

Report

Team __

2024-01-28

NA's

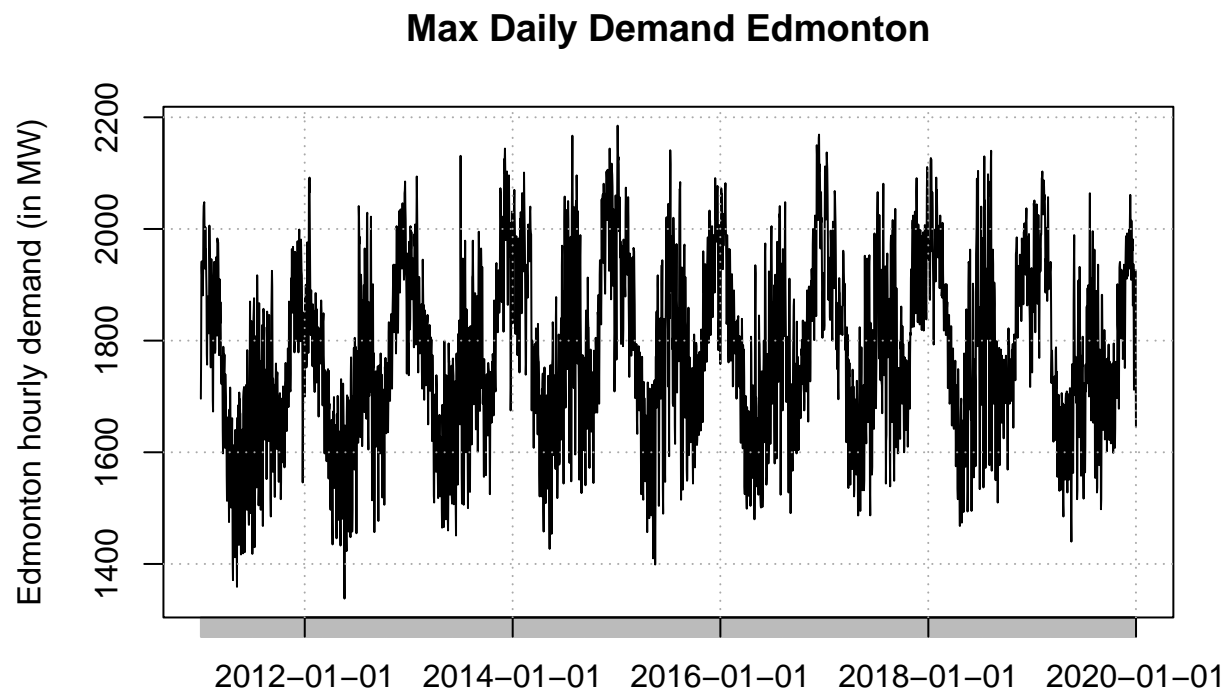
These are all the records with missing values:

##		DT_MST	South	Northwest	Northeast	Edmonton	Calgary	Central
##	SOUTH	2011-03-13	NA	NaN	NaN	NaN	NaN	NaN
##	52610	2012-03-11	NA	NaN	NaN	NaN	NaN	NaN
##	52611	2013-03-10	NA	NaN	NaN	NaN	NaN	NaN
##	52612	2014-03-09	NA	NaN	NaN	NaN	NaN	NaN
##	52613	2015-03-08	NA	NaN	NaN	NaN	NaN	NaN
##	52614	2016-03-13	NA	NaN	NaN	NaN	NaN	NaN

This link shows daylight savings dates in Canada, they coincide with the missing values: 'https://www.timetemperature.com/canada/daylight_saving_time_canada.shtml'

The rest of the data does not have missing values. These data points were removed from the dataset for further steps as those timestamps do not really exist.

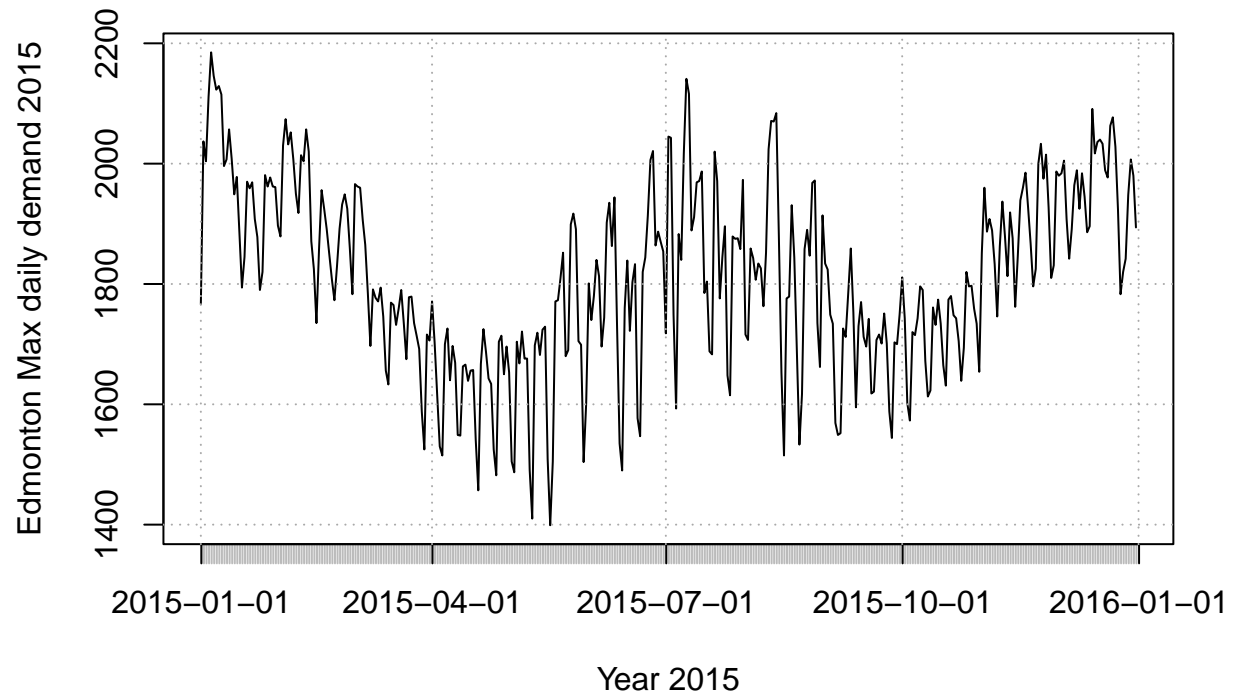
Timeseries Plots



Observations:

- Upward trend from 2011 to 2015
- Outliers to investigate: One point in january 2012

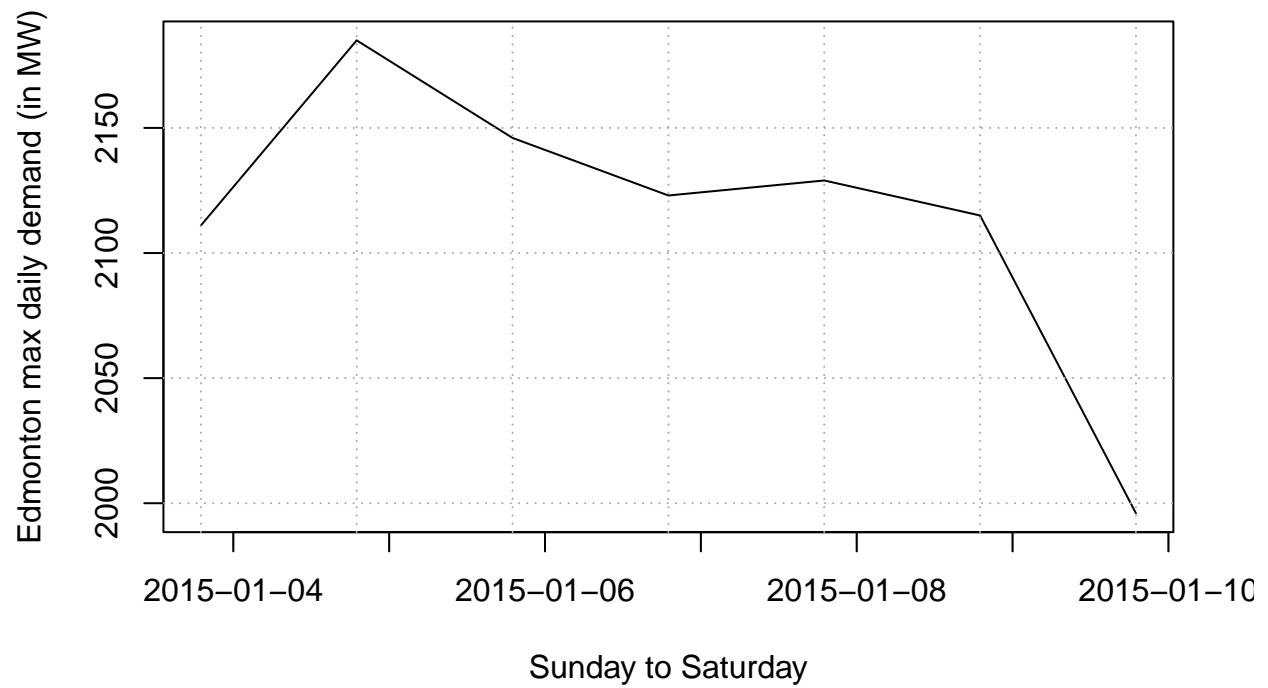
Max Daily Demand Edmonton 2015



Observations:

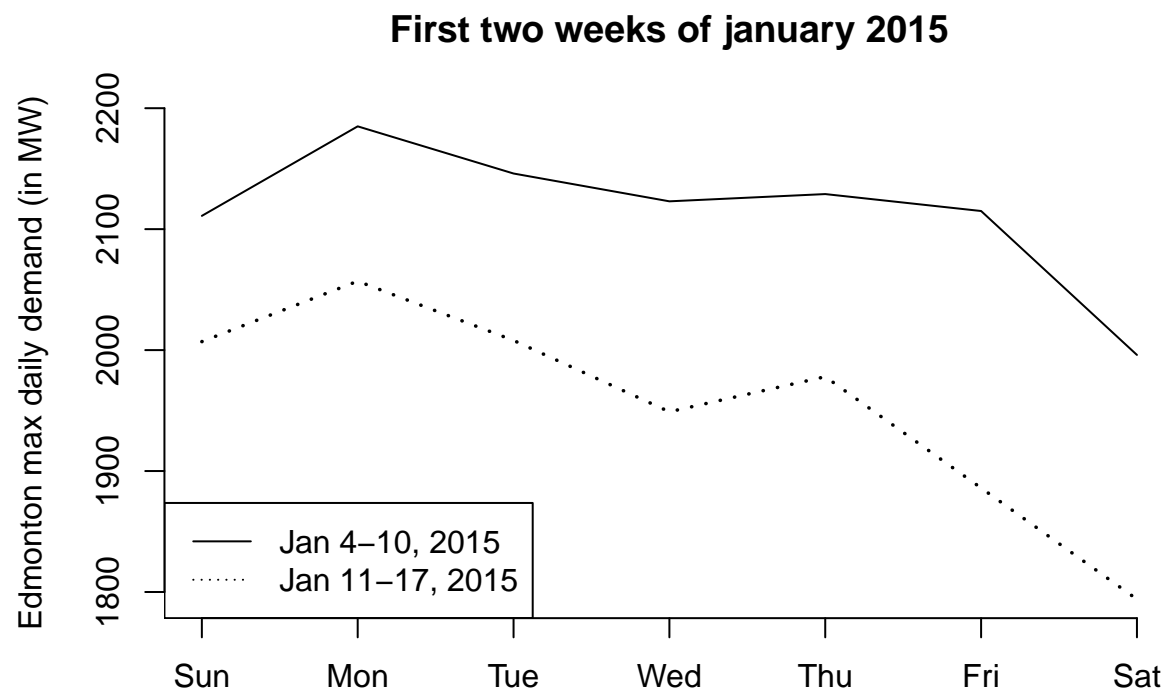
- Seasonal pattern
- Highest max consumption in summer and winter
- Highest variability in summer

First week of 2015



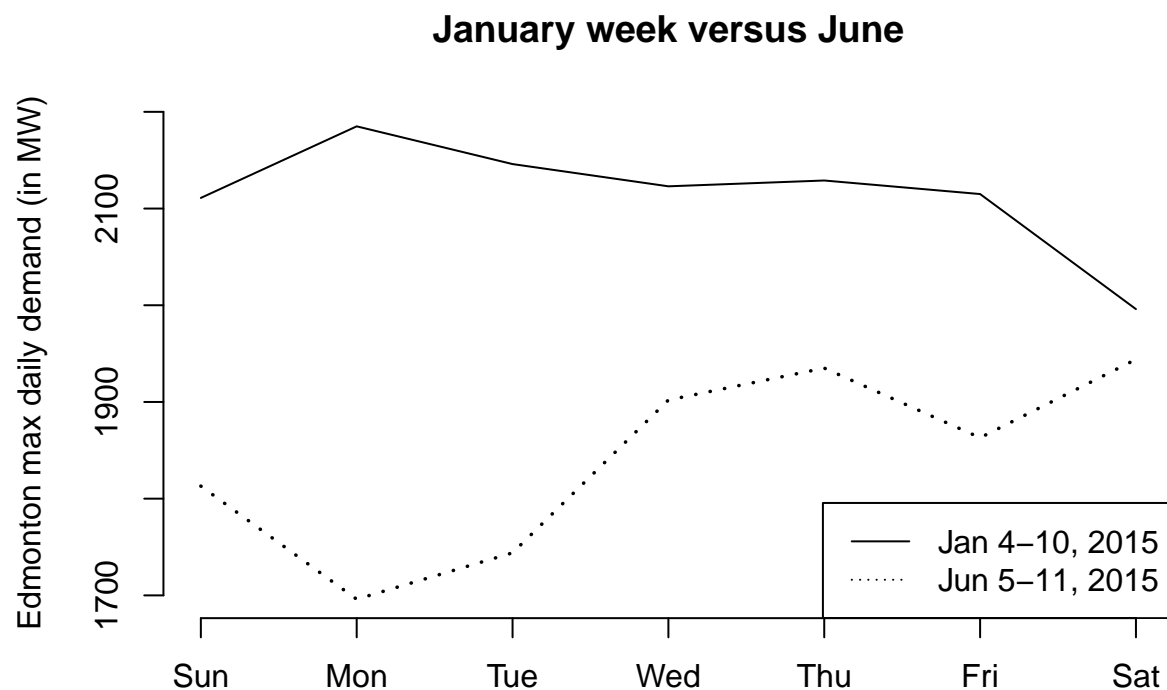
Observations:

- High max consumption on monday
- Low ma consumption on saturday



Observations:

- Similar patterns for both weeks



Observations:

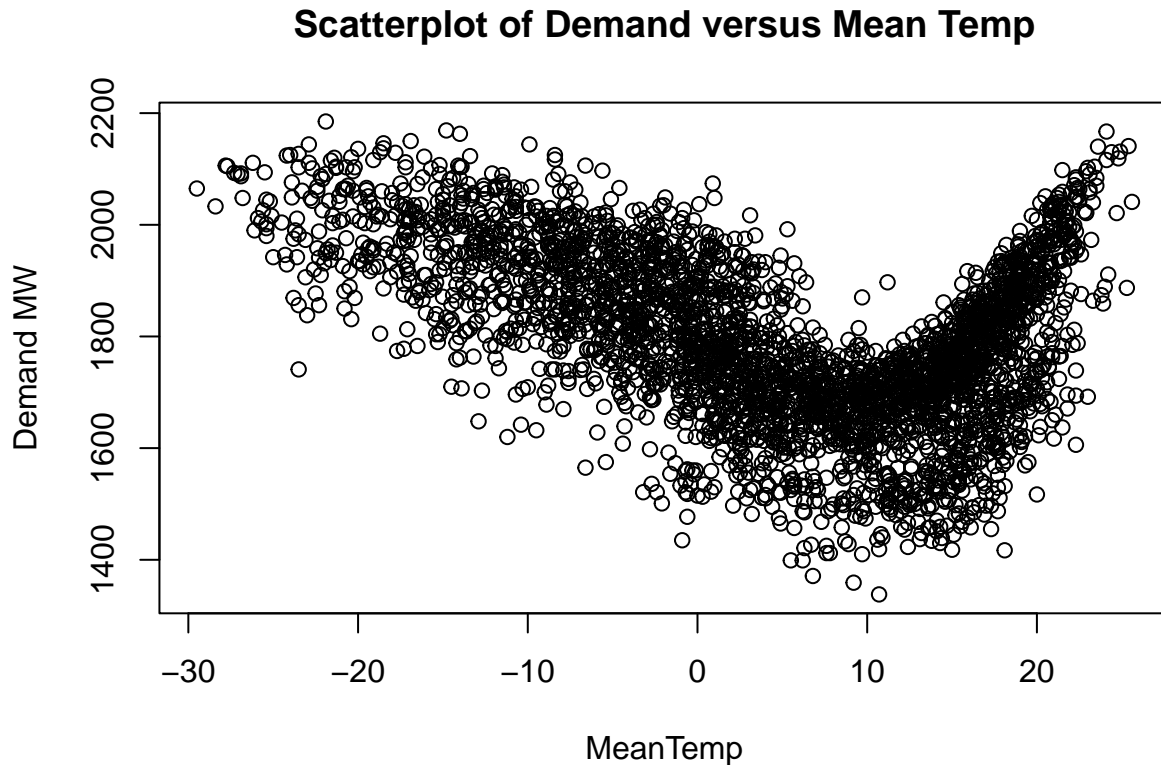
- Opposite patterns between summer and winter

Plots covariates

##	Date	Edmonton	MaxTemp	MinTemp	MeanTemp	HDD	CDD	Month	Year	Dow	ConsHDD
## 1	2011-01-01	1696	-1.5	-19.8	-10.7	20.7	0	1	2011	Sat	1
## 2	2011-01-02	1771	-1.1	-7.2	-4.2	14.2	0	1	2011	Sun	1
## 3	2011-01-03	1931	-1.9	-7.4	-4.7	14.7	0	1	2011	Mon	1
## 4	2011-01-04	1942	0.2	-6.5	-3.2	13.2	0	1	2011	Tue	1
## 5	2011-01-05	1919	1.9	-6.3	-2.2	12.2	0	1	2011	Wed	1
## 6	2011-01-06	1932	2.5	-4.8	-1.2	11.2	0	1	2011	Thu	1
##	ConsCDD	HDD_1	HDD_2	CDD_1	CDD_2	IsBusinessDay					
## 1	0	15.0	15.0	0	0	0					
## 2	0	20.7	15.0	0	0	0					
## 3	0	14.2	20.7	0	0	1					
## 4	0	14.7	14.2	0	0	1					
## 5	0	13.2	14.7	0	0	1					
## 6	0	12.2	13.2	0	0	1					

This is an overview of the covariates that could be useful in our analysis. The data was retrieved from Environment Canada API. The data was collected by the Blatchford weather station in Edmonton. Missing values were imputed using nearby weather station (15 km radius). Some variables were engineered like HDD, CDD and other variables derived from them, as well as business day.

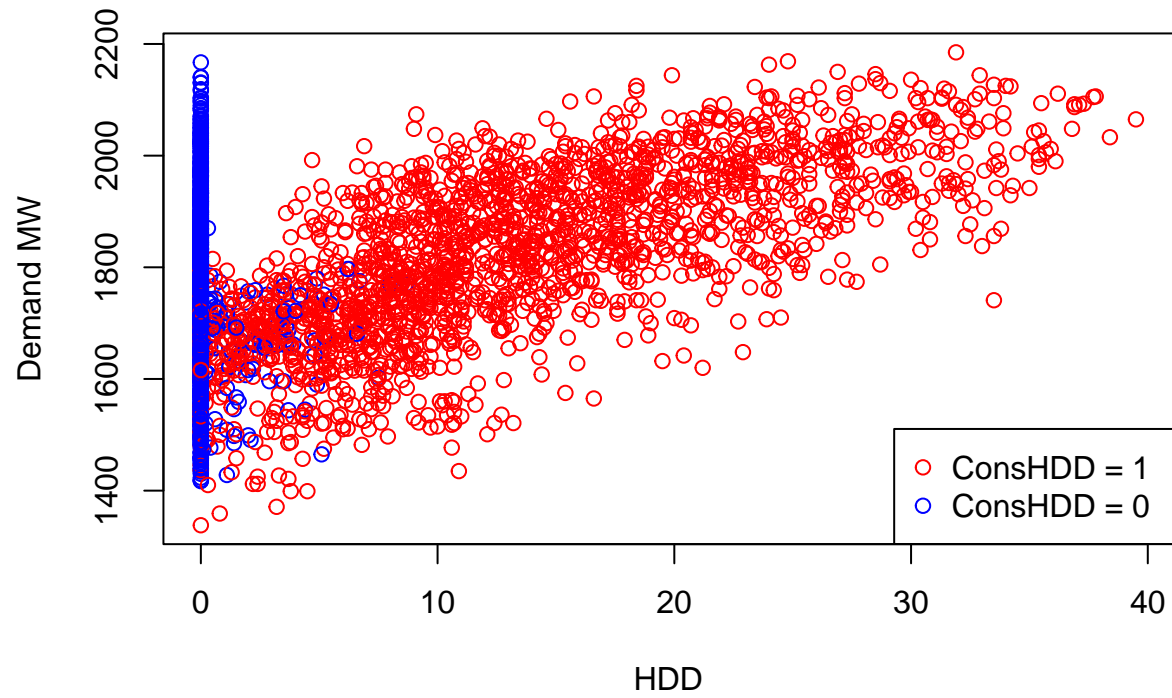
- Max, Min and Mean temperature for a given day retrieved from blatchford station.
- HDD and CDD (Heating Degree Days and Cooling Degree Days) with 10 degrees as a reference point. The reference point was determined from the plot below. The variables with subscript n ($_n$) represent the measures taken at a lag(n). The prefix 'Cons' refers to consecutive heating degree days or cooling degree days.
- Temporal data like Month, Year, DoW (Day of Week) and Business Day. The DoW indicator was derived from a canadian holiday calendar (not specific to Alberta).



Observations:

- We clearly see a change of effect of temperature on demand at around 10 degrees.
- We also observe that there are two different groups of points one over the other. This could be explained by another binary variable.

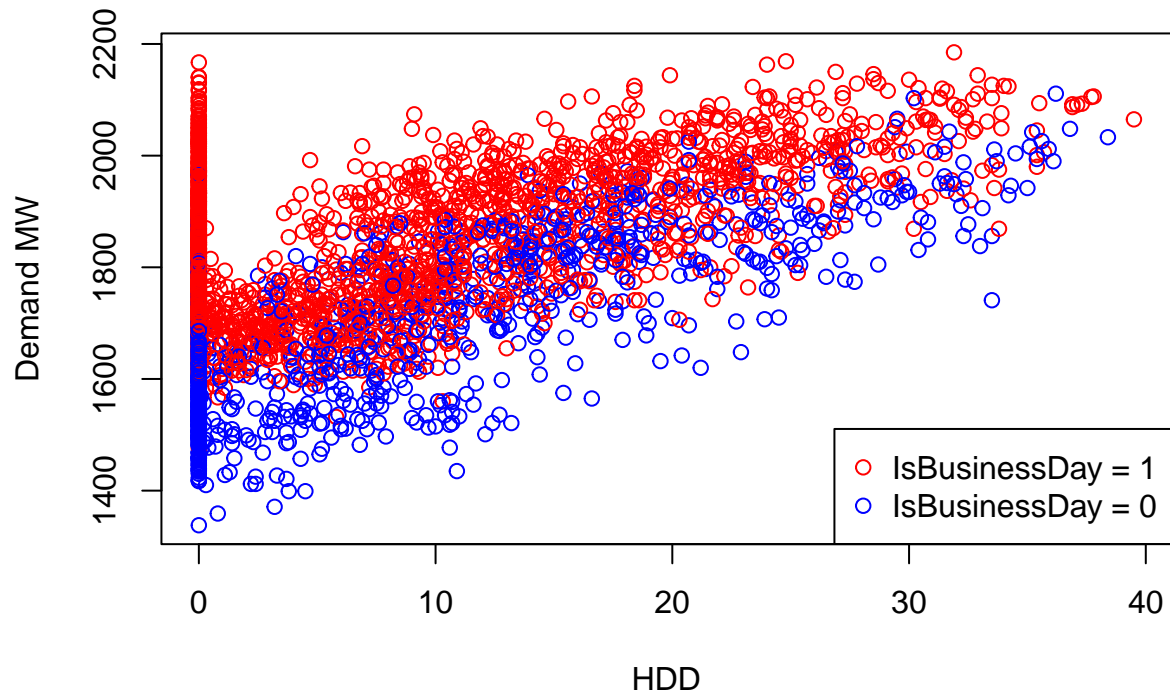
Scatterplot of Demand versus HDD (ConsHDD)



Observations:

- Consecutive Heating degree days does not seem to provide much additional information on this graph.

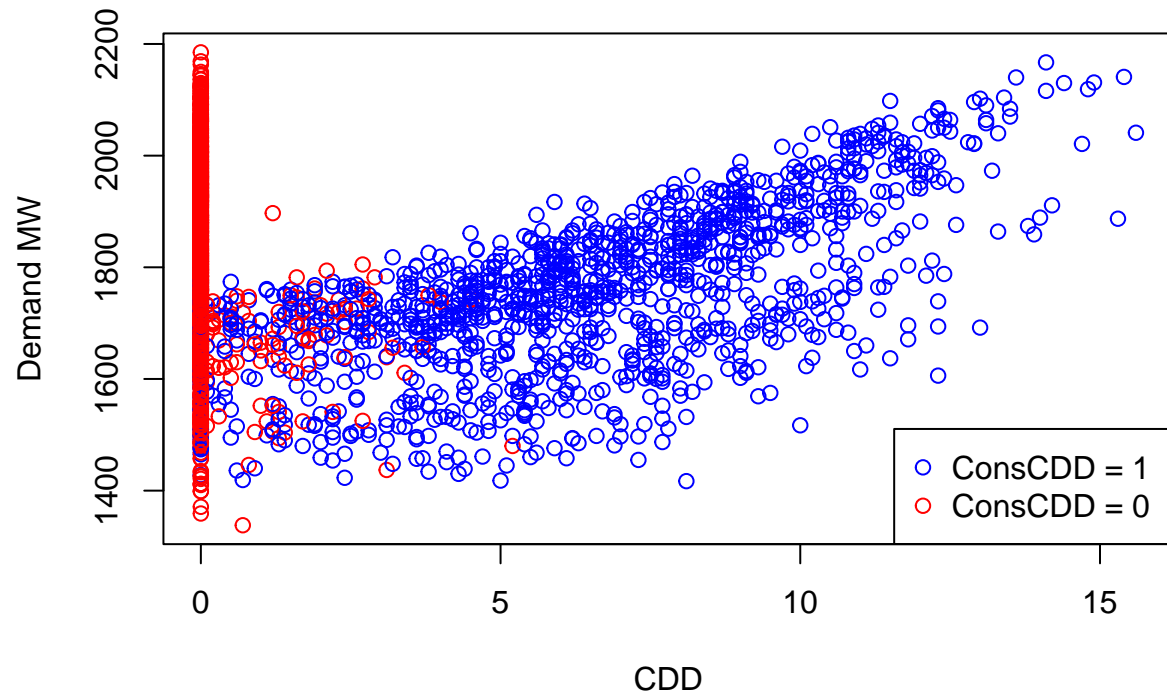
Scatterplot of Demand versus HDD (BusinessDay)



Observations:

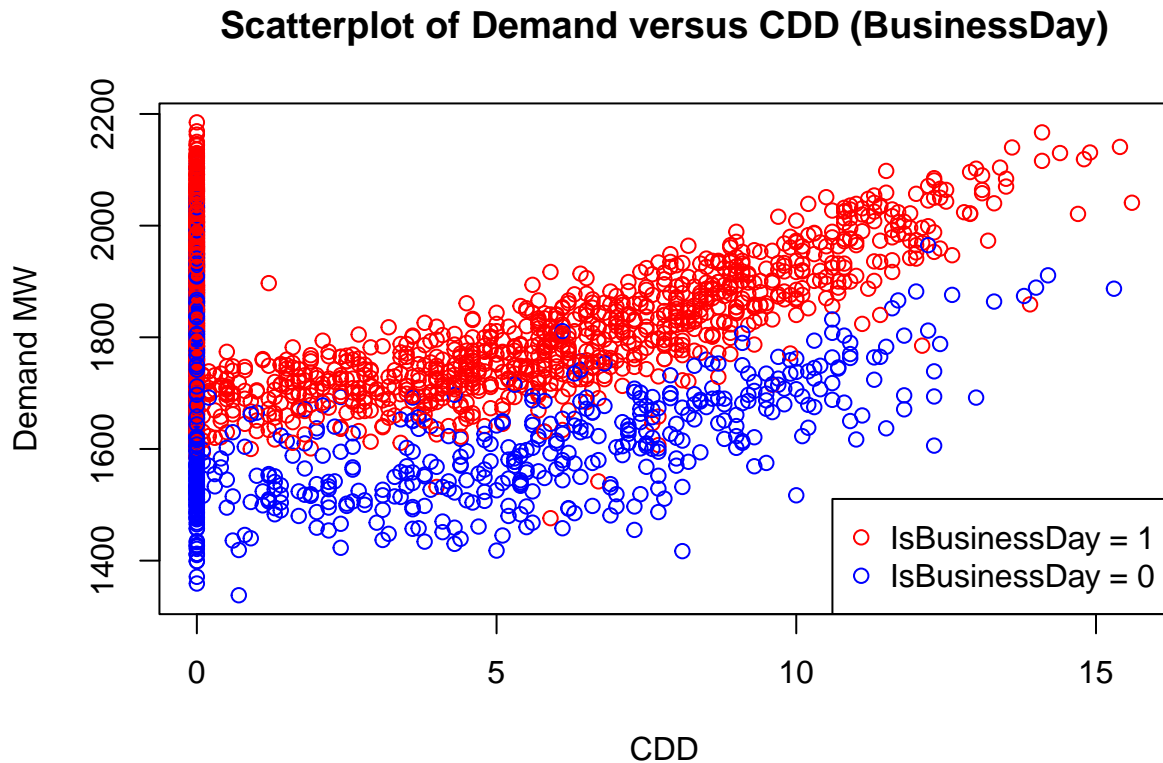
- Business day seems to be a very good variable that segregates between low and high demand for a given temperature.

Scatterplot of Demand versus CDD (ConsCDD)



Observations:

- Here we clearly see the two curves one over the other.



Observations:

- Business day explains very well the differences between the curves.

TO DO

- an introduction with a good description of the zone under study
- exploratory data analysis
- evaluation of naive methods
- description of possible explanatory variables, clearly indicating the source of the data, and providing some summary statistics

Format

Part I of the report, in .pdf format: – with AT MOST 12 PAGES. Includes all submitted pages except cover page2 ; – in Times New Roman font, 12 pt, double-spaced, “letter” page format with minimum 1 inch margins.