Hourly Energy Consumption

...

Time Series Analysis

Why care?

Energy Demand vs. Energy Supply

Imbalance could lead to energy waste, inconvenience, nationwide blackouts

Energy efficiency

The Data

Collected by PJM - regional transmission organization

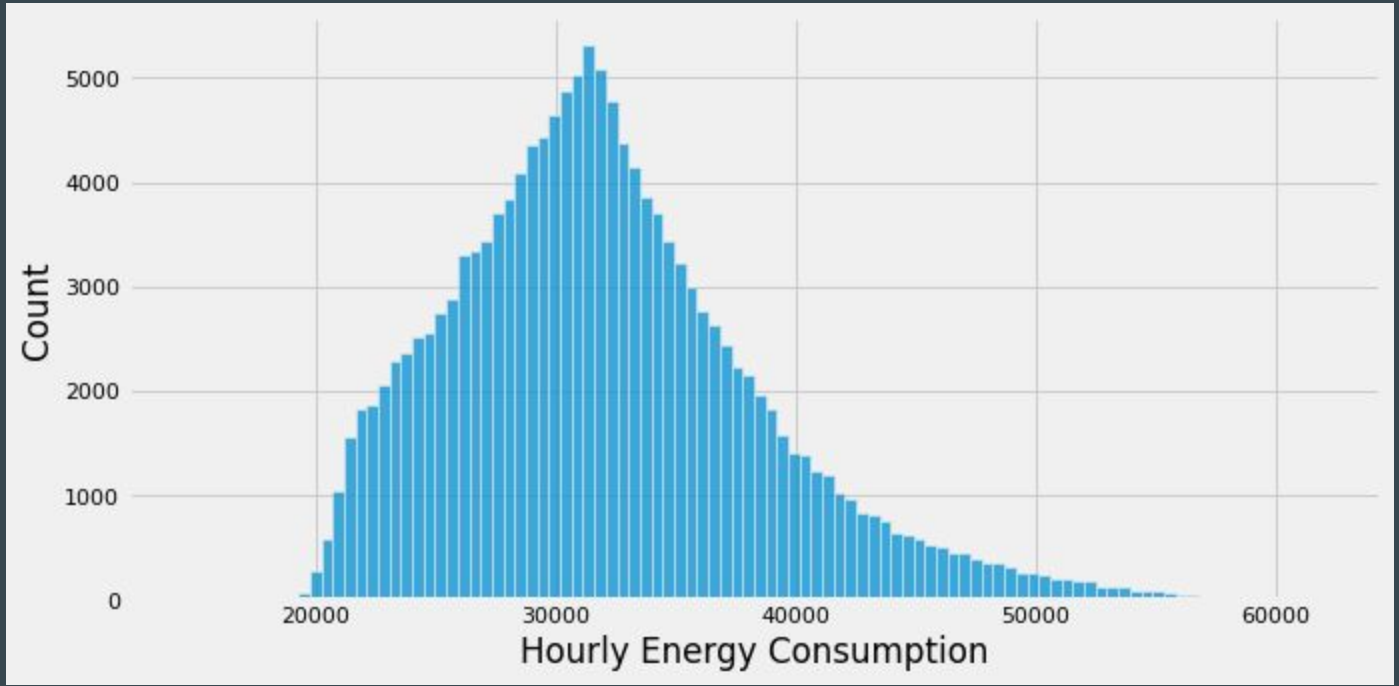
Timeframe : 2002 - 2018

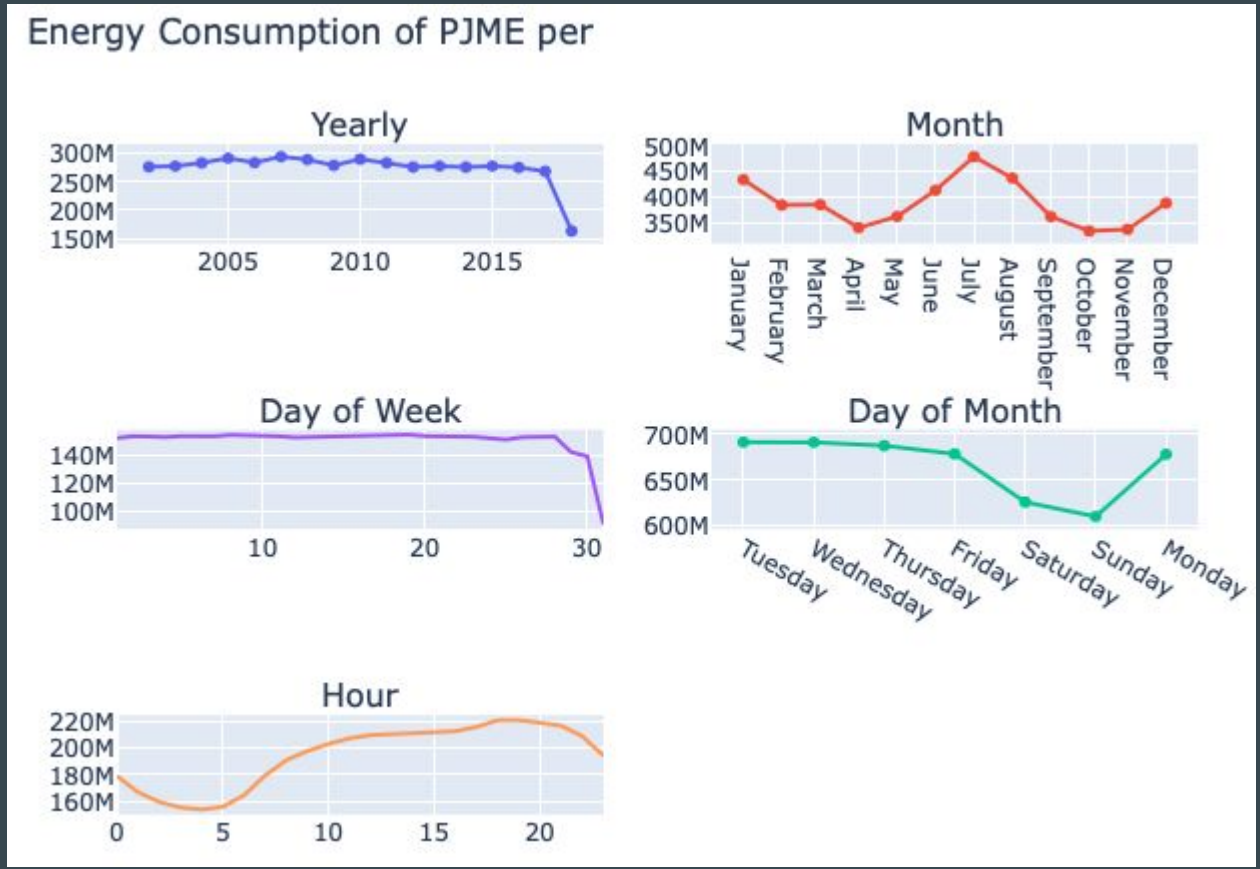
Focused on the East coast

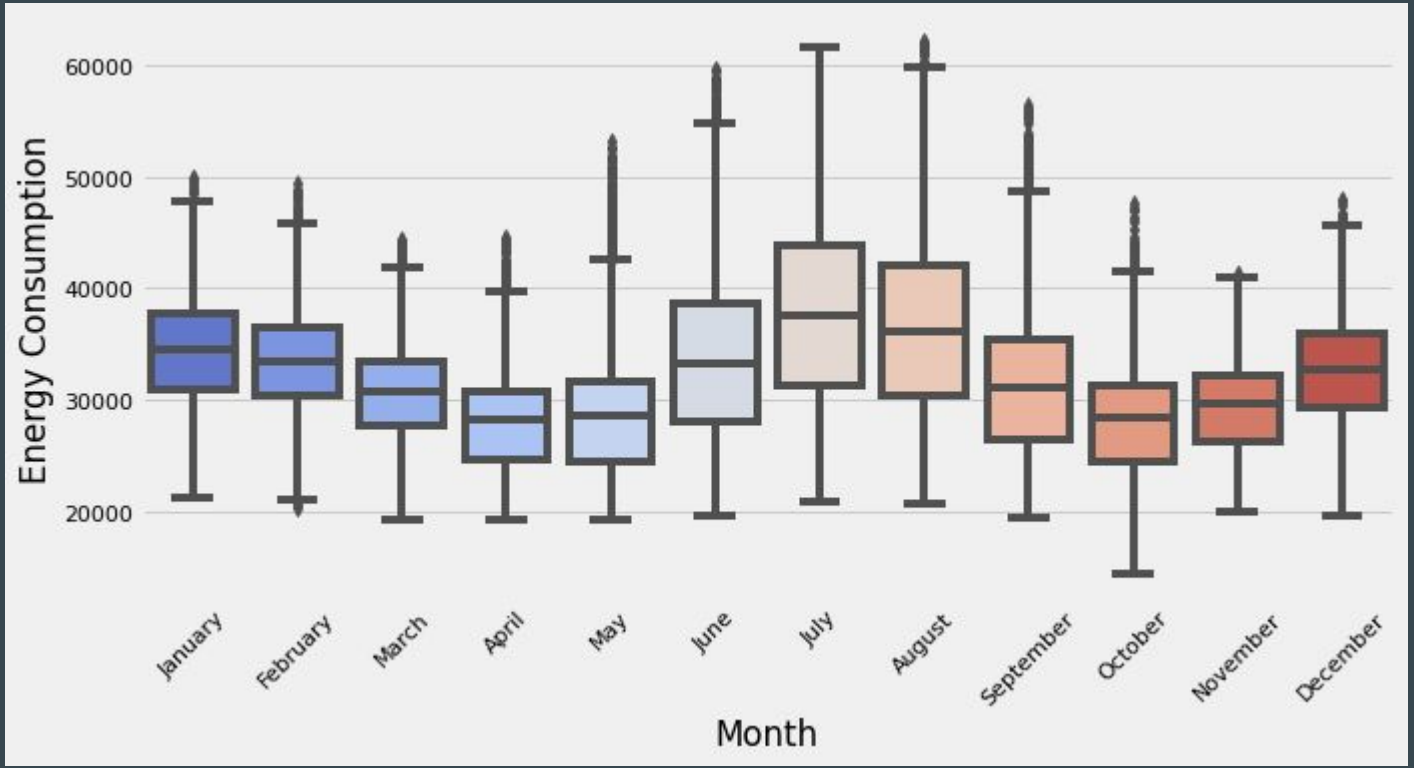
Data Preparation

We’ll skip that for the purpose of this presentation

Exploratory Data Analysis

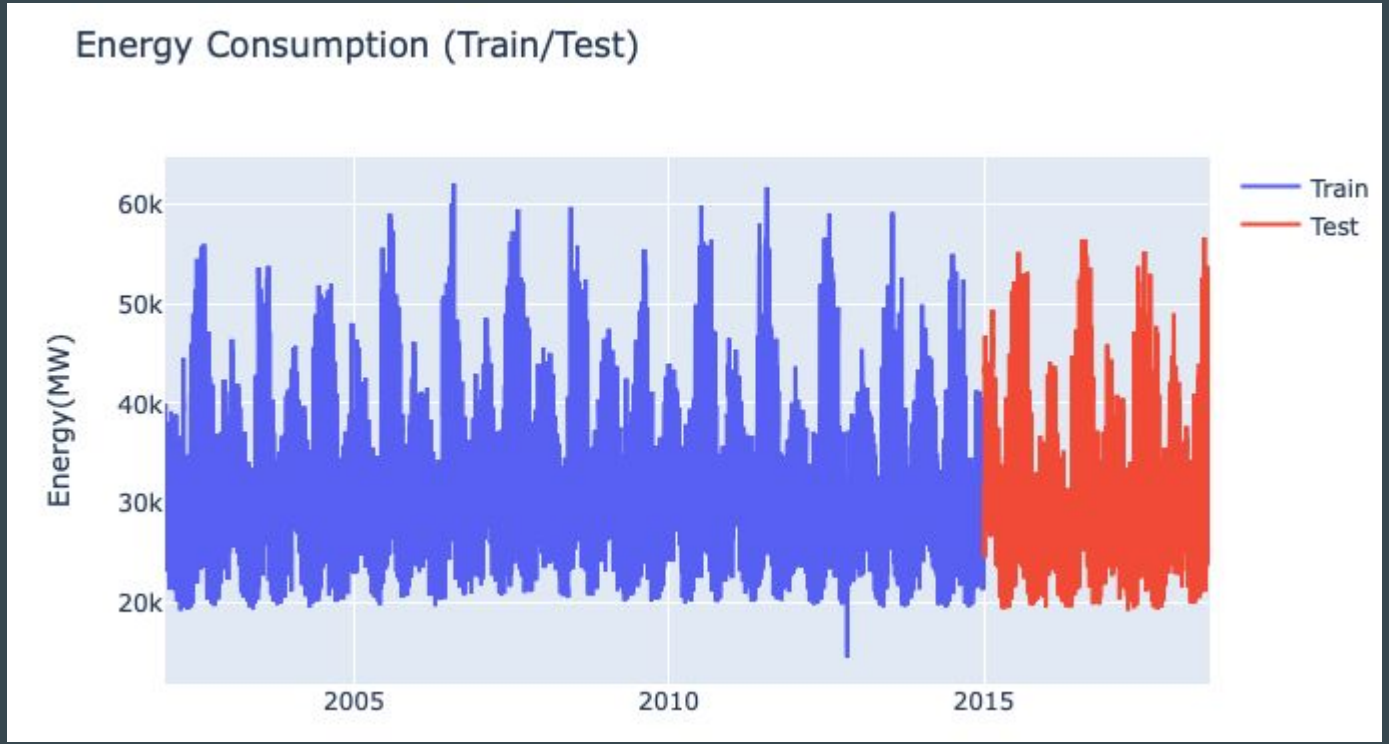

5000
4000
3000
2000
1000
0
Count
20000
30000
40000
50000
60000
Hourly Energy Consumption


Energy Consumption of PJME per
Yearly
300M
250M
200M
150M
2005
2010
2015
Month
500M
450M
400M
350M
January
February
March
April
May
June
July
August
September
October
November
December
Day of Week
140M
120M
100M
10
20
30
Day of Month
700M
650M
600M
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Monday
Hour
220M
200M
180M
160M
0
5
10
15
20

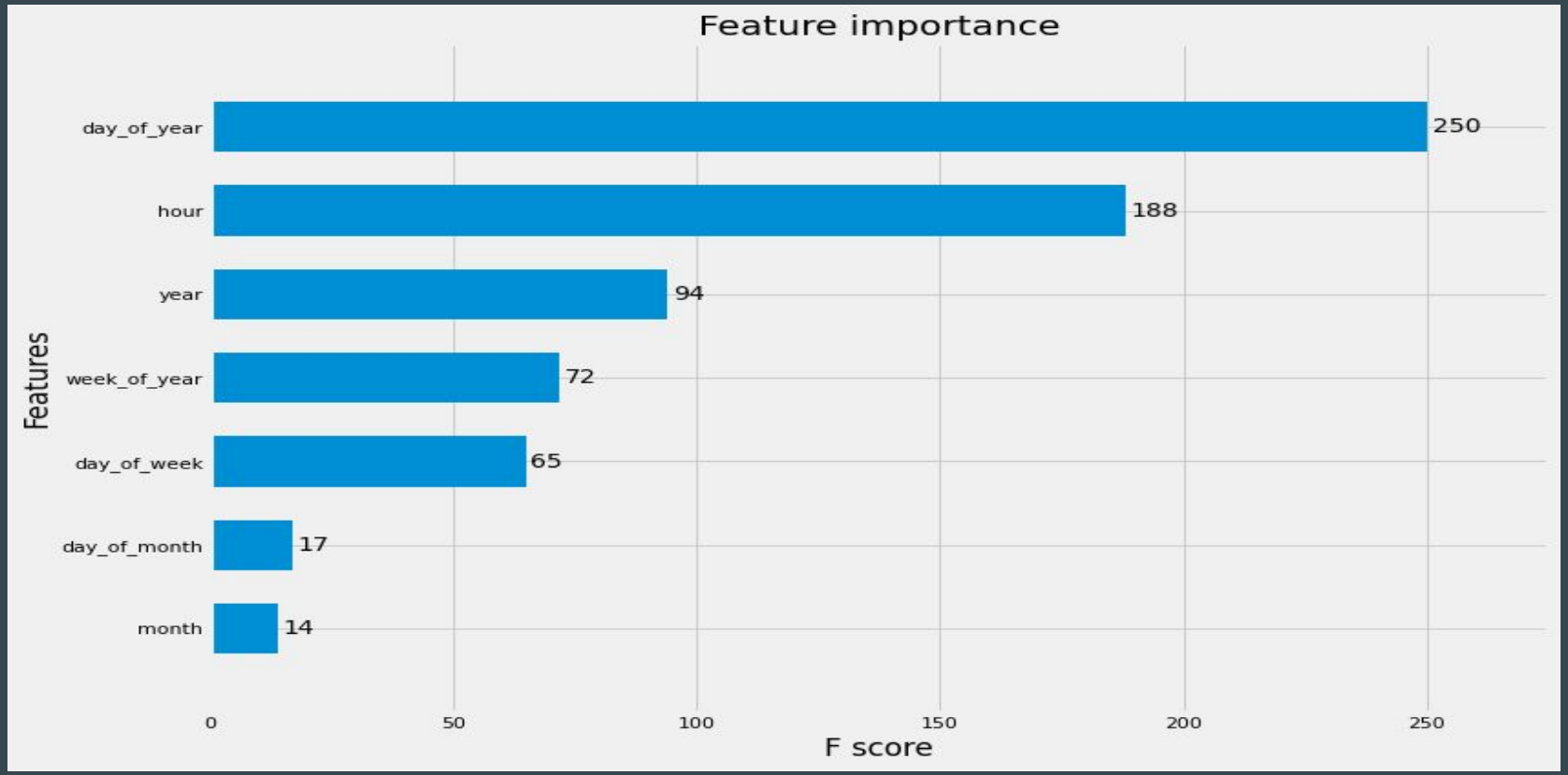

60000
50000
40000
30000
20000
Energy Consumption
January
February
March
April
May
June
July
August
September
October
November
December
Month

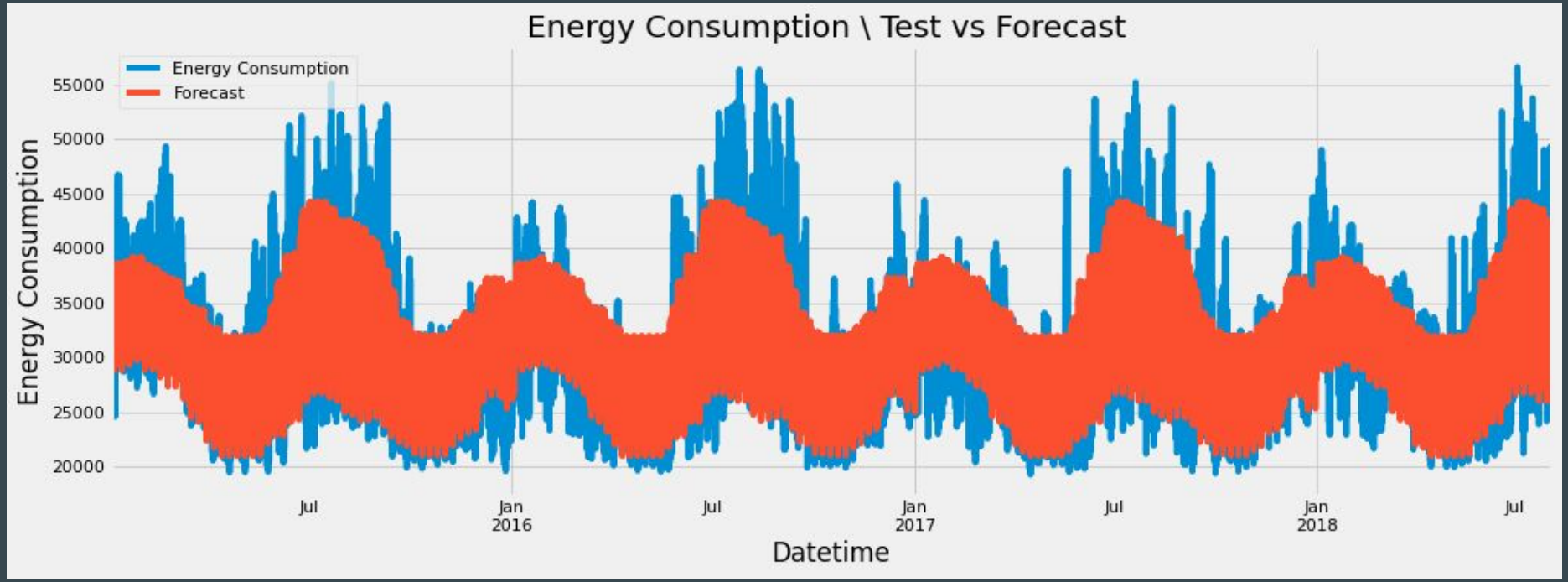
Forecasting

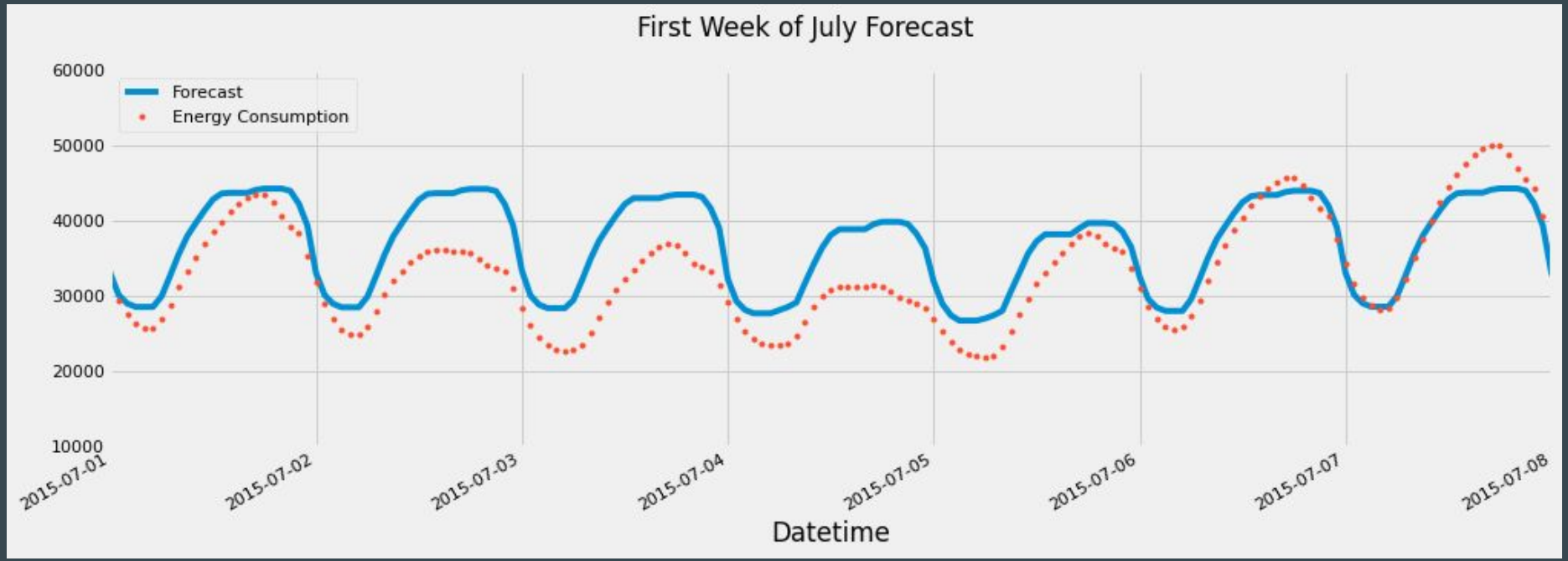
The data Test/Train



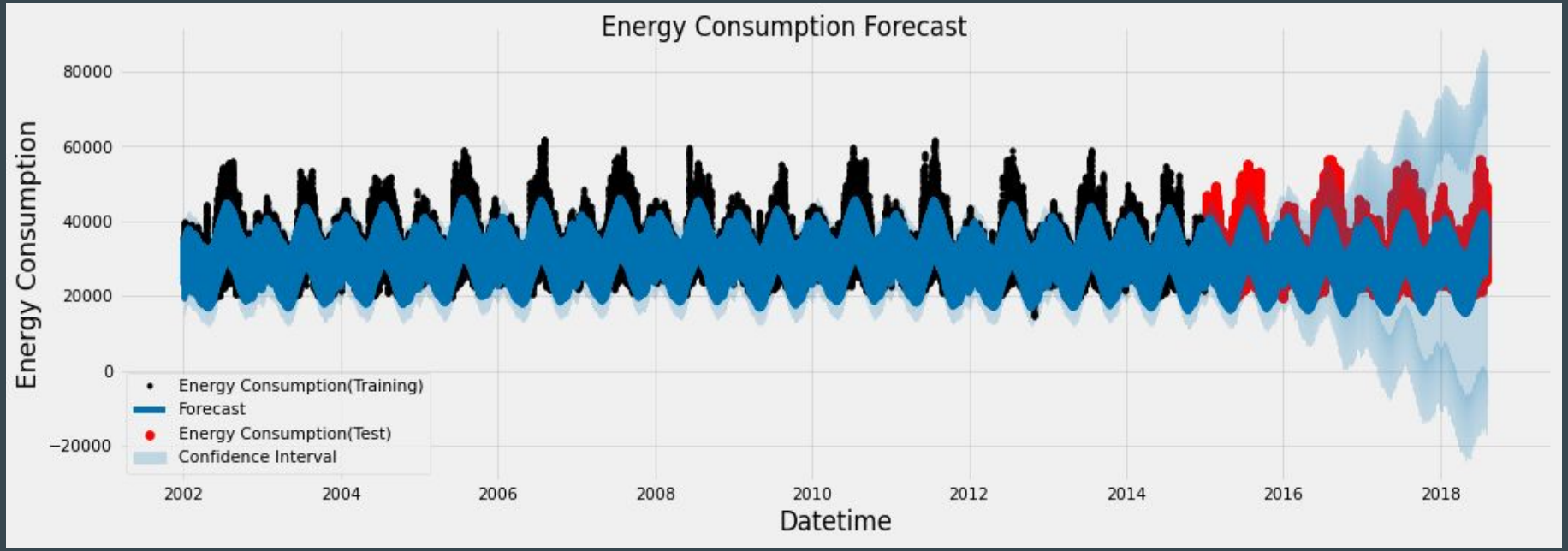
XGBoost

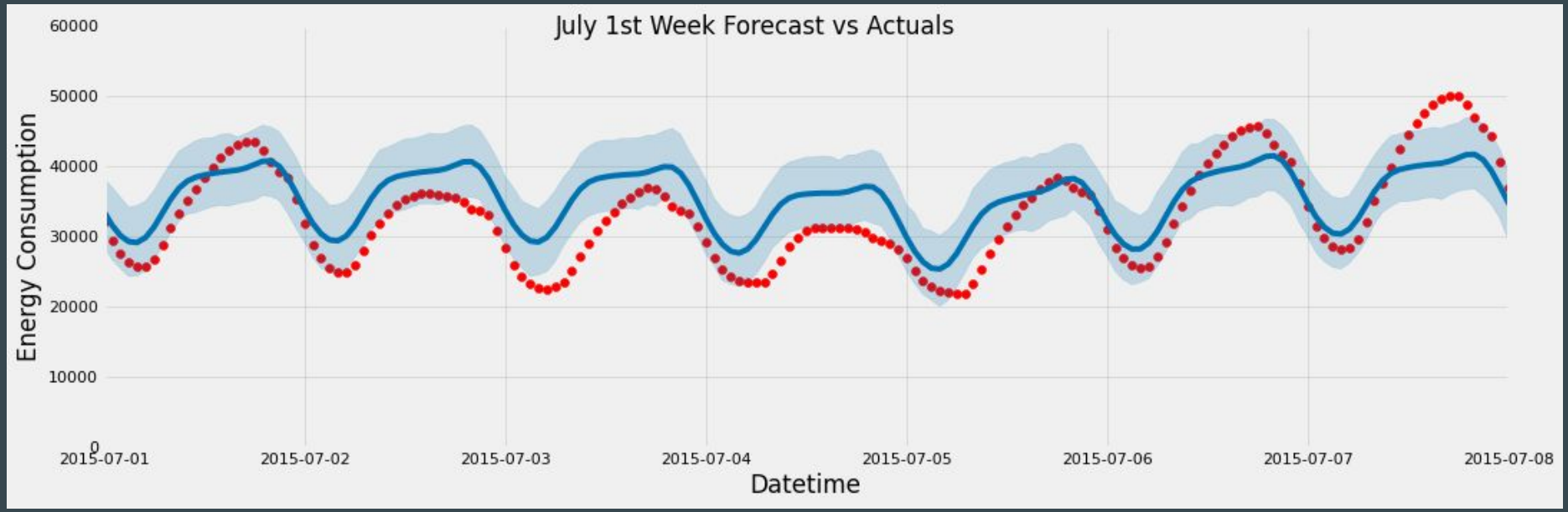

Feature importance
day_of_year
hour
year
week_of_year
day_of_week
day_of_month
month
250
188
94
72
65
17
14
Features
0
50
100
150
200
250
F score


Energy Consumption \ Test vs Forecast
Energy Consumption
Forecast
55000
50000
45000
40000
35000
30000
25000
20000
Energy Consumption
Jul
Jan
2016
Jul
Jan
2017
Jul
Jan
2018
Jul
Datetime


First Week of July Forecast
Forecast
Energy Consumption
60000
50000
40000
30000
20000
10000
2015-07-01
2015-07-02
2015-07-03
2015-07-04
2015-07-05
2015-07-06
2015-07-07
2015-07-08
Datetime

Facebook's Prophet




July 1st Week Forecast vs Actuals
Energy Consumption
60000
50000
40000
30000
20000
10000
0
2015-07-01
2015-07-02
2015-07-03
2015-07-04
2015-07-05
2015-07-06
2015-07-07
2015-07-08
Datetime

Takeaways

Both models do a decent job at forecasting

Models do not capture the “peaks”

Metrics :

Root Mean Squared Error :

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{N}}$$

RMSD = root-mean-square deviation
i = variable i
N = number of non-missing data points
x_i = actual observations time series
$\hat{x}_i$ = estimated time series

Mean Absolute Percentage Error:

$$M = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

M = mean absolute percentage error
n = number of times the summation iteration happens
A_t = actual value
F_t = forecast value

How well are these models doing?

Prophet:

RMSE : 4195.90

MAPE : 9.86 %

XGBoost

RMSE : 3745.90

MAPE : 9.108%

XGBoost does slightly better

Improvements

Add lag variables

Add Holiday Indicators

Add weather source data

Log/other transformation to minimize the spikes

Weighted MAPE evaluation

VotingRegressor - combine both models