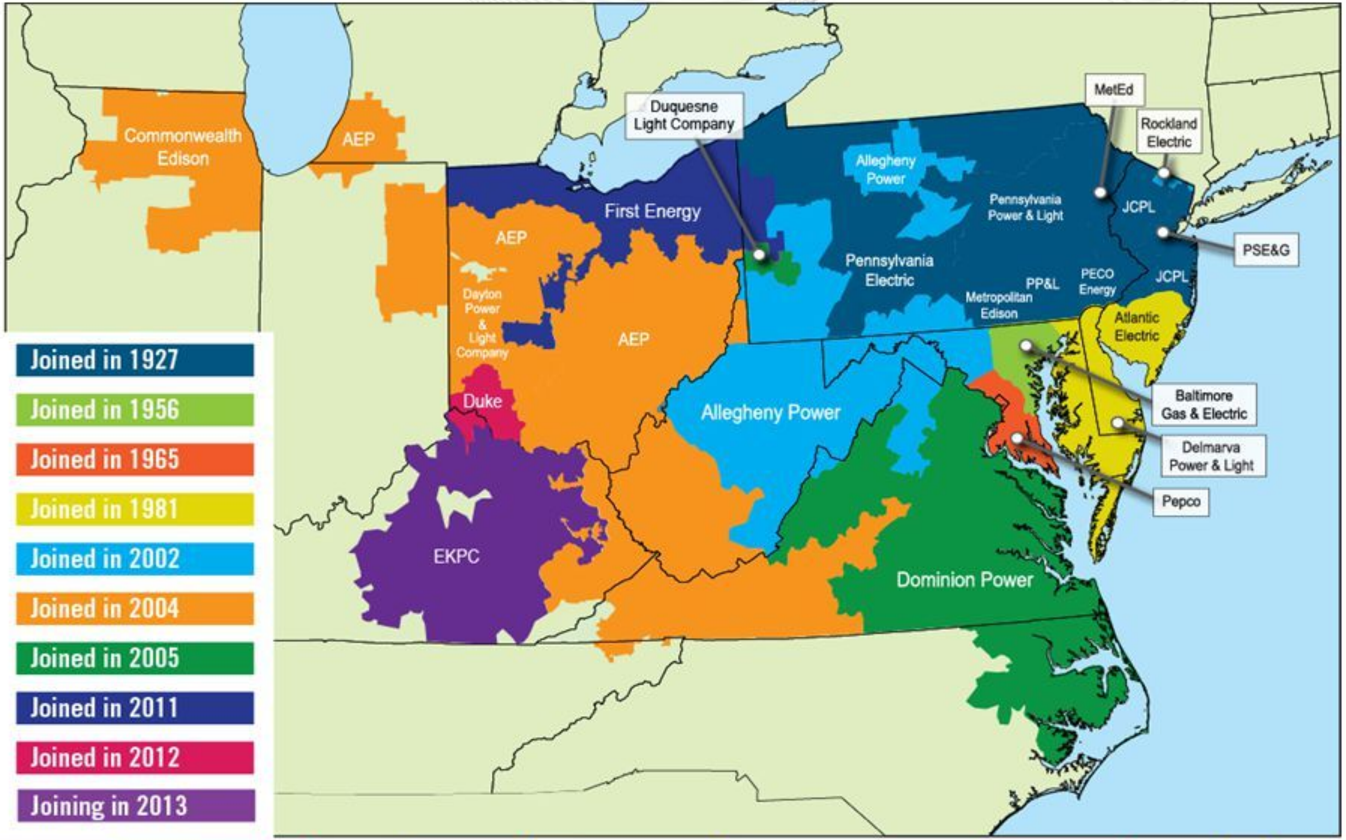
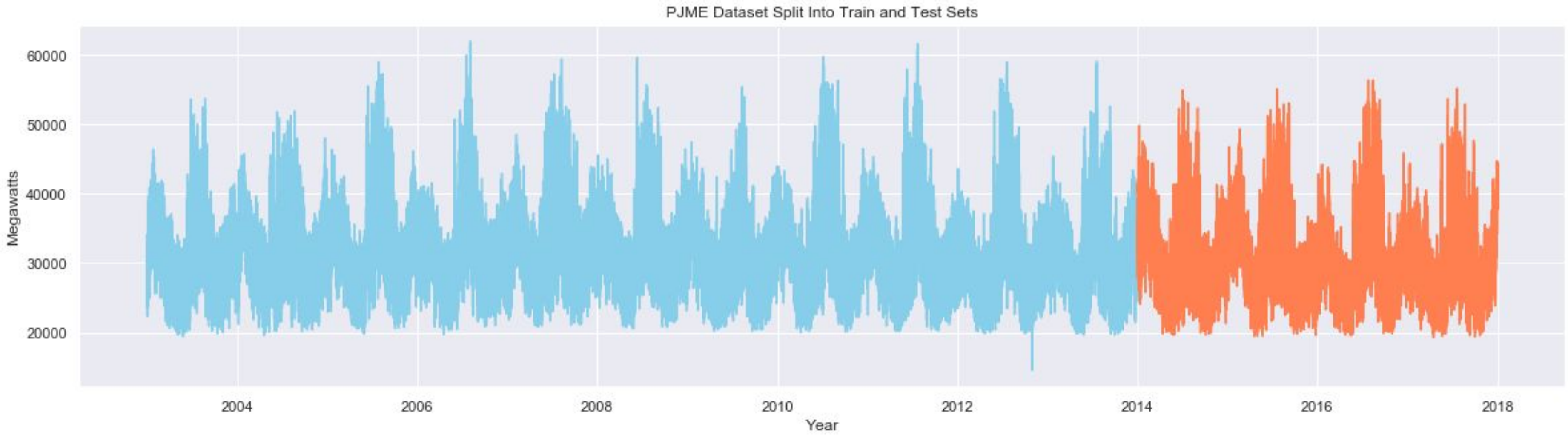
Weather and Power Consumption

PJM Evolution





The Target dataset contains 131548 entries for the PJME grid, represented in form of megawatts measurements are hourly and taken between 2003 through the end of 2017

Arports:

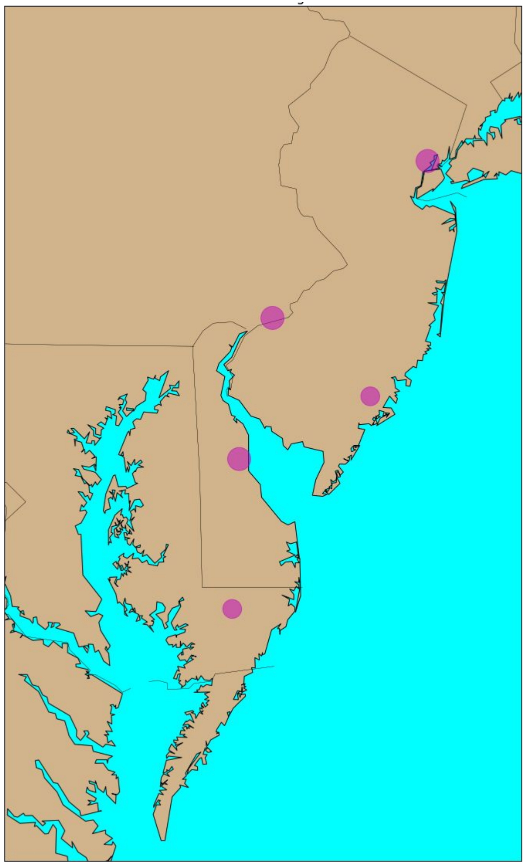
Philadelphia

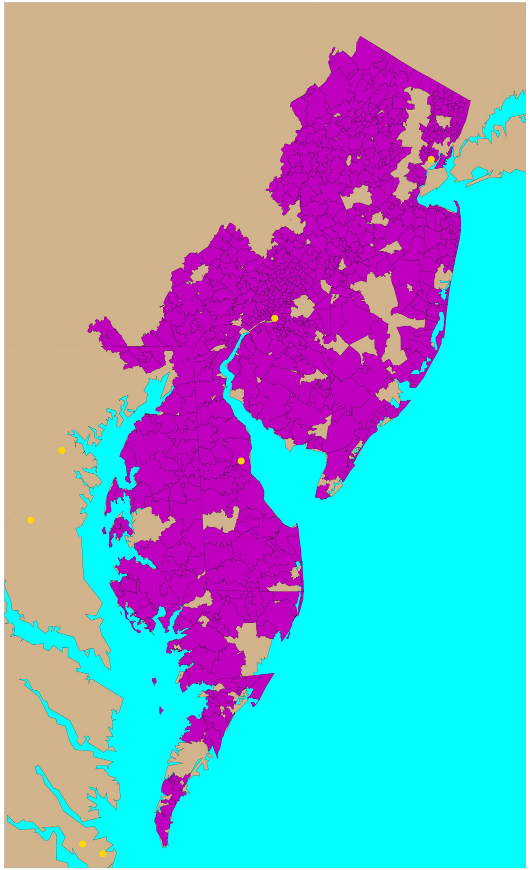
Newark

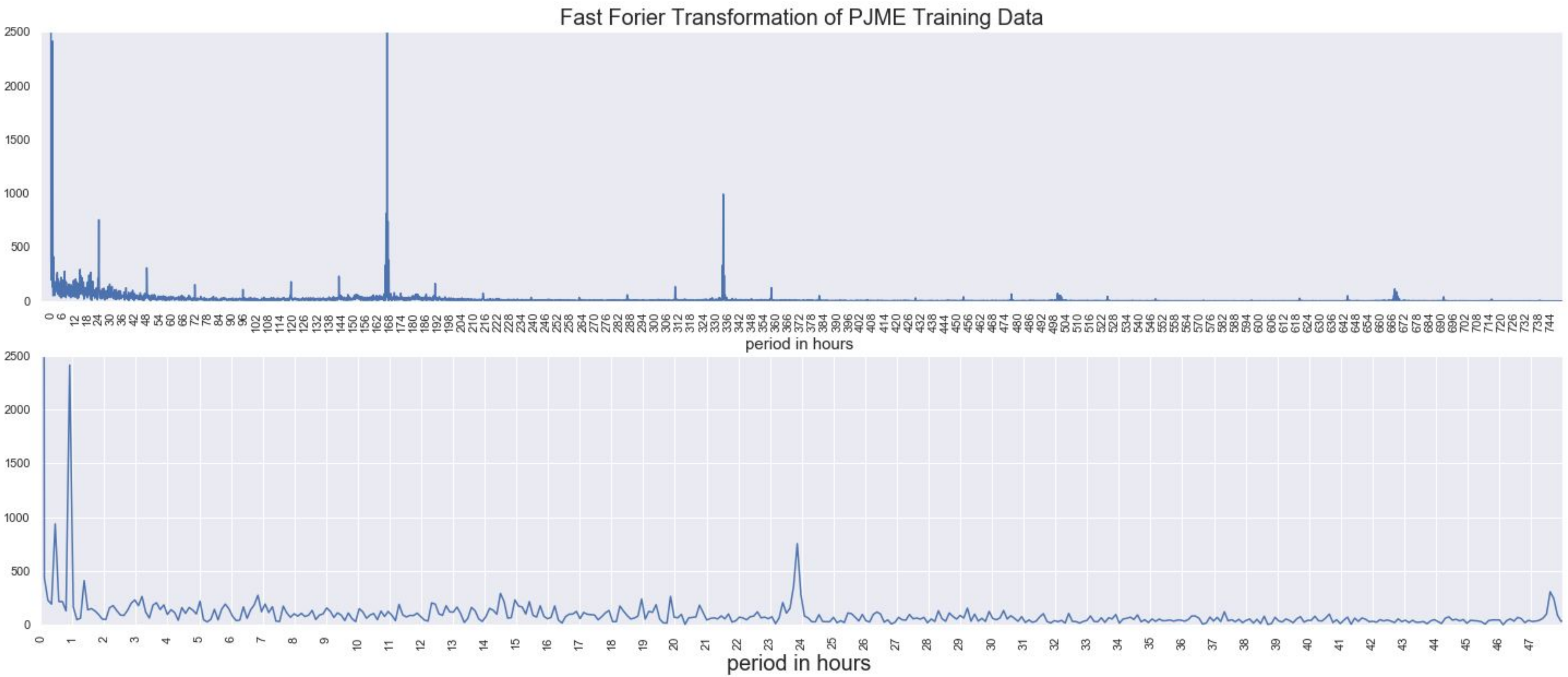
Atlantic City

Dover

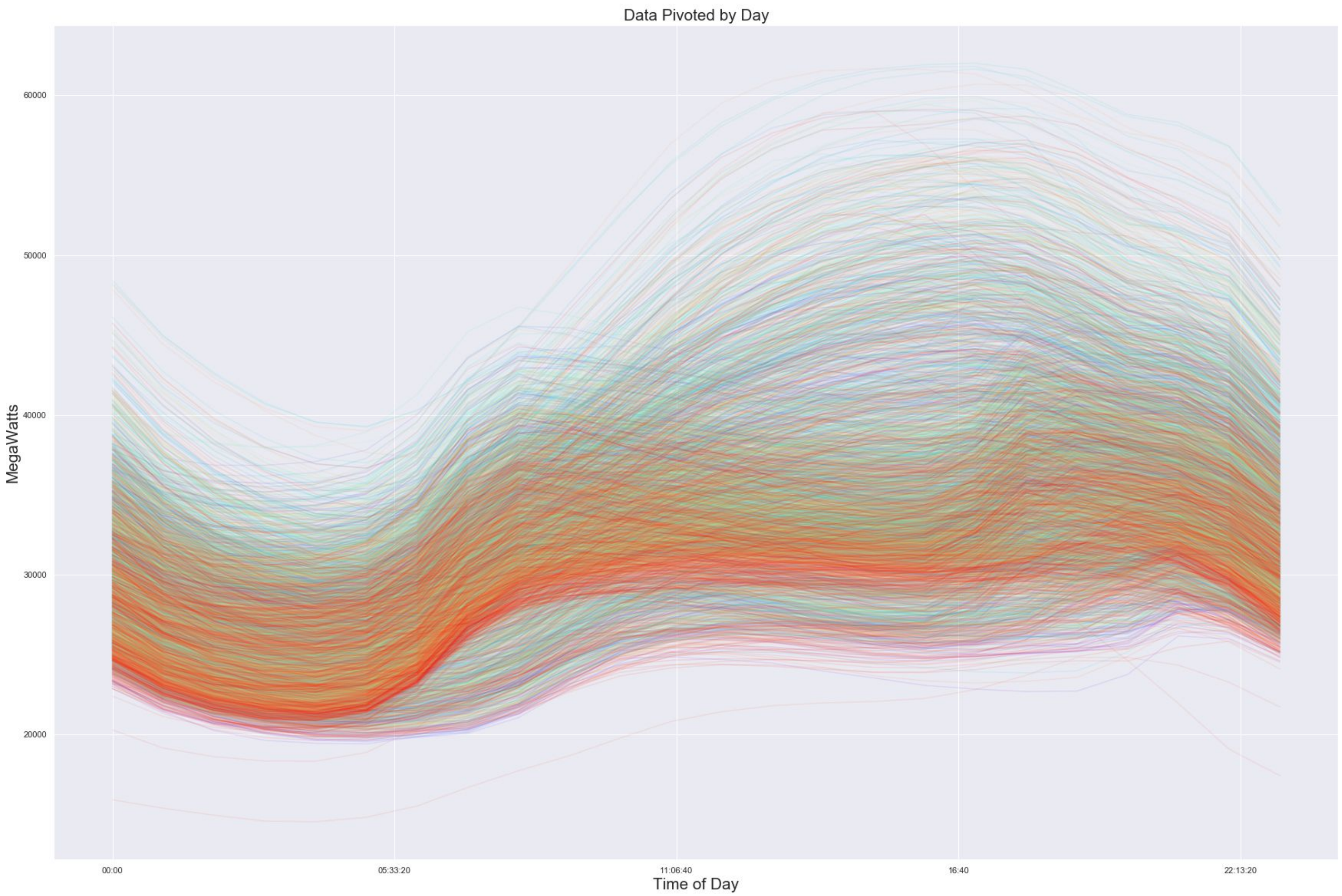
Salisbury Ocean City





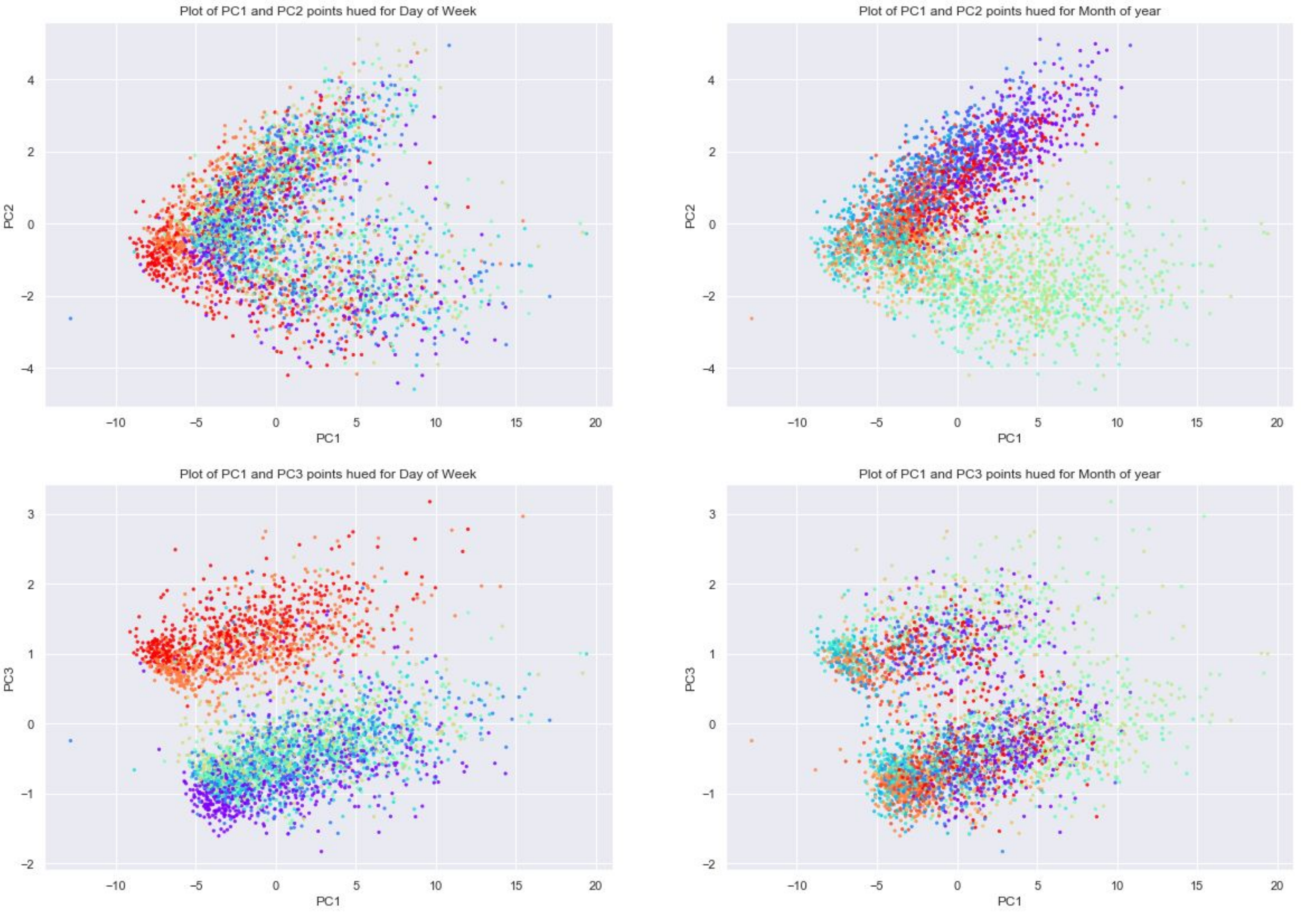



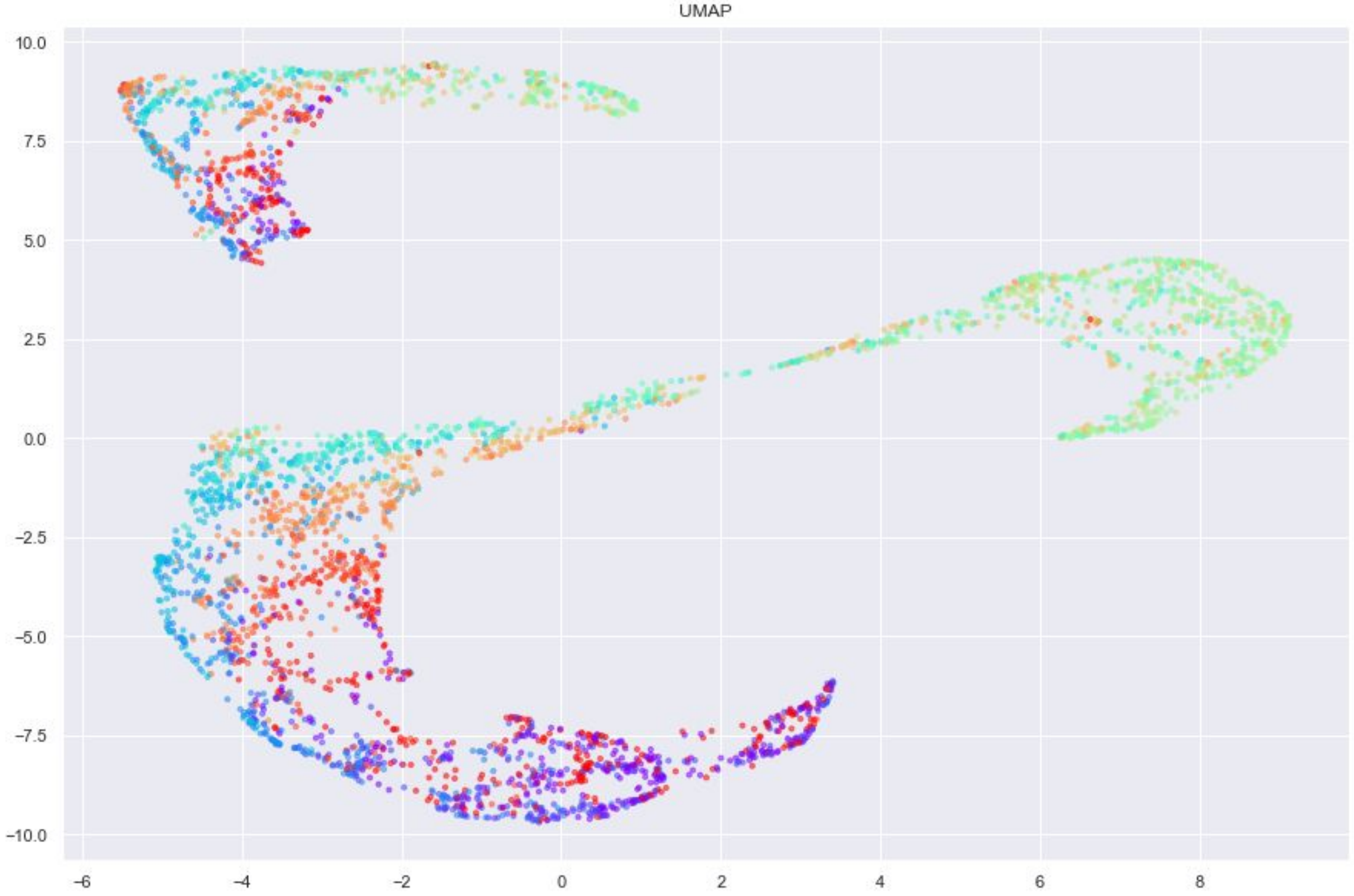
Fast Forier Transformation of PJME Training Data
period in hours
period in hours



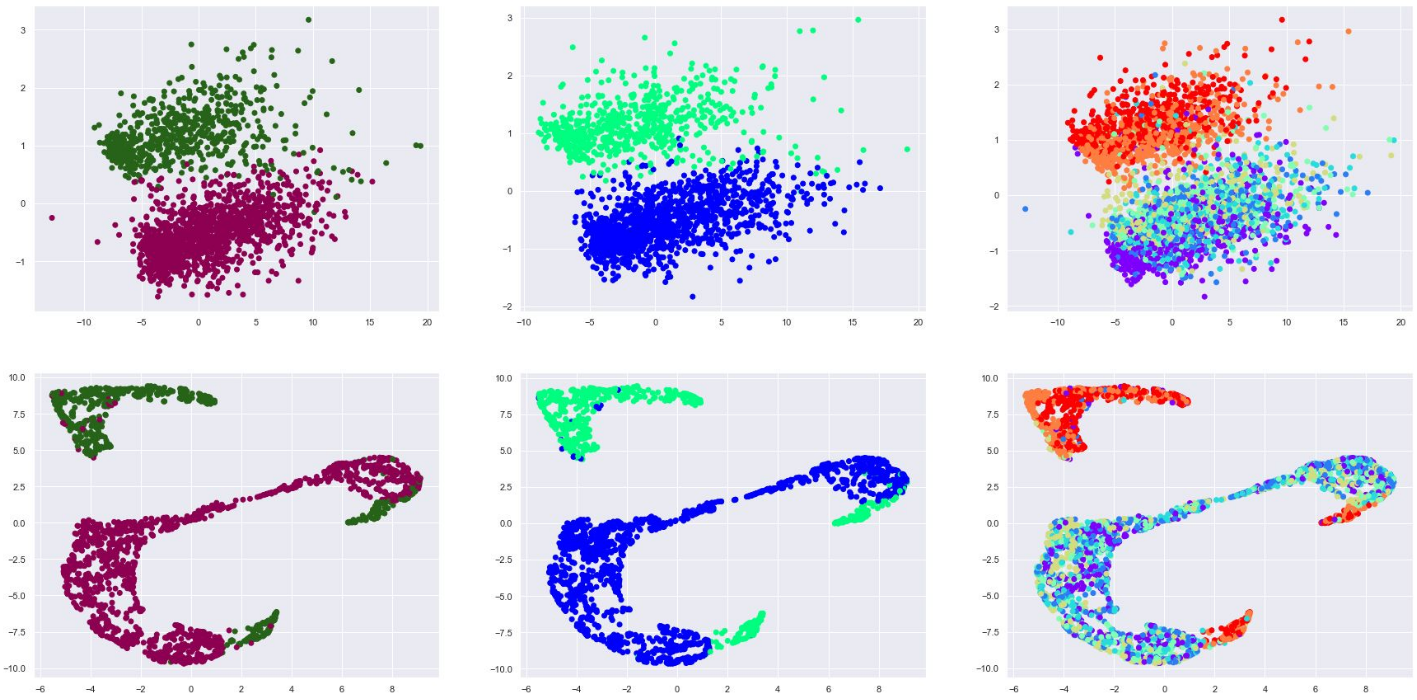
Data Pivoted by Day
MegaWatts
60000
50000
40000
30000
20000
00:00
05:33:20
11:06:40
16:40
22:13:20
Time of Day

Unsupervized Section

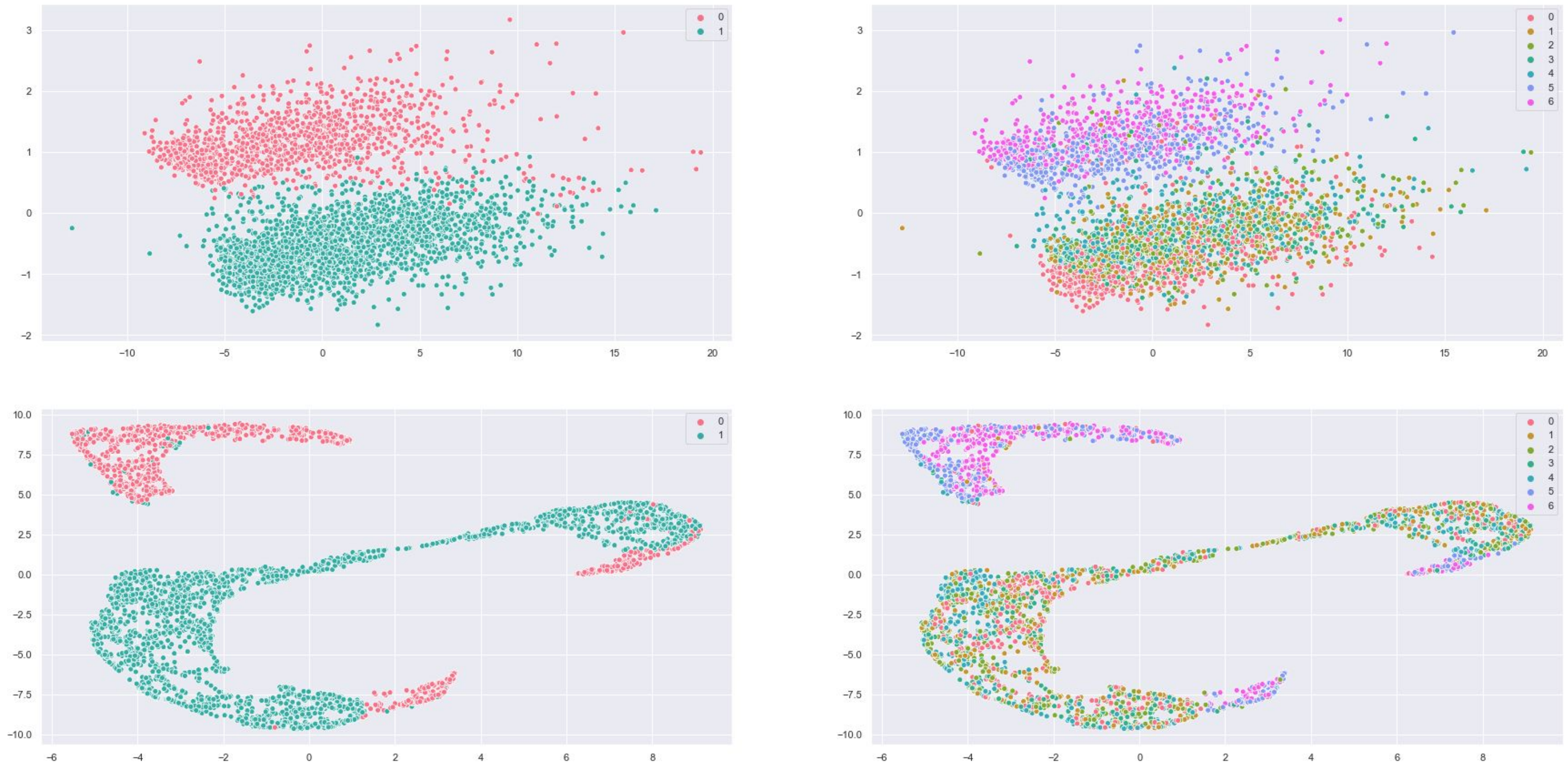

Plot of PC1 and PC2 points hued for Day of Week
Plot of PC1 and PC2 points hued for Month of year
Plot of PC1 and PC3 points hued for Day of Week
Plot of PC1 and PC3 points hued for Month of year
PC1
PC2
PC3
−10
−5
0
5
10
15
20
4
2
0
−2
−4
3
1
−1


UMAP
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−6
−4
−2
0
2
4
6
8

Consistency Check



the silhouette score for group 1: 0.3118673511408589
the silhouette score for group 2: 0.3106149257529404



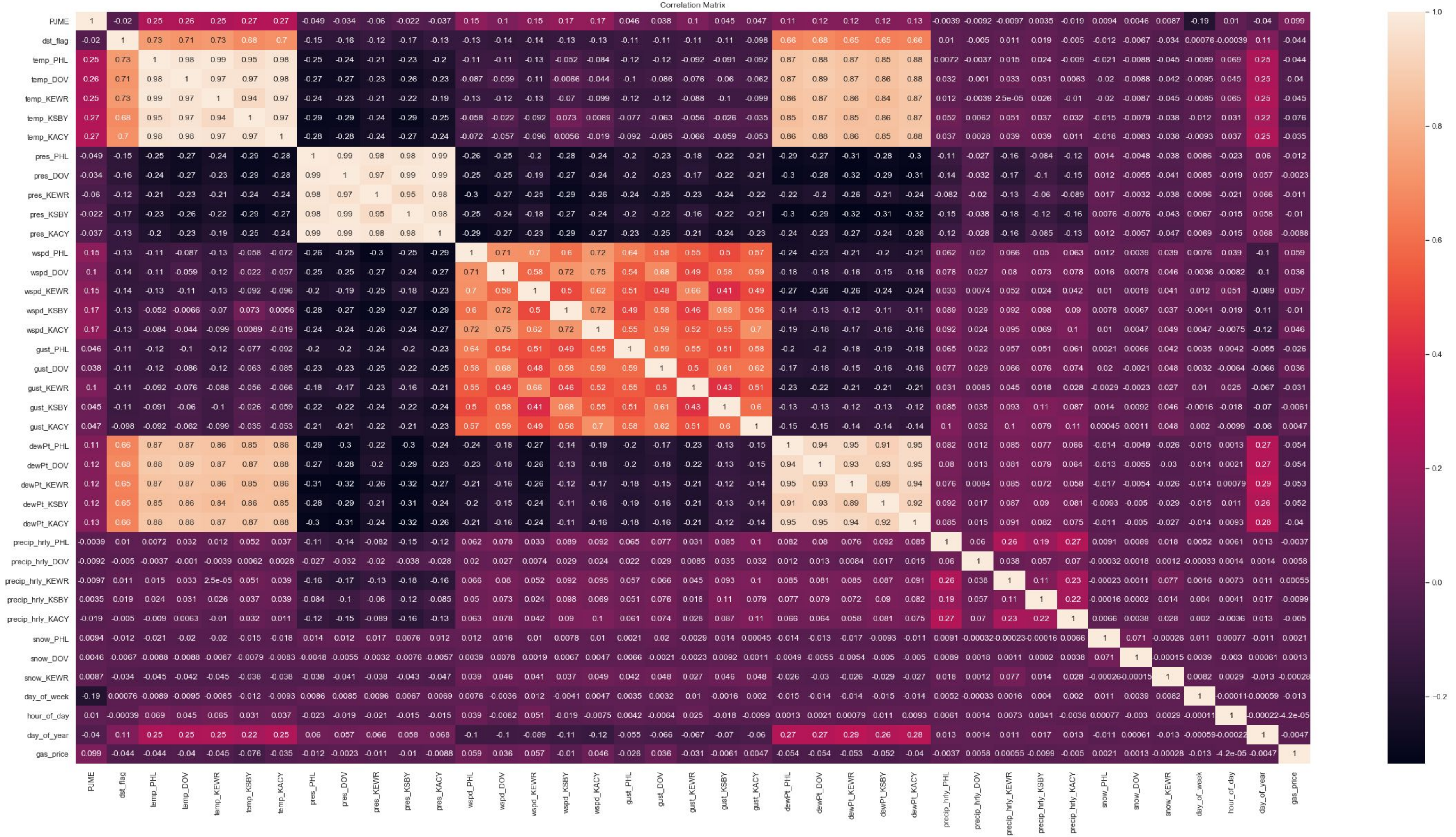
0
1
0
1
2
3
4
5
6

Days in GMM Cluster 0

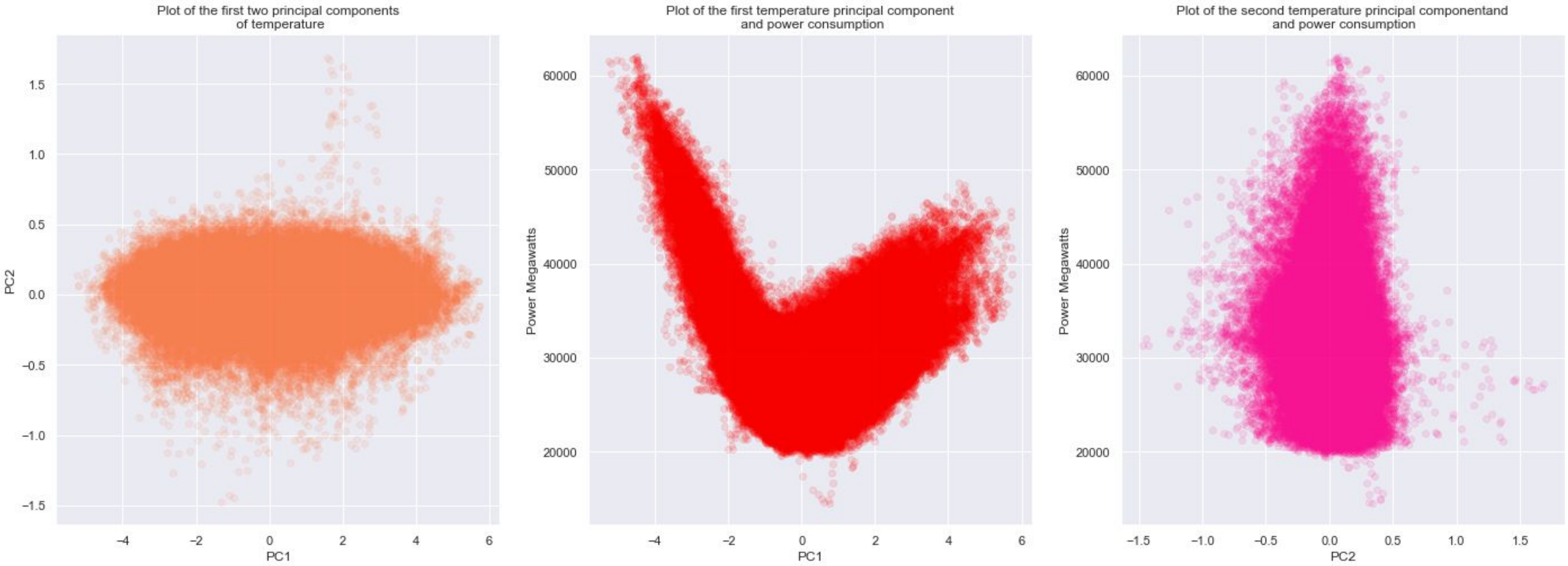
- Christmas
- The day after Christmas
- July 4th
- The two days after July 4th
- The two days before July 4th
- Labor Day
- Memorial Day
- New Years Day
- the two days after New Years Day
- New Years Eve
- Presidents Day
- Thanksgiving
- The day before Thanksgiving

Weekday Holidays in cluster1:

- MLK day
- Columbus Day
- Veterans Day



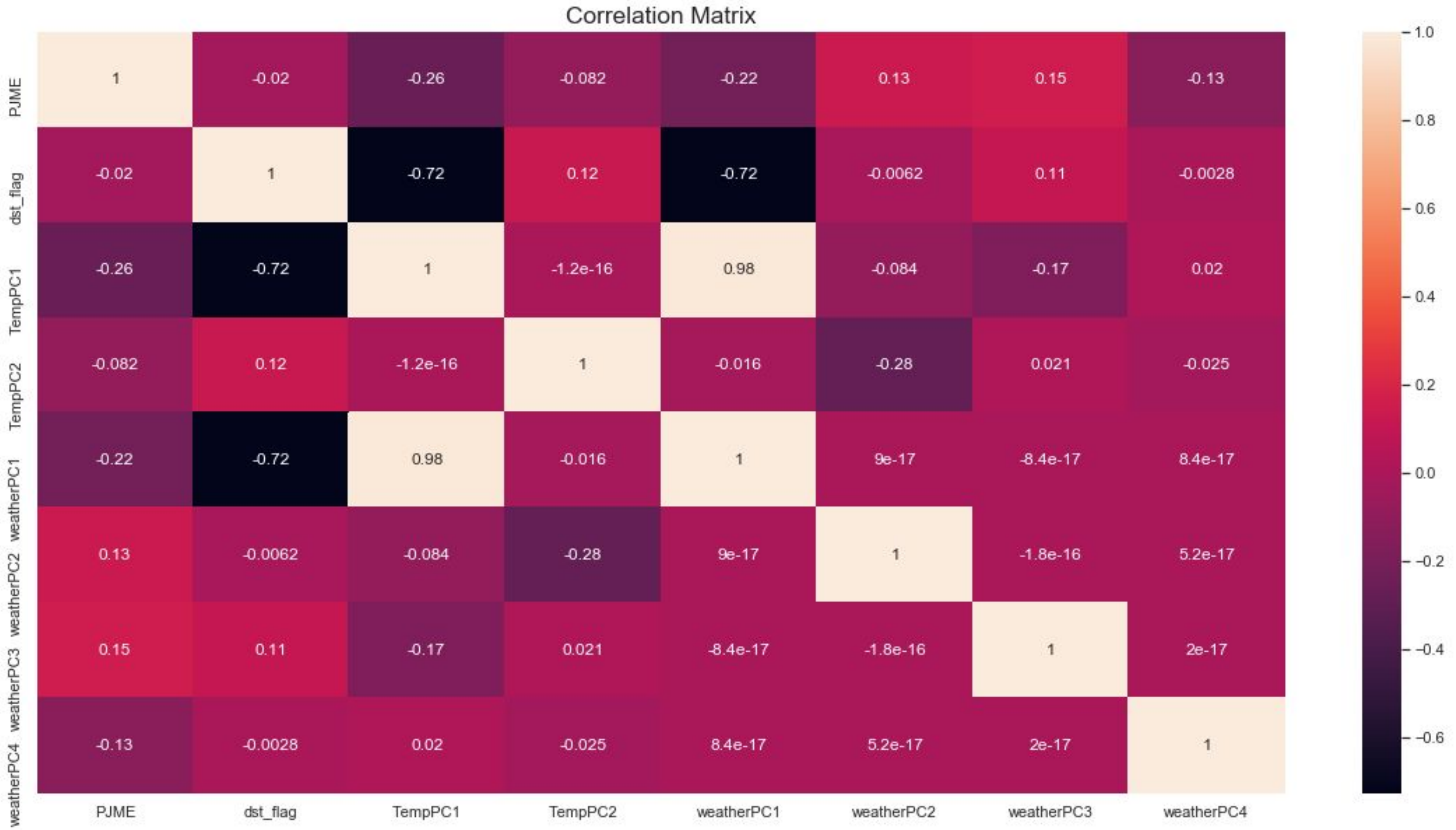
Correlation Matrix
PJME
dst_flag
temp_PHL
temp_DOV
temp_KEWR
temp_KSBY
temp_KACY
pres_PHL
pres_DOV
pres_KEWR
pres_KSBY
pres_KACY
wspd_PHL
wspd_DOV
wspd_KEWR
wspd_KSBY
wspd_KACY
gust_PHL
gust_DOV
gust_KEWR
gust_KSBY
gust_KACY
dewPt_PHL
dewPt_DOV
dewPt_KEWR
dewPt_KSBY
dewPt_KACY
precip_hrly_PHL
precip_hrly_DOV
precip_hrly_KEWR
precip_hrly_KSBY
precip_hrly_KACY
snow_PHL
snow_DOV
snow_KEWR
day_of_week
hour_of_day
day_of_year
gas_price
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

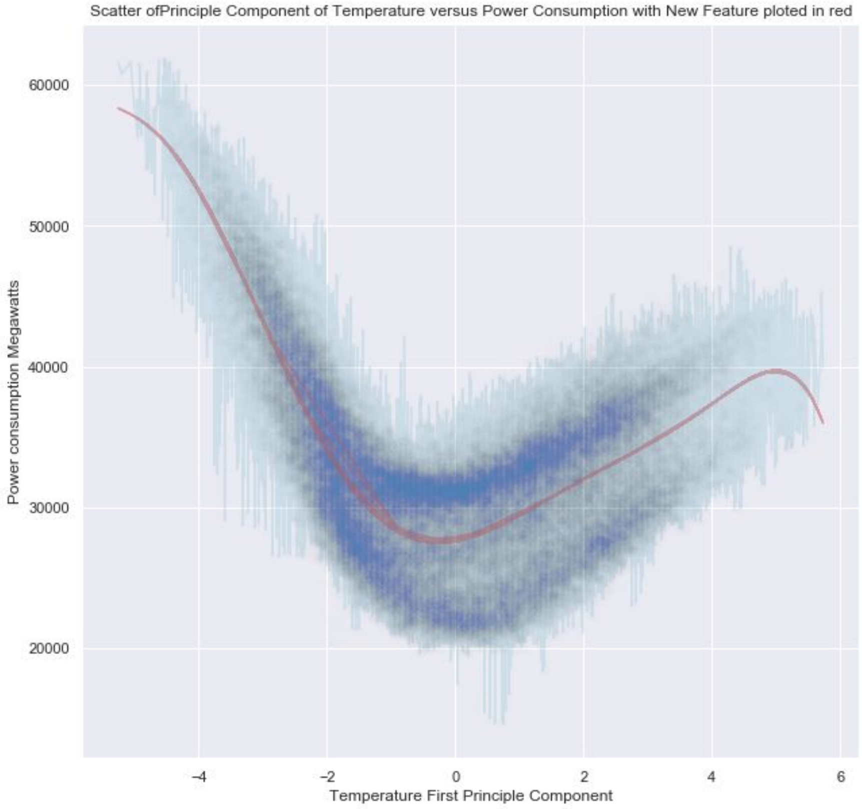


Plot of the first two principal components
of temperature
PC1
PC2
Plot of the first temperature principal component
and power consumption
PC1
Power Megawatts
Plot of the second temperature principal componentand
and power consumption
PC2
Power Megawatts

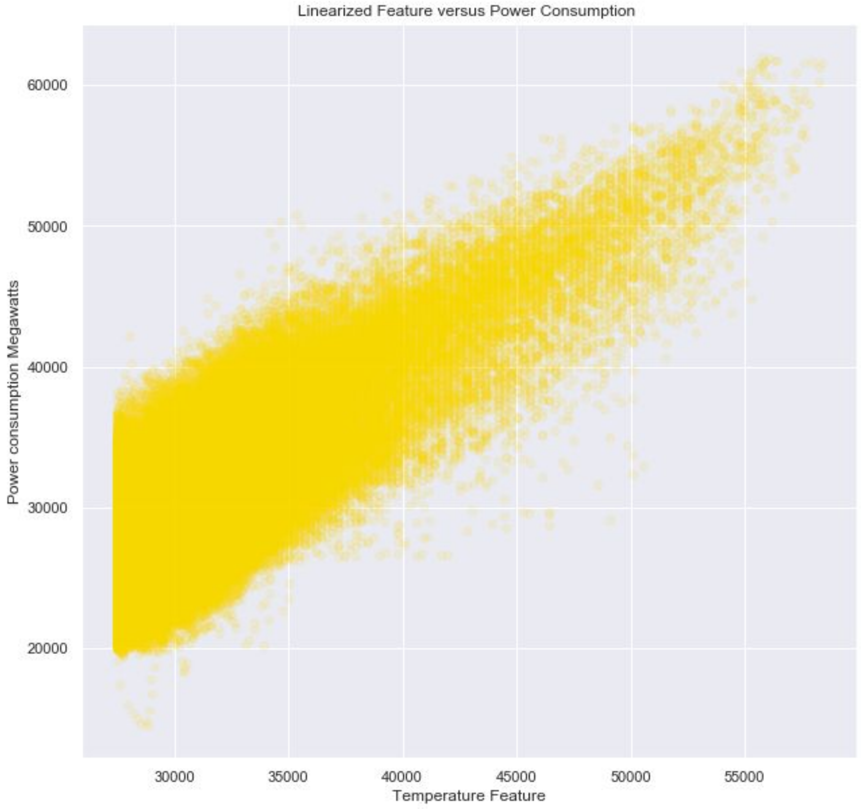

70
80
90
60
70
80
90
HONEYWELL


70°F
71°
PROG
MAN
SET
PROG/MAN
COOL - OFF - HEAT
AUTO - ON


Correlation Matrix
PJME
dst_flag
TempPC1
TempPC2
weatherPC1
weatherPC2
weatherPC3
weatherPC4
1 -0.02 -0.26 -0.082 -0.22 0.13 0.15 -0.13
-0.02 1 -0.72 0.12 -0.72 -0.0062 0.11 -0.0028
-0.26 -0.72 1 -1.2e-16 0.98 -0.084 -0.17 0.02
-0.082 0.12 -1.2e-16 1 -0.016 -0.28 0.021 -0.025
-0.22 -0.72 0.98 -0.016 1 9e-17 -8.4e-17 8.4e-17
0.13 -0.0062 -0.084 -0.28 9e-17 1 -1.8e-16 5.2e-17
0.15 0.11 -0.17 0.021 -8.4e-17 -1.8e-16 1 2e-17
-0.13 -0.0028 0.02 -0.025 8.4e-17 5.2e-17 2e-17 1
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6



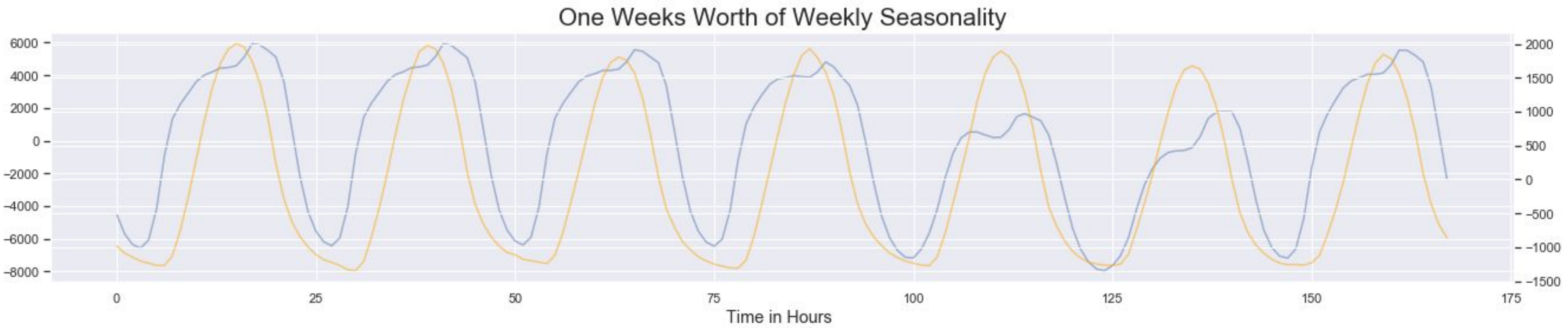
Scatter ofPrinciple Component of Temperature versus Power Consumption with New Feature ploted in red
60000
50000
40000
30000
20000
Power consumption Megawatts
−4
−2
0
2
4
6
Temperature First Principle Component

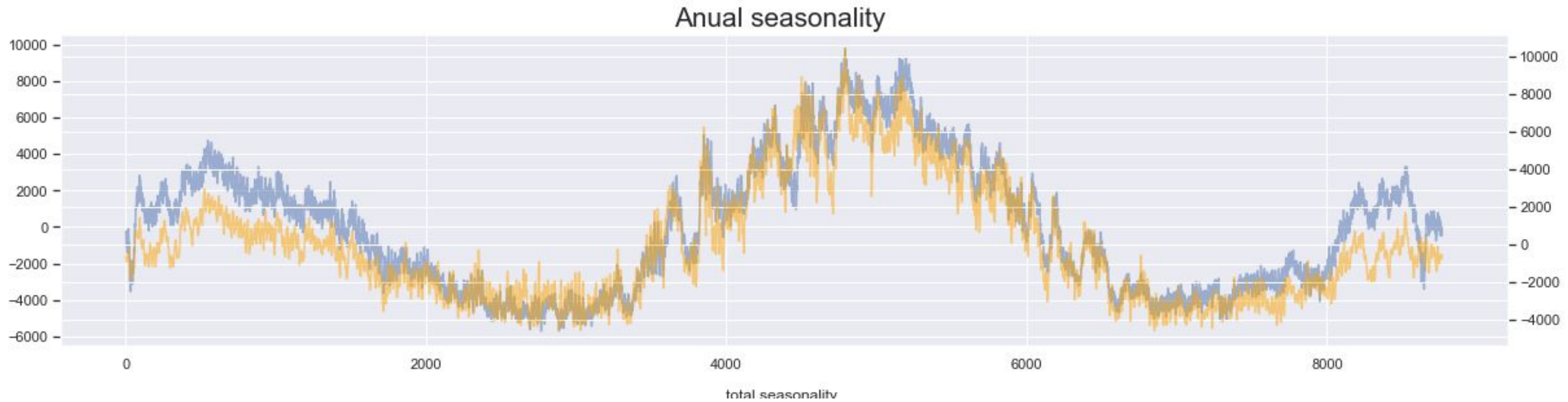


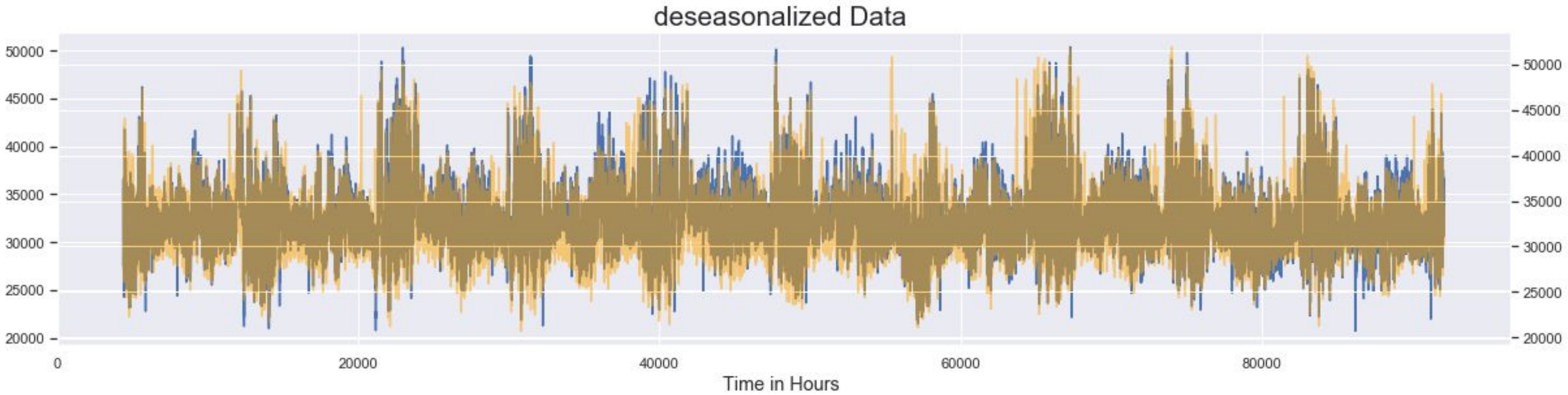
Linearized Feature versus Power Consumption
60000
50000
40000
30000
20000
Power consumption Megawatts
30000
35000
40000
45000
50000
55000
Temperature Feature

Models

Vector Autoregrssion


One Weeks Worth of Weekly Seasonality
Time in Hours


Anual seasonality
total seasonality



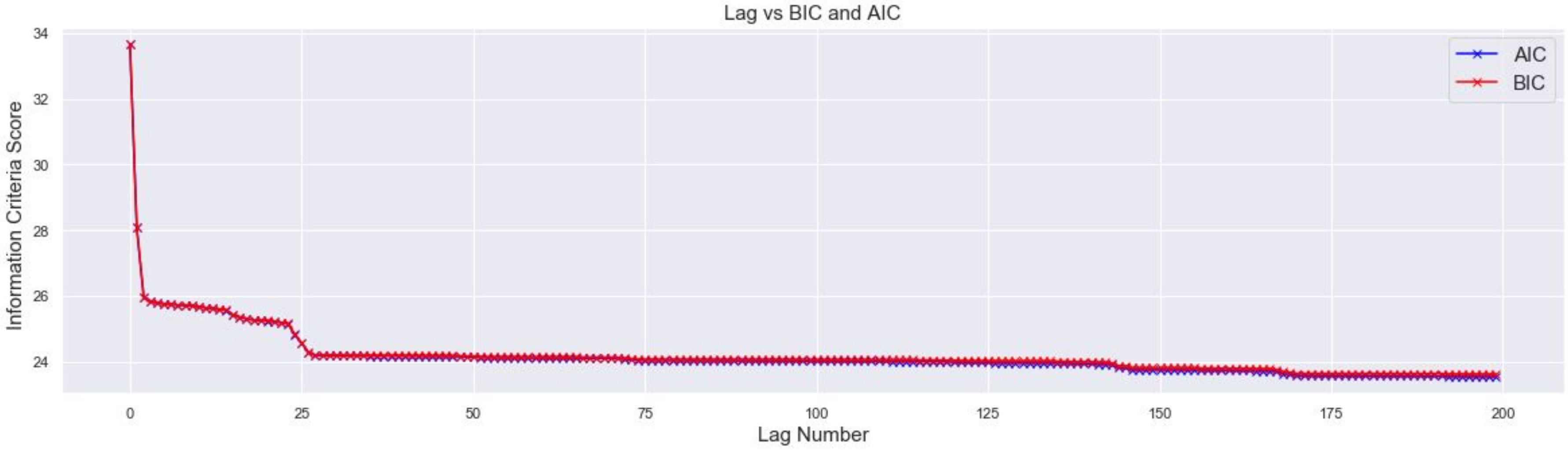
ADF Statistic for power consumption :
-14.058557,   p-value: 0.000000
Critical Values:
1%: -3.430
5%: -2.862
10%: -2.567

ADF Statistic for Temperature : -14.408056
p-value: 0.000000
Critical Values:
1%: -3.430
5%: -2.862
10%: -2.567

H_0 = there is a unit root, a sign that the dataset is not stationary

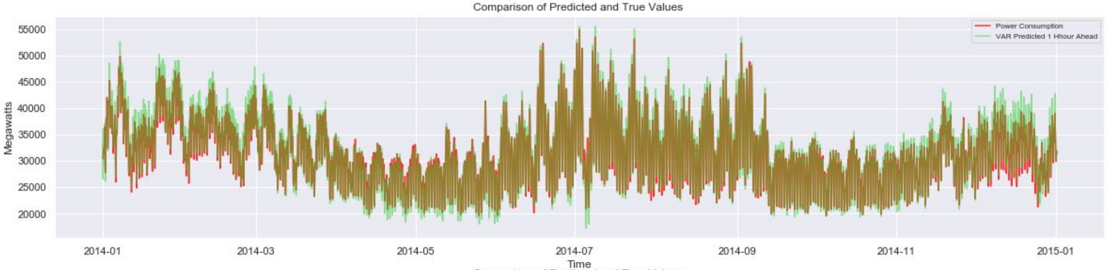
We reject the null hypothesis that the time series is not stationary

Lag Selection

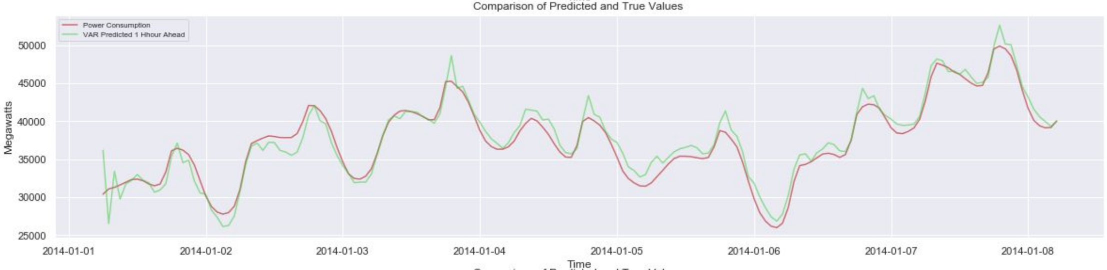


Lag chose = 4

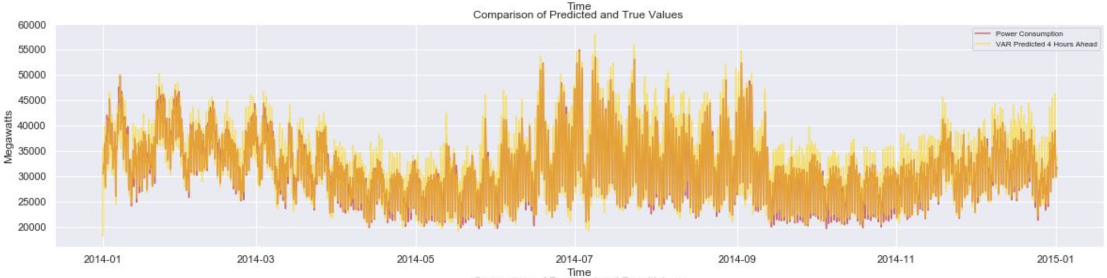
Beyond Lag 4 Coefficients began to lose significance and the residuals showed a high level of Leptokurtosis



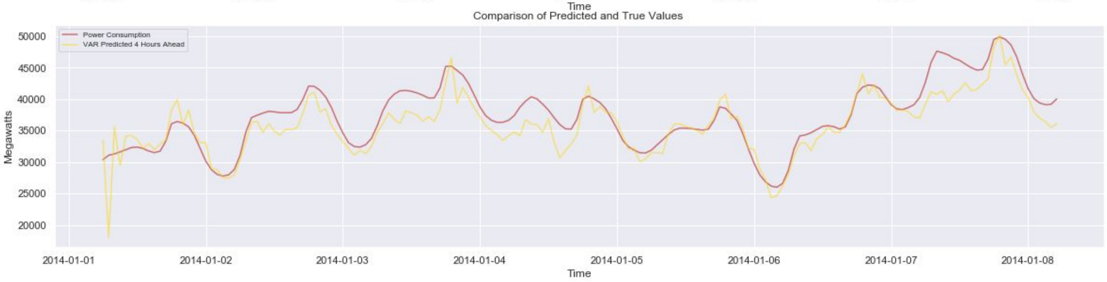
Comparison of Predicted and True Values
Power Consumption
VAR Predicted 1 Hhour Ahead
Megawatts
55000
50000
45000
40000
35000
30000
25000
20000
2014-01
2014-03
2014-05
2014-07
2014-09
2014-11
2015-01
Time



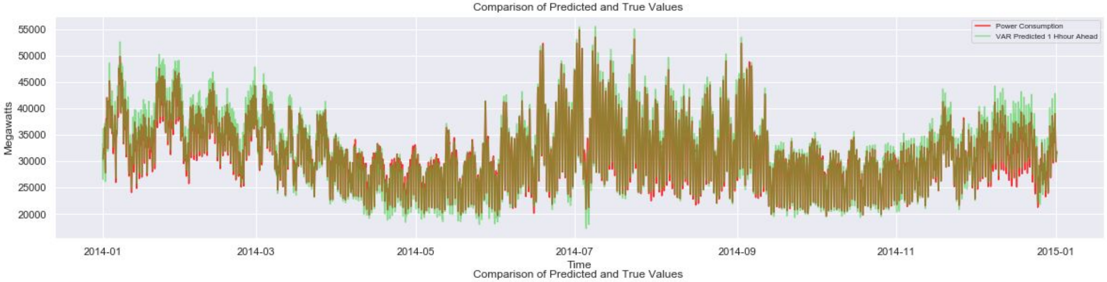
Comparison of Predicted and True Values
Power Consumption
VAR Predicted 1 Hhour Ahead
Megawatts
50000
45000
40000
35000
30000
25000
2014-01-01
2014-01-02
2014-01-03
2014-01-04
2014-01-05
2014-01-06
2014-01-07
2014-01-08
Time

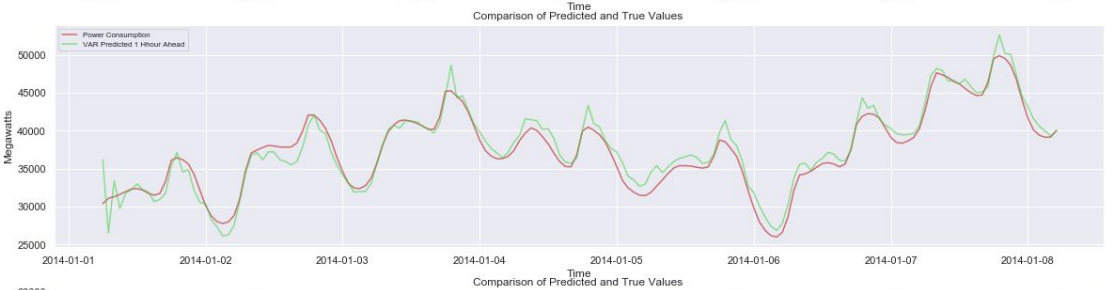


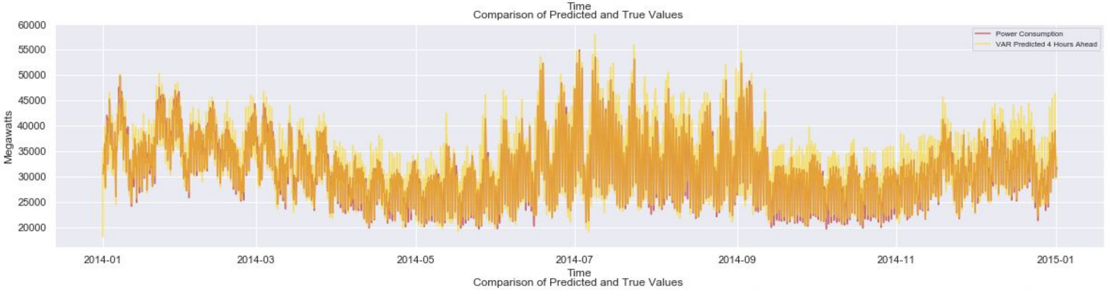
Comparison of Predicted and True Values
Power Consumption
VAR Predicted 4 Hours Ahead
Megawatts
60000
55000
50000
45000
40000
35000
30000
25000
20000
2014-01
2014-03
2014-05
2014-07
2014-09
2014-11
2015-01
Time

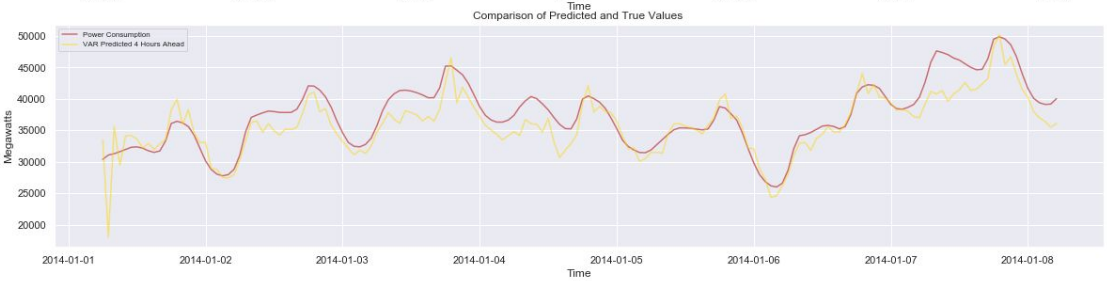


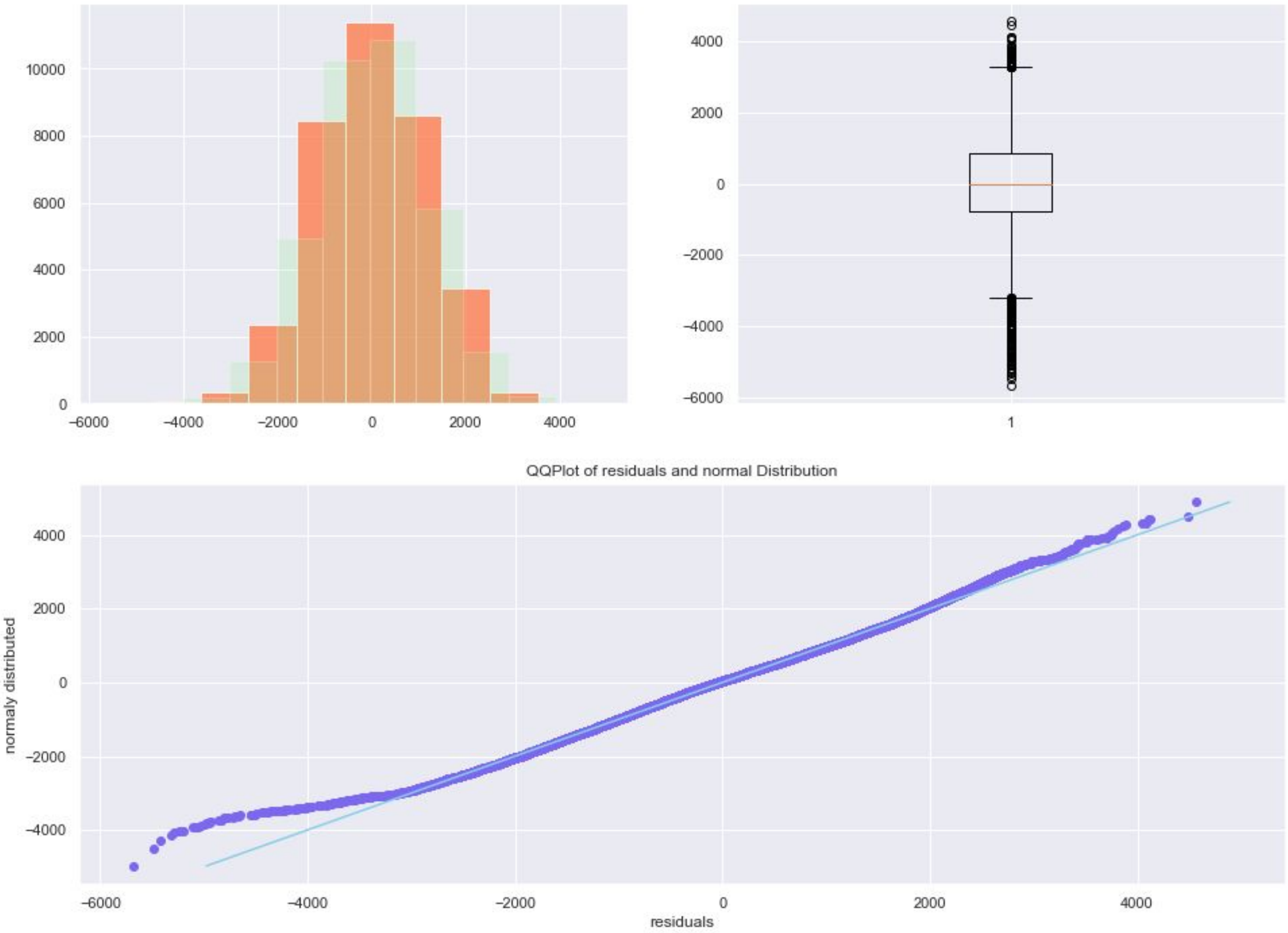
Comparison of Predicted and True Values
Power Consumption
VAR Predicted 4 Hours Ahead
Megawatts
50000
45000
40000
35000
30000
25000
20000
2014-01-01
2014-01-02
2014-01-03
2014-01-04
2014-01-05
2014-01-06
2014-01-07
2014-01-08
Time


Comparison of Predicted and True Values
Power Consumption
VAR Predicted 1 Hhour Ahead
Megawatts
Time

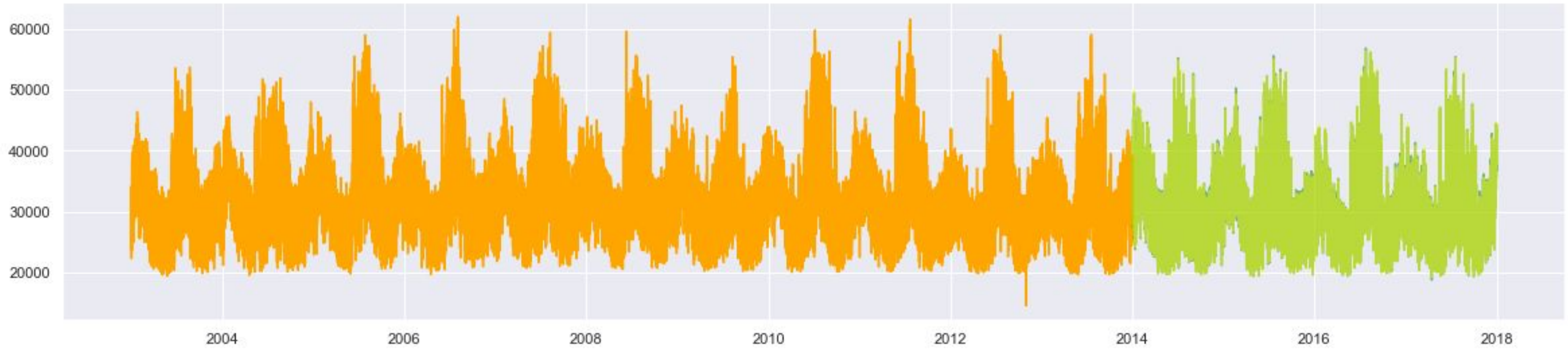

Comparison of Predicted and True Values
Power Consumption
VAR Predicted 1 Hhour Ahead
Megawatts
Time

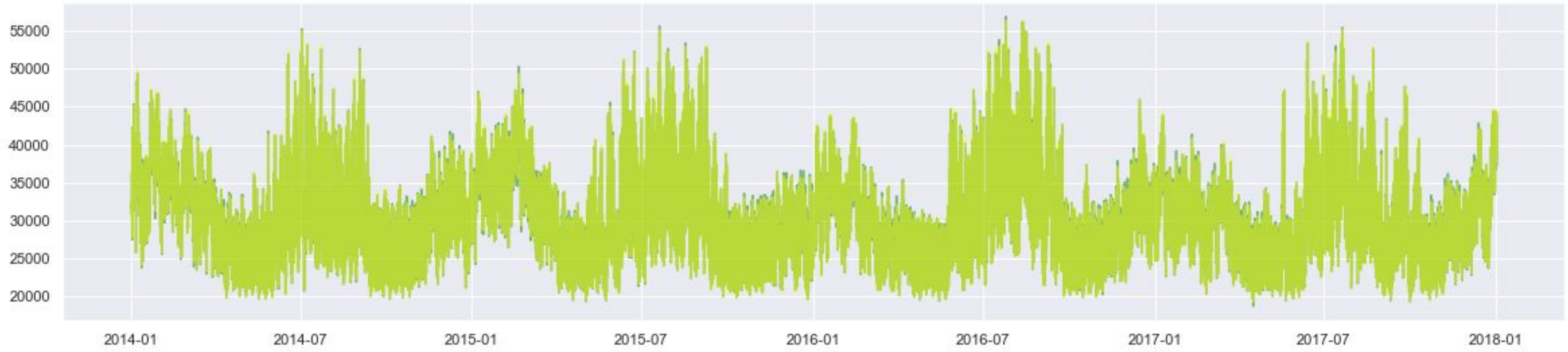

Comparison of Predicted and True Values
Power Consumption
VAR Predicted 4 Hours Ahead
Megawatts
Time

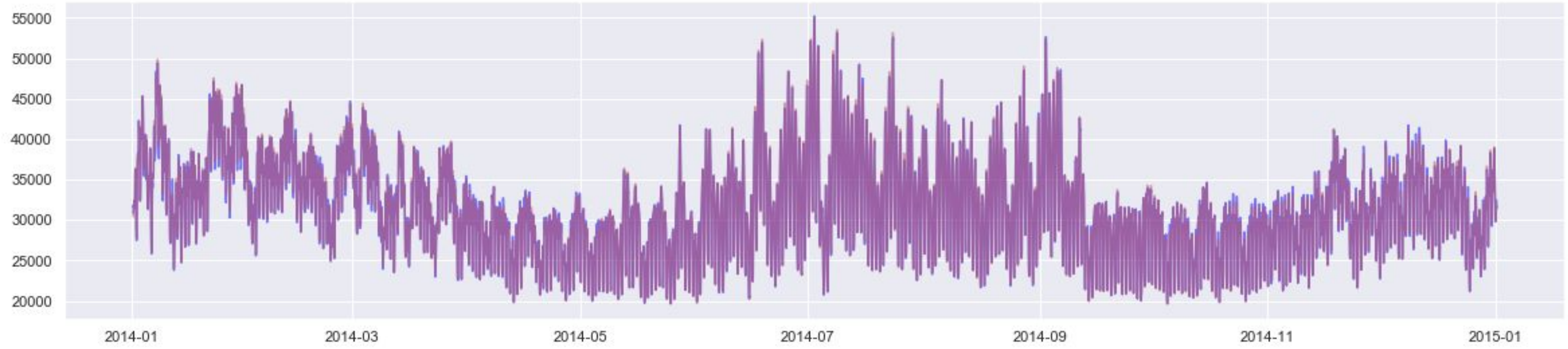

Comparison of Predicted and True Values
Power Consumption
VAR Predicted 4 Hours Ahead
Megawatts
Time

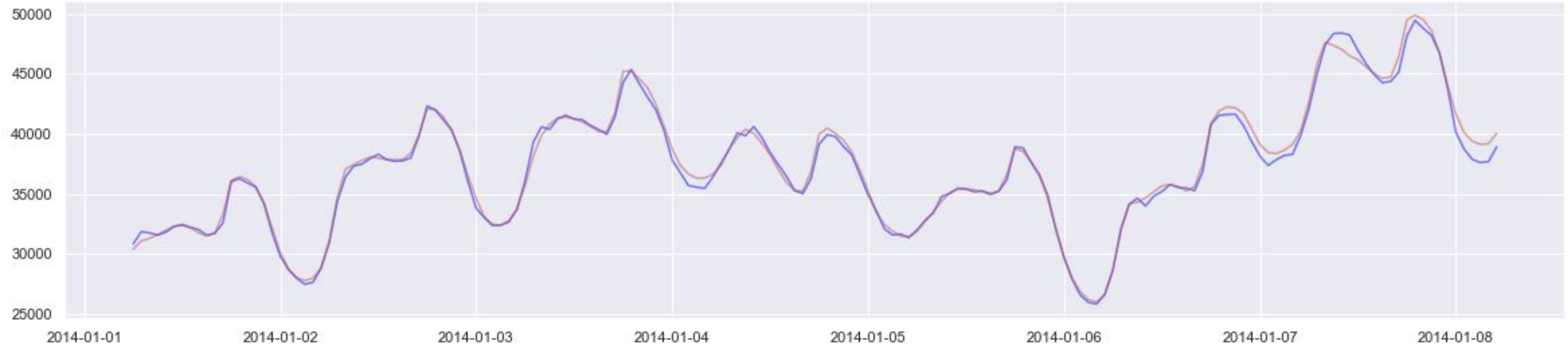

QQPlot of residuals and normal Distribution
normaly distributed
residuals

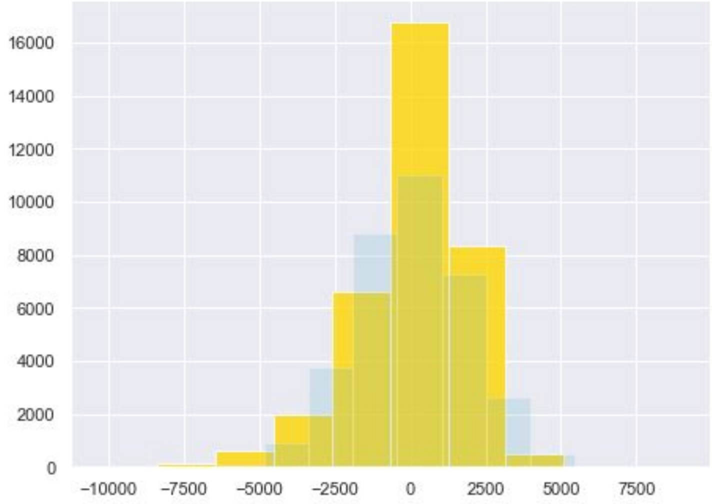
Boosting Model

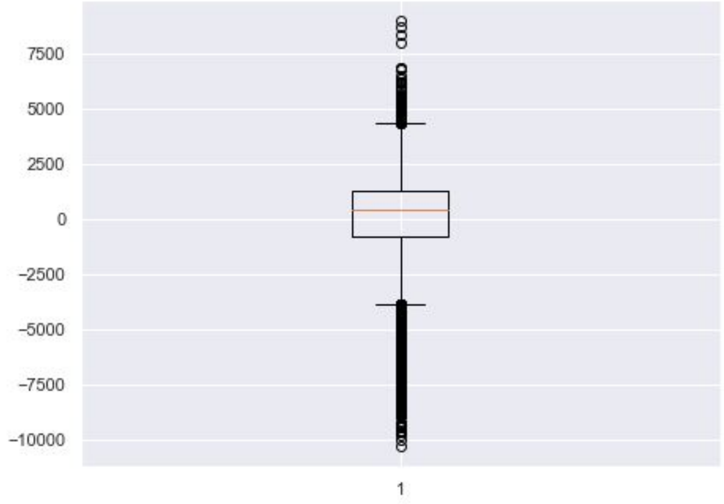

60000
50000
40000
30000
20000
2004
2006
2008
2010
2012
2014
2016
2018

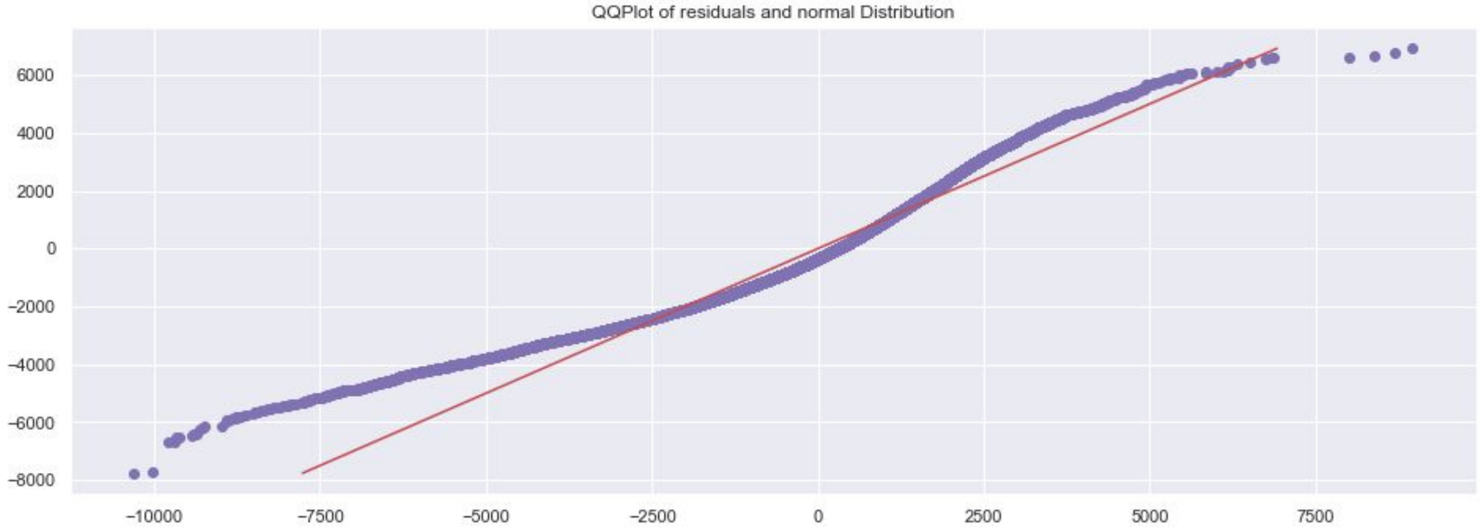

55000
50000
45000
40000
35000
30000
25000
20000
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01


55000
50000
45000
40000
35000
30000
25000
20000
2014-01
2014-03
2014-05
2014-07
2014-09
2014-11
2015-01

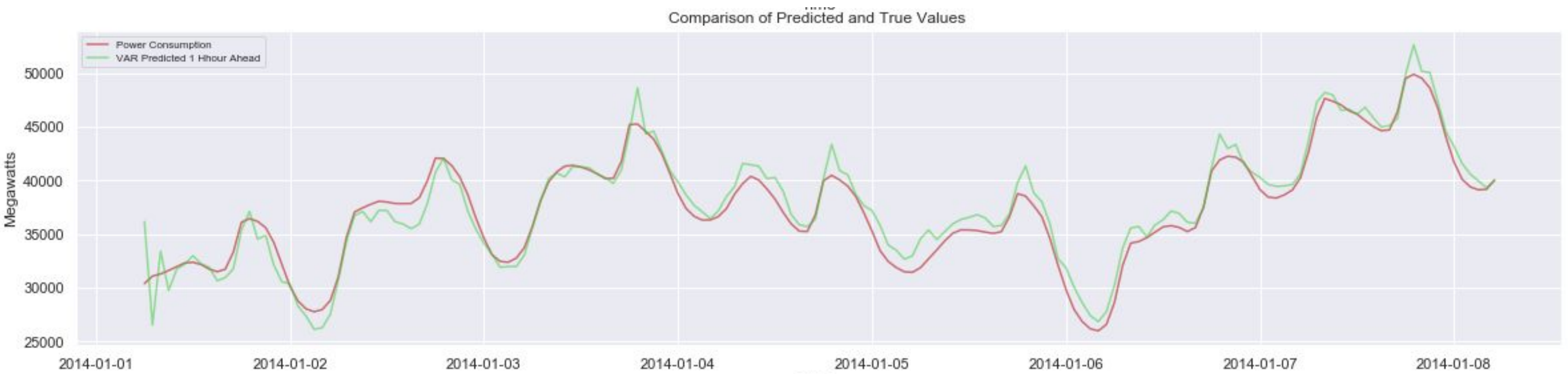

50000
45000
40000
35000
30000
25000
2014-01-01
2014-01-02
2014-01-03
2014-01-04
2014-01-05
2014-01-06
2014-01-07
2014-01-08


16000
14000
12000
10000
8000
6000
4000
2000
0
−10000 −7500 −5000 −2500 0 2500 5000 7500

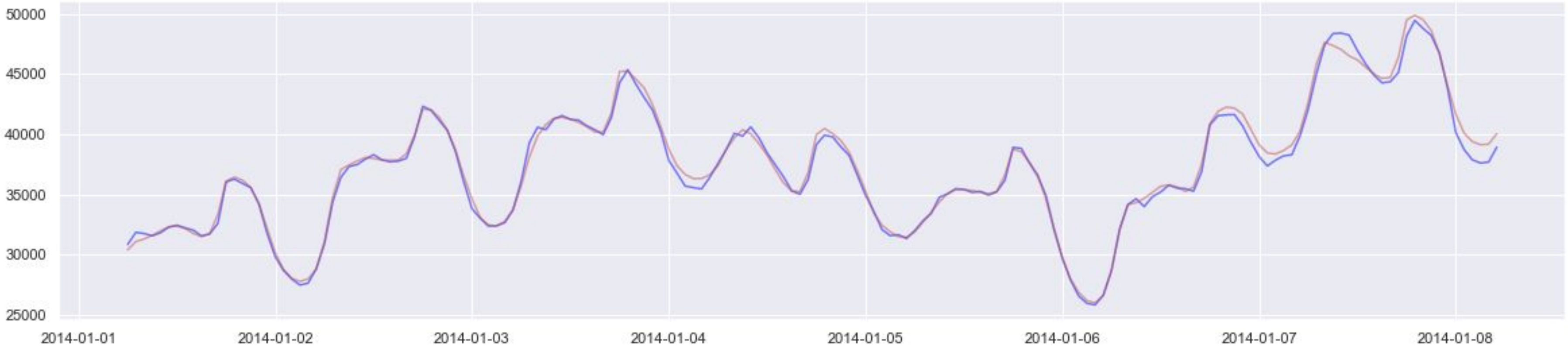

7500
5000
2500
0
−2500
−5000
−7500
−10000
1


QQPlot of residuals and normal Distribution
6000
4000
2000
0
−2000
−4000
−6000
−8000
−10000 −7500 −5000 −2500 0 2500 5000 7500

Model	R-squared	MAPE	RMSE	MSE
VAR 1 hour ahead	0.965	3.11%	1194.70	1427321.67
VAR 2 hours ahead	0.950	3.59%	1431.07	2047961.98
VAR 3 hours ahead	0.902	4.89%	1997.65	3990622.22
VAR 4 hours ahead	0.902	6.46%	2655.94	7054022.48
Boosting With Lag -1	0.997	0.76%	336.40	113168.20
Boosting with no Lags	0.920	4.30%	1810.86	3279224.79



Comparison of Predicted and True Values
Power Consumption
VAR Predicted 1 Hhour Ahead
Megawatts
50000
45000
40000
35000
30000
25000
2014-01-01
2014-01-02
2014-01-03
2014-01-04
2014-01-05
2014-01-06
2014-01-07
2014-01-08



50000
45000
40000
35000
30000
25000
2014-01-01
2014-01-02
2014-01-03
2014-01-04
2014-01-05
2014-01-06
2014-01-07
2014-01-08

Future Improvements on Model

Find a better way to predict the probability and scale of power outages

Try making an ensemble model from the VAR and Gradient Boosting.

Incorporate More Econometric Data such as electricity prices and

Try LSTM Model

Future Experimentation

Use Rooftop data and PVlib to Determine capacity for solar Power Generation

Compare performance on COVID 19 Power consumption

Try LSTM

Use the model to make finer scale predictions on economic data by removing weather related changes