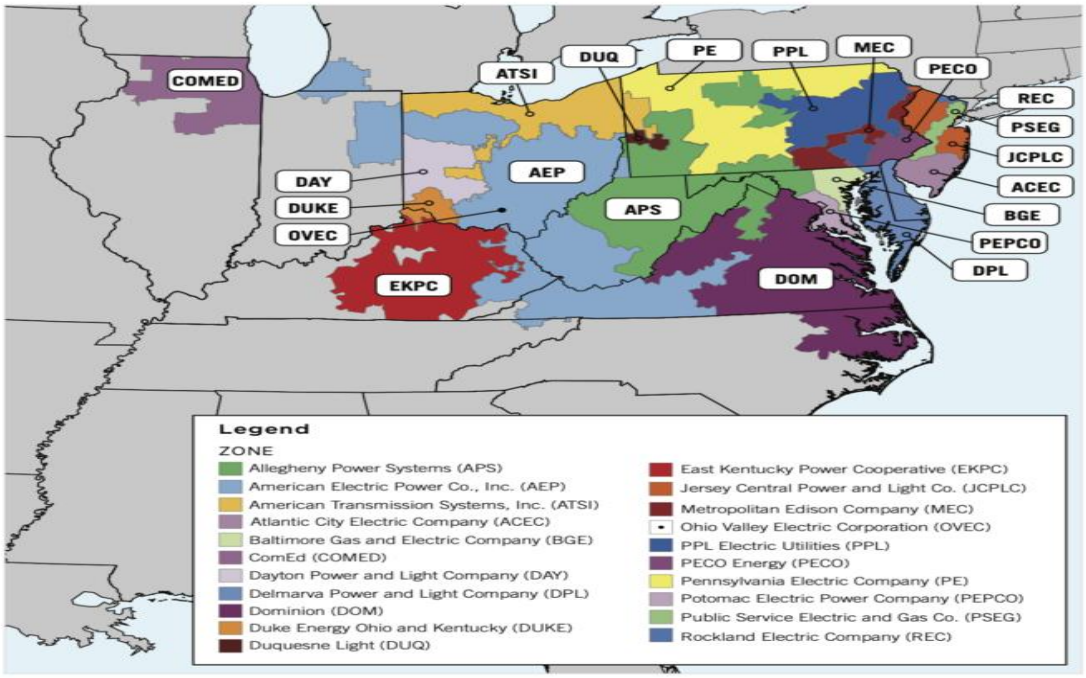
PJM Hourly Energy Consumption



PJM Interconnection LLC (PJM) is a regional transmission organization (RTO) in the United States. It is part of the Eastern Interconnection grid operating an electric transmission system serving all or parts of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, and the District of Columbia.The regions have changed over the years so data may only appear for certain dates per region

Data

The hourly power consumption data comes from PJM's website and are in megawatts (MW). This dataset contains over 10 years of hourly energy consumption data from PJM in Megawatts

Kaggle dataset

Data Cleaning

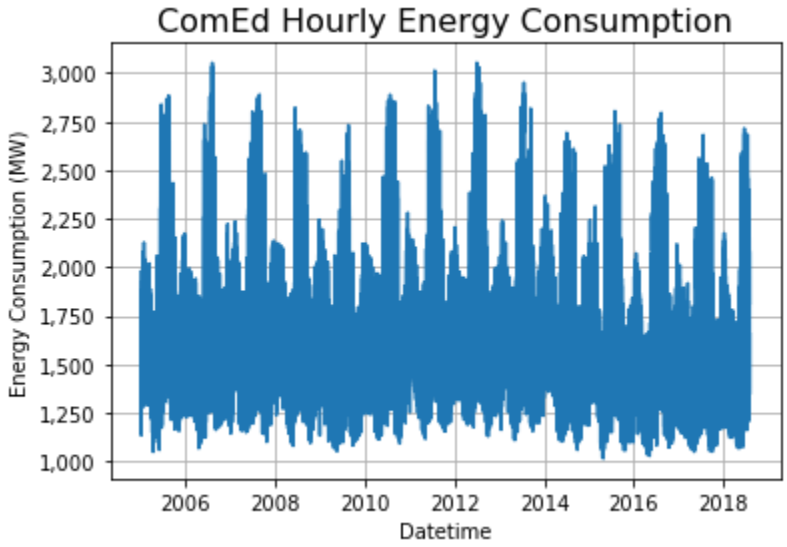
	Date time	AEP_Mw
0	2004-12-31 01:00:00	13478.0
1	2004-12-31 02:00:00	12865.0
2	2004-12-31 03:00:00	12577.0
3	2004-12-31 04:00:00	12517.0
4	2004-12-31 05:00:00	12670.0

1 There were some data which was empty that has been removed

2 convert the Datetime into proper date format

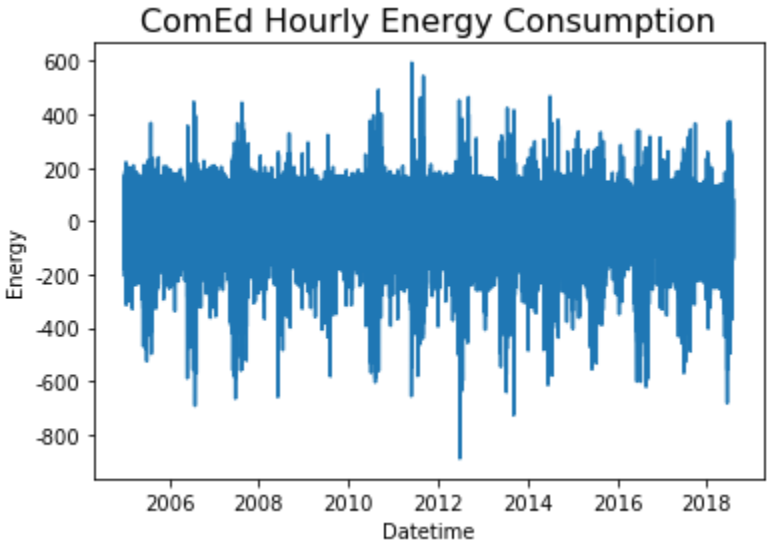
Exploratory Data Analysis

The energy consumption in is shown blow



I convert the Un stationary data to stationary data

Then plot the stationary dataset



Algorithm and Machine Learning

ML Notbook

I chose to work with the Python surprise library scikit for training predicting the future value. I have used another model SARIMAX from python library. I run both the model on the pjm electric consumption. The Result came out as below

ARIMAX Results

Dep. Variable:	DUQ_MW	**No. Observations:**	96387
Model:	ARIMA(3, 0, 4)	**Log Likelihood**	-512411.168
Date:	Sun, 18 Sep 2022	**AIC**	1024840.337
Time:	23:14:16	**BIC**	1024925.622
Sample:	0	**HQIC**	1024866.262
	- 96387		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0023	0.108	0.021	0.983	-0.209	0.213
ar.L1	2.2803	0.024	94.943	0.000	2.233	2.327
ar.L2	-1.6861	0.046	-36.856	0.000	-1.776	-1.596
ar.L3	0.3672	0.023	15.779	0.000	0.322	0.413

ma.L1	-1.9067	0.024	-79.571	0.000	-1.954	-1.860
ma.L2	0.9952	0.037	26.859	0.000	0.923	1.068
ma.L3	-0.1054	0.010	-10.542	0.000	-0.125	-0.086
ma.L4	0.0388	0.007	5.496	0.000	0.025	0.053
sigma2	2444.5145	4.964	492.473	0.000	2434.786	2454.243

Ljung-Box (L1) (Q):	0.21	**Jarque-Bera (JB):**	555429.62
Prob(Q):	0.65	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.01	**Skew:**	-0.08
Prob(H) (two-sided):	0.43	**Kurtosis:**	14.76

SRIMAX MODEL:

Dep. Variable:	DUQ_MW	**No. Observations:**	96387
Model:	SARIMAX(3, 1, 4)	**Log Likelihood**	-523052.785
Date:	Sun, 18 Sep 2022	**AIC**	1046123.571
Time:	23:28:56	**BIC**	1046208.856
Sample:	0	**HQIC**	1046149.496
	- 96387		
Covariance Type:	opg		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
intercept	0.0008	0.002	0.484	0.628	-0.002	0.004
ar.L1	-0.6233	0.005	-130.452	0.000	-0.633	-0.614
ar.L2	-0.0738	0.005	-14.575	0.000	-0.084	-0.064
ar.L3	0.6515	0.004	144.866	0.000	0.643	0.660
ma.L1	0.2320	0.005	43.030	0.000	0.221	0.243
ma.L2	-0.3347	0.003	-112.943	0.000	-0.340	-0.329
ma.L3	-0.9383	0.003	-303.119	0.000	-0.944	-0.932
ma.L4	0.0421	0.005	8.833	0.000	0.033	0.051
sigma2	3278.3741	7.654	428.335	0.000	3263.373	3293.375

Ljung-Box (L1) (Q):	24.73	**Jarque-Bera (JB):**	384242.38
Prob(Q):	0.00	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.00	**Skew:**	-0.23
Prob(H) (two-sided):	0.89	**Kurtosis:**	12.77

Hyper parameter tuning:

Used auto arima function to find the best hypermeter for the Model

```
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(0,0,0)[0] intercept   : AIC=1268297.534, Time=111.71 sec
 ARIMA(0,0,0)(0,0,0)[0] intercept   : AIC=1353873.238, Time=2.05 sec
 ARIMA(1,0,0)(0,0,0)[0] intercept   : AIC=1289412.107, Time=2.58 sec
 ARIMA(0,0,1)(0,0,0)[0] intercept   : AIC=1312404.737, Time=19.92 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=1353871.238, Time=1.04 sec
 ARIMA(1,0,2)(0,0,0)[0] intercept   : AIC=1288083.908, Time=20.86 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=1288325.900, Time=33.53 sec
 ARIMA(3,0,2)(0,0,0)[0] intercept   : AIC=1268799.873, Time=180.29 sec
 ARIMA(2,0,3)(0,0,0)[0] intercept   : AIC=1263531.142, Time=212.37 sec
 ARIMA(1,0,3)(0,0,0)[0] intercept   : AIC=1287706.831, Time=63.49 sec
 ARIMA(3,0,3)(0,0,0)[0] intercept   : AIC=inf, Time=240.32 sec
 ARIMA(2,0,4)(0,0,0)[0] intercept   : AIC=1261738.234, Time=160.88 sec
 ARIMA(1,0,4)(0,0,0)[0] intercept   : AIC=inf, Time=123.83 sec
 ARIMA(3,0,4)(0,0,0)[0] intercept   : AIC=1261587.732, Time=215.24 sec
 ARIMA(4,0,4)(0,0,0)[0] intercept   : AIC=1266048.083, Time=244.14 sec
 ARIMA(3,0,5)(0,0,0)[0] intercept   : AIC=1278114.065, Time=231.61 sec
```

```
 It shows ARIMA(3,0,5)(0,0,0) is the best suitable
```

Prediction:

Result came out as

67.48176385315581 For Arima model

68.515717572227 For SRIMAX model

Future Improvements

In the future, I would love to spend more time create a future prediction model.

This could be improve better in terms of accuracy by applying LSTM model