



# Series Temporales

## EXAMEN: Demanda Energía

Marisa Alonso González-Noriega

CUNEF UNIVERSIDAD  
MASTER CIENCIA DE DATOS

Enero de 2022

# Índice general

<b>1. Introducción</b>	<b>2</b>
1.1. Objetivo . . . . .	2
<b>2. Desarrollo</b>	<b>3</b>
2.1. Datos horarios . . . . .	3
2.1.1. Modelo a batir: Naïve estacional . . . . .	5
2.1.2. ARIMA . . . . .	6
2.1.3. ETS . . . . .	6
2.1.4. Prophet . . . . .	6
2.1.5. Comparación modelos . . . . .	7
2.1.6. Predicciones . . . . .	8
2.1.7. Anexo: Librería AutoTS. Predicciones horarias diciem- bre . . . . .	9
2.2. Datos diarios . . . . .	10
2.2.1. Modelo a batir: Naïve estacional . . . . .	11
2.2.2. ARIMA . . . . .	12
2.2.3. ETS . . . . .	12
2.2.4. Prophet . . . . .	13
2.2.5. Comparación modelos . . . . .	14
2.2.6. Predicciones . . . . .	14
2.2.7. Anexo: Librería AutoTS. Predicciones Diarias diciembre	16
<b>3. Conclusiones</b>	<b>17</b>

# Capítulo 1

## Introducción

### 1.1. Objetivo

El objetivo de este informe es realizar un modelo predictivo sobre los datos de demanda de energía en California.

Estos datos están disponibles tanto en periodicidad horaria como diaria, datan del 1 de julio del 2015 hasta el 30 de noviembre de 2021.

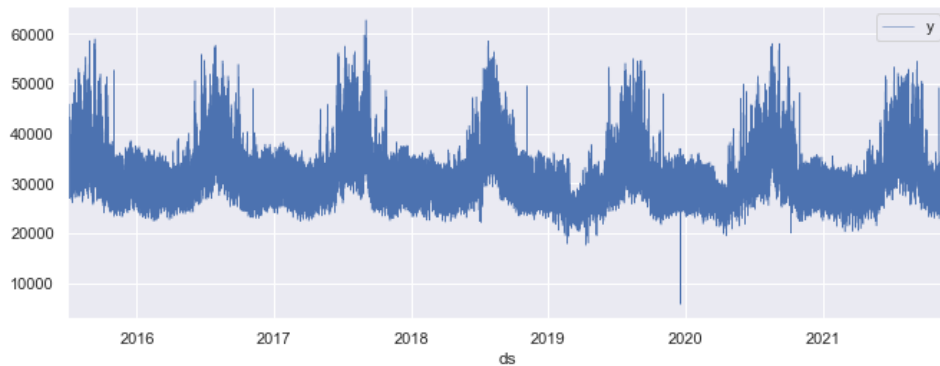
Nuestro objetivo final es realizar la predicción de los primeros 7 días de diciembre de 2021 en periodicidad horaria y la predicción diaria de los 15 primeros días de diciembre 2021.

## Capítulo 2

# Desarrollo

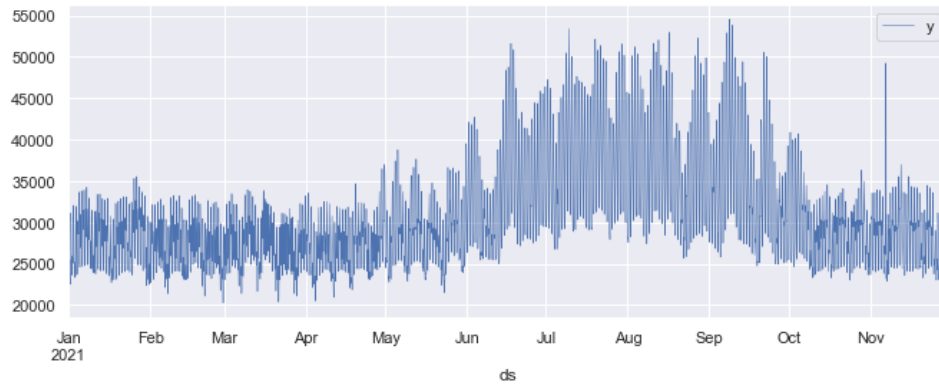
### 2.1. Datos horarios

Comenzamos tomando los datos horarios(hora local), tenemos una gran cantidad pues a una media de 8760h por año y teniendo en cuenta que en nuestro dataset tenemos datos de 6 años y 4 meses resulta muy complicado la interpretación gráfica de nuestros datos.



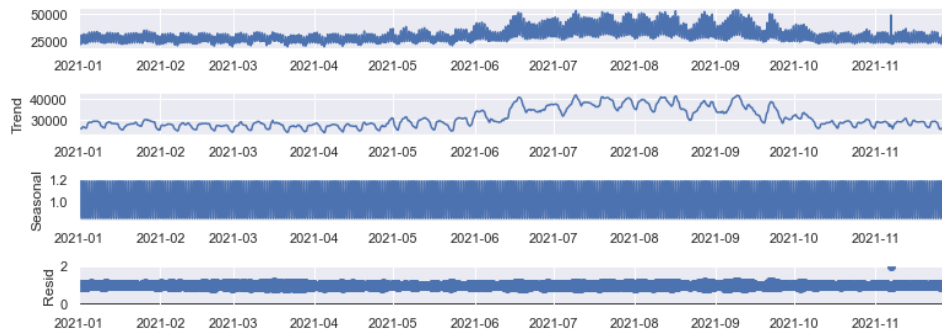
(a) Datos horarios.

Observamos que la serie es estacional, se observa que la demanda aumenta en los meses de verano y disminuye en el resto. Por todo lo descrito anteriormente decidimos quedarnos únicamente con los datos de 2021.



(b) Datos horarios 2021.

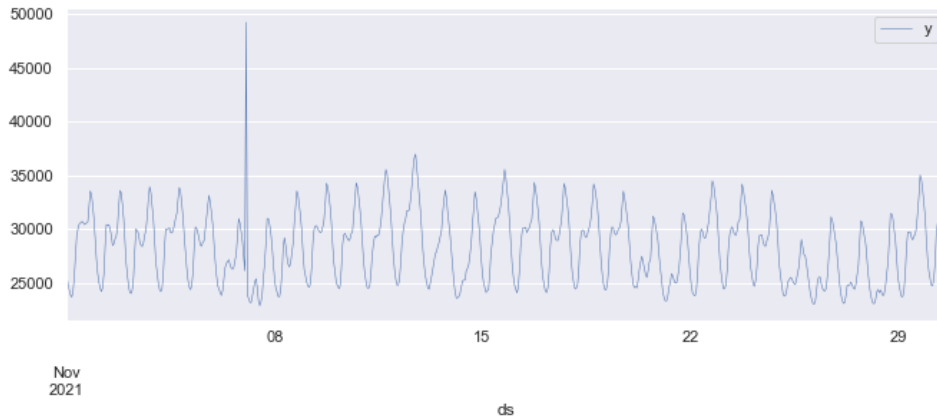
Vamos a observar más en detalle los patrones componentes de estos datos, pero antes de eso observamos que falta un valor el día 14 de Marzo a las 02 a.m.; como es únicamente un dato decidimos sustituirlo por el valor medio de la hora anterior y la hora posterior. Ahora ya si, después de esta pequeña modificación de nuestros datos, observamos la descomposición de las componentes:



(c) Patrones componentes 2021.

Observamos que la serie no tiene tendencia, pero poca más información nos da el gráfico al tener un gran número de observaciones.

Como quitando los meses de verano nuestra serie sigue el mismo patrón y al ser datos horarios tenemos mucha información, decidimos realizar las predicciones tomando únicamente los datos de noviembre.



(d) Noviembre 2021.

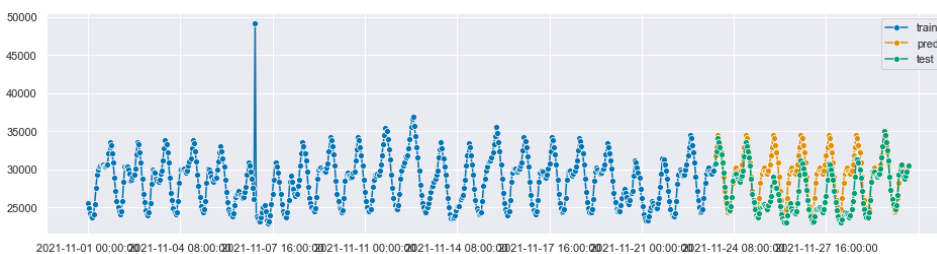
A simple vista no podemos determinar si nuestra serie es estacionaria o no, realizamos por ello la prueba de Dickey Fuller aumentada [1] y concluimos que la serie es estacionaria.

Observamos un valor atípico que corresponde a un impulso, el día 7 de Noviembre a la 1 a.m. Decidimos afrontar esto de dos formas distintas: utilizando el dataset original y modificando ese dato por la mediana de dicho día. Se obtuvieron resultados muy similares y concluimos que este valor atípico no afecta a nuestras predicciones, por tanto continuamos con el dataset original.

### 2.1.1. Modelo a batir: Naïve estacional

Como nuestros datos son estacionales utilizamos como modelo de referencia el método ingenuo o Naïve estacional, así cada predicción sera igual al último valor observado teniendo en cuenta la estacionalidad, en este caso la misma hora del día anterior.

Como podemos observar en la siguiente representación obtenemos una predicción bastante buena.

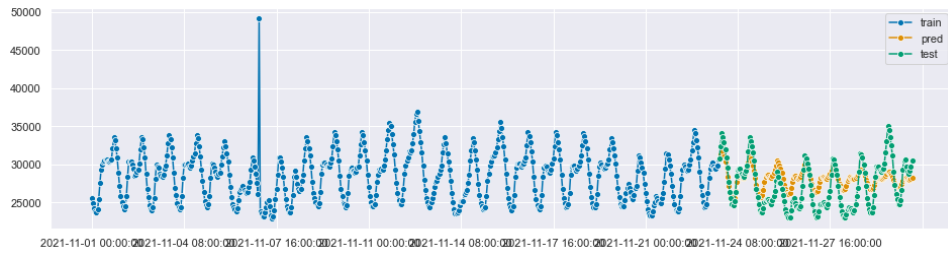


(e) NAIVE datos horarios.

### 2.1.2. ARIMA

Comenzamos probando con un modelo ARIMA, como la serie es estacionaria no realizamos en principio ninguna transformación en los datos, pero los resultados no fueron del todo buenos y decidimos finalmente tomar logaritmos para obtener una serie más estacionaria y por tanto, en este caso, una predicción más precisa.

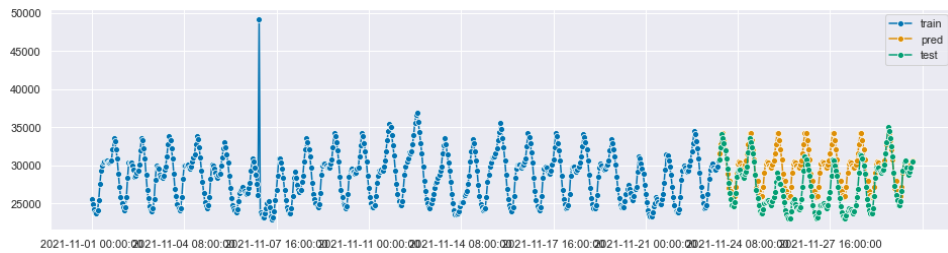
Tras realizar el autoarima para encontrar el mejor modelo, obtenemos un SARIMAX (Seasonal ARIMA)  $(1, 0, 1) \times (2, 0, [1, 2], 24)$ , es decir, con auto-regresivo de orden 1, la serie depende del último retardo y media móvil de orden 1, es decir, la serie depende del error ocurrido en el retardo anterior, la periodicidad es 24 porque tenemos datos horarios.



(f) AUTOARIMA datos horarios.

### 2.1.3. ETS

Probamos ahora con un modelo ETS, es decir un modelo de suavizado exponencial. Tras realizar el autoETS obtenemos un ETS(MNM), es decir, un modelo con Error y estacionalidad multiplicativos y sin tendencia.

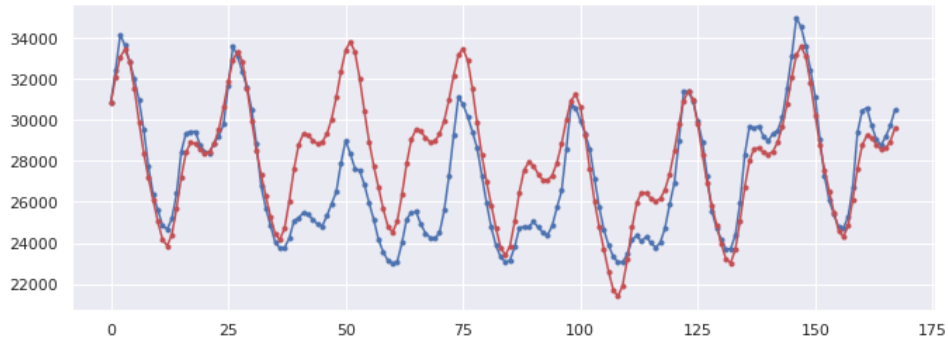


(g) AUTOETS datos horarios.

### 2.1.4. Prophet

Probamos también con el paquete Prophet [2] que es un algoritmo creado por Facebook para series temporales.

Probamos el modelo base sin realizar búsqueda de hiperparámetros, obtenemos a simple vista unas predicciones muy buenas, pero no profundizamos más en la búsqueda de los hiperparámetros debido al gran coste computacional.



(h) Prophet horarias.

### 2.1.5. Comparación modelos

Para decidir cuál de nuestros modelos realiza mejores predicciones y así elegir el mejor modelo para obtener la predicción para los primeros 7 días de diciembre calculamos los errores que cometen.

Calculamos el error cuadrático medio, ya que las unidades son las mismas y este error no nos da en este caso ningún problema y también el MAPE es decir el error absoluto medio en porcentaje.

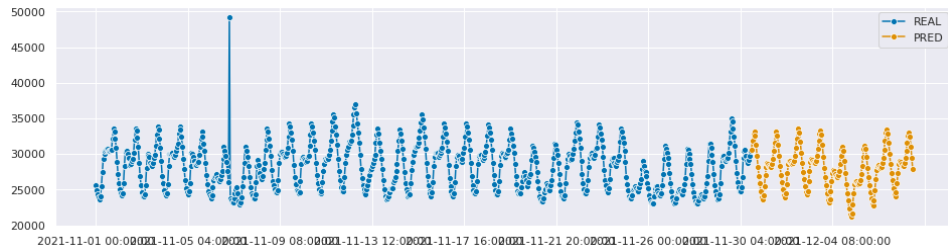
El modelo que comete menor error y por tanto mejores predicciones es el ARIMA.

Modelo	RMSE	MAPE
NAIVE	2888,6755	0,0773
ARIMA	2437,6322	0.0765
ETS	3189,2504	0,0926
PROPHET	2056,3165	0,0556



### 2.1.6. Predicciones

Debido a que es el modelo que menor error comete, realizamos las predicciones con el modelo Prophet explicado anteriormente, obteniendo los siguientes resultados:



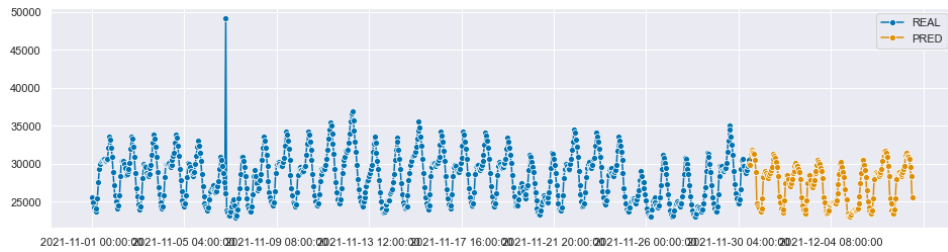
(i) Representación gráfica predicción.

Mostramos a continuación las predicciones correspondientes a las primeras horas del 1 de diciembre y parece que son unas predicciones bastante buenas.

Date	Demand
2021-12-01 00:00:00	26937.6307
2021-12-01 01:00:00	25855.0534
2021-12-01 02:00:00	24804.5616
2021-12-01 03:00:00	23935.3261
2021-12-01 04:00:00	23625.4823
2021-12-01 05:00:00	24167.1539
2021-12-01 06:00:00	25458.5025
2021-12-01 07:00:00	26993.087
2021-12-01 08:00:00	28171.3648
2021-12-01 09:00:00	28675.1868

### 2.1.7. Anexo: Librería AutoTS. Predicciones horarias diciembre

Posteriormente a la realización de este informe, descubrí la librería AutoTS [3], que consiste en un paquete de series temporales para Python, diseñado para realizar predicciones con alta precisión. Incluye modelos Naive, estadísticos, de machine learning y de deep learning. Utilizo este paquete para realizar las mismas predicciones obteniendo los siguientes resultados:



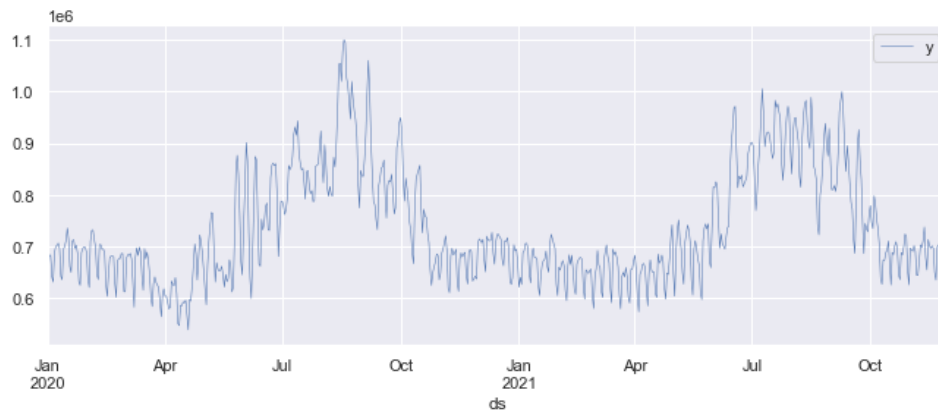
(j) Representación gráfica predicción.

Date	Demand
2021-12-01 00:00:00	24717.7586
2021-12-01 01:00:00	24353.5816
2021-12-01 02:00:00	24141.6505
2021-12-01 03:00:00	23743.1554
2021-12-01 04:00:00	23656.0608
2021-12-01 05:00:00	24166.2612
2021-12-01 06:00:00	25475.175
2021-12-01 07:00:00	28203.82
2021-12-01 08:00:00	28984.3911
2021-12-01 09:00:00	29028.1751

El modelo seleccionado en este caso ha sido un DatepartRegression model, y las predicciones son bastante similares a las obtenidas con el modelo Prophet.

## 2.2. Datos diarios

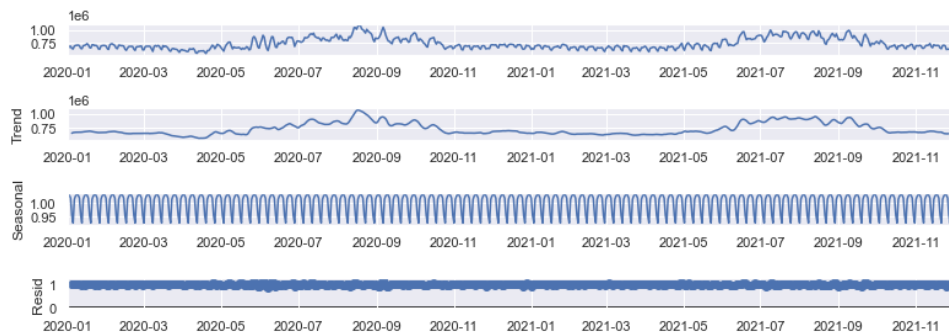
Tomamos ahora los datos diarios, en este caso, decido tomar únicamente los datos de los dos últimos años, es decir desde el 1 de enero de 2020.



(k) Datos diarios 2020-2021.

Observamos que, como habíamos analizado con los datos horarios la serie es estacional; se observa que la demanda aumenta en los meses de verano y disminuye en el resto.

Vamos a observar más en detalle la descomposición de las componentes:

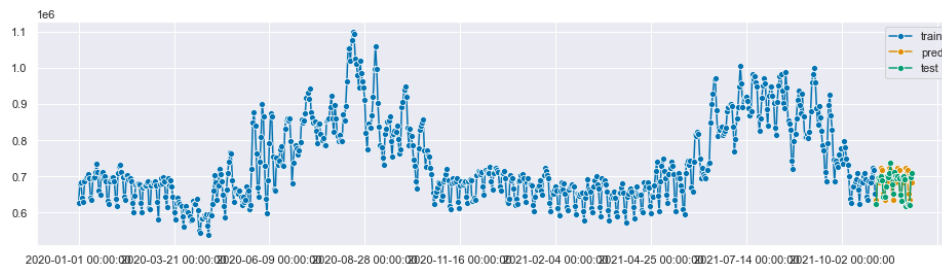


(l) Patrones componentes 2020-2021.

Observamos que la serie no tiene tendencia y que es estacional. Realizamos la prueba de Dickey fuller para confirmar que la serie es estacionaria.

### 2.2.1. Modelo a batir: Naïve estacional

Como nuestros datos son estacionales utilizamos también en este caso como modelo de referencia el método ingenuo o Naïve estacional. Como podemos observar en la siguiente representación obtenemos una predicción bastante buena.

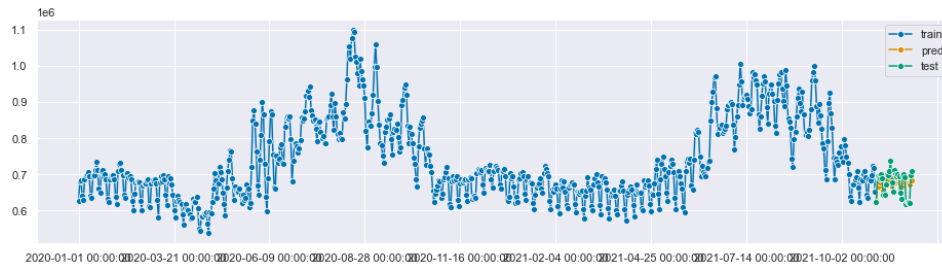


(m) NAIVE datos diarios.

### 2.2.2. ARIMA

Comenzamos probando con un modelo ARIMA; decidimos hacerlo tanto con los datos originales como transformándolos a logaritmo, ya que en el caso de los datos horarios esta transformación mejoró nuestras predicciones. En este caso obtenemos mejores resultados con los datos originales.

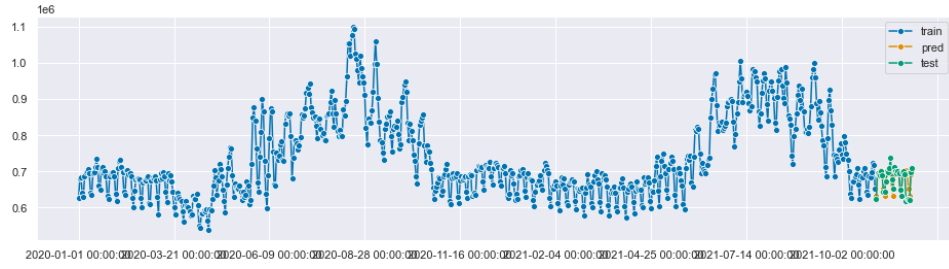
Realizando un autoarima, con los datos originales, obtenemos un SARIMAX(1, 1, 2)x(2, 0, [], 7), es decir, con autoregresivo de orden 1 (la serie depende del último retardo) y media móvil 2, es decir, la serie depende del error ocurrido en los últimos 2 retardos, la periodicidad es 7 porque tenemos datos diarios.



(n) AUTOARIMA datos diarios.

### 2.2.3. ETS

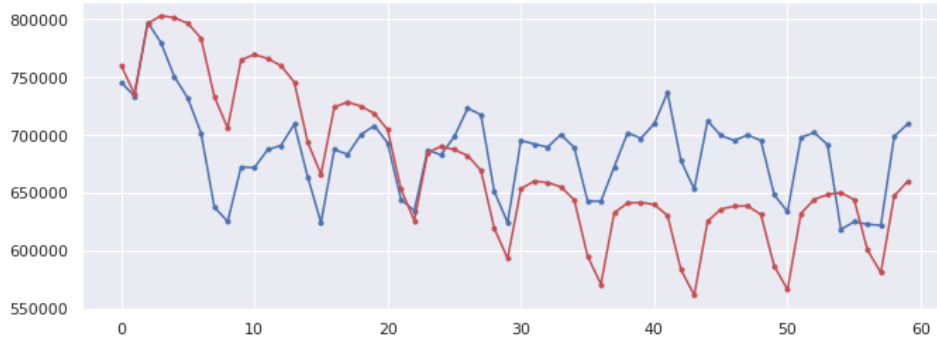
Probamos ahora con un modelo ETS, es decir un modelo de suavizado exponencial. Tras realizar el autoETS obtenemos un ETS(MNM), es decir, un modelo con Error y estacionalidad multiplicativos y sin tendencia.



(ñ) AUTOETS datos diarios.

#### 2.2.4. Prophet

Al igual que en el caso de las predicciones horarias probamos el modelo base sin realizar búsqueda de hiperparámetros y en este caso observamos que las predicciones son peores que en los otros modelos, por tanto, no profundizamos más en buscar los hiperparámetros.



(o) Prophet diarias.

### 2.2.5. Comparación modelos

Para decidir cuál de nuestros modelos realiza mejores predicciones y así poder elegir el mejor modelo para obtener la predicción para los primeros 15 días de diciembre, calculamos los errores de los modelos.

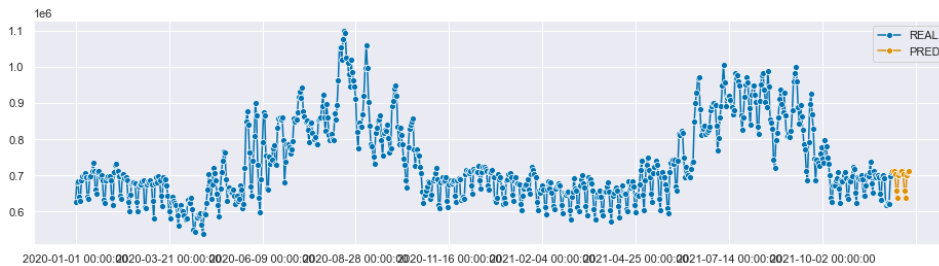
Calculamos el error cuadrático medio, ya que las unidades son las mismas y este error no nos da en este caso ningún problema y también el MAPE es decir el error absoluto medio en porcentaje.

El modelo que comete menor error y por tanto mejores predicciones, es en este caso el ETS.

Modelo	RMSE	MAPE
NAIVE	29984,3047	0,0297
ARIMA	28831,4736	0,0329
ETS	23951,0247	0,0218
PROPHET	54769,0651	0,0707

### 2.2.6. Predicciones

Debido a que es el modelo que menor error comete realizamos las predicciones con el modelo ETS explicado anteriormente, obteniendo los siguientes resultados, mostramos unicamente los datos de 2021 para que se aprecie mejor la predicción.



(p) Representación gráfica predicción.

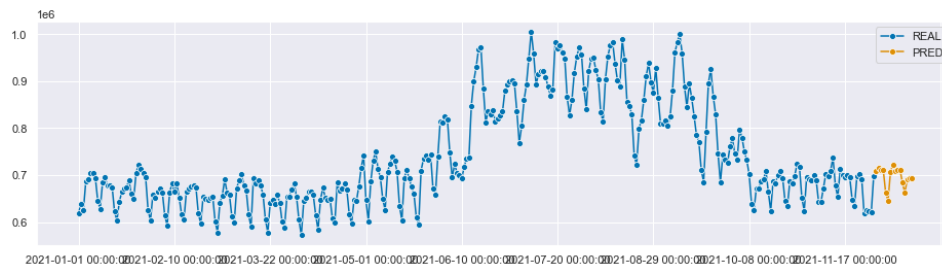
Mostramos a continuación los datos obtenidos.

Date	Demand
2021-12-01	711712.7248
2021-12-02	711035.2463
2021-12-03	703271.5141
2021-12-04	659341.0542
2021-12-05	638934.3987
2021-12-06	699973.5031
2021-12-07	709567.8965
2021-12-08	711712.7248
2021-12-09	711035.2463
2021-12-10	703271.5141
2021-12-11	659341.0542
2021-12-12	638934.3987
2021-12-13	699973.5031
2021-12-14	709567.8965
2021-12-15	711712.7248



## 2.2.7. Anexo: Librería AutoTS. Predicciones Diarias diciembre

Al igual que en el caso horario introduzco este paquete para realizar las predicciones obteniendo los siguientes resultados:



(q) Representación gráfica predicción.

Date	Demand
2021-12-01	714629.7371
2021-12-02	710998.0236
2021-12-03	710847.8027
2021-12-04	662492.9659
2021-12-05	645994.3133
2021-12-06	706500.8694
2021-12-07	722724.1037
2021-12-08	709291.6513
2021-12-09	710370.0071
2021-12-10	711130.3477
2021-12-11	684279.3292
2021-12-12	662477.3471
2021-12-13	690664.7879
2021-12-14	694536.567
2021-12-15	694642.432

Obtenemos predicciones que parecen bastante buenas a simple vista, el modelo seleccionado en este caso ha sido el MultivariateMotif.

## Capítulo 3

# Conclusiones

Tras realizar el estudio tanto de los datos horarios como de los datos diarios podemos concluir que hemos conseguido unas predicciones bastante buenas.

En resumen, nos quedaremos en el caso de los datos horarios con las predicciones obtenidas con el modelo Prophet, y en el caso de los datos diarios, con el modelo ETS.

En ambos casos, las predicciones son muy similares a las obtenidas utilizando el paquete Autots; pero debido a que es un paquete muy reciente y aún hay poca documentación, decidimos quedarnos con los modelos vistos en clase.

# Bibliografía

- [1] Prueba Dickey Fuller

Ejemplos de aplicación prueba de Dickey Fuller <https://quantspace.es/2020/08/01/analisis-de-series-temporales-con-python-parte-2/>

<https://medium.com/datos-y-ciencia/introducci%C3%B3n-al-an%C3%A1lisis-de-series-cronol%C3%B3gicas-con-python-y-pandas-99fc8d4bb56d#:~:text=Se%20puede%20decir%20que%20una,son%20constantes%20en%20el%20tiempo%20.>

- [2] Predicciones utilizando Prophet

<https://facebook.github.io/prophet/#:~:text=Prophet%20is%20a%20forecasting%20procedure%20implemented%20in%20R%20and%20Python.&text=Prophet%20is%20a%20procedure%20for,daily%20seasonality%2C%20plus%20holiday%20effects.>

Caso particular aplicado a la energía:

[https://github.com/abhinav-codes/EnergyPredictions/blob/master/EnergyPrediction\\_v2\\_fbprophet.ipynb](https://github.com/abhinav-codes/EnergyPredictions/blob/master/EnergyPrediction_v2_fbprophet.ipynb)

- [3] Paquete AutoTS

<https://github.com/winedarksea/AutoTS>

- [4] Predicción de consumo energético horario

<https://towardsdatascience.com/part-2-time-series-analysis-predicting-hourly-en>