Here we perform some initial exploratory data analysis (EDA) for both the Seattle and Los Angeles weather/electricity data sets. One of the first questions we ask is what is the relationship between the independent variables (the weather features) and the dependent variable (electricity demand measured in MWh). Pandas has a nice method (the .corr() method) for finding the Pearson correlation coefficients between all the different variables. **Figure 1** shows the sorted Pearson correlation coefficients for both Seattle and Los Angeles. The results from the Pearson correlation coefficients between Seattle and Los Angeles are interesting. Namely that cooling degree days (CDD) are a significant, positive correlation and heating degree days (HDD) are a significant, negative correlation for LA, whereas the exact opposite is true for Seattle. **Figure 2** shows how the electricity demand varies in time for both LA and Seattle and sheds light on why we observe this behavior. Seattle's electricity demand spikes in the winter months, indicating that the majority of power is used for heating, whereas LA's electricity demand spikes in the summer months, indicating that the majority of power is used for cooling. This is consistent with the computed Pearson correlation coefficients, but a more thorough multivariate regression would be informative.

DEMAND CORRELATIONS (PEARSON) FOR LA	
dailycoolingdegreedays	0.572275
hourlydewpointtempf	0.383226
hourlyrelativehumidity	0.365116
hourlywetbulbtempf	0.302208
hourlycoolingdegrees	0.192912
hourlydrybulbtempf	0.044278
hourlyvisibility	0.009670
hourlyprecip	-0.022645
dailyheatingdegreedays	-0.221475
hourlywindspeed	-0.232047
hourlystationpressure	-0.265702
hourlysealevelpressure	-0.266134
hourlyaltimetersetting	-0.267439
hourlyheatingdegrees	-0.290998

DEMAND CORRELATIONS (PEARSON) FOR SEATTLE	
hourlyheatingdegrees	0.582695
dailyheatingdegreedays	0.574442
hourlyrelativehumidity	0.269787
hourlyprecip	0.061526
hourlywindspeed	0.030342
hourlysealevelpressure	0.005554
hourlyaltimetersetting	-0.027498
hourlystationpressure	-0.034864
hourlyvisibility	-0.068125
hourlycoolingdegrees	-0.079621
dailycoolingdegreedays	-0.115990
hourlydewpointtempf	-0.519638
hourlydrybulbtempf	-0.552038
hourlywetbulbtempf	-0.580742

Figure 1. Pearson correlation coefficients for both LA (left) and Seattle (right). As discussed, CDD are positive correlations for LA, while they are negative in Seattle, and vice versa for HDD.

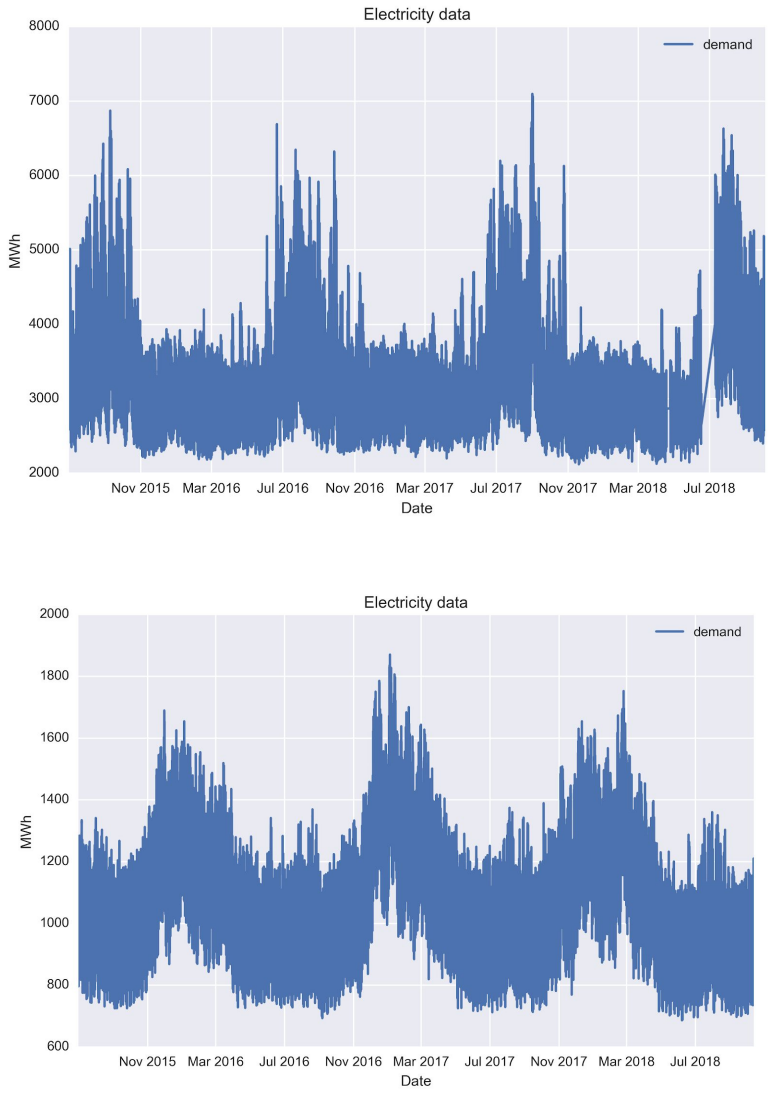

Figure 2. Electricity demand versus time for both LA (top) and Seattle (bottom). Demand peaks in the winter in Seattle, whereas it peaks in the summer in LA.

Next, we are interested in the significance of the CDD and HDD correlations since they are the strongest. Using bootstrapping statistics with ten thousand simulations, we compute p-values of exactly zero for all HDD and CDD features, indicating strong confidence ($p < 10^{-4}$) that the correlations are significant.

Additionally, **Figure 3** shows computed r^2 values for both the Seattle and Los Angeles data sets. Here we can see the relative power of each of the features for both data sets. Once again, HDD is a significant feature for Seattle while CDD is a significant feature for LA.

DEMAND CORRELATIONS (r^2) FOR LA

	col	r^2
9	dailycoolingdegreedays	0.327499
7	hourlydewpointtempf	0.146862
13	hourlyrelativehumidity	0.133310
5	hourlywetbulbtempf	0.091330
1	hourlyheatingdegrees	0.084680
0	hourlyaltimetersetting	0.071524
3	hourlysealevelpressure	0.070827
12	hourlystationpressure	0.070597
11	hourlywindspeed	0.053846
10	dailyheatingdegreedays	0.049051
6	hourlycoolingdegrees	0.037215
4	hourlydrybulbtempf	0.001961
8	hourlyprecip	0.000513
2	hourlyvisibility	0.000094

DEMAND CORRELATIONS (r^2) FOR SEATTLE

	col	r^2
1	hourlyheatingdegrees	0.339533
5	hourlywetbulbtempf	0.337261
10	dailyheatingdegreedays	0.329984
4	hourlydrybulbtempf	0.304746
7	hourlydewpointtempf	0.270023
13	hourlyrelativehumidity	0.072785
9	dailycoolingdegreedays	0.013454
6	hourlycoolingdegrees	0.006340
2	hourlyvisibility	0.004641
8	hourlyprecip	0.003785
12	hourlystationpressure	0.001215
11	hourlywindspeed	0.000921
0	hourlyaltimetersetting	0.000756
3	hourlysealevelpressure	0.000031

Figure 3. r^2 values for both LA (left) and Seattle (right).