Boston Power Consumption

Assignment 3

Team-6

INFO 7390 Spring 2016

By
Nikita Khamkar
Rashmi Yadav
Prateek Mane

Table of contents

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.

```
Agassiznew <- cbind(data2,data1$value+data4$value)
Agassiznew <- cbind(Agassiznew,data3$value+data5$value)
```

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.

```
data1 <- Agassizmdata1[seq(1,length(Agassizmdata1$X),5),]
data2 <- Agassizmdata1[seq(2,length(Agassizmdata1$X),5),]
data3 <- Agassizmdata1[seq(3,length(Agassizmdata1$X),5),]
data4 <- Agassizmdata1[seq(4,length(Agassizmdata1$X),5),]
data5 <- Agassizmdata1[seq(5,length(Agassizmdata1$X),5),]

Agassiznew <- cbind(data2,data1$value+data4$value)
Agassiznew <- cbind(Agassiznew,data3$value+data5$value)
Agassiznew$X <- seq(1,length(Agassiznew$Account))
colnames(Agassiznew)[9] <- "kWH"
colnames(Agassiznew)[10] <- "Power Factor"
colnames(Agassiznew)[11] <- "kVARh"
Agassiznew$Channel <-  as.integer("507115429")
Agassiznew$Units <- NULL

write.csv(Agassiznew, "/Users/rashmiyadav/Desktop/DataScience/csvFolder/finalcsv/Agassiznew.csv", row.names=FALSE)
```

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

```
df = data.frame(as.Date(Dataframe$Date, format ="%m/%d/%y"))
df$date <- day(as.Date(df$as.Date.Dataframe.Date..format.....m..d..y..,format = "%m/%d/%y"))
df$month <- month(as.Date(df$as.Date.Dataframe.Date..format.....m..d..y..,format = "%m/%d/%y"))
df$year <- year(as.Date(df$as.Date.Dataframe.Date..format.....m..d..y..,format = "%m/%d/%y"))
df$day <- weekdays(as.Date(df$as.Date.Dataframe.Date..format.....m..d..y..,format = "%m/%d/%y"),)
df$season <- ifelse(df$month >=7 & df$month <= 9,2,
                    ifelse(df$month >= 10 & df$month <= 12, 3,
                           ifelse(df$month >=1 & df$month <= 2,4,1)))


df$weekday <- ifelse(df$day == 'Saturday' | df$day == 'Sunday',1,0)
View(df)
View(Dataframe)
```

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.

	X	Account	Date	Area	Building.Type	Channel	variable	kWH	Power.Factor	kVARh	df$date	df$month	df$year	df$day	df$season	weekday
0	1	2.83E+10	1/1/2014	Hemenwa	Department of Ne	605105418	X0.05	0.612	0.904819	0.288	1	1	2014	Wednesday	4	0
1	2	2.83E+10	1/2/2014	Hemenwa	Department of Ne	605105418	X0.05	0.504	0.934488	0.192	2	1	2014	Thursday	4	0
2	3	2.83E+10	1/3/2014	Hemenwa	Department of Ne	605105418	X0.05	0.894	0.789067	0.696	3	1	2014	Friday	4	0
3	4	2.83E+10	1/4/2014	Hemenwa	Department of Ne	605105418	X0.05	0.492	0.872143	0.276	4	1	2014	Saturday	4	1
4	5	2.83E+10	1/5/2014	Hemenwa	Department of Ne	605105418	X0.05	0.456	0.884918	0.24	5	1	2014	Sunday	4	1
5	6	2.83E+10	1/6/2014	Hemenwa	Department of Ne	605105418	X0.05	0.42	0.889288	0.216	6	1	2014	Monday	4	0
6	7	2.83E+10	1/7/2014	Hemenwa	Department of Ne	605105418	X0.05	0.39	0.857493	0.234	7	1	2014	Tuesday	4	0
7	8	2.83E+10	1/8/2014	Hemenwa	Department of Ne	605105418	X0.05	0.456	0.870349	0.258	8	1	2014	Wednesday	4	0
8	9	2.83E+10	1/9/2014	Hemenwa	Department of Ne	605105418	X0.05	0.468	0.870978	0.264	9	1	2014	Thursday	4	0
9	10	2.83E+10	1/10/2014	Hemenwa	Department of Ne	605105418	X0.05	0.42	0.868243	0.24	10	1	2014	Friday	4	0
10	11	2.83E+10	1/11/2014	Hemenwa	Department of Ne	605105418	X0.05	0.456	0.865426	0.264	11	1	2014	Saturday	4	1
11	12	2.83E+10	1/12/2014	Hemenwa	Department of Ne	605105418	X0.05	0.384	0.888803	0.198	12	1	2014	Sunday	4	1
12	13	2.83E+10	1/13/2014	Hemenwa	Department of Ne	605105418	X0.05	0.39	0.886098	0.204	13	1	2014	Monday	4	0
13	14	2.83E+10	1/14/2014	Hemenwa	Department of Ne	605105418	X0.05	0.408	0.883788	0.216	14	1	2014	Tuesday	4	0
14	15	2.83E+10	1/15/2014	Hemenwa	Department of Ne	605105418	X0.05	0.588	0.941742	0.21	15	1	2014	Wednesday	4	0

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.

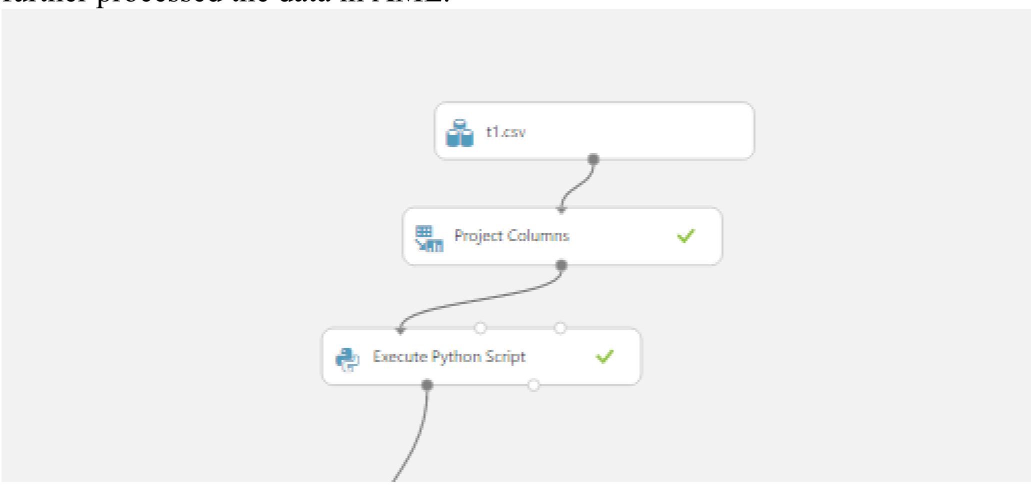
Technical Tools that were used in the implementation of web service from raw data is as follow:

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.



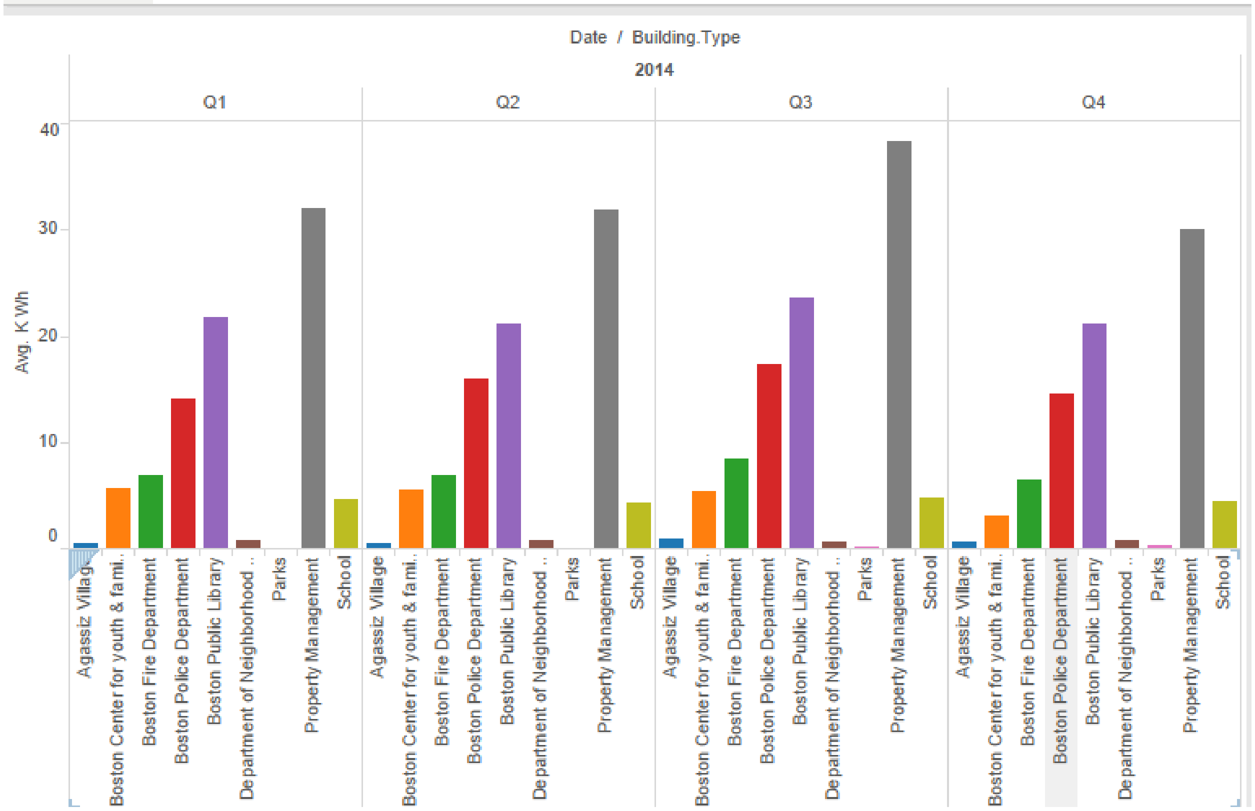
NOTE: t1 represents as Boston Property Management

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.

Title **Consumption by quarter by building type**



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

Year of ..	Quarter..	12 AM	1 AM	2 AM	3 AM	4 AM	5 AM	6 AM	7 AM	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM
		Variable													
2014	Q1	6.659	6.599	6.651	6.669	6.717	7.117	9.287	10.394	11.246	11.862	12.009	12.016	11.906	11.764
	Q2	6.492	6.425	6.384	6.408	6.413	6.807	8.815	10.157	11.159	11.907	12.140	12.235	12.194	12.123
	Q3	7.543	7.448	7.480	7.456	7.532	8.038	10.085	11.559	12.671	13.524	13.631	13.695	13.719	13.689
	Q4	6.304	6.283	6.254	6.251	6.276	6.736	8.738	10.051	10.956	11.461	11.538	11.551	11.463	11.441

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

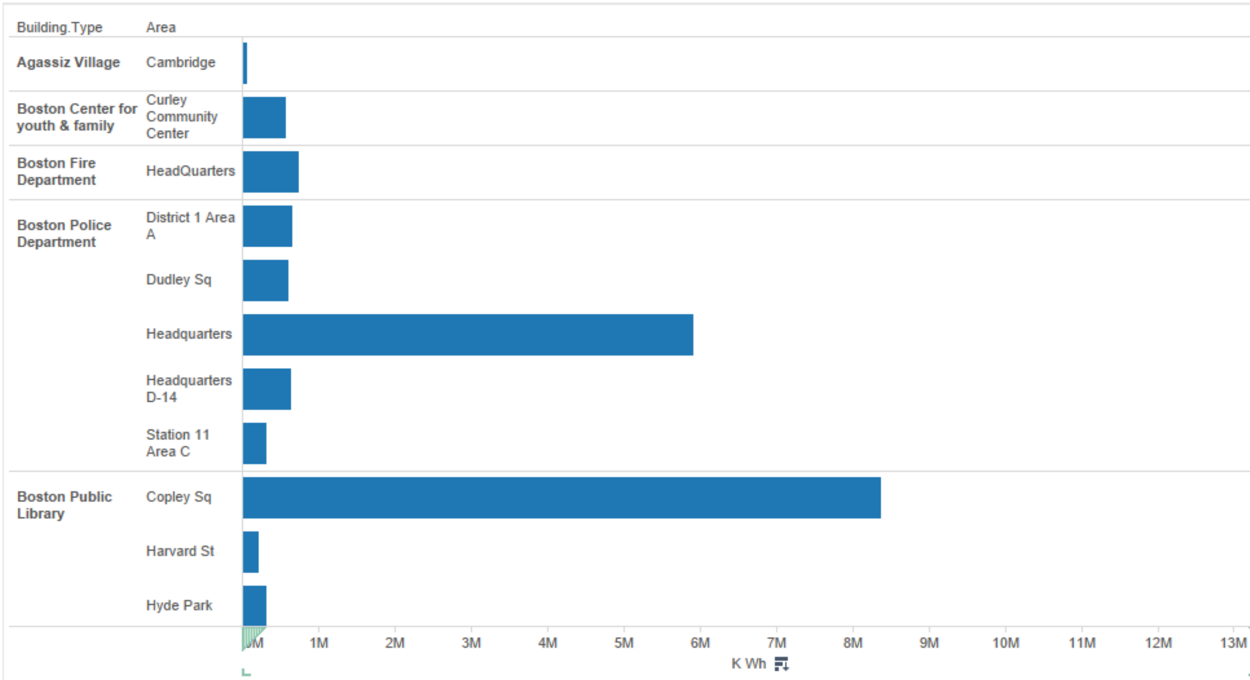


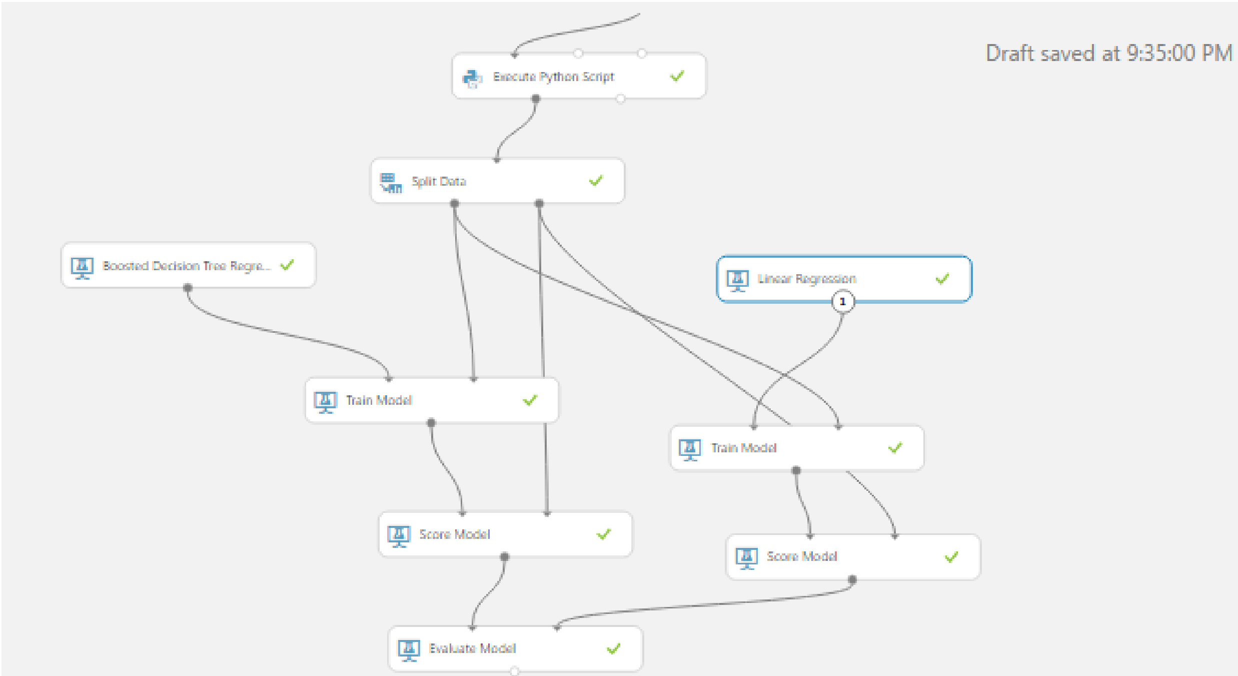
Figure 1

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:

Experiment created on 4/15/2016 › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	0.762523
Root Mean Squared Error	1.099787
Relative Absolute Error	0.428025
Relative Squared Error	0.229925
Coefficient of Determination	0.770075

Metrics

Mean Absolute Error	1.043092
Root Mean Squared Error	1.49031
Relative Absolute Error	0.585516
Relative Squared Error	0.422205
Coefficient of Determination	0.577795

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.

2) Boston Police Department:

Boston Property Management › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	1.072382
Root Mean Squared Error	1.902364
Relative Absolute Error	0.065708
Relative Squared Error	0.008494
Coefficient of Determination	0.991506

Metrics

Mean Absolute Error	15.67348
Root Mean Squared Error	19.75186
Relative Absolute Error	0.960353
Relative Squared Error	0.915638
Coefficient of Determination	0.084362

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.

3) Boston Fire Department:

Boston Fire Department › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	0.904223
Root Mean Squared Error	1.283198
Relative Absolute Error	0.575852
Relative Squared Error	0.373119
Coefficient of Determination	0.626881

Metrics

Mean Absolute Error	1.175639
Root Mean Squared Error	1.778813
Relative Absolute Error	0.748703
Relative Squared Error	0.717003
Coefficient of Determination	0.282997

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.

4) Boston Library:

PROP_MGT › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	5.995819
Root Mean Squared Error	11.117066
Relative Absolute Error	0.185186
Relative Squared Error	0.061736
Coefficient of Determination	0.938264

Metrics

Mean Absolute Error	1.58834
Root Mean Squared Error	3.248638
Relative Absolute Error	0.049057
Relative Squared Error	0.005272
Coefficient of Determination	0.994728

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.

5) School:

SCL › Evaluate Model › Evaluation results

Metrics

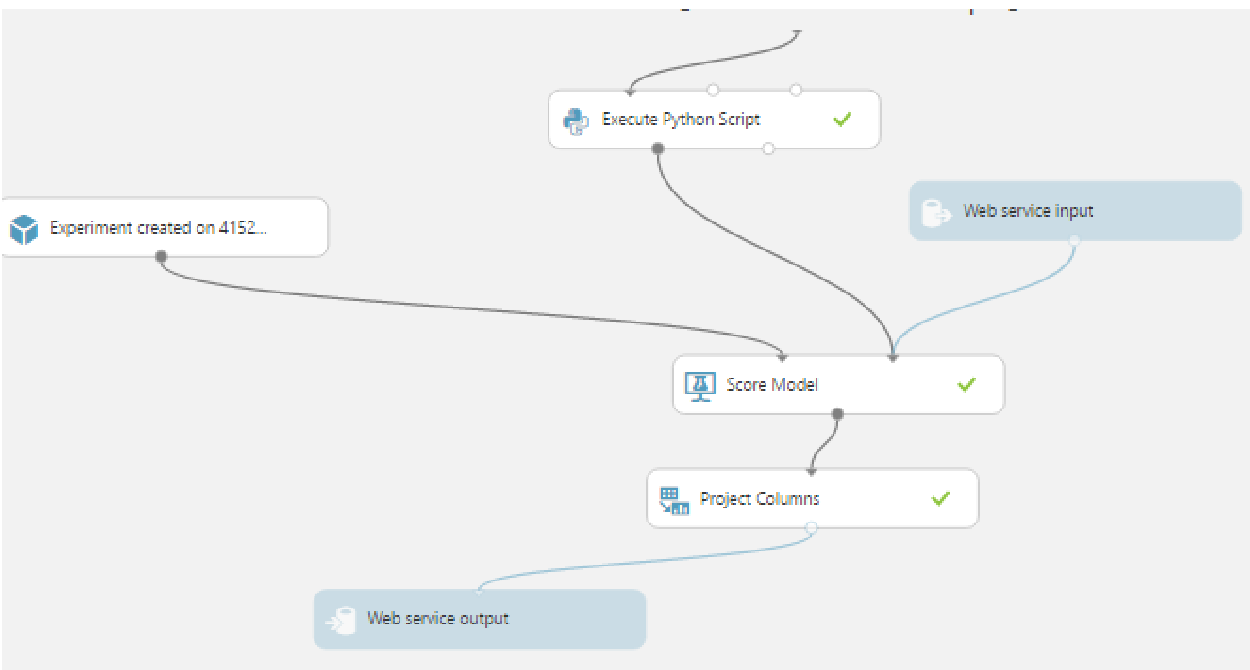
Mean Absolute Error	4.100983
Root Mean Squared Error	6.301432
Relative Absolute Error	0.711033
Relative Squared Error	0.610011
Coefficient of Determination	0.389989

Metrics

Mean Absolute Error	2.365437
Root Mean Squared Error	3.88717
Relative Absolute Error	0.410122
Relative Squared Error	0.232127
Coefficient of Determination	0.767873

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.

API key

ddMZyeiL/6a14Zlh2BlOfUkkf76A70/0DgtGuWVuXBy9RLploLlDAuAGe5PIP7cLmqcvRC805h7CjxV2ysijZw==

Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	Test	Excel 2013 or later | Excel 2010 or earlier workbook	4/17/2016 10:23:16 PM
BATCH EXECUTION		Excel 2013 or later workbook	4/17/2016 10:23:16 PM

Amazon web service:

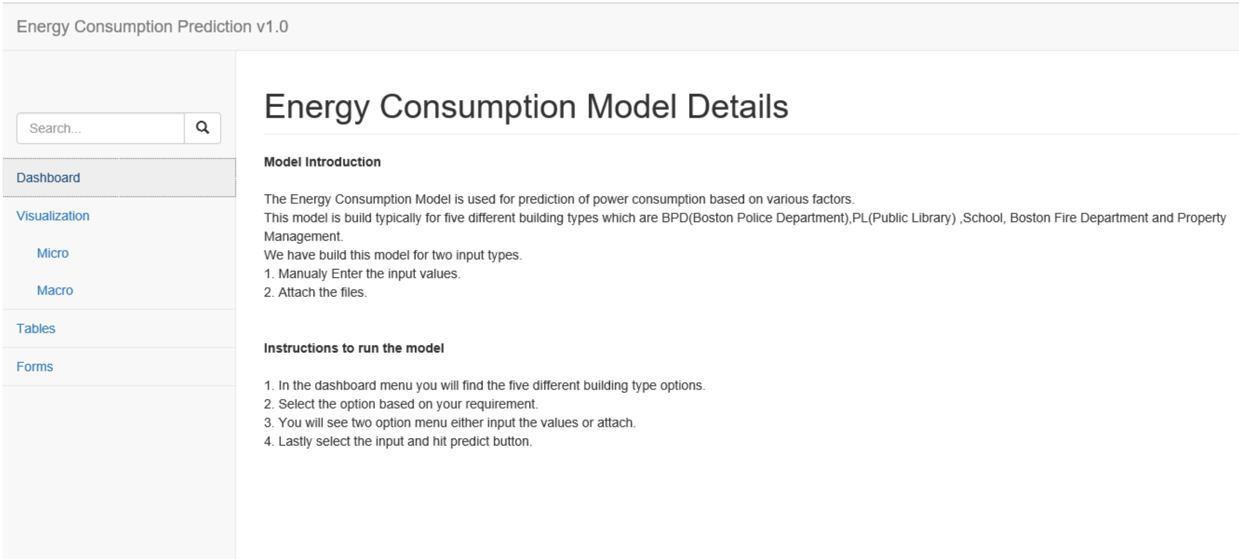
We used AWS for using the API for prediction. The following is the snippet that shows the configuration for prediction of power consumption by School.

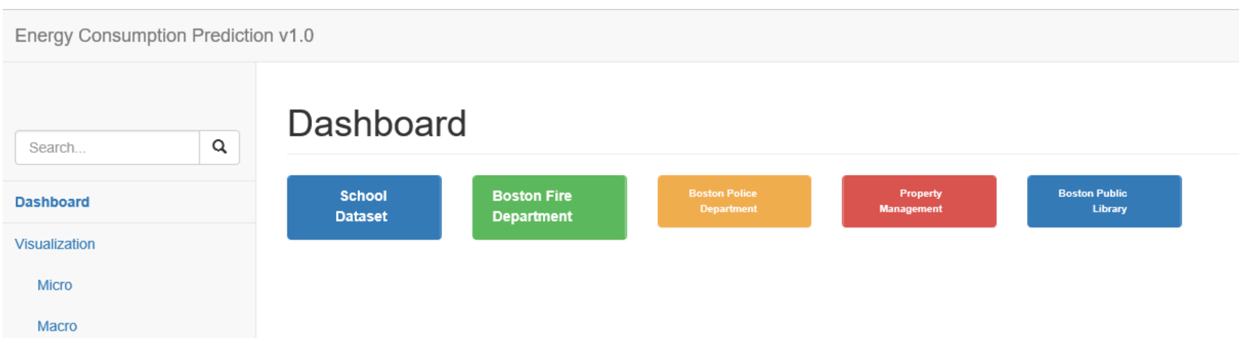
	Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
	test	i-3b415fe1	t1.micro	us-west-2c	stopped		None		
	ADS	i-58e3c380	m4.large	us-west-2b	running	2/2 checks ...	None	ec2-52-38-108-155.us-w..	52.38.108.155

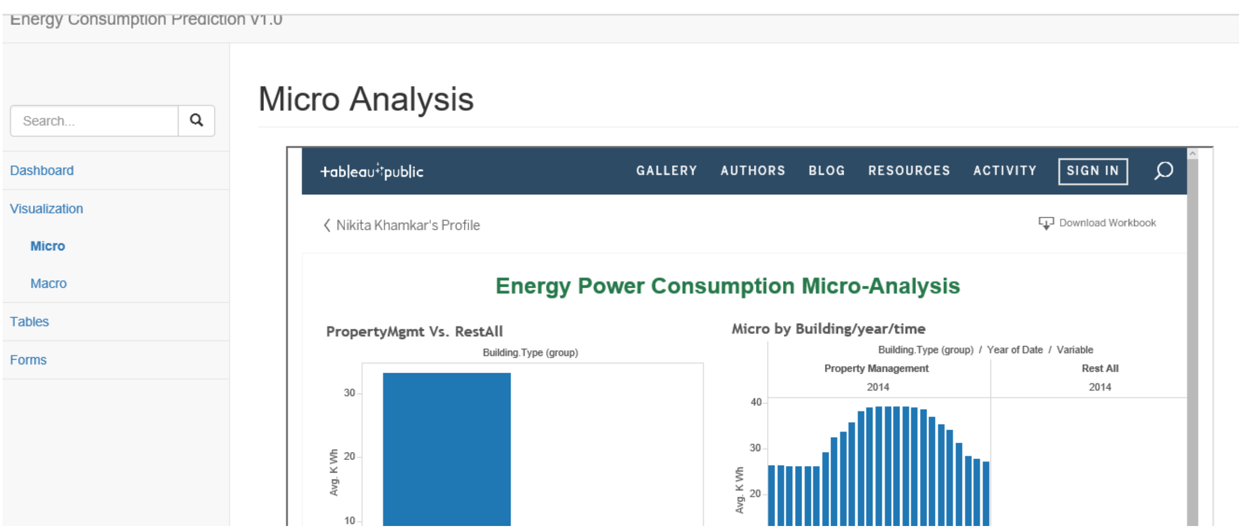
After creating the API connection, we build the front end in the HTML and CSS. We have published our tableau visualization in public tableau and used that link to show the visualization in our front end. Following is the snippet of tableau:

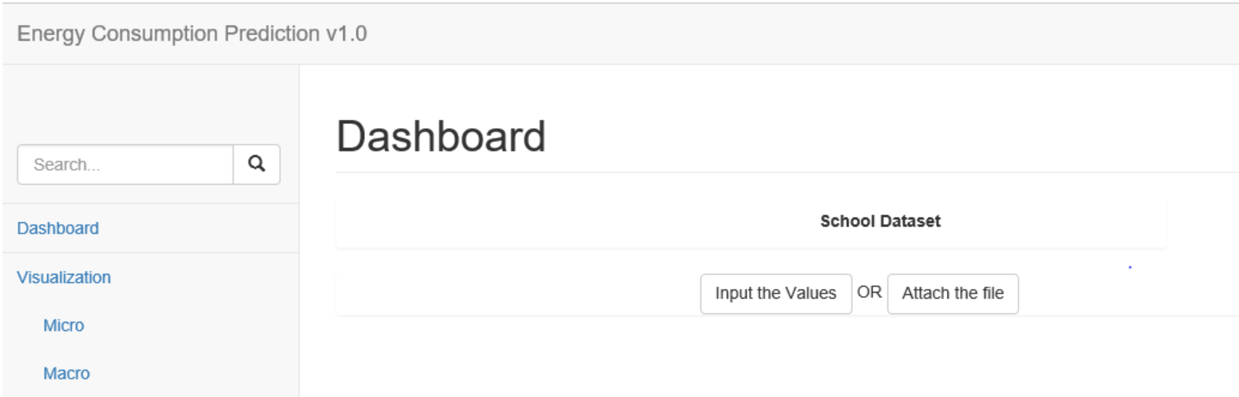
```
<!-- Page Content -->
<div id="page-content-wrapper">
    <div class="container-fluid">
        <div class="row">
            <div class="col-lg-12">
                <h1>Assignment 3</h1>
                <asp:ContentPlaceHolder ID="Details" runat="server">
                <iframe src="https://public.tableau.com/profile/shrenik.jain#!/vizhome/ElectricityConsumption_2/ElectricityConsumption" width="1100" height="1100">
                </iframe>
                </asp:ContentPlaceHolder>
            </div>
        </div>
    </div>
</div>
```

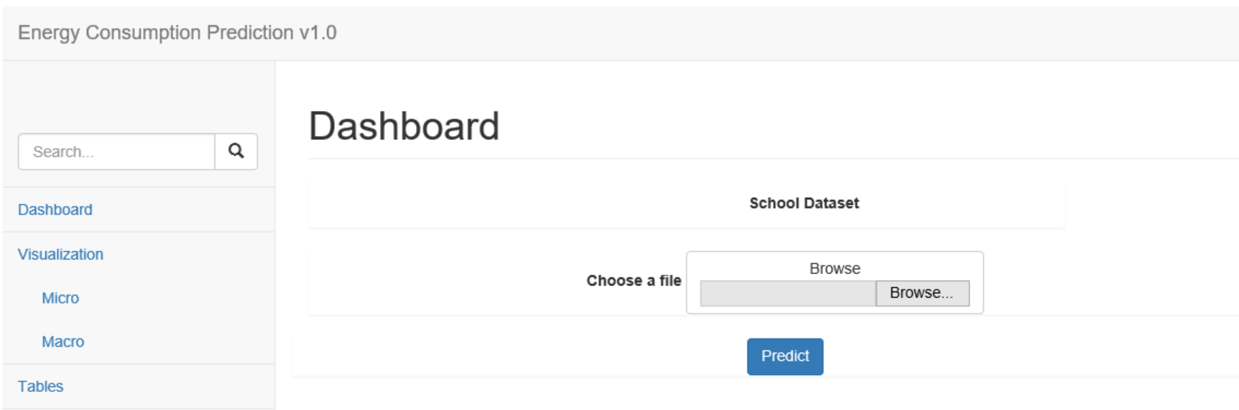
The following are the screenshots from our website that we created:

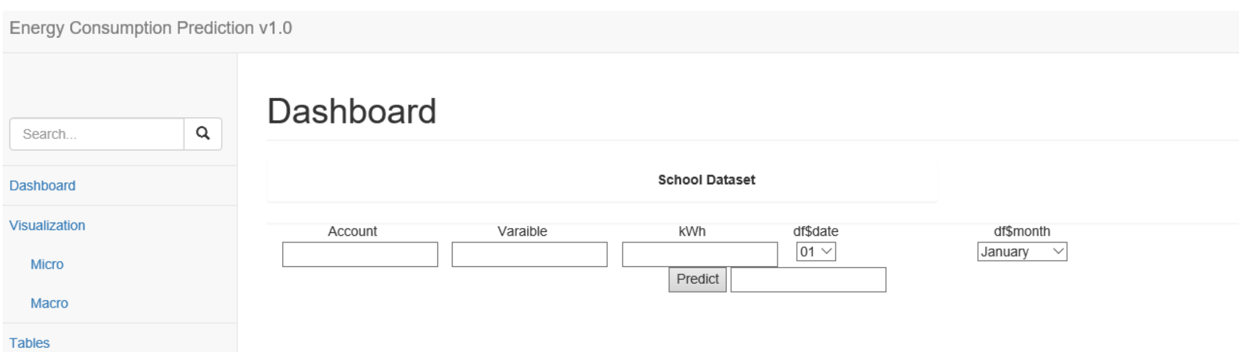

Energy Consumption Prediction v1.0
Search...
Dashboard
Visualization
Micro
Macro
Tables
Forms
Energy Consumption Model Details
Model Introduction
The Energy Consumption Model is used for prediction of power consumption based on various factors.
This model is build typically for five different building types which are BPD(Boston Police Department),PL(Public Library) ,School, Boston Fire Department and Property Management.
We have build this model for two input types.
1. Manualy Enter the input values.
2. Attach the files.
Instructions to run the model
1. In the dashboard menu you will find the five different building type options.
2. Select the option based on your requirement.
3. You will see two option menu either input the values or attach.
4. Lastly select the input and hit predict button.


Energy Consumption Prediction v1.0
Search...
Dashboard
Visualization
Micro
Macro
Dashboard
School Dataset
Boston Fire Department
Boston Police Department
Property Management
Boston Public Library


Energy Consumption Prediction v1.0
Search...
Dashboard
Visualization
Micro
Macro
Tables
Forms
Micro Analysis
tableau public
GALLERY
AUTHORS
BLOG
RESOURCES
ACTIVITY
SIGN IN
Nikita Khamkar's Profile
Download Workbook
Energy Power Consumption Micro-Analysis
PropertyMgmt Vs. RestAll
Building.Type (group)
Avg. K Wh
Micro by Building/year/time
Building.Type (group) / Year of Date / Variable
Property Management
Rest All
2014
2014
Avg. K Wh







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/

Github Link: https://github.com/prateekmane99/ADS

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.