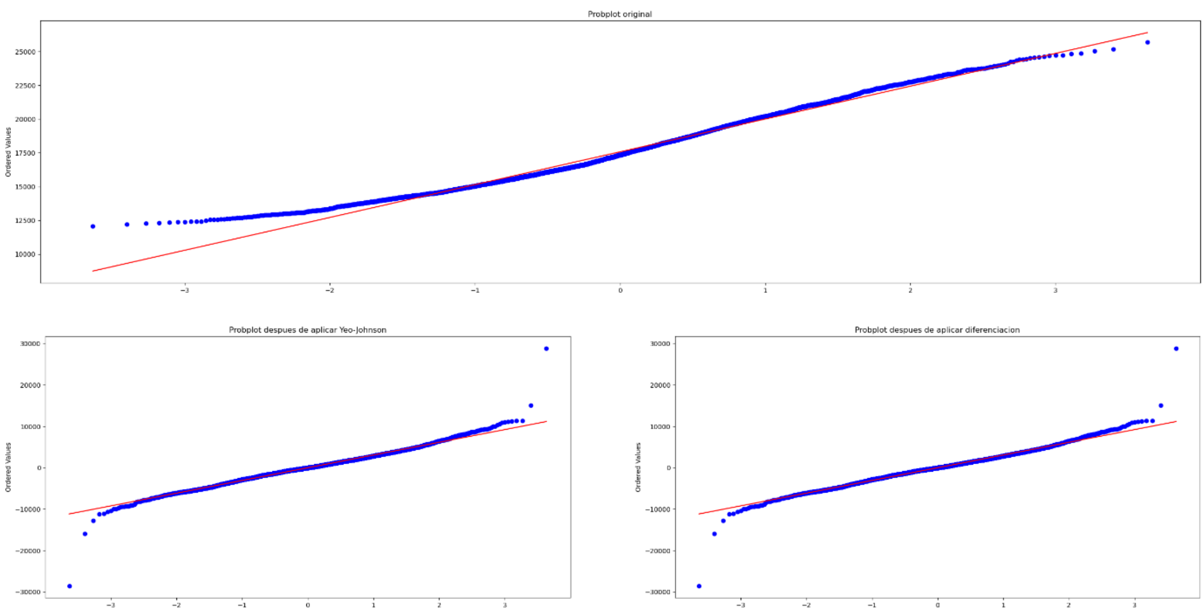
Al haber aplicado diferenciación en nuestro dataset ahora podemos llegar a tener valores negativos, por lo tanto, para poder normalizar nuestra data usamos Yeo-Johnson.

Histogramas:



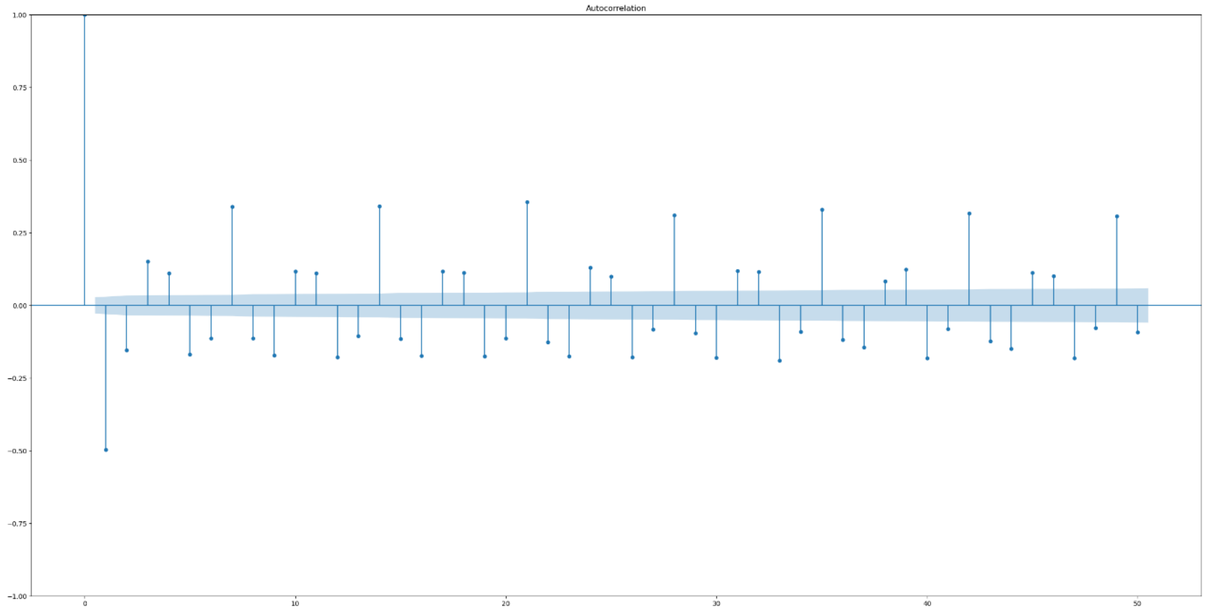
Aquí vemos los histogramas de nuestra serie de tiempo, primero sin ninguna transformación, segundo después de aplicar diferenciación y finalmente habiendo aplicado Yeo-Johnson.

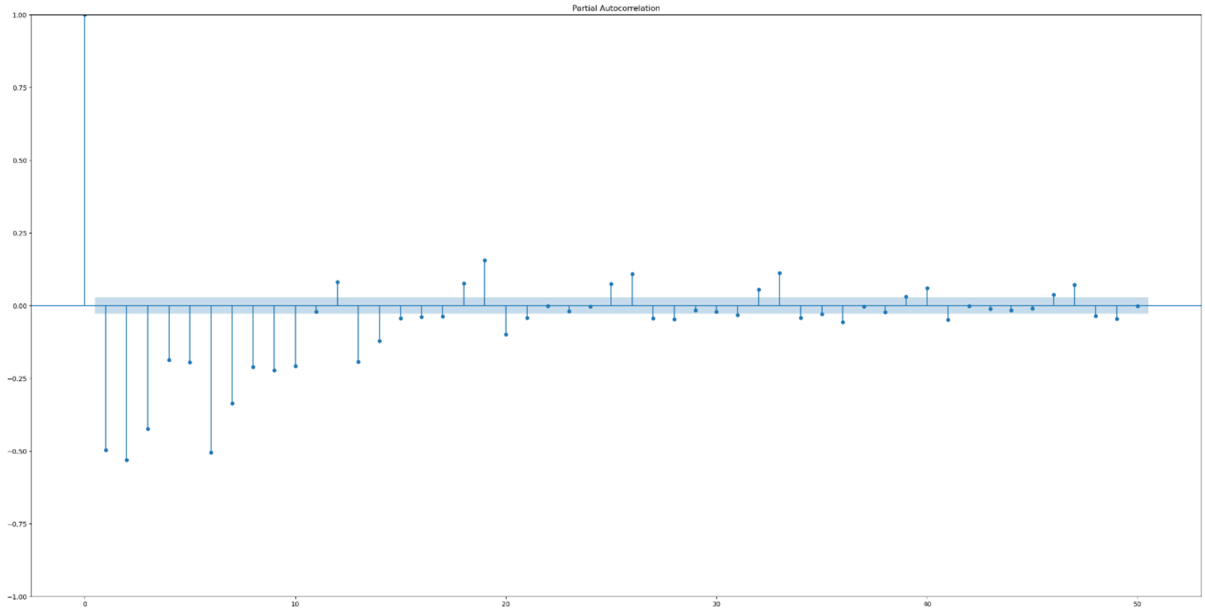
ProbPlots:



Ahora viendo toda la información podemos notar que, una vez aplicada la diferenciación, Yeo-Johnson no tiene que hacer prácticamente nada, ya que en los probplot podemos notar que son prácticamente iguales, por lo tanto, el dataset esta tan normalizado como se podría.

6- Autocorrelación y autocorrelación parcial





Usando este grafico de autocorrelación y autocorrelación parcial podemos notar que la serie de tiempo no es aleatoria y hay correlación en la misma a través del tiempo.