Case 1: Energy Forecasting

Abstract: The city of Boston needs to build a forecasting model to predict their energy usage. The Power usage model needs to build on the multiple parameters i.e. temperature, day of week, month, weekday, hour of day.

Propose Solution:

Built a forecasting model using R in RStudio:

1. Data wrangling and cleansing:

1. The following parameter needs to be derived in following formats from the RawData.csv file

▪ kWh	Kwh is aggregated hourly. Sum of 12 observations (5 min intervals rolled up to hourly)
▪ month	1-12 => Jan-Dec – Derived from dates
▪ day	1-31 – Derived from dates
▪ year	Derived from dates
▪ hour	0-23 – Derived for each record corresponding to the hour of observation
▪ Day of Week	0-6 – Sun-Sat – Derived from dates
▪ Weekday	1- Yes 0- No – Derived from dates
▪ Peakhour	7AM-7PM – 1 ; 7PM-7AM – 0

2. You will have to get temperature data for Boston (KBOS) for the time period. For that review https://cran.r-project.org/web/packages/weatherData/weatherData.pdf

2. Multiple-Linear Regression:

Designed a Regression Model using the following parameters:
y=Kwh # this is the dependent variable which we are trying to predict.

- x1=constant
- x2=month
- x3=day
- x4=hour
- x5=day of week

- x6=WeekDay
- x7=PeakHour

3. Forecast.

RMSE – This parameter gives the value which defines the model is correctly forecasting the KWH energy consumption; if the RMSE value is lower than the minimum value of the range of power consumption.
As per the predicted model the RMSE value (**39.67 KWH**) which is very near to the Range of independent variables (**33 – 228 KWH**).

Predictive Model implementation Steps

Flow of Implementation:

- Gathered relevant data
- Transformed data into required regression format
- Cleansed data
- Data Quality check
- Segregated data into training and test
- Computed calculated variables
- Figured correlation between the independent variables
- Ran the regression model to understand the relevant variable
- Determined relevant variables for the regression
- Ran the predictive model over the test data
- Evaluate model over the evaluation matrices on test set.