TEAM PROJECT

Part 1

Team 3
Yang Yu, Chaoyang Zheng, Meng Peng & Katherine Gimon-Valencia
Feb. 22/2019

Introduction

This report is the first in a three-part series which will systematically explore various forecasting methods in order to accurately predict daily electricity demand for h=1 from data available on day t. This report specifically analyzes the electricity demand of the Rockland Electric Company (RECO), a subsidiary of the Orange & Rockland Utilities Inc., which serves Northern New Jersey[1]. It forms part of the larger PJM Interconnection, "a regional transmission organization (RTO) that operates a competitive wholesale electricity market and manages the high-voltage electricity grid in [various states]"[2]. RECO is the retailer who sells electricity to the end-user.

Figures 1 and 2 show the areas covered by RECO. These include portions of the Sussex, Bergen and Passaic counties in Northern New Jersey. Overall it is a relatively small area covering the regions surrounding the cities of Alpine, Saddle River, Mahwah, Franklin Lakes and West Milford.

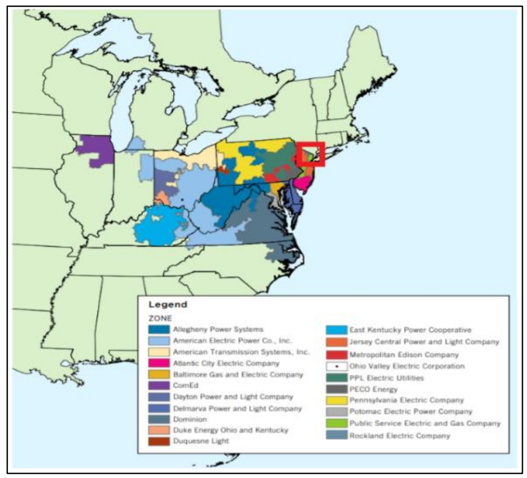


Figure 1

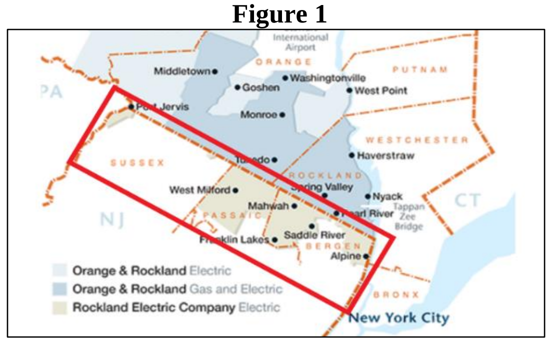


Figure 1

While the small area in Sussex county is predominantly rural and includes public lands such as High Point State Park[3], the rest of the regions within Bergen and Passaic are mainly residential, commercial office spaces and warehouses, including some rural patches. Major

[1] Orange & Rockland, "About Orange & Rockland", Accessed Feb 12, 2019: https://www.oru.com/en/about-us/company-information

[2] PJM, "Who We Are", Accessed Feb 13, 2009: https://www.pjm.com/about-pjm/who-we-are.aspx

[3] Wikipedia, "Montague Township, New Jersey", Accessed Feb. 12, 2019: https://en.wikipedia.org/wiki/Montague_Township,_New_Jersey

industries include transportation and warehousing, life sciences and finance/insurance/administration (see Appendix A)[4]. For example, the economy of Saddle River specializes in management and finance and employs "2.72 and 2.62 times more people than what would be expected in a location of this size"[5]. The area of Alpine, NJ is particularly well-off due to its proximity to Manhattan (20 minutes) and its significant number of millionaires; as such, residential houses are particularly luxurious[6]. In fact, the median household income for Saddle River, Franklin Lakes and Alpine, NJ respectively are $132K, $160K and $115K[7].

In total, there are 30 weather stations in Bergen county, 25 in Passaic county and 31 in Sussex county. The large majority are smaller stations who do not measure all available weather data types (computed relative humidity for example is *only* available at airports). Furthermore, a significant number of stations do not cover the time period from 2005 to present day. This is significant because this is the time period for which historical energy demand is available for RECO.

As such, the only weather stations we could consider were those of Teterboro Airport located in Bergen county (13 miles southwest from the RECO border), and Sussex Airport located within Sussex county (located about 14 miles southeast from the RECO border). As Appendix B illustrates, both airports reported similar weather trends with Sussex in general being slightly colder. However, in terms of proximity to consumers, few are located in Sussex since the area is mainly national parks. As such, because of its closer proximity to electricity users, only the data from Teterboro Airport was considered.

[4] The Official Site of the State of New Jersey, "New Jersey Key Industry Clusters", Accessed Feb 12, 2019: https://www.state.nj.us/state/planning/docs/dfplan_industrysectors.pdf
[5] Data USA, "Saddle River, NJ", Accessed Feb. 12, 2019: https://datausa.io/profile/geo/saddle-river-nj/
[6] NJ.com, "New Alpine Mansion Listed for $25M, Third Most Expensive in N.J." Accessed Feb. 12, 2019: https://www.nj.com/entertainment/index.ssf/2017/04/most_expensive_homes_for_sale_nj_alpine.html
[7] Data USA, Accessed Feb. 12, 2019: https://datausa.io/

Exploratory Data Analysis

Past electricity demand for RECO was available from the PJM website only from 2005 until present day. The decision was made to use the entire data set to maximize the number of data points analyzed.

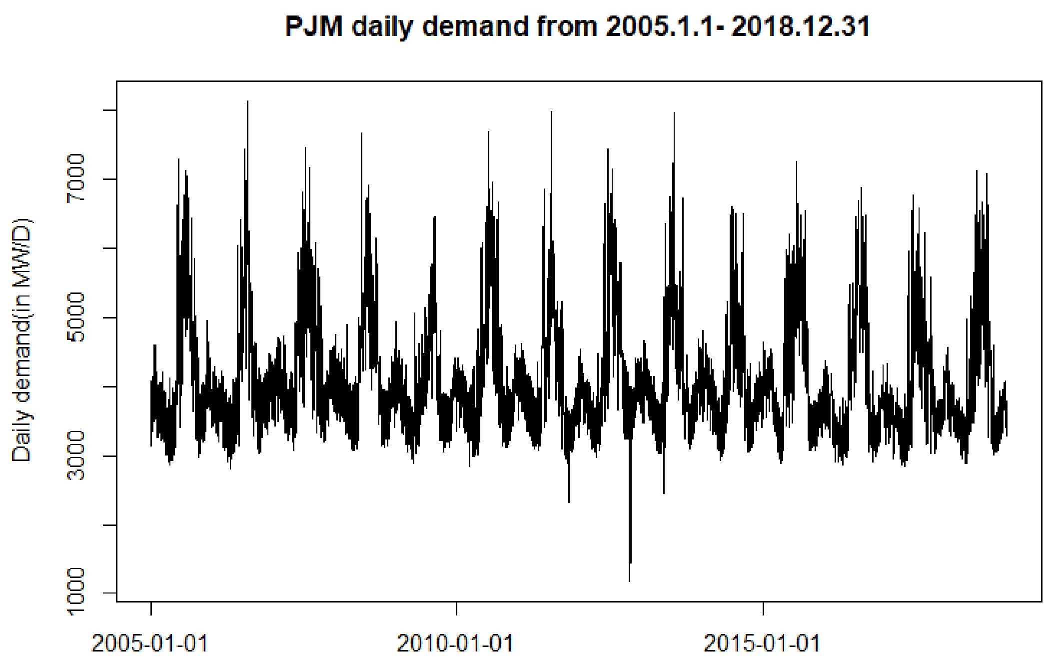


Figure 3

Looking at Figure 3, we can clearly see that there is a stable annual seasonality in the demand. However, as time goes on, there are fewer peaks and the data is slightly more compressed (the highest demands from period 2013-2018 are smaller than 2005-2012). The data shows some particularly low data points, the largest taking place from October 30th to November 3rd 2012 where daily demand dropped to around 800 MW/h. This was not due to a decrease in demand but rather an inability to supply electricity as a result of the broad power outages caused by Hurricane Sandy [8]. At the time, 2.4 million New Jersey households were out of power with the

[8] Dan Eggen, William Branigin (2012, October 30). "Powerful Storm Devastates New York, New Jersey". The Washington Post

outage lasting on average for about 5 days. The second dip in the data was also caused by another outage which took place from October 29th to November 3rd 2011. During that time, heavy wet snow aided by strong winds resulted in tree and power line damage across northern New Jersey causing widespread power outages[9]. Another decrease was noted towards the end of January 2010 when another winter storm caused scattered power outages in Northern New Jersey[10]. Lastly, in May 2013 there was another decrease in demand however we have not been able to identify the cause. For those decreases in demand which were a result of supply-side issues, we propose for Parts 2 & 3 of the project to replace these data points with the average demand of the same day of the previous three years (same day of week).

In looking at a sample of daily demand over a full year (Figure 4), it is clear that a yearly seasonality is present. There is a significant surge in demand between August and September. The lowest demand takes place from March-May and again in October-November. We also note a slight increase between December and February. It

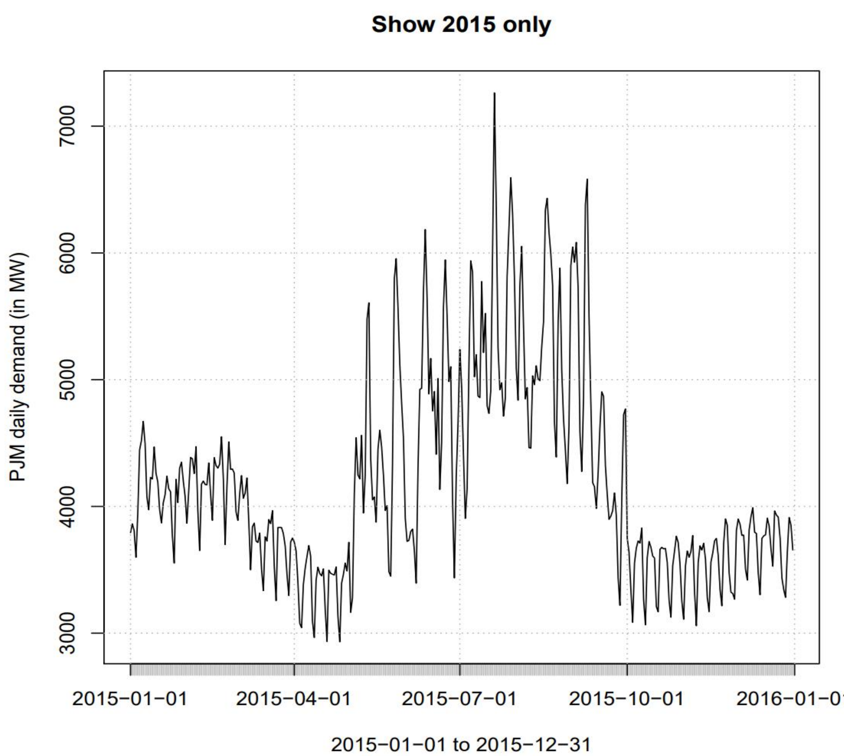


Figure 4

[9] National Weather Service, "Winter Storm Summary for Oct 29, 2011 Event". Accessed Feb. 19, 2019: https://www.weather.gov/phi/10292011wss?fbclid=IwAR3hbl5PhgZnOb7kzpQTPQgFevIlmFr2TwtrAAiaAj2aIuoAPwMP-beJr1A

[10] National Weather Service, "Winter Storm Summary for Jan 30 2010 Event". Accessed Feb 19, 2019: https://www.weather.gov/phi/10292011wss?fbclid=IwAR3hbl5PhgZnOb7kzpQTPQgFevIlmFr2TwtrAAiaAj2aIuoAPwMP-beJr1A

appears that demand for electricity is highest summer, is lower during winter and at its lowest in the spring and fall. This trend implies that residents have a lower tolerance for heat and perhaps a higher tolerance for the cold.

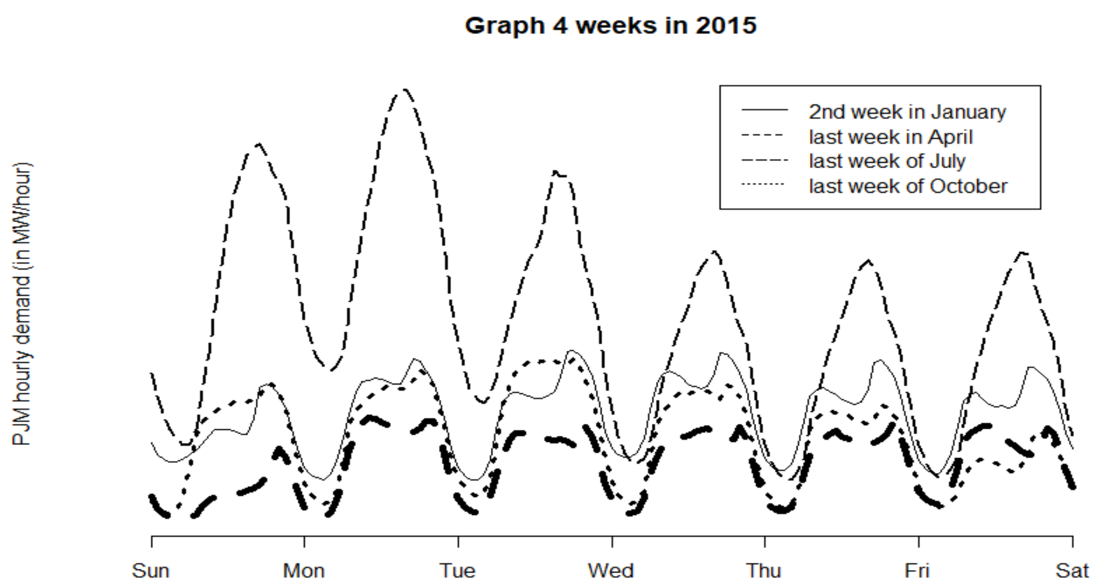


Figure 5

4 weeks from different seasons (when there were no holidays) for the same year were graphed in Figure 5. The first element we note is the presence of a weekly seasonal trend. The demand starts to rise when the day begins and reaches a peak at noon. After that, the demand starts to fall dropping to its lowest in the middle of the night when consumers are sleeping (majority of consumers are residents and commercial office spaces). The demand is slightly higher during weekdays although there are variations depending on the season (summer shows much higher demand from Sundays through Wednesdays, winters show a double peak during the day while remaining relatively stable regardless of the day).

We will divide the entire dataset from January 1st, 2005 to December 31st, 2018 chronologically into training, validation, and testing data sets in order to be able to propose and test various forecasting methods later in the project. We propose splitting the data from January 1st, 2005 to

December 31st, 2013 into the training dataset, and this represents 64% of the data. We feel this amount would provide sufficient data to capture all the different elements (trends, seasonality, peaks) which are underlying. The validation set wold be selected from January 1st 2014 to December 31st 2016, and this representing 21% of the data. Lastly, the test set would be split from January 1st 2017 until December 31st 2018, therefore using the most recent data to evaluate the performance of the models and representing 7% of the data.

Evaluation of Naïve Methods

Three models using Naïve methods were calculated and will be used as a benchmark to evaluate other forecasting methods later in the project- no change, seasonal no change and a rolling three-day moving average. Below is a summary of the results (to see the graphs please see Appendix C).

Method	Entire Data Set (2005-2018)	Validation Data Set (2014-2016)
No-Change Model	bias= -0.0156 %bias= 0.512 mape= 7.32	bias= 0.527 %bias= 0.459 mape= 6.95
Seasonal No-Change Model	bias= 12.4 %bias= 1.64 mape= 12.3	bias= 49.8 %bias= 2.49 mape= 11.7
Rolling Three-Day Moving Average	bias= -0.0111 %bias= 1.02 mape= 9.87	bias= 1.52 %bias= 0.929 mape= 9.29

The No-Change Model was built by using the electricity demand for yesterday to predict the electricity demand for today. In looking at the graph, the predicted value generally aligns with the observed value. Calculated over the entire data set, the bias is -0.0156 and MAPE of 7.32.

The Seasonal No-Change Model was built using the demand from the same date of the previous year to predict the demand for today. Overall we see that most of the observed data and the predictions share the similar pattern generally sharing the same peaks and bottoms. However, the forecasts are systematically lower than the observed data points. The bias and MAPE of this method are 12.4, 12.3 showing that this method is not very precise which might be explained by the variance in temperatures over the two years.

The Rolling 3-Day Moving Average Model was built by using the average demand for electricity of today, the day before, and two days before to predict the electricity demand for tomorrow. It provides a bias of -0.0111 % and an MAPE of 9.87. We see that the predicted value has less variation and is smoother than the observed value but fails to capture the peaks and valleys at the appropriate moments (peaks are not high enough and come too late). Thus it generates a higher MAPE.

Comparing all three methods, the no-change forecast is the most precise with the lowest MAPE. The reason could be that there is a strong correlation between the day and the previous day. On the other hand, seasonal no-change forecast method is the worst likely due to the changes in temperature and also the day of the week of the previous year is not the same.

Explanatory Variables

Holidays and Days Close to Holidays

For the complete list of Federal and State Holidays please see Appendix D. When considering how Holidays can affect demand for electricity, we will consider both fixed and variable Holidays and days around them. For example, we mark December 31 and January 2 as the days around New Year's Day and December 24 and December 26 as the days around Christmas. We considered the day of week as well. For example for Good Friday we considered that a number of companies would also consider Easter Monday as a Holiday.

In addition, conditions were set for variable Holidays such as Independence Day, because in addition to the variability of when companies may provide the day off, employees will often try to group Holidays with vacation to get an extra long weekend. For example, if the Holiday is a Monday, Wednesday, and Friday, we only mark those days, but if it falls on a Tuesday, we also the Monday, if it falls on a Thursday we also mark the Friday and if it is a weekend, we mark the Friday and Monday.

Heating Degree Days and Cooling Degree Days

The National Weather Center recommended the use of 65 Fahrenheit as the reference temperature to calculate HDD and CDD. However, as we see from the graph in Figure 6, it appears that there is a portion in the middle where there is not much demand for electricity

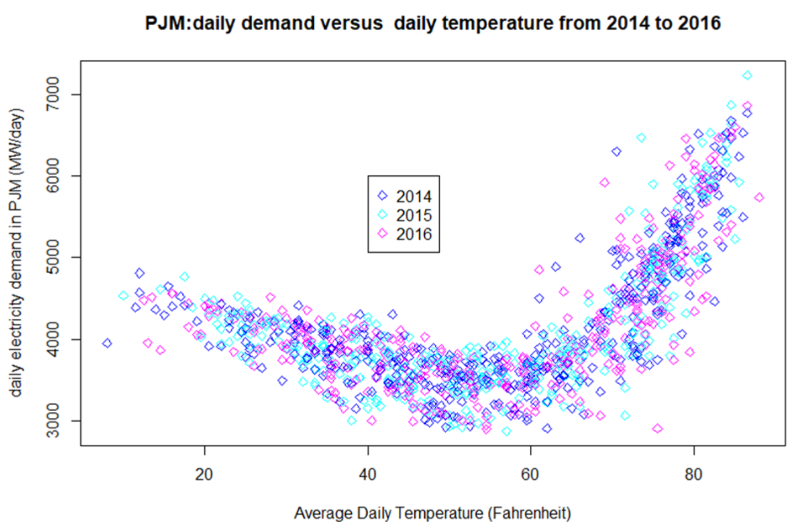


Figure 6

and that in fact the temperature for our zone may be different based on consumer behaviour. As such, in later parts of this report, further analysis will be conducted to determine what this ideal temperature is.

Windchill

We will consider windchill as an explanatory variable. For now, we will use 65 Fahrenheit as the reference temperature but as further analysis is conducted on the ideal HDD temperature for our particular area, this may change. Already we are seeing in the graph that demand for electricity and wind chill are positively related when wind chill is greater than approximately 50 °F. However, when wind chill is below 50, the trend is absent. There are some outliers shows that the demand is abnormally high when wind chill is low. The outlier will be identified and treated in the further analysis.

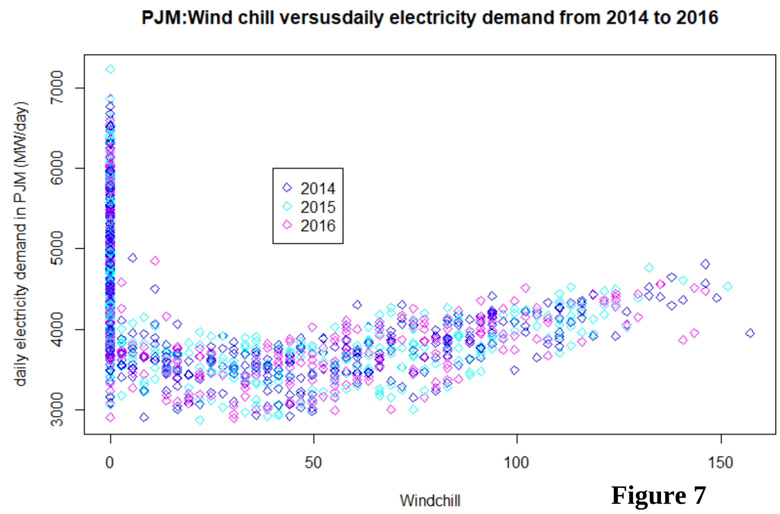
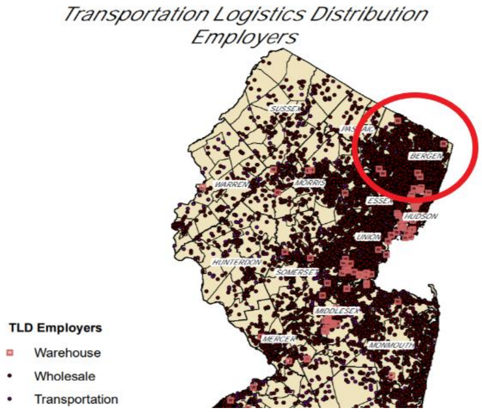


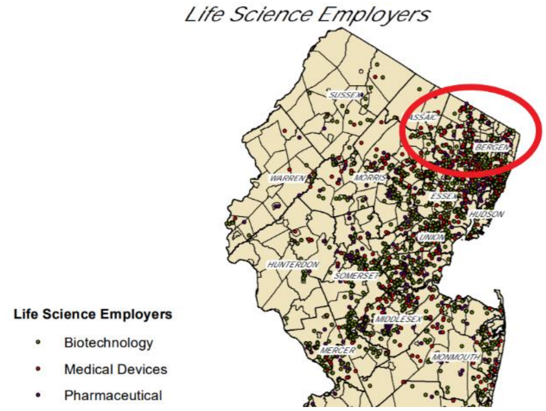
Figure 7

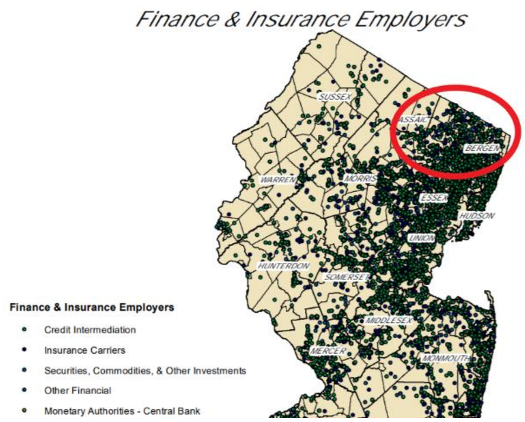
Relative humidity

The relative humidity was obtained hourly however it is still unclear at this point what its relationship is to temperature, dew point and electricity. We propose calculating a weighted average of the relative humidity over daytime hours (since this is when people are awake and consuming the most electricity) however this will be further explored in parts 2 and 3.

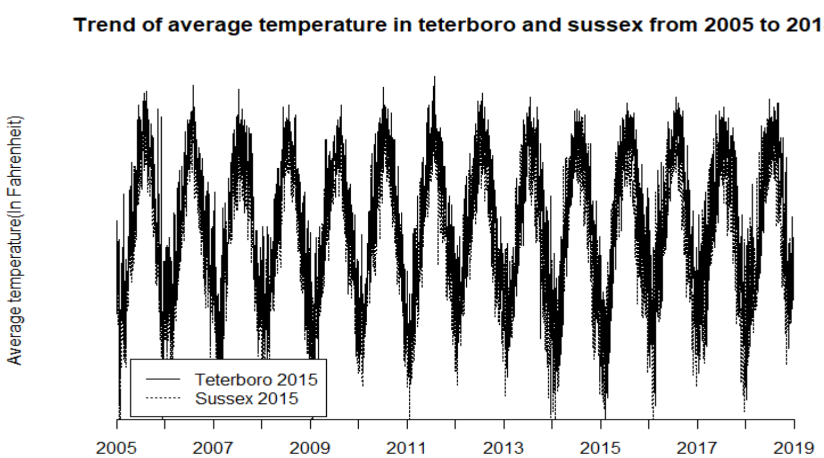
Appendix A: Major Industries in New Jersey



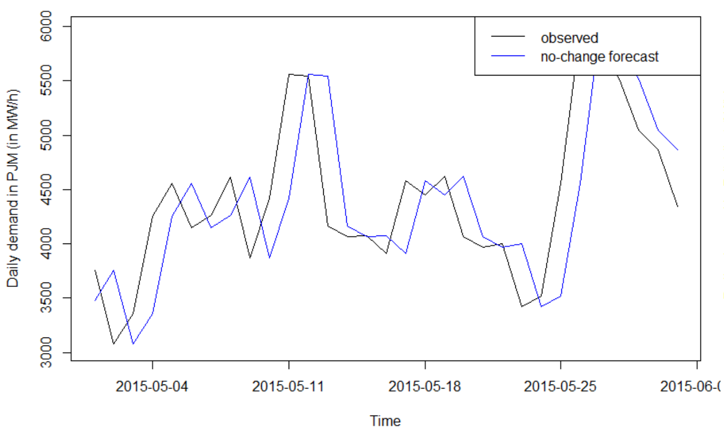


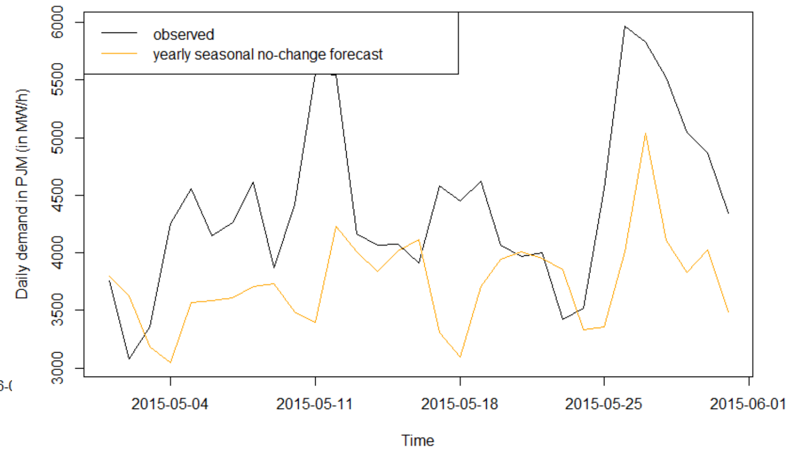


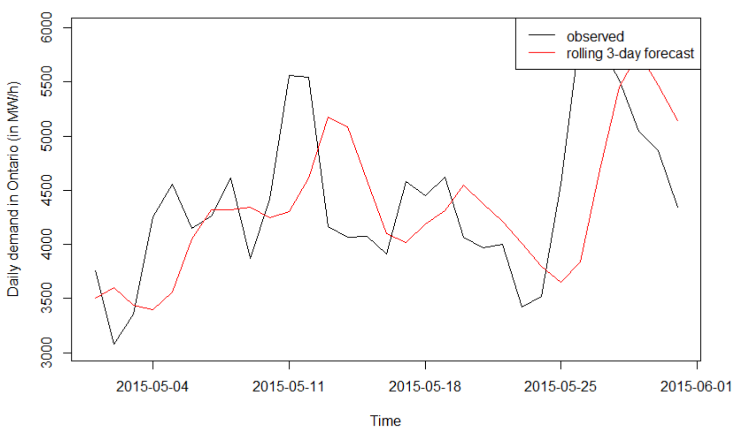
Appendix B- Average Temperature: Teterboro vs Sussex Airports

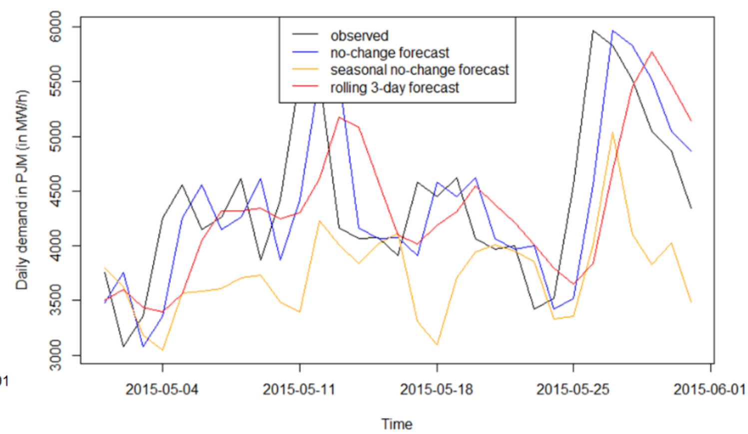


Appendix C- Naïve Model Graphs for Sample 2015









Appendix D- New Jersey State and Federal Holidays[11]

- January 1 - New Year's Day
- 3rd Monday in January - Martin Luther King Day
- 3rd Monday in February - Washington's Birthday (called "Presidents Day" in New Jersey)
- variable date - Good Friday
- Fourth Monday in May - Memorial Day
- July 4 - Independence Day
- 1st Monday in September - Labor Day
- 2nd Monday in October - Columbus Day
- Tuesday after 1st Monday in November - Election Day
- November 11 - Veterans Day
- 4th Thursday in November - Thanksgiving
- Friday following 4th Thursday in November - Day After Thanksgiving (this used to be a state holiday for all branches of government; and it is sometimes still proclaimed as a holiday for the Judicial branch of government, usually not until November.)
- December 25 - Christmas Day

[11] NJ Gov, “State Holidays”. Accessed February 20, 2019: https://www.nj.gov/nj/about/facts/holidays/