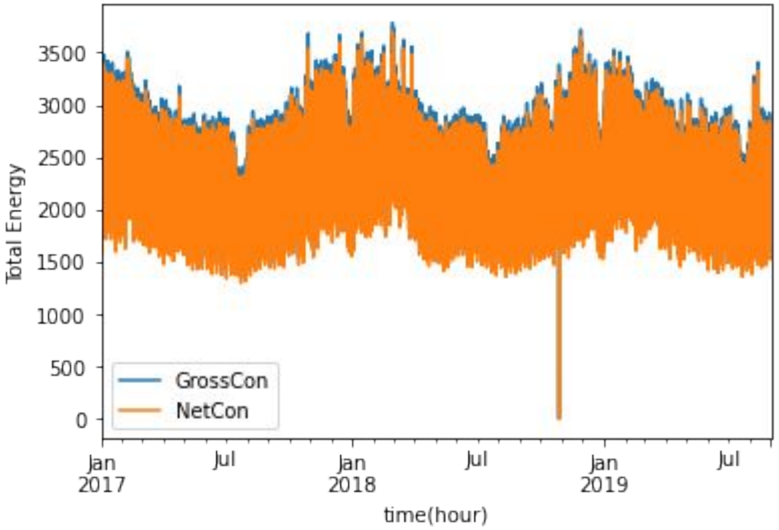
Forecasting Energy Consumption in Denmark: A Data Science Approach

Avesta Narimani
1/10/2024

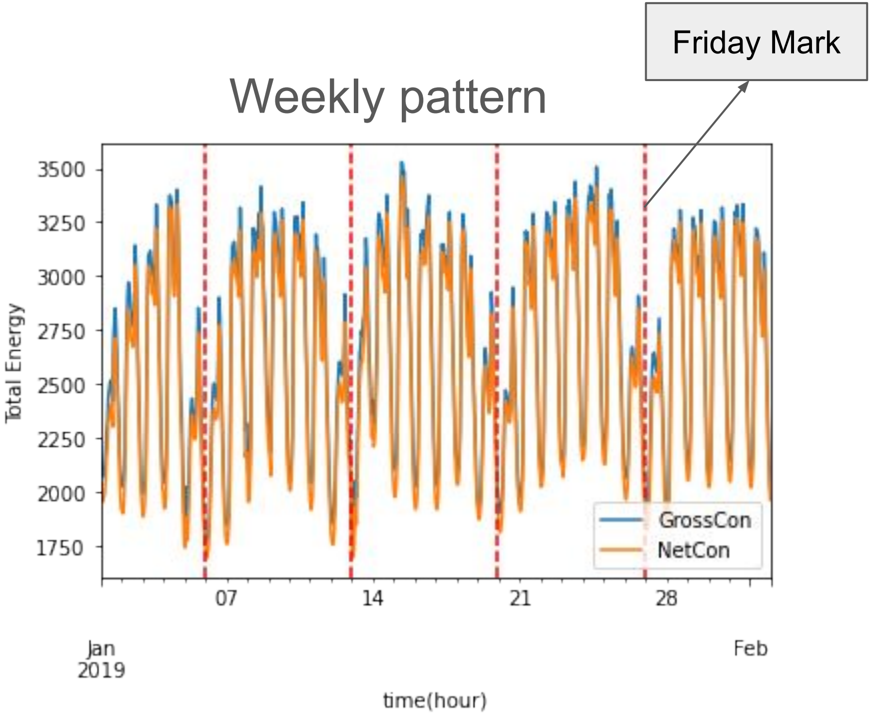
Part One: EDA

Gross Electricity Consumption vs Net Electricity Consumption:

Yearly pattern



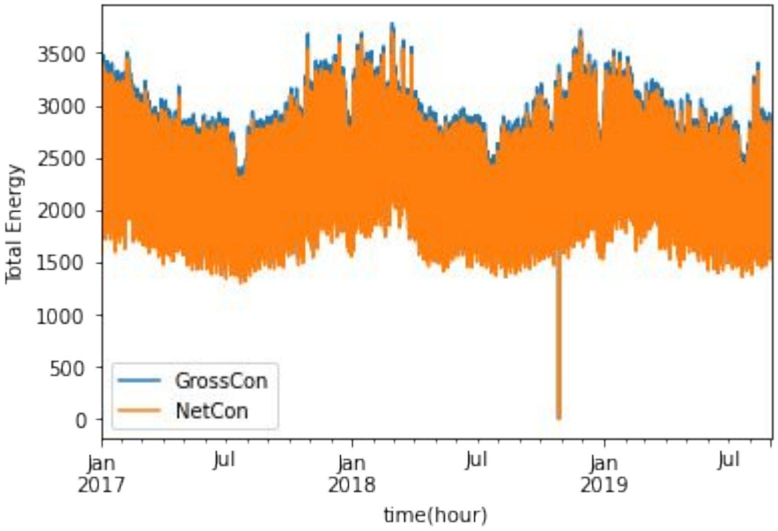
Weekly pattern



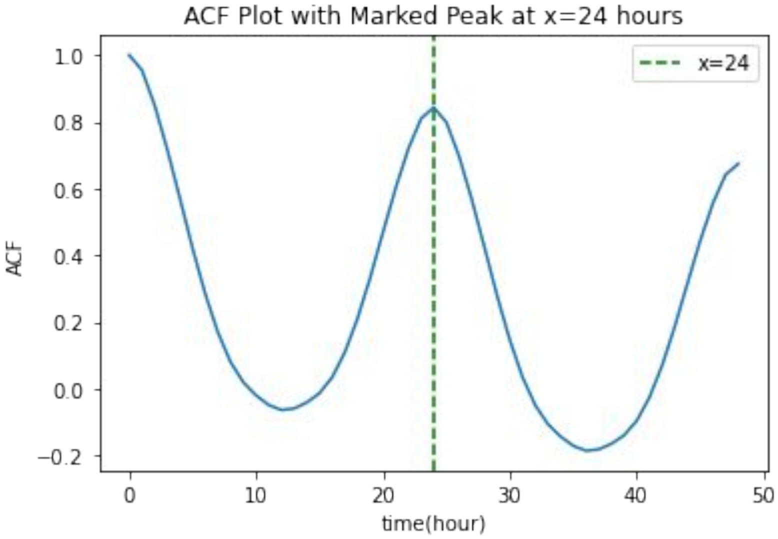
Autocorrelation function:

Gross Electricity Consumption:

Yearly pattern

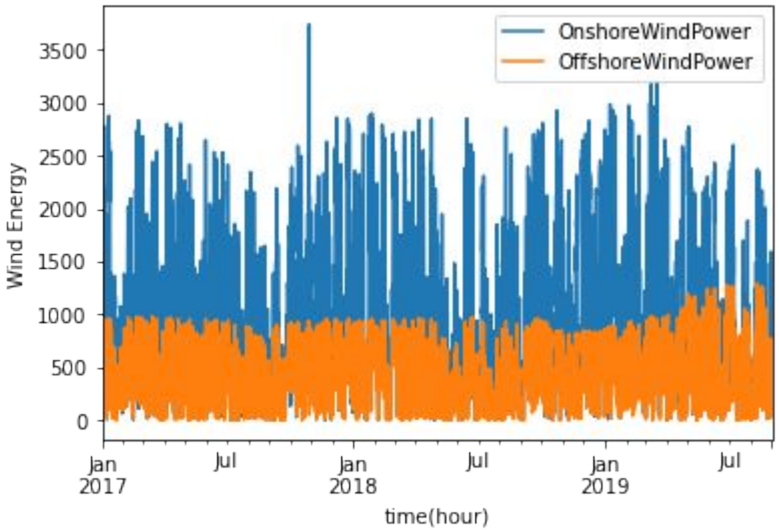


Autocorrelation function

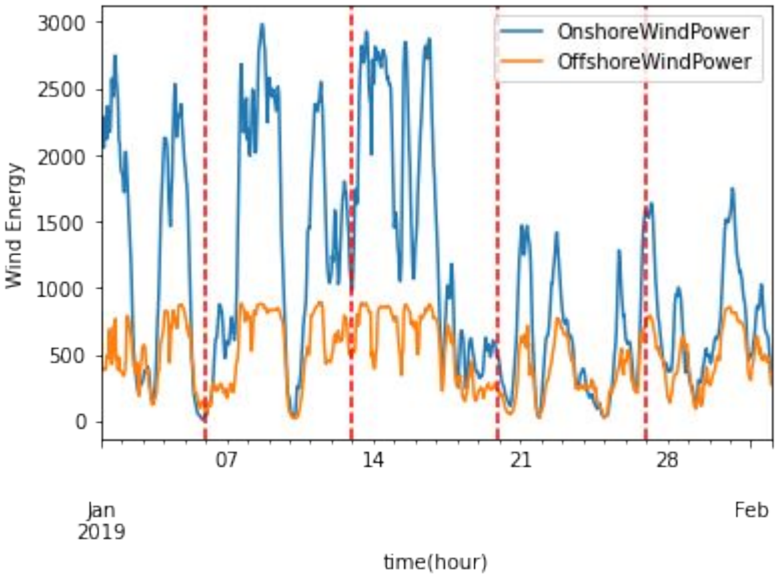


Onshore Wind Power and Offshore Wind Power:

Yearly pattern

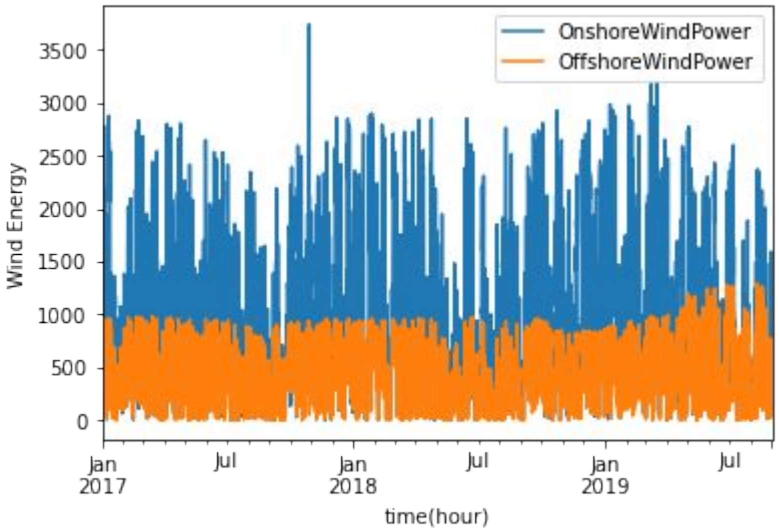


Weekly pattern

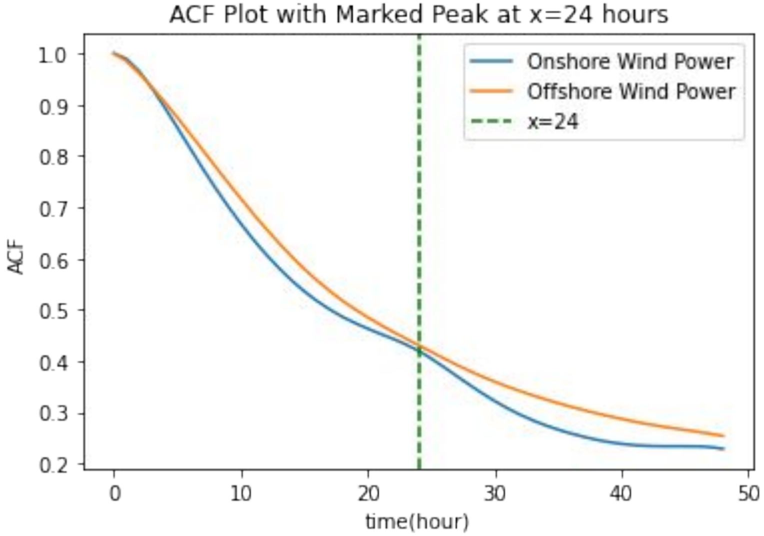


Wind power autocorrelation:

Yearly pattern

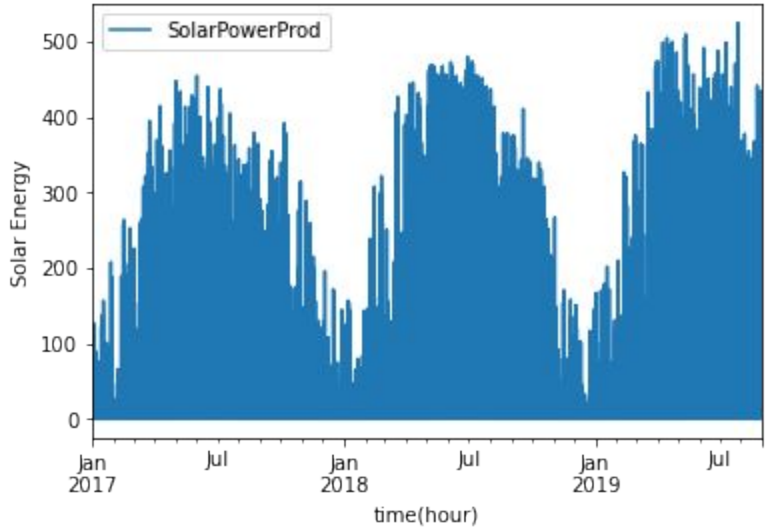


Autocorrelation function

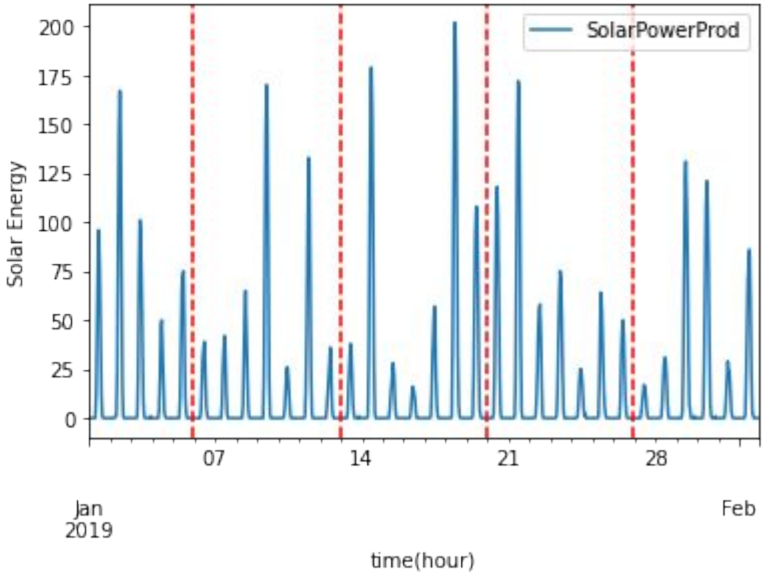


Solar power:

Yearly pattern

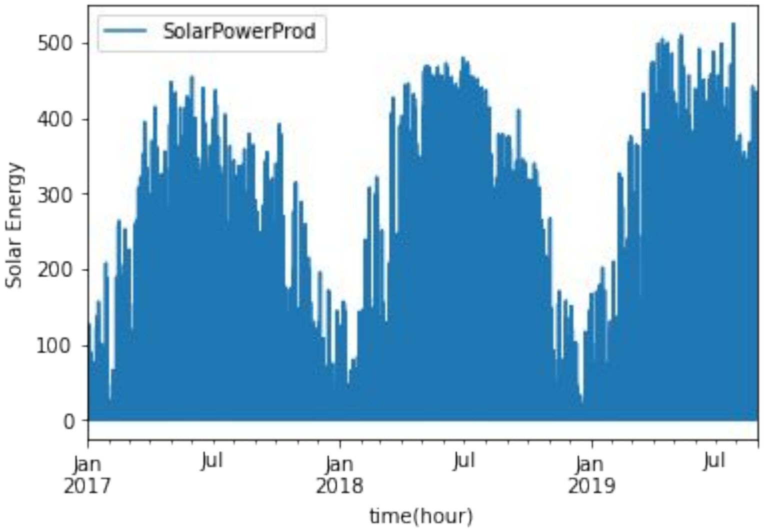


Weekly pattern

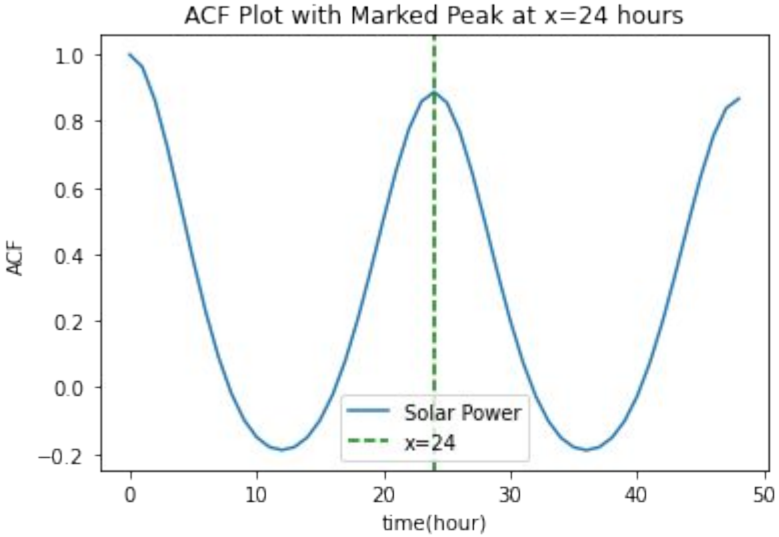


Wind power autocorrelation:

Yearly pattern

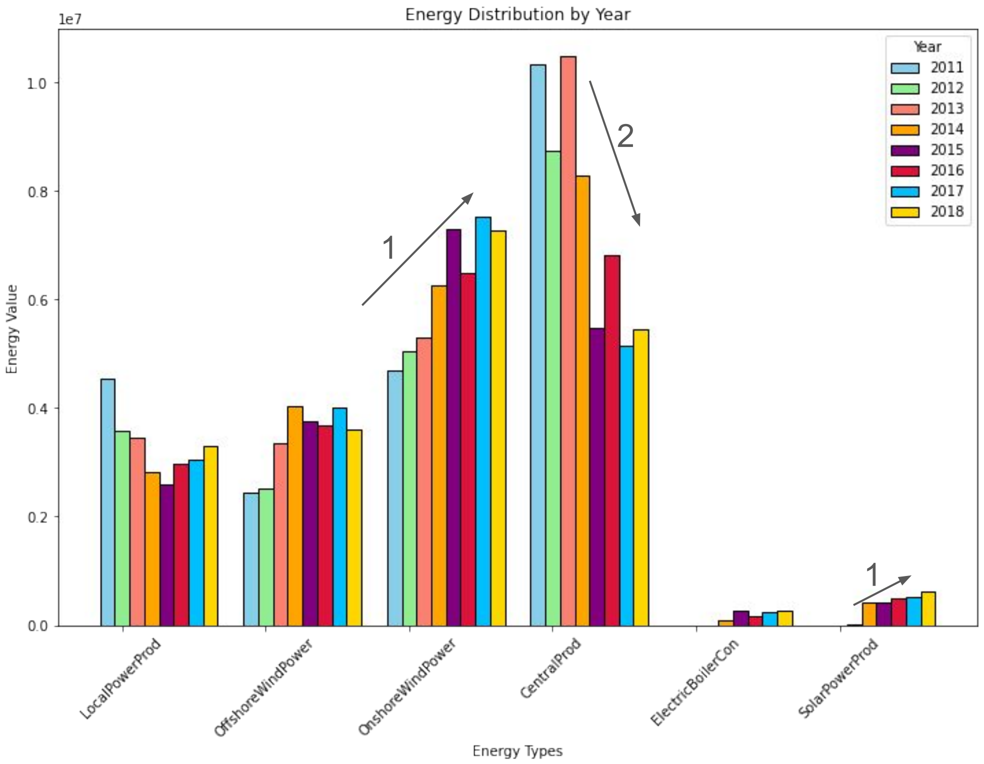


Autocorrelation function



Bar plot all energies:

1. Onshore wind power and solar energy increased over the last few years
2. Central Production decreased over the last few years



Future work:

Forecast the timeline for Denmark to achieve complete electricity generation from renewable sources and eliminate the use of fossil fuels (both locally and centrally produced).

Part Two: forecasting models

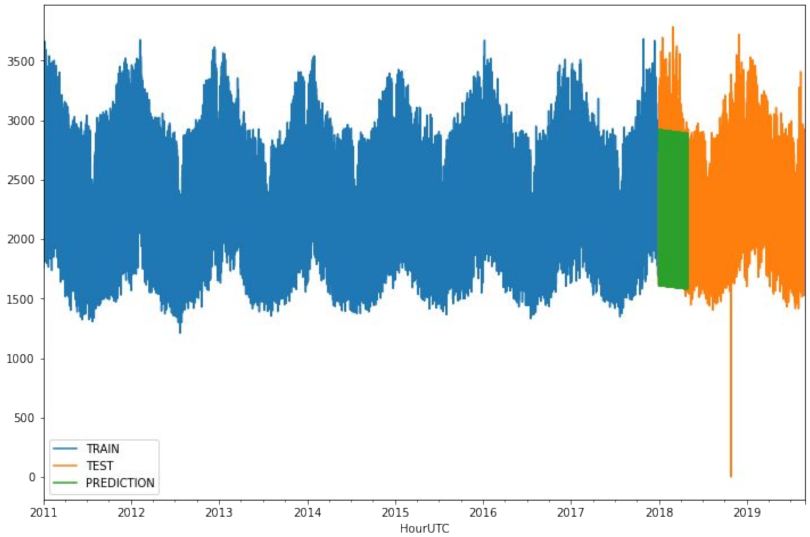
Is data stationary?

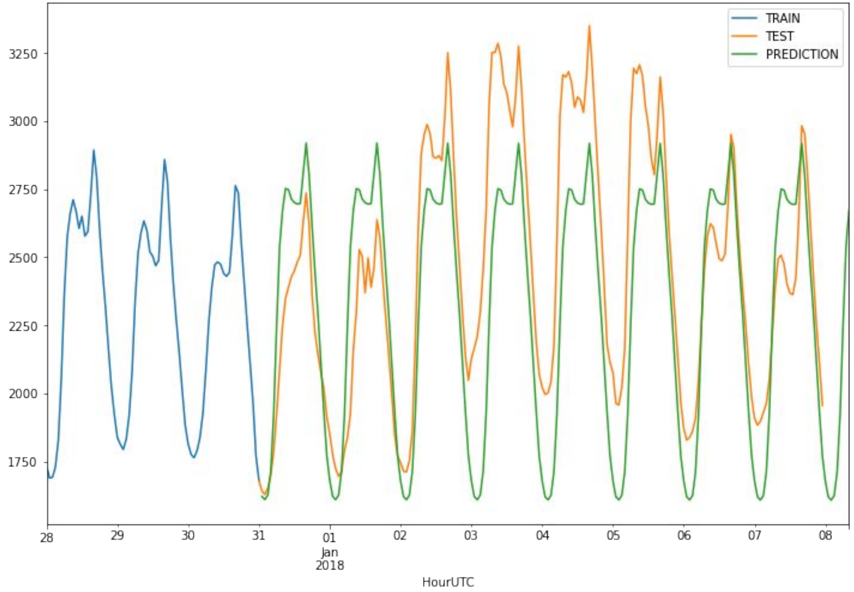
Augmented Dickey-Fuller Test

Test results:

```
Augmented Dickey-Fuller Test:
ADF test statistic          -21.634155
p-value                       0.000000
# lags used                  63.000000
# observations            75895.000000
critical value (1%)          -3.430436
critical value (5%)          -2.861578
critical value (10%)         -2.566790
Strong evidence against the null hypothesis
Reject the null hypothesis
Data has no unit root and is stationary
```

First Model: Exponential Smoothing

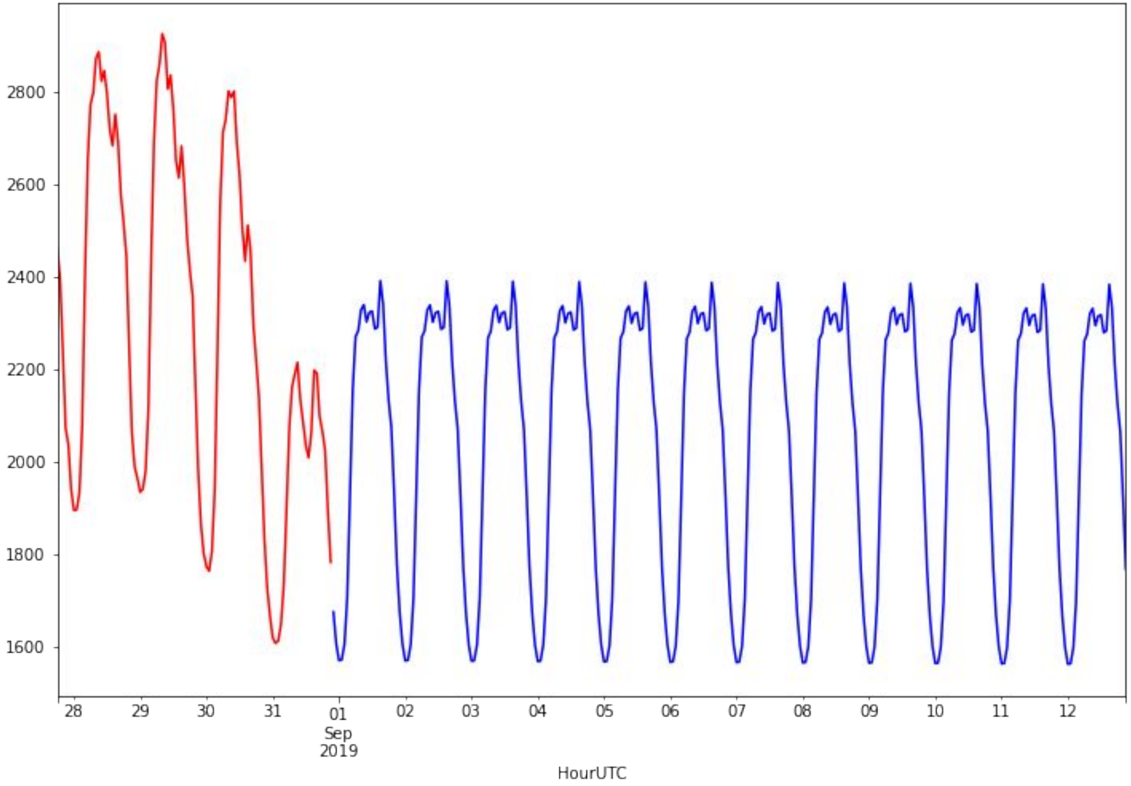




This model caters the 24 hour frequency but the amplitude does not match the expected values!

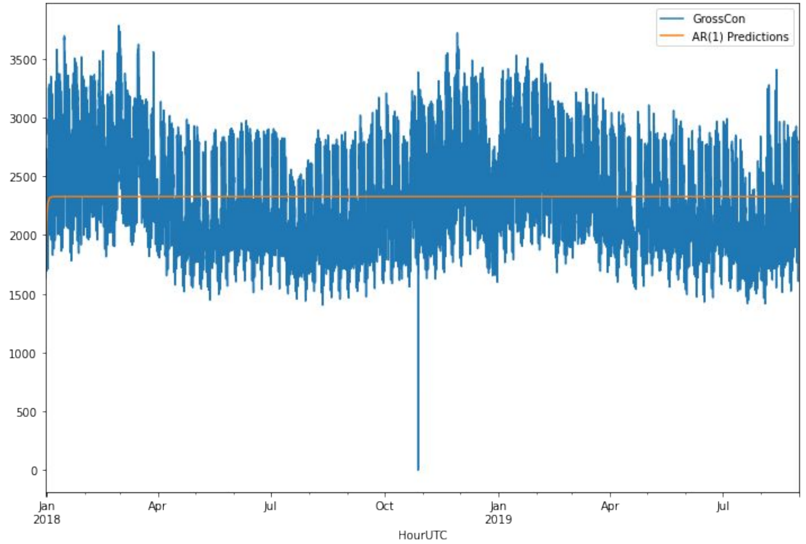
Mean Absolute Error: 379.4
Mean Squared Error: 193060.1
Root Mean Squared Error: 439.3

Forecasting into the Future using Exponential Smoothing



Second model: Autoregression (AR) Model

This model accurately predicts the average value but struggles to capture the oscillations.

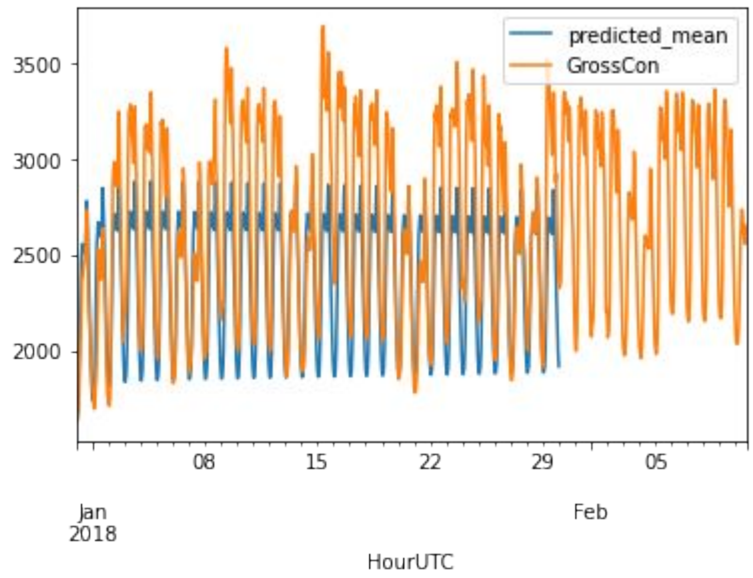


Mean Absolute Error: 390.7
Mean Squared Error: 215660.9
Root Mean Squared Error: 464.4

Third model:
Automated ARIMA Model Selection with `pmdarima`

ARIMA(1,0,2)(1,0,2)[24] is the best model.

This model successfully captures the 24-hour frequency of the data but does not accurately predict the amplitude.

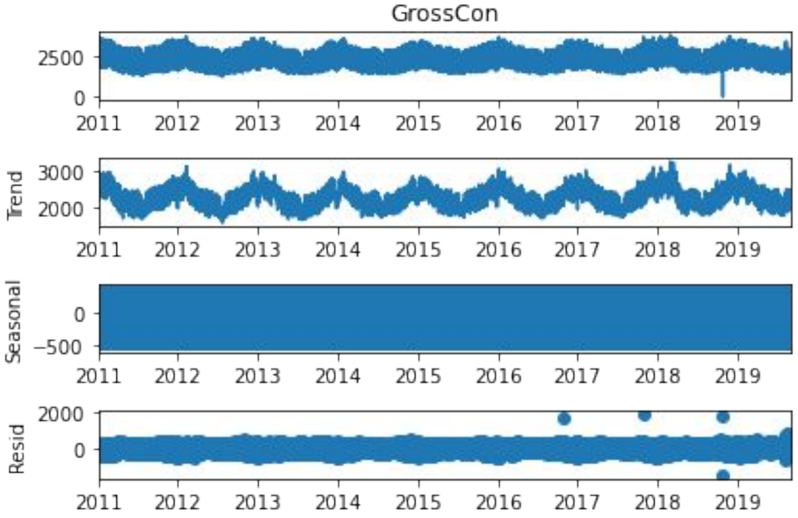


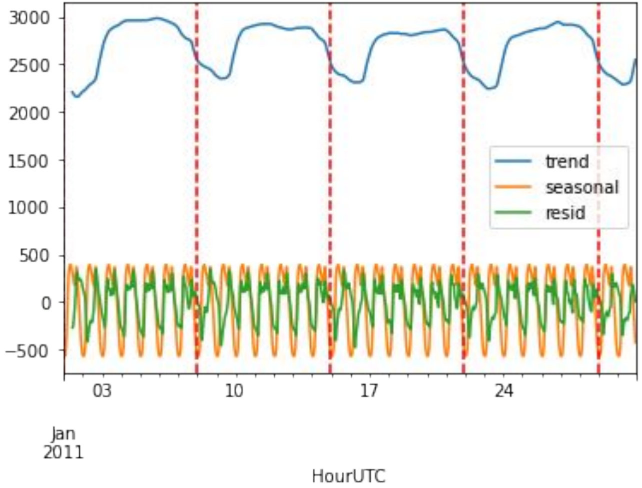
Mean Absolute Error: 317.8
Mean Squared Error: 147733.2
Root Mean Squared Error: 384.3

Fourth model:

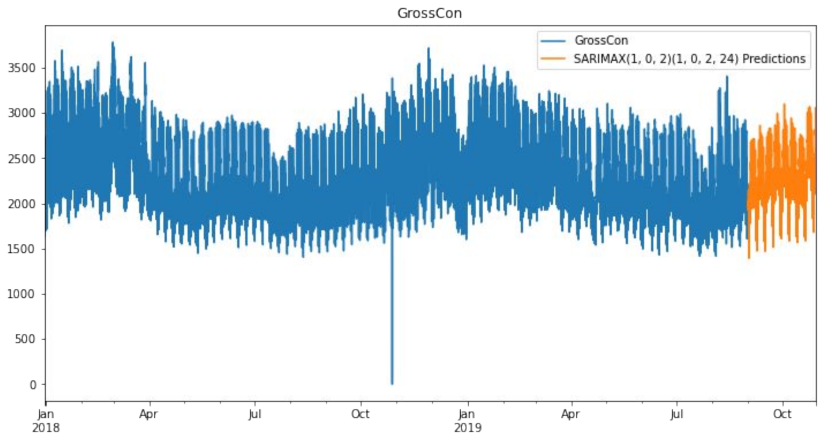
Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

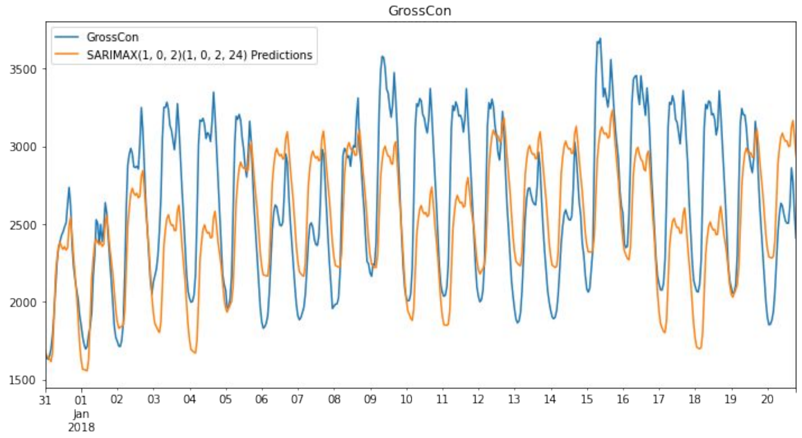
First decompose the data and use that as exogenic data in the model:





Forecasting based on test and train data:





Mean Absolute Error: 351.0
Mean Squared Error: 176078.1
Root Mean Squared Error: 419.6

Forecasting the future:

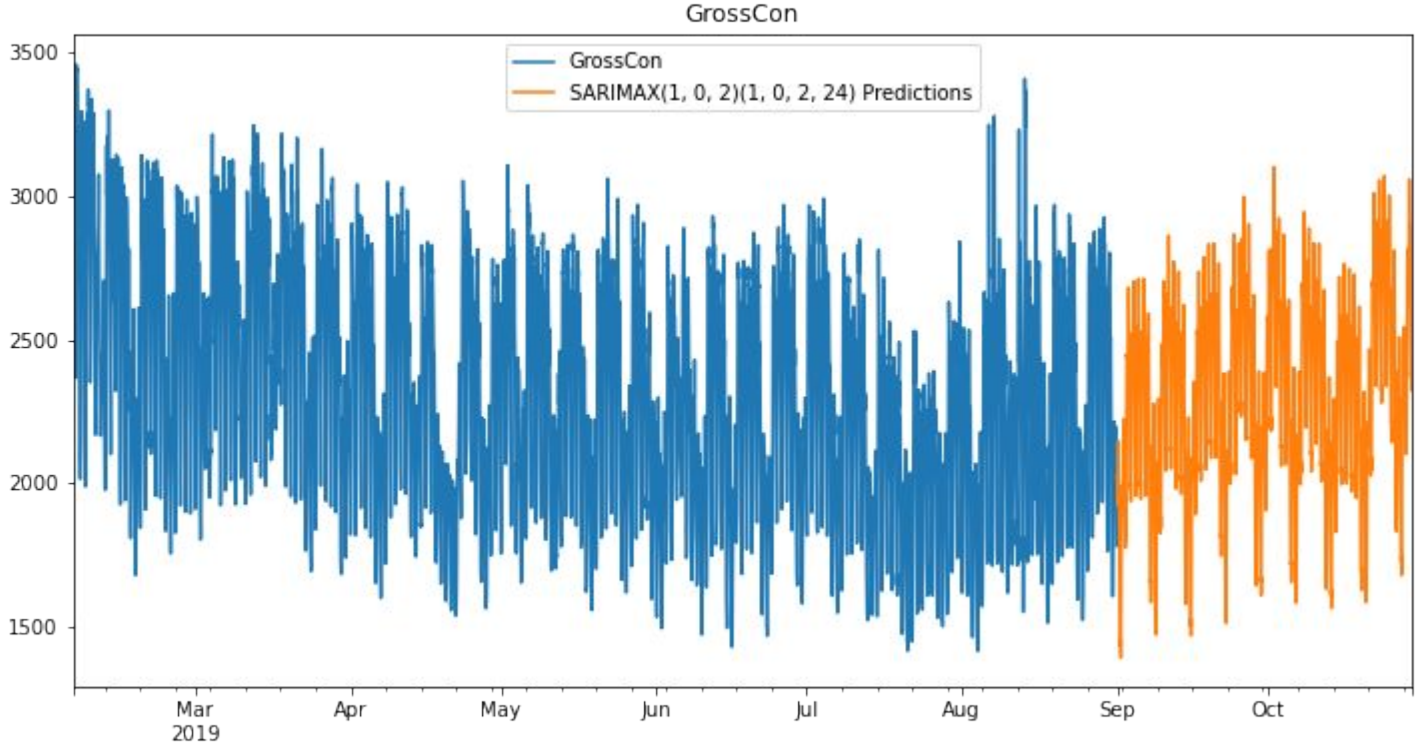


Table for different metrics for each model:

Model	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error
Exponential Smoothing	379.4	193060.1	439.4
AR	390.7	215660.9	464.4
ARIMA	317.8	147733.2	384.4
SARIMA	351.0	176078.1	419.6

Thank you!