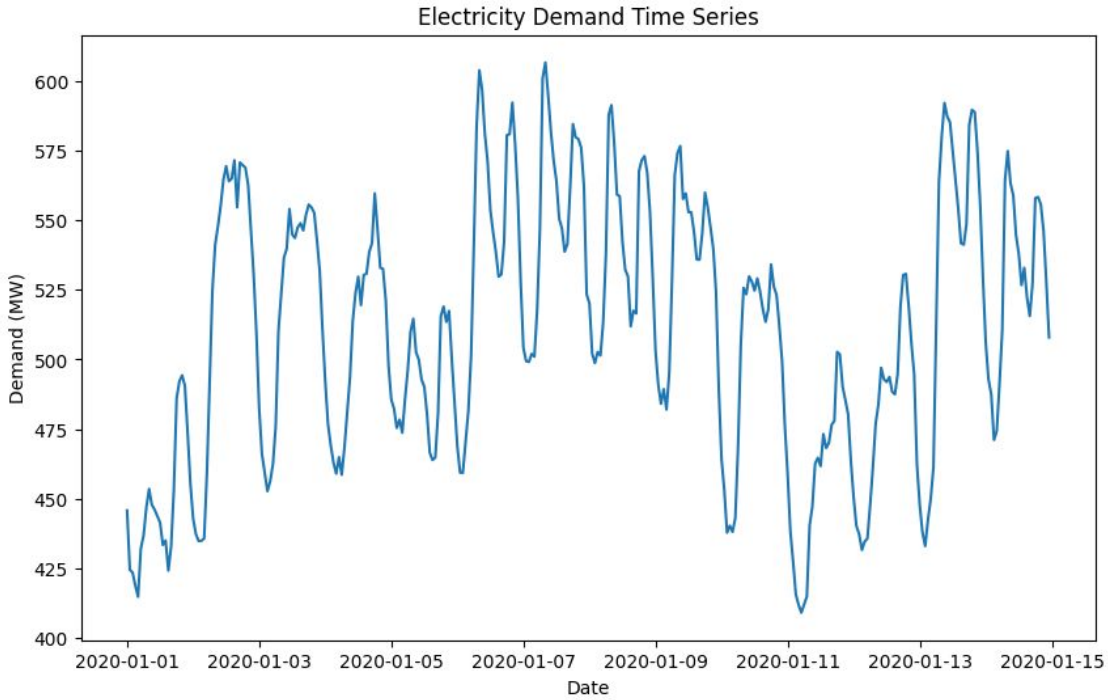
Time Series Analysis and Forecasting :

Univariate Random Forest Regressor

Nirajan Bekoju

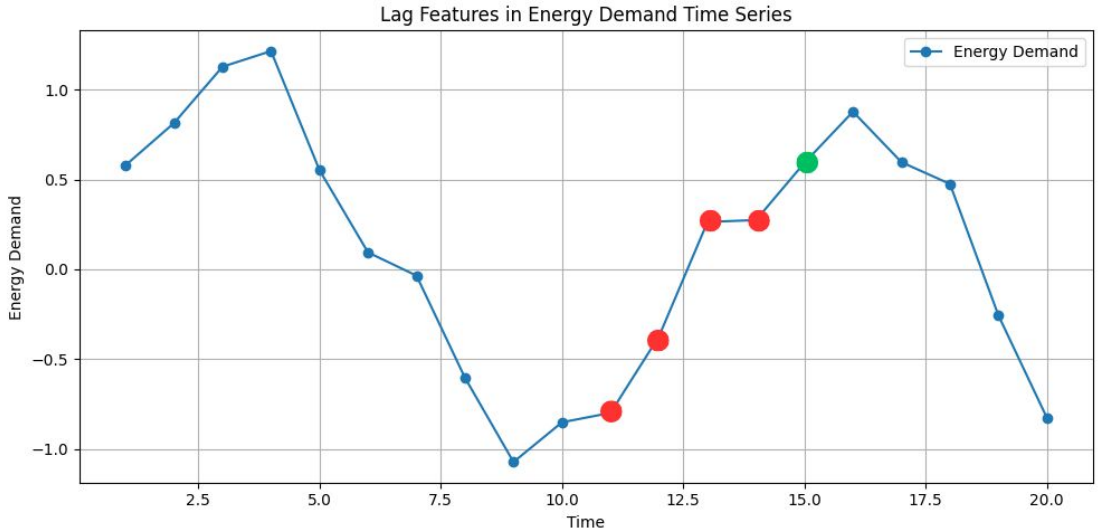
Energy Demand - Time series data

datetime	Demand (MW)
2020-01-01 00:00:00	445.8
2020-01-01 01:00:00	424.5
2020-01-01 02:00:00	423.5
2020-01-01 03:00:00	418.8
2020-01-01 04:00:00	414.8



Lags Feature

Cycles is associated with how values in a series at one time **depends** on the values in the previous time.



Stationarity

The Stationarity data is the time series data in which characteristics of the data such as **mean, variance, and autocorrelation** do not change over time.

ADF Test for Stationarity

H0: The time series is non-stationary. In other words, it has some time-dependent structure and doesn't have constant variance over time.

H1: The time series is stationary

ADF Test for Stationarity

-10.35
test statistics

2.45 * 10^(-18)
p-value

Conclusion: From the above ADF test, we can observe p-value < 0.05, hence Null Hypothesis is rejected.

Lags Feature

datetime	Demand (MW)	lag_1	lag_2	lag_3	lag_4
2020-01-01 00:00	445.8	NaN	NaN	NaN	NaN
2020-01-01 01:00	424.5	445.8	NaN	NaN	NaN
2020-01-01 02:00	423.5	424.5	445.8	NaN	NaN
2020-01-01 03:00	418.8	423.5	424.5	445.8	NaN
2020-01-01 04:00	414.8	418.8	423.5	424.5	445.8

Depends on

Current Demand

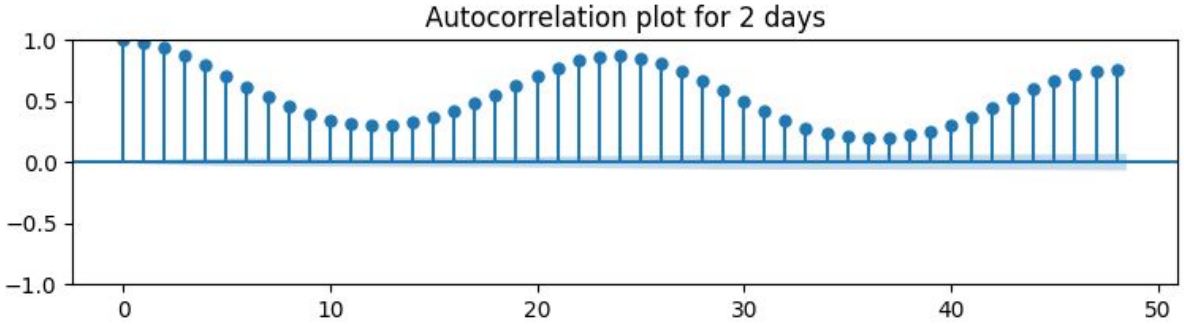
2020-01-01 03:00	418.8	423.5	424.5	445.8	NaN

Lags Feature

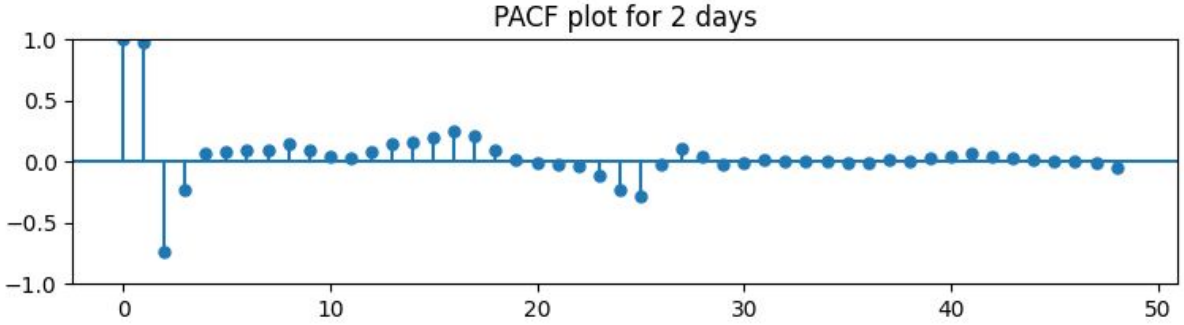
datetime	Demand (MW)	lag_1	lag_2	lag_3	lag_4	lag_5	lag_6	lag_7	lag_8	lag_9	...
2020-01-22 00:00	592.5	619.9	638.9	665.4	661.5	661.4	640.0	611.1	590.5	596.2	...
2020-01-22 01:00	588.1	592.5	619.9	638.9	665.4	661.5	661.4	640.0	611.1	590.5	...
2020-01-22 02:00	587.4	588.1	592.5	619.9	638.9	665.4	661.5	661.4	640.0	611.1	...
2020-01-22 03:00	587.0	587.4	588.1	592.5	619.9	638.9	665.4	661.5	661.4	640.0	...
2020-01-22 04:00	590.3	587.0	587.4	588.1	592.5	619.9	638.9	665.4	661.5	661.4	...

3 week lags =
3 * 7 * 24 =
504 lags features

ACF and PACF

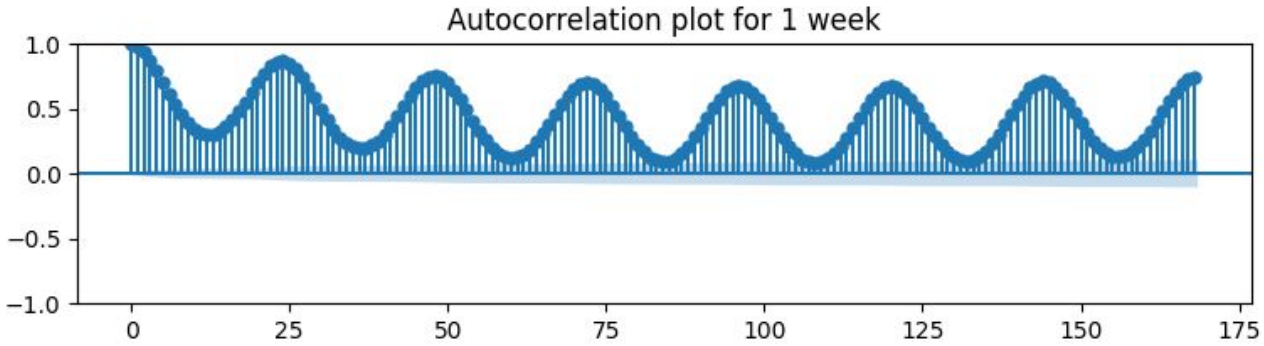


Highly correlated with lag 24

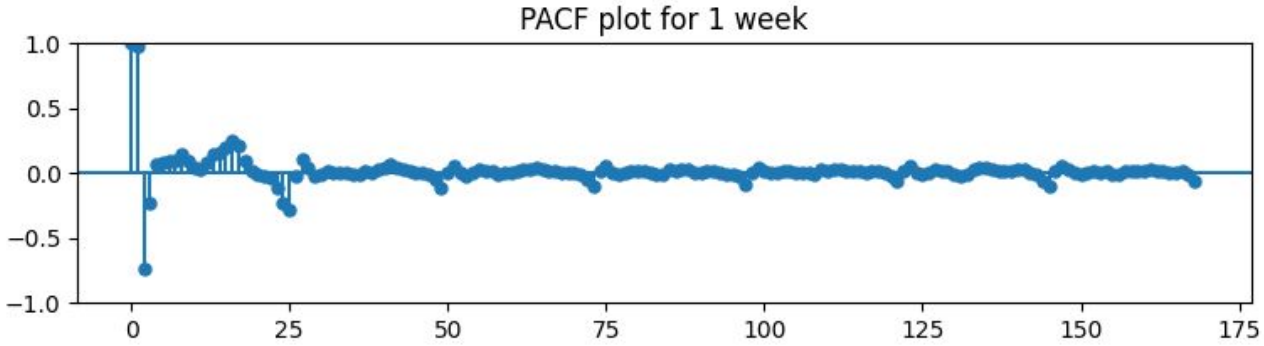


Almost all lags are significant as they are outside the confidence interval

ACF and PACF



Highly correlated with lag 24

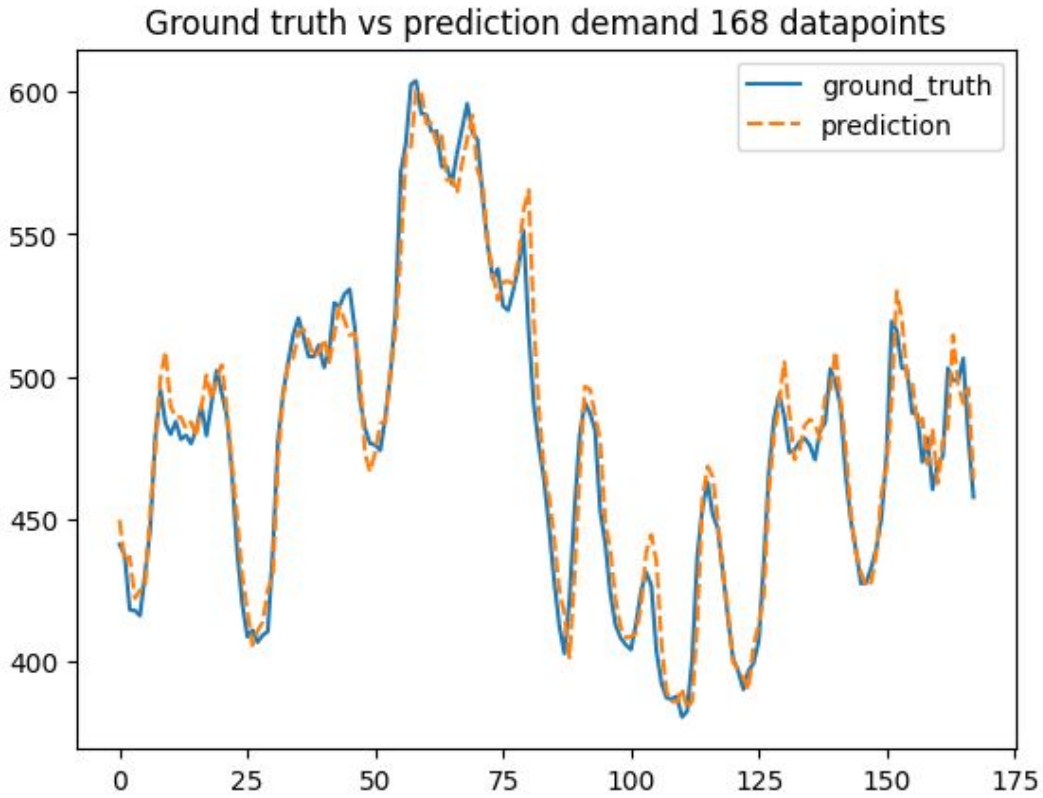


Almost all lags are significant as they are outside the confidence interval

Random Forest Regressor

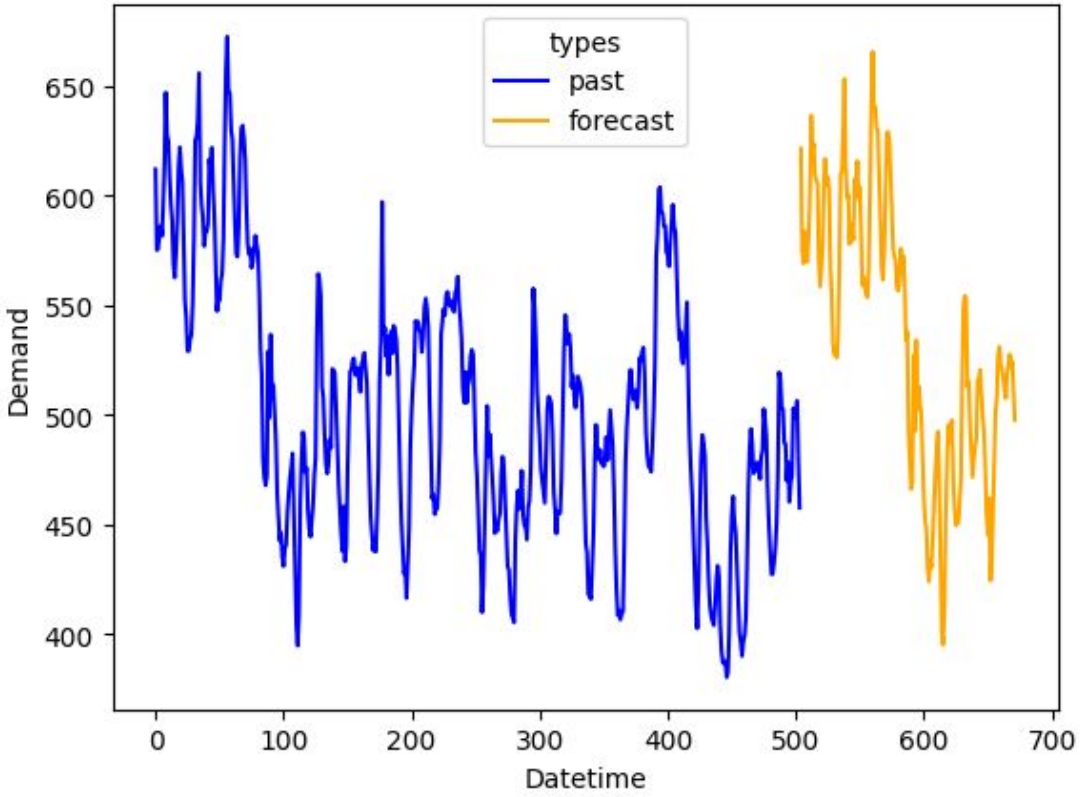
Train Test Split

Validation



train mse	18.53
val mse	124.63

1 Week Forecast



Key Takeaways