As Los Angeles develops both its infrastructure and its population, it will become increasingly important to prepare for the energy needs of one of America's largest cities. The Los Angeles Department of Water and Power (LADWP) is one of the major providers of electricity in the LA metropolitan area and every hour they have to be ready to supply the power that allows their customers to operate their lives.

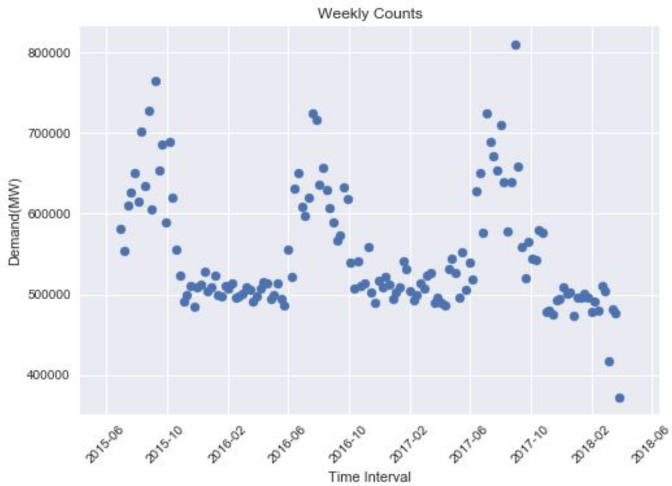
The Energy Information Administration (EIA) collects demand data and predictions from the LADWP on an ongoing basis and makes it publicly available through their website and API. Using this data, I set off to create a model that predicts hourly demand with accuracy that is equal to or greater than the predictions of the LADWP. Such a model would allow the LADWP to prepare for daily demand more accurately and prevent power outages or unnecessary expenditure of resources for a lower demand day.

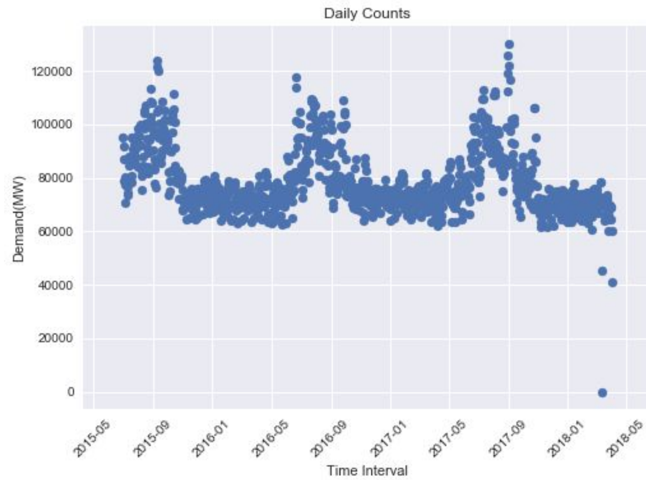
Aside from the raw demand data provided by the EIA, which contains hourly demand for energy in Megawatts (MW) as well as the LADWP Forecast figures for demand, I engineered features to account for factors such as day of the week, holidays, daylight, and temperature. Additionally, lag features were utilized given the temporal nature of the data to account for the pattern of demand leading up to the hours being predicted.

The raw data was downloaded in bulk from the EIA website and the day of the week feature was created from this data using the 'Datetime' library. Holidays were gathered using the 'holidays' library from pipy which accepts a datetime object and outputs whether or not that date is a holiday in the specified region. Weather and daylight data was scraped from the Weather Underground website which provides historical records of these features.

My approach to the data was to group the hourly data into days and weeks to extract the features that are constant throughout the day, like day of the week. These features would be propagated through each hour of the day in order to map to the original data. Then, using the days grouping queried the hourly weather and daylight data from the Weather Underground site as the applet only accepts date and not time.

With the data wrangled, my first thought was to look at the trend of demand across time from different perspective to see if there was a pattern. It did indeed have a cyclical, seemingly seasonal pattern with consistent peaks and valley like below:

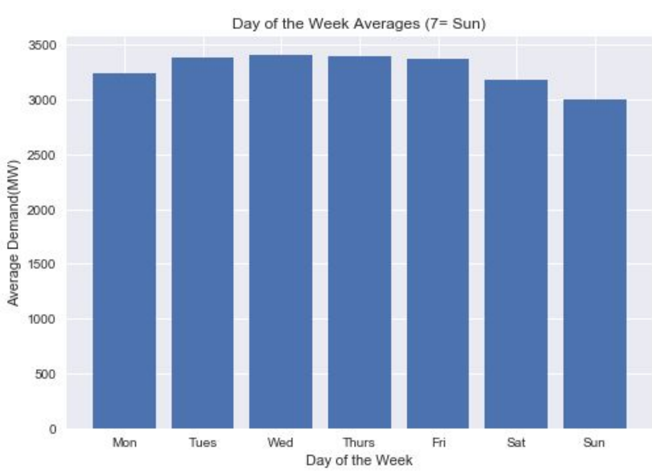


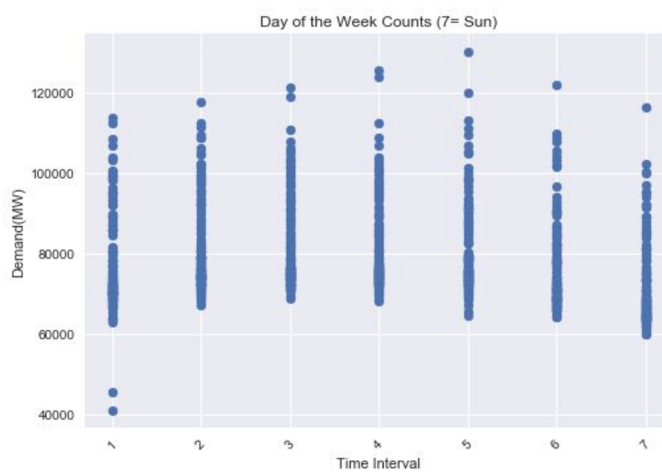


The first chart above, 'Weekly Counts', displays the total Demand of each week over the 2+ year time span that the data covers. We can see that there is a clear cyclical pattern with Demand rising considerably in the Summertime.

The second chart, ‘Daily Counts’ displays total Demand of each day which illustrates the same pattern as the Weekly Counts but at a higher resolution. We can see that there are some outliers however the same general pattern is present with Demand being higher in Summer time.This lead me to believe that temperature may indeed play an important role in predicting demand, however I will cover the importance features when discussing the modeling phase.

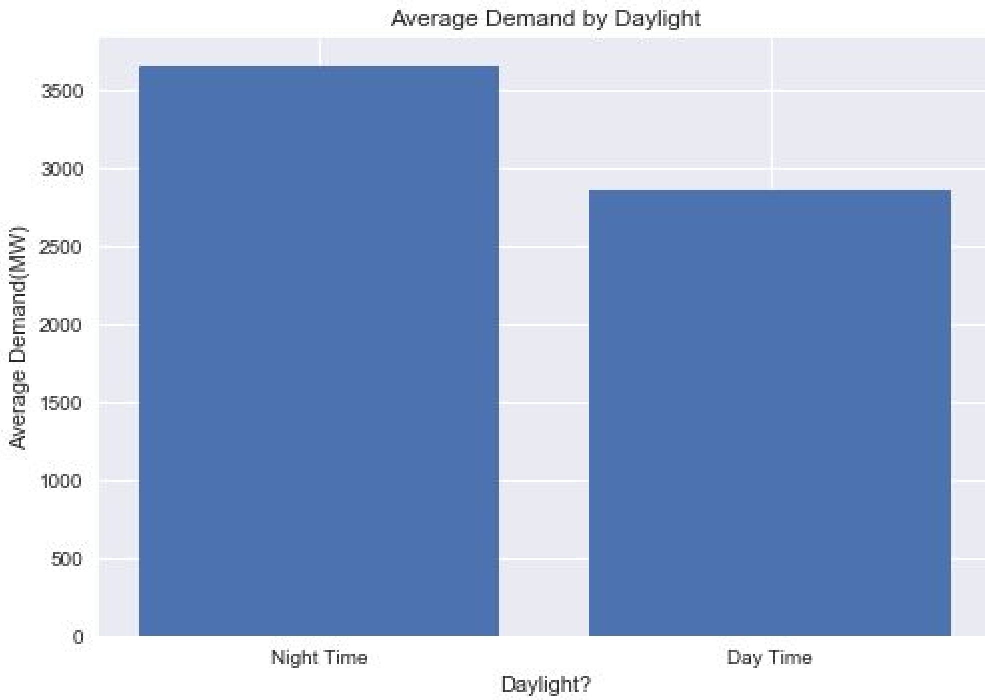
Next, I examined the average demand of each day of the week from 2 perspectives, on average and as a whole:





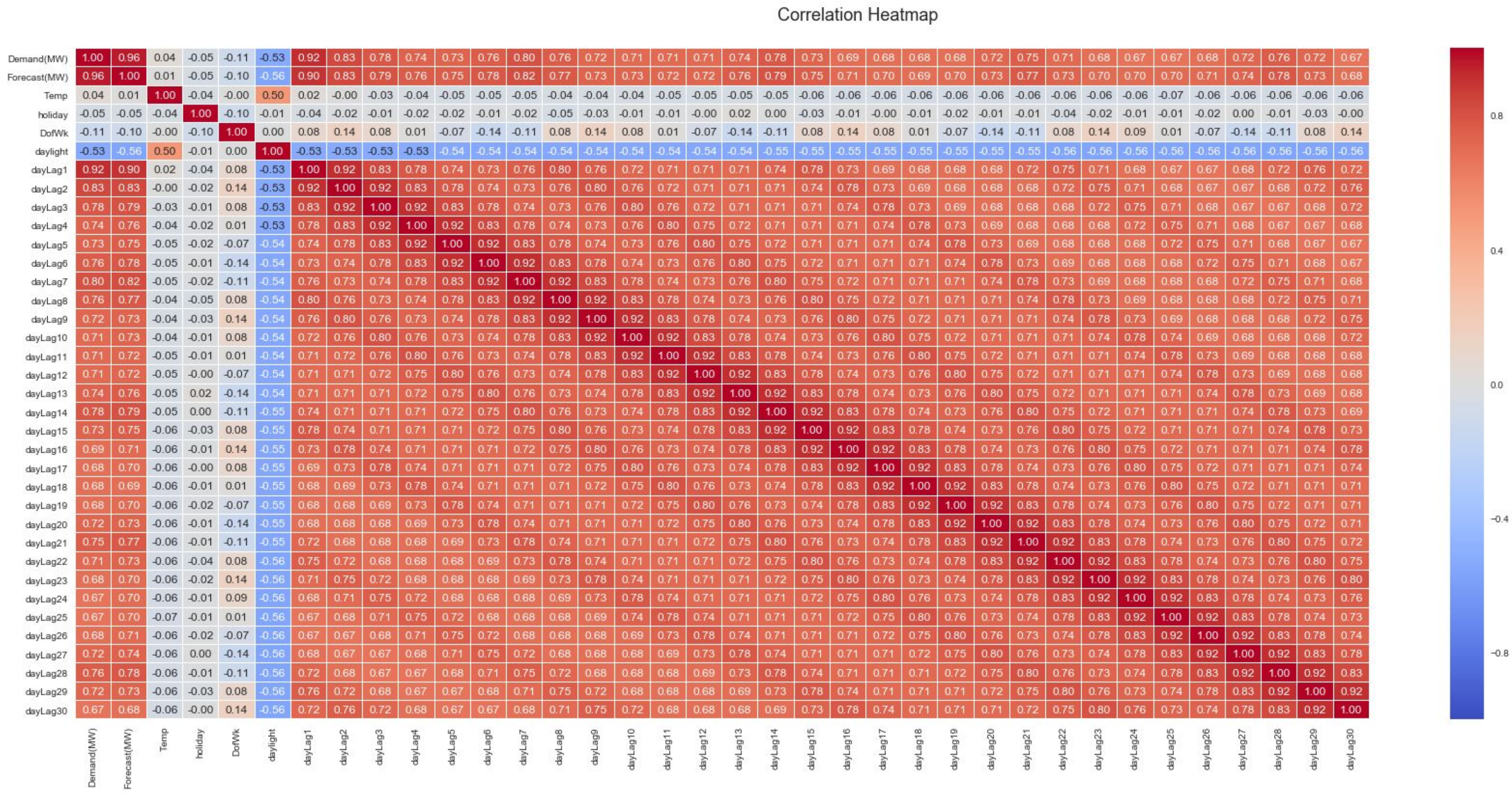
we can see that Demand is relatively even across the 7 days of the week. Sunday has the most outliers and Friday seems to have the highest Demand, but the concentration of data is very similar across all days of the week.

As for the scraped feature, Temperature appeared to have no correlation with Demand and holidays had little impact (however Demand went down on holidays which was contradictory to my original thoughts). What did have a noticeable impact was Daylight:



You can see that there is an intuitive Demand pattern here with a higher average Demand occuring at night when lights and heating is needed. These features provide some insight into the nature of electricity Demand, however for predictive purposes we must consider as much data as possible, namely the past.

Lag features were created on a DAY-1 basis such that the first hour of yesterday is a predictor of the first hour of today, and so on and so forth all the way back 30 days. These lag features were then correlated with Demand and sure enough proved to be very valuable:



You can see by the coloration that the lag features are highly correlated with Demand which is the top row and furthest left column. This makes it clear that there is a pattern that comes through across at least a 30 day time span, though it is likely that is may extend even further than that. For now, using 30 days of lag features should prove to be sufficient and coherent for predicting the next 24 hours, however further research into the value of extending the time span of the lag features may provide some additional predictive power.