The analysis is performed on the year 2019. All the below mentioned datasets are merged on Date column. Since the load data, is recorded for every 5 minutes the average value of actual load is computed by pivoting the tables in excel. The load data is web scrapped using the code in webscrapping\_noaa.ipynb. Total 5 different features are used in predicting the RTD actual load.

**List of features used in predicting the Actual Load**:

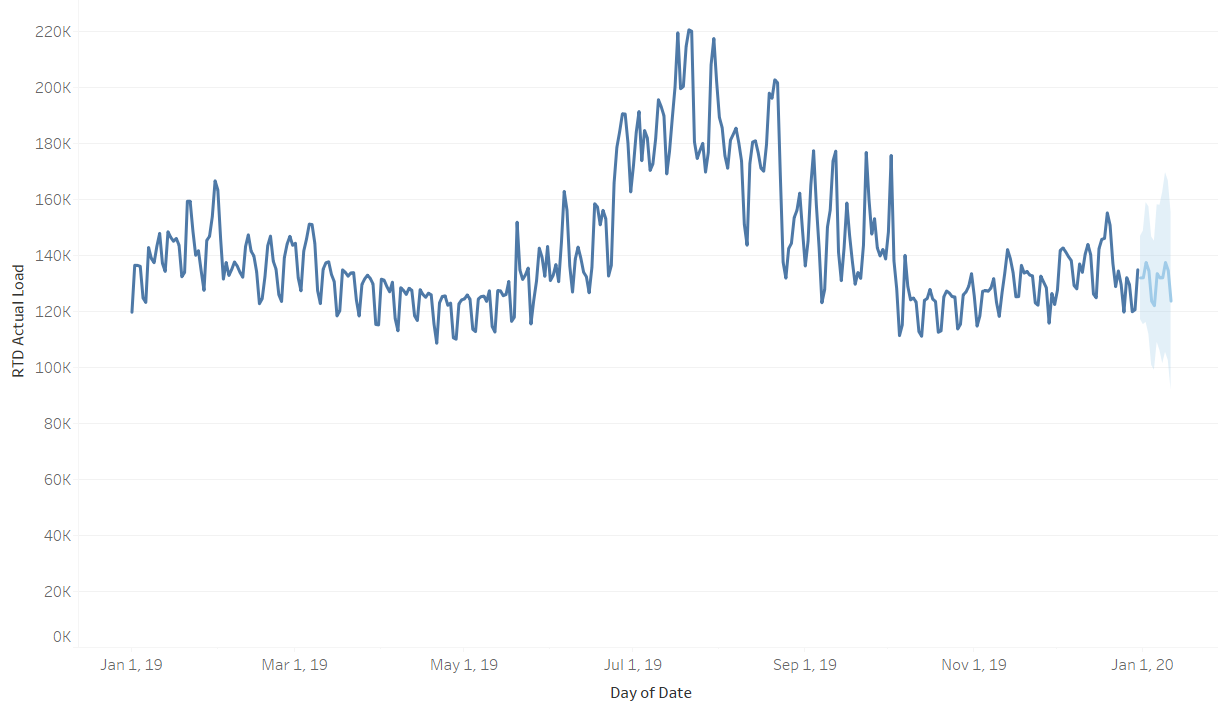
Load Data – Average of RTD Actual Load

Historical weather data– HLY-TEMP-NORMAL, HLY-PRES-NORMAL, HLY-DEWP-NORMAL

Energy-market-operational-data – DAM ZONAL LBMP, DAM Zonal Congestion

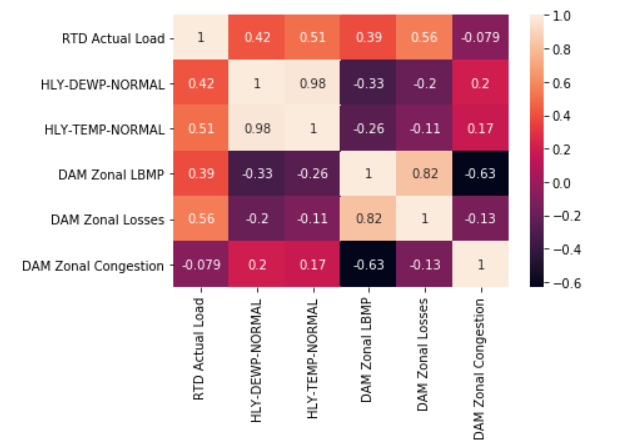
**Data Visualization**

The below graph shows the RTD actual load across different months of 2019. A forecast for 1-12 Jan. 2020 is shown below:



**Co-relational plot:**

From the below plot, DAM Zonal Losses and HLY\_TEMP\_NORMAL are highly corelated with RTD Actual Load, because the correlational value is greater than 0.5. Out of 5 features DAM Zonal Congestion is negatively corelated with RTD Actual Load.



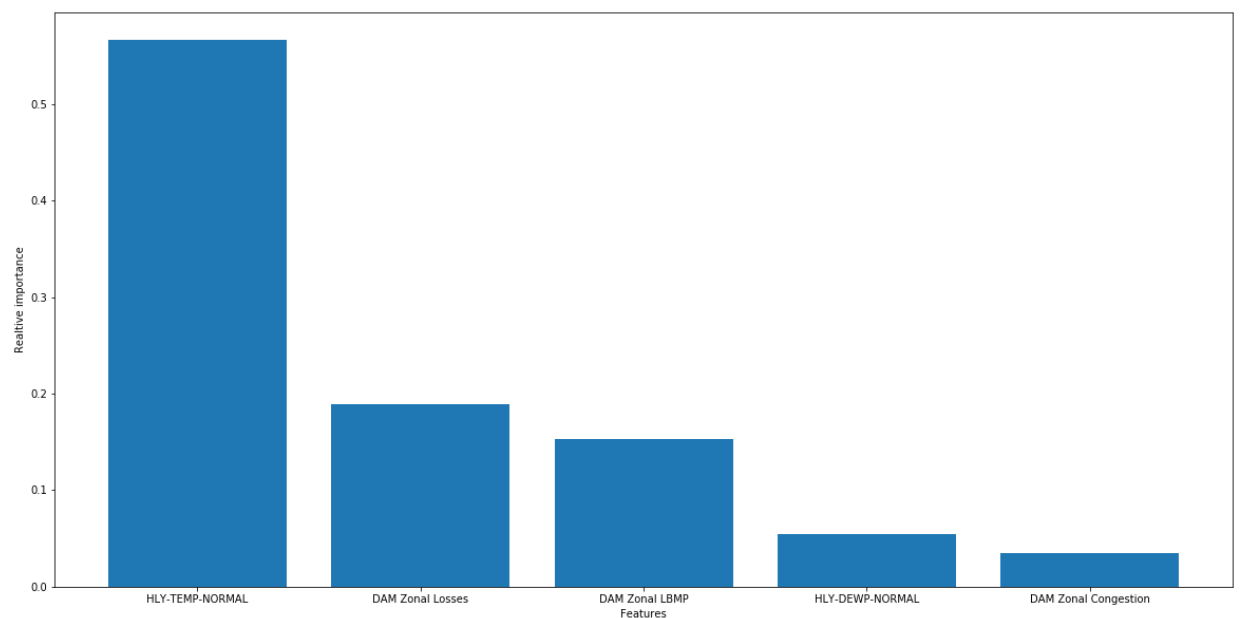
**Model Training**

Four different shallow learning models are trained on to predict the Actual Load. Random Forest Regressor outperformed all other models with a r square of 91%.

|  |  |  |  |
| --- | --- | --- | --- |
| Model Name | R-square | MAE | MSE |
| Linear Regression | 67% | 555 | 480523 |
| Decision Tree Regressor | 83% | 349.83 | 246150 |
| Gradient Boosting Regressor | 86% | 349.91 | 131616 |
| Random Forest Regressor | 91% | 269 | 199054 |

**Interpreting Random Forest Results**

The important feature in predicting the random forest models is HLY-TEMP-NORMAL with DAM Zonal Losses, DAM Zonal LBMP coming next. The below plots represent the features importance with their relative importance values. The features are represented on x-axis and their relative importance values are represented on y-axis.



The below image represents how the R-square value changes with number of top features selected. The r-square obtained in predicting the Actual Load with HLY TEMP NORMAL is about 55%. This implies that temperature plays an important role in predicting the target variable.

