Boston Power Consumption

Assignment 3

Team-6

INFO 7390 Spring 2016

Table of contents

|  |  |  |
| --- | --- | --- |
| Sr.No | Topic | Page.No |
| 1. | Executive Summary | 3 |
| 2. | Data Preparation | 3 |
| 3. | Data Analysis | 5 |
| 4. | Data Visualization | 6 |
| 5. | Technical Implementation | 8 |
| 6. | Conclusion | 14 |

Executive Summary:

In this assignment we had different datasets of electricity consumption of different departments of Boston Area. We have tried to predict the electricity consumption by department wise for different locations which can play an important role in their current and future Electricity Consumption and make the city aware of making proper utilization of power in order to save power.

We had put together the analytical strategies followed by successful development of model for prediction of power consumption based on different input actions.

This document covers the cleaning of datasets of Boston power consumption, and technical aspects of using Azure Machine Learning (AML), Microsoft Visual Studio for using the restful API generated in AML for the implementation and deployment of Electricity Consumption Prediction. The document consists of two main parts:

1. Data Preparation
2. Visualization
3. Technical implementation

Data Preparation:

It summaries the kind of data that is needed for creating the data model for power consumption prediction. After doing analysis on raw data provided, which consists of basically five different categories of building types (Department) for example BPD(Boston Police Department),Pl(Public Library), BFD(Boston Fire Department), PM(Property Management) and School.

**Data-Preprocessing**:

We had around 75 different csv files for different locations of various departments.

Below are the steps that we followed for data cleaning in R and Python

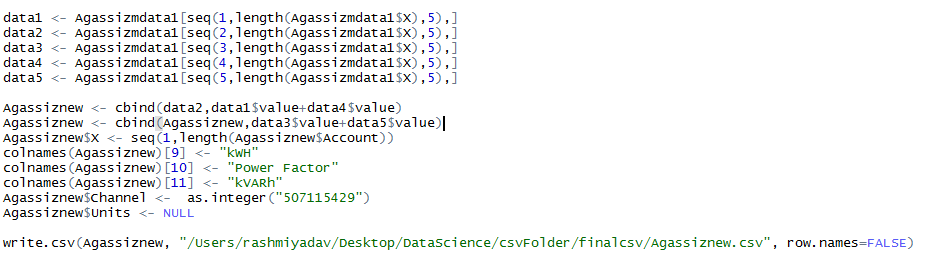
STEP 1: Combined the data of same departments using cbind() in R.



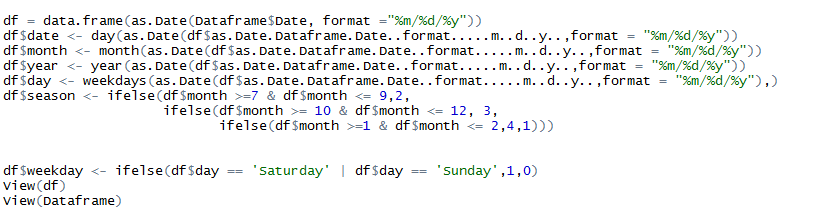
STEP 2: Transposed the data from 4th column to last column which were time intervals for every 5 min, in order to get the time value, power factor, kWh in the columns.

STEP 3: Handled duplicate values in channel column in order to get correct consumption rate.

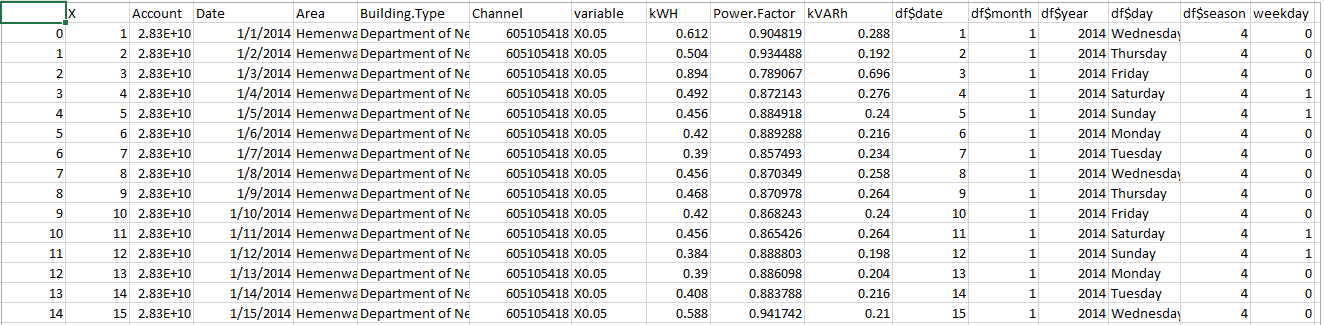
STEP 4: Also in column channel there were different channels for same building for which readings were taken so we added those values. The below scripts is used for adding the power consumption values.



STEP 5: Derived a column from date whether a specific day is weekday or weekend



STEP 6: After completing the above steps we imported the file as a csv format and repeated the process for all the other department. In next page there is a sample of how our dataset look after pre-processing.



**NOTE**: Handled Missing values and additional characters added in dataset after running R scripts with Python Script directly in assure.

Features of dataset – We have 11 features in all common in all the five files which are

1. Account No.

2. Area

3. Building Type (Department)

4. Channel

5. Variables

6. kWH

7. Power Factor

8. kVARh

9. Weekend

10. Season

Records – Range 120000 – 140000 in each file

Response Variable – kWH (Kilo watt per hour)

Column Name and Description

Account No. – It describes the account number of meter for each department

Area - It describes the building exists in which area

Building Type - Basically a department for example library, police department etc.

Variable - It describes the time interval for which reading is taken, it sis every 5 min

kWH – Kilo watts per hour utilization for each building

kVARh – It is reactance power for every 5 min interval

Power Factor – It is ratio of power utilization and reactance power for each interval for each building

Data Analysis:

Analysis were performed on data sets using business intelligence tool tableau and Microsoft excel. The tableau dashboards and charts describes the analysis performed on the various departments’ power consumption in different months and trend of power consumption variation on every 5 min interval.

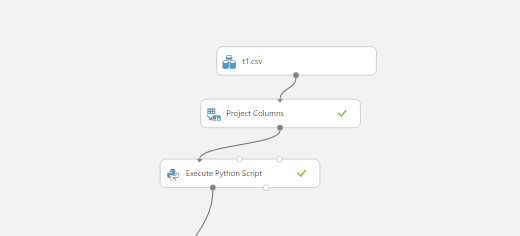
The technical implementation covers the concept of web services required for deployment of predictive analytics solutions. We also outline a typical architecture of an end-to-end operationalized solution.

Technical Tools that were used in the implementation of web service from raw data is as follow:

1. R & Python -> Data Cleaning
2. Tableau -> Visualization
3. Azure Machine Learning -> Predictive Model and restful API generation
4. Amazon web service(AWS) -> Deploying website
5. Html, CSS , Bootstrap-> Front End
6. Node js, Python -> Utilization of restful API for Prediction

**Data-Modeling and Handling in Azure**:

After creating all the datasets, we uploaded the datasets in Azure Machine Learning and we further processed the data in AML.

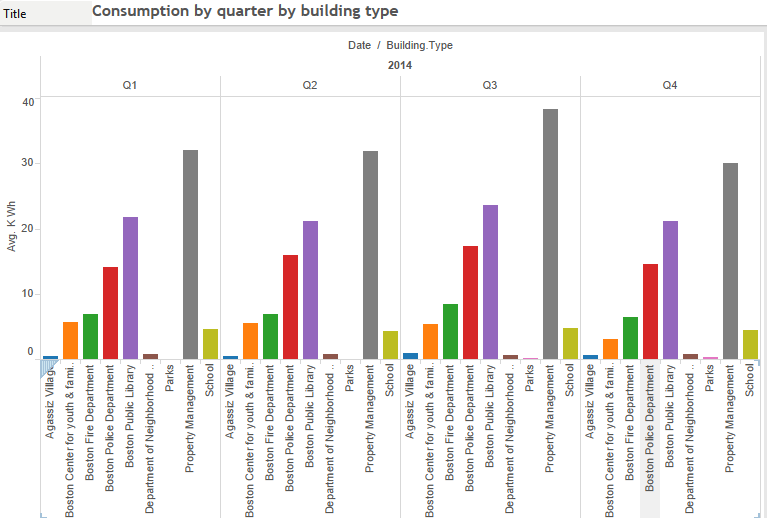


**NOTE**: t1 represents as Boston Property Management

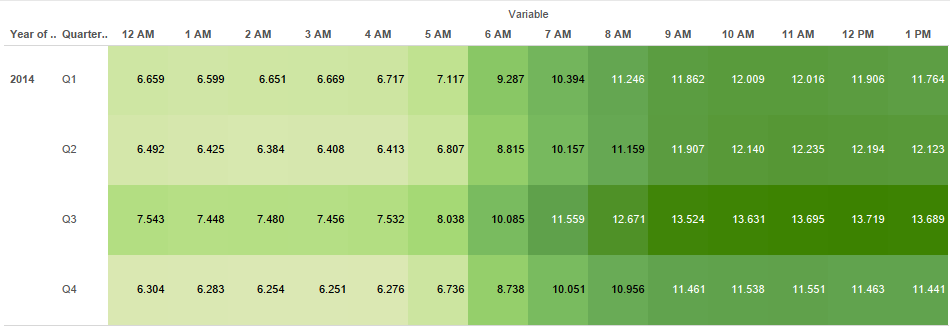
**VISUALIZATION**

We did the analysis on 6 different datasets that we created. The analysis is as follows:

1. The first visualization we have done is the average power consumption for each building by year and then we can further drill down and see quarter and month.



2. In the Below chart we have shown the actual power consumption for each hour by year/quarter



3. From the chart below we understand exactly in which department in which area the power consumption is more or less.

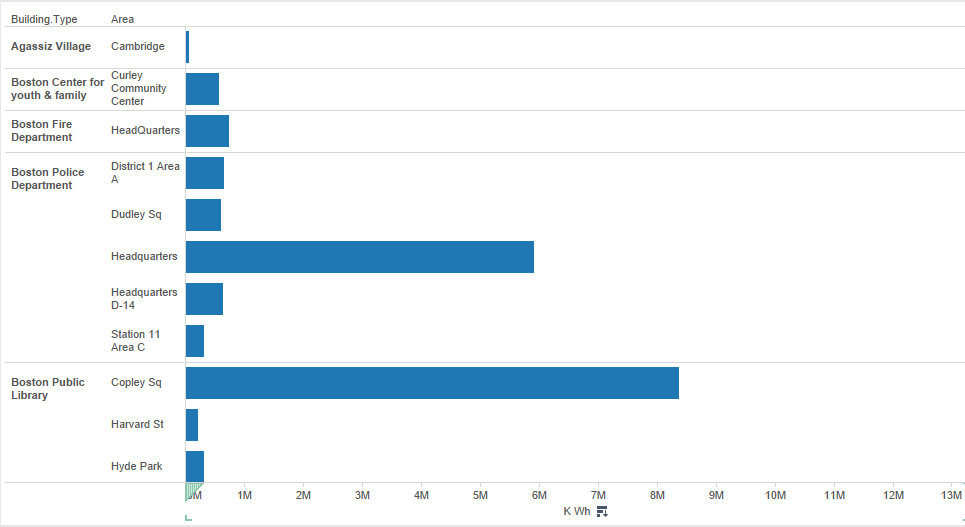


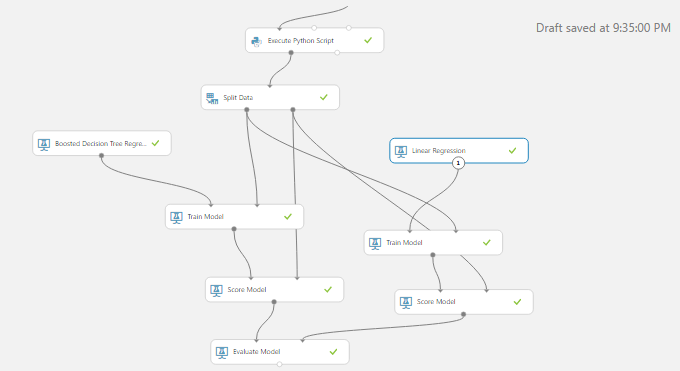
Figure 1

In above analysis we can see the trend of power consumption in a year and can determine which department utilizes the most power.

**TECHNICAL IMPLEMENTATION**

**Modeling:**

We created our predictive model in Azure Machine Learning. We used two predictive model Linear Regression and Boosted Decision Tree Regression for our prediction. We created predictive model for all five departments. Below is the screenshot of the steps that we followed after the data cleaning. We compared the 2 model for each of the department and suggested the model accordingly for departments.

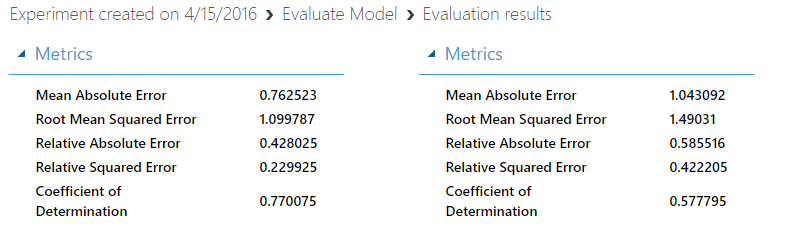


The following were the important factors that were affecting the power consumption of different locations at different time:

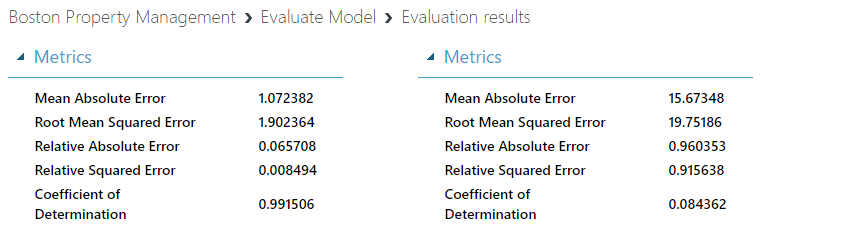
Location(Categorical), Month, Day, Time, Power Factor

The following are the results of predictive model for each department:

1. Boston Property Management:

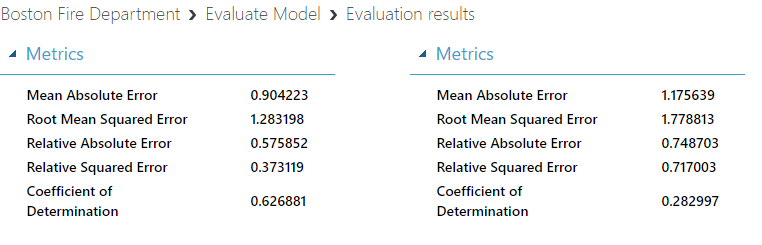
In this case we suggested the boosted tree regression model as the Co-efficient of determination and RMSE for it was better than the linear regression.

1. Boston Police Department:



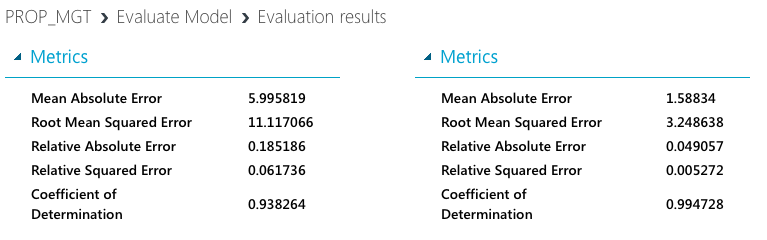
In this case again we suggested the boosted tree regression model as the Co-efficient of determination and RMSE for it was better than the linear regression.

1. Boston Fire Department:



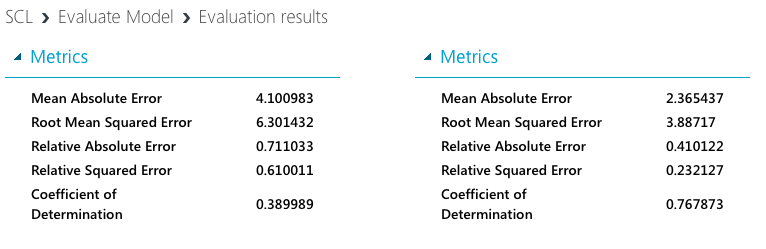
In this case we suggested the Linear regression model as the Co-efficient of determination for Boosted Tree Regression was close to 1 and it seems to be the case of over fitting.

1. Boston Library:



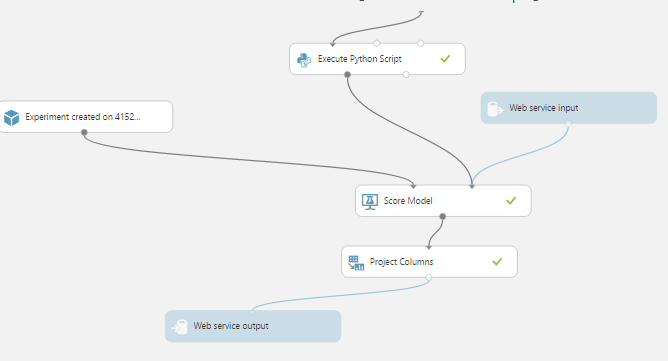
In this case again we suggested the Linear regression model as the Co-efficient of determination for Boosted Tree Regression was close to 1 and it seems to be the case of over fitting.

1. School:



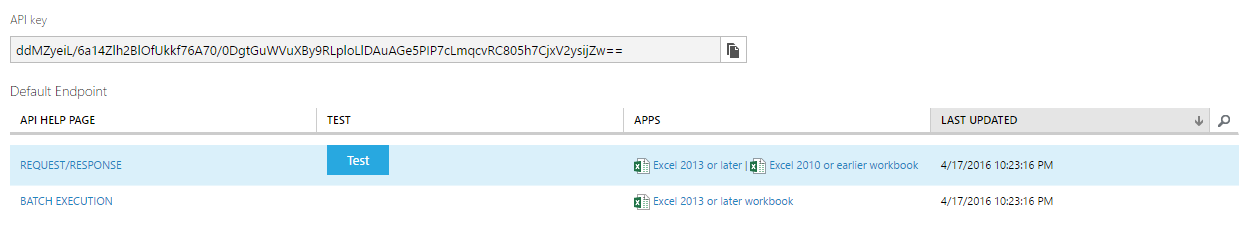
In this case we suggested the boosted tree regression model as the Co-efficient of determination and RMSE for it was better than the linear regression.

**Web Service Deployment:**



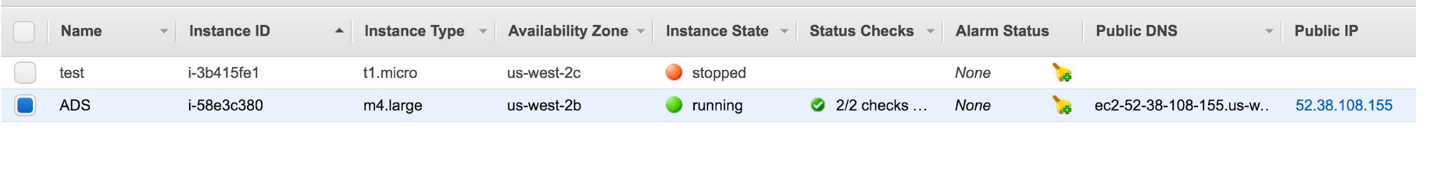
After modeling we deployed it as a web service in AML to generate the restful API so that we can use the API in the frontend.

This is the result that we got from the web service deployment. The following is the screenshot of API that we got from web service.

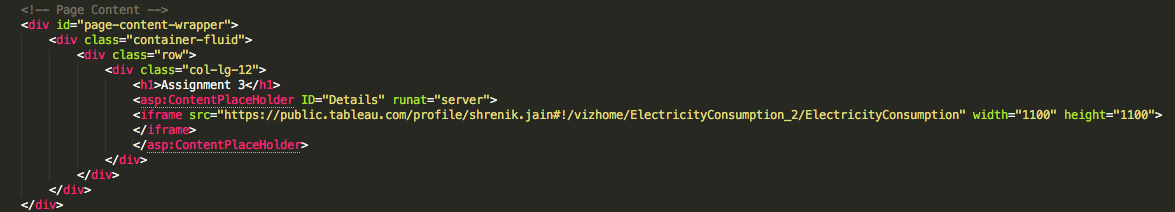


**Amazon web service:**

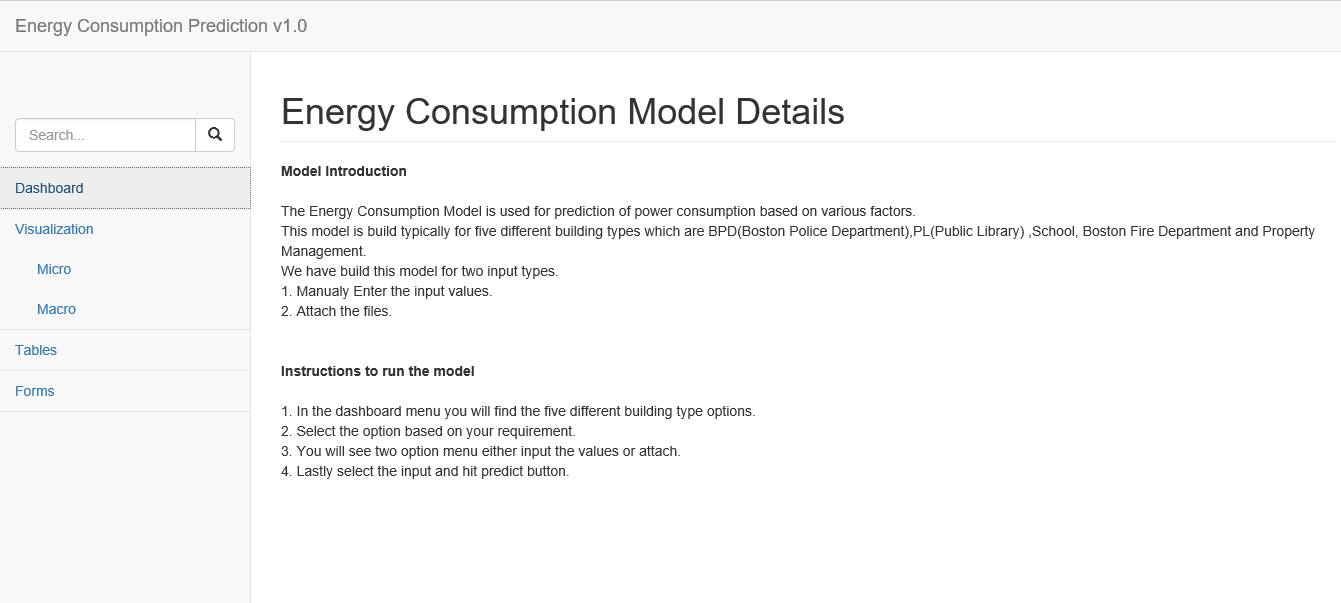
We used AWS for using the API for prediction. The following is the snippet that shows the configuration for prediction of power consumption by School.

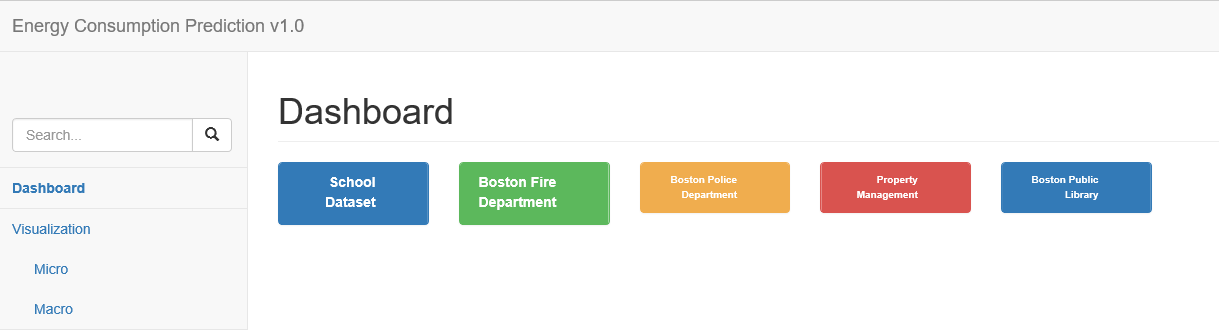


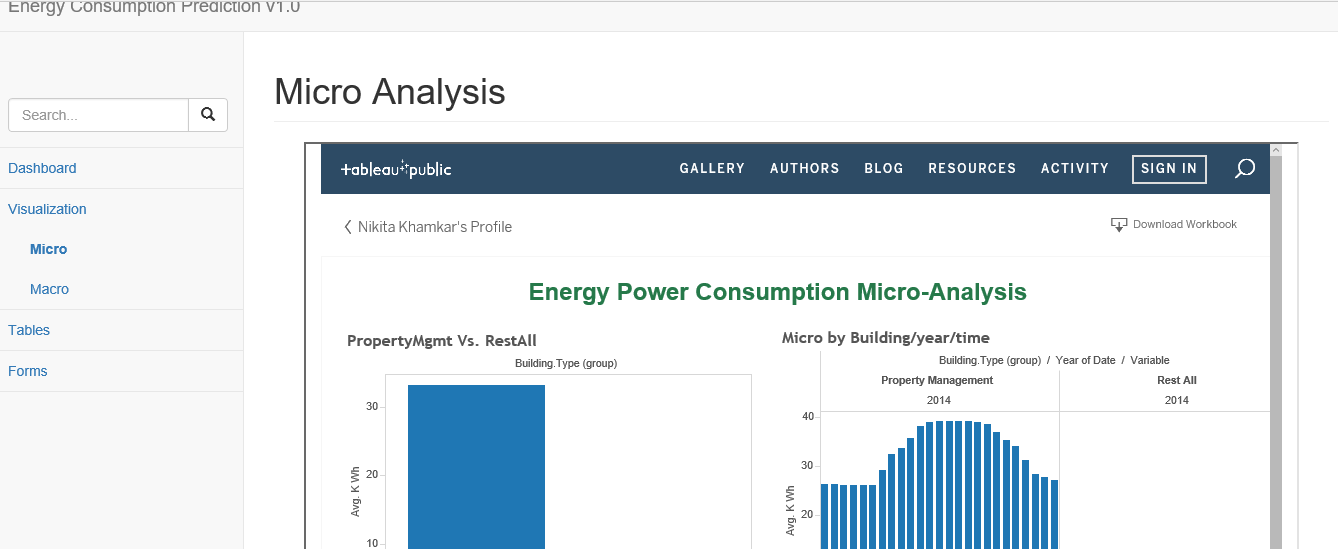
After creating the API connection, we build the front end in the HTML and CSS. We have published our tableau visualization in public tableau and used that link to show the visualization in our front end. Following is the snippet of tableau:

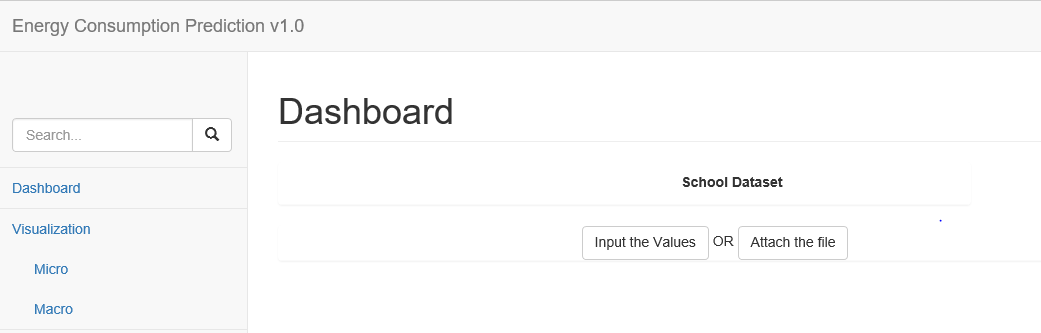


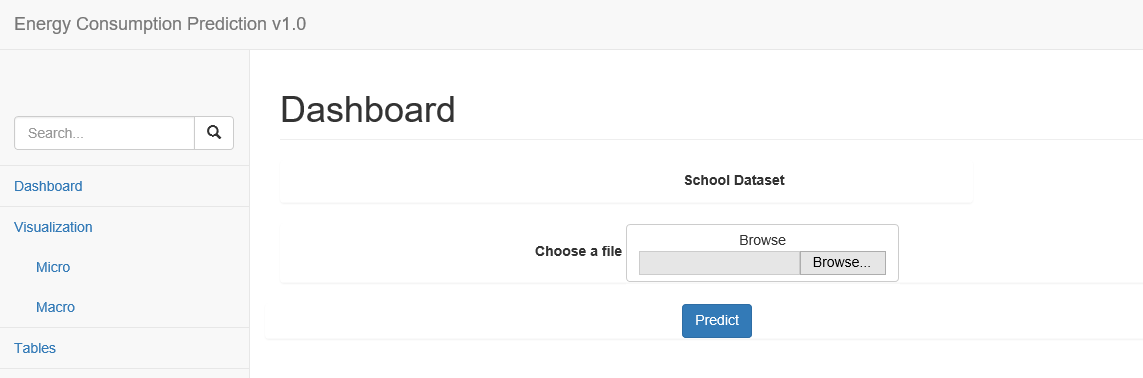
The following are the screenshots from our website that we created:

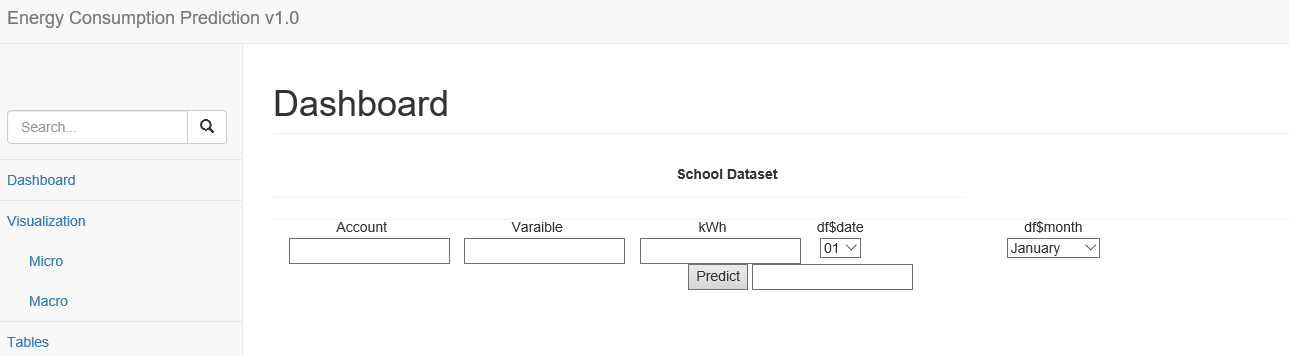












We also published our website using Visual Studio and have also got a new domain for our website from GoDaddy.com. The following is the link of our website:

<http://52.38.108.155:4030/>

**CONCLUSION:**

* From our website we can Forecast the electricity consumption department wise for different locations which can play an important role in their current and future Electricity Consumption using AWS Framework for integrating the restful API generated from Azure Machine Learning
* We have developed a user friendly website giving business an ability to forecast electricity consumption by different departments. Also view the trend of power consumption variation using tableau interface.