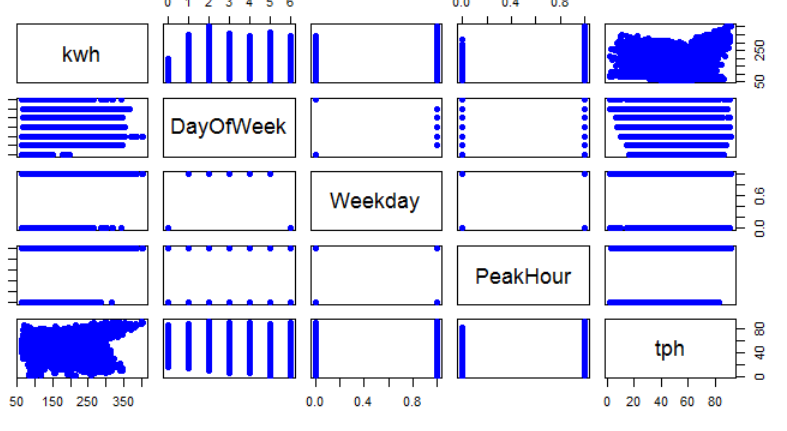
**Assignment -1-Energy Forecast**

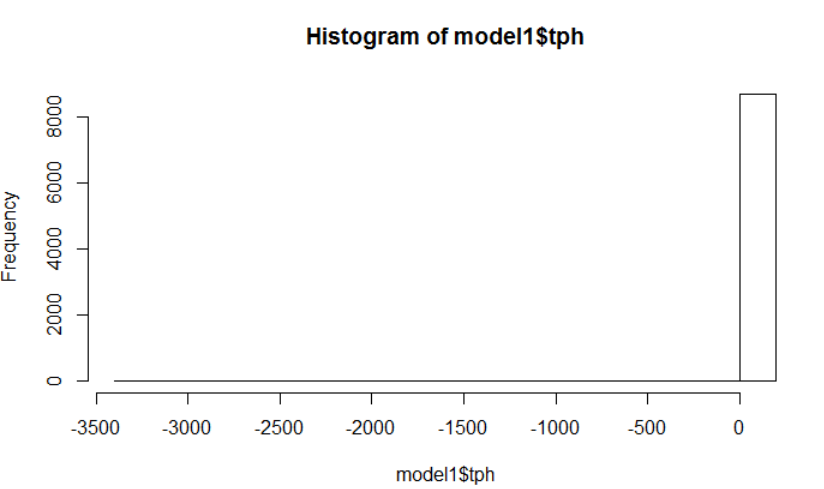
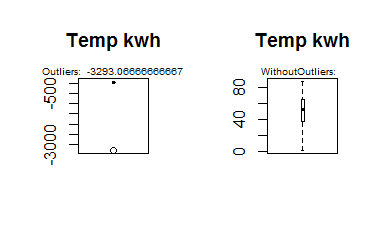
**Team 9**

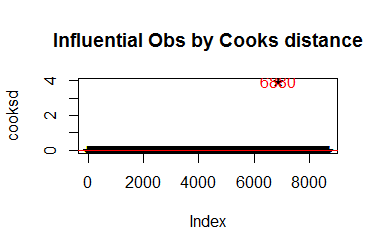
**Part1: Data wrangling and cleaning**

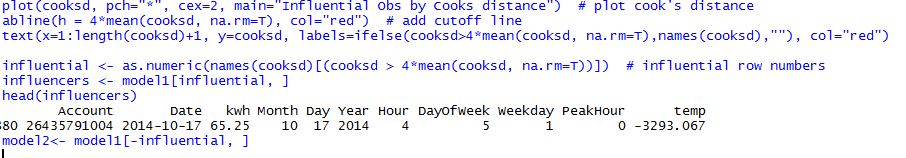
1. **Data is cleansed and temperature is merged using R into final file sampleformat.csv**
2. **Following Methods were used to transform the data and remove outliners**

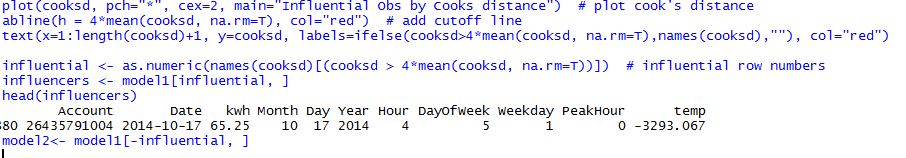
* ***Histogram* of all variables were plotted to see the dependency of each predictors on output.**
* ***Scatter plot* was plotted to see the behavior of Kwh with all predictors.**
* ***BoxPlot* was plotted to see outliners and how the value varies.**
* **We also performed Influential of all variables through *Cook’s* *distance* so that we get a clear idea of the outliners and extremes in dataset.**

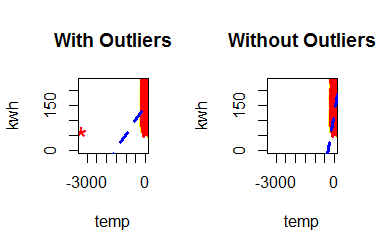
****

****

****

****

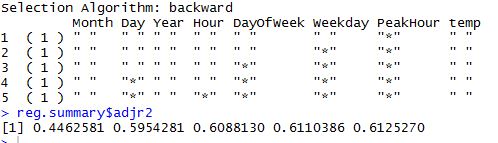
****

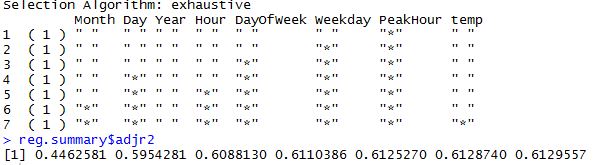
****

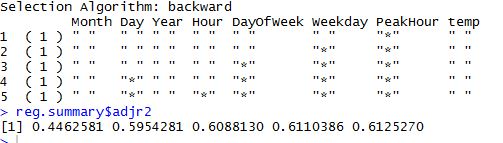
1. **Next we made subset of the dataset to get consistent data without outliners.**

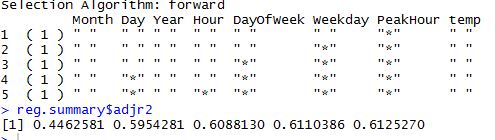
**Part2: Multiple Linear Regression**

1. **Feature Selection was done for all the variables and we concluded that DayofWeek, PeakHour, Week are the variable with maximum influence to Kwh.**
2. **We used forward selection, backward Selection, Exhaustive search to infer this conclusion.**
3. **It was also observed that Temperature does not play much role when it comes to calculation but logically it plays an important factor to predict Kwh. *We Squared the value of Temperature to make it efficient predictor***
4. **Linear regression model was constructed with variables DayofWeek, PeakHour, Week and Temperature.**

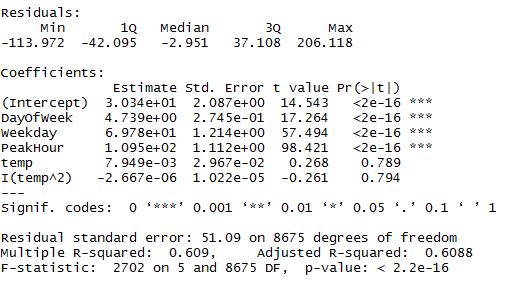
****

****

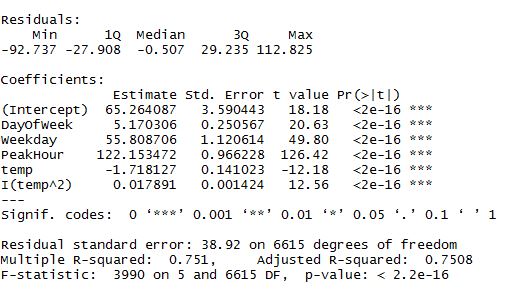
****

****

**Original Coefficients:**

****

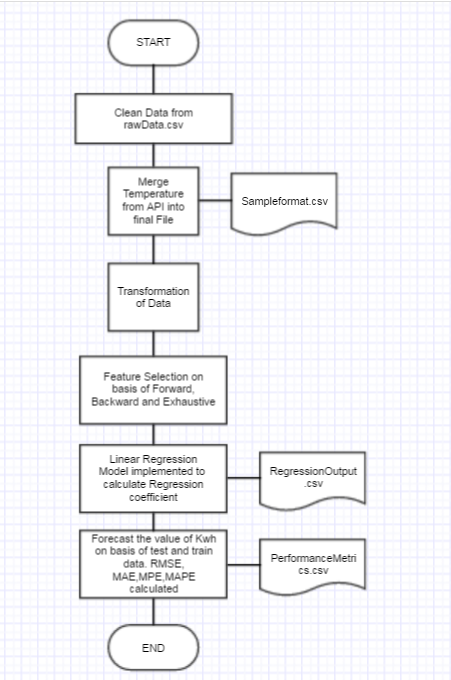
**Coefficients After Transformation:**

****

**Part 3: Forecast**

1. **The dataset was divided into 75% train and 25% test data to predict the Kwh.**
2. **Model was run on Test data and Kwh was forecasted using libraries.**
3. **The same was also forecasted against the data given by Professor.**

**FLOWCHART**

****