**Overview**

The dataset provided to us included multiple csv files containing power consumption data belonging to 10 different departments in the city of Boston, MA.The dataset had the following attributes – Time, Day, Month, Year, kWh, kVARh and Power Factor.Each department had multiple accounts and channels. kWh gives the power consumption through each of the channel per account and department.

**Approach**

Our approach towards analysis of Boston power consumption is to categorize the data depending on various departments like Boston Police Department, Boston Fire Department, Boston Public Library, etc. and publish a website to predict the power consumption based on different departments. Web Application features:

* Dashboard for easy online access and visualization of electricity consumption for different departments in Boston
* Enabling user to predicted power consumption value in kWh by selecting the time, day and month
* Enabling user to choose between multiple departments, timeframe and unit to visualize the prediction data

**Technologies**

* Data Preprocessing – R
* Prediction – Azure Machine Learning Studio
* Visualization – Tableau
* Front-End – HTML, CSS, Bootstrap, JavaScript

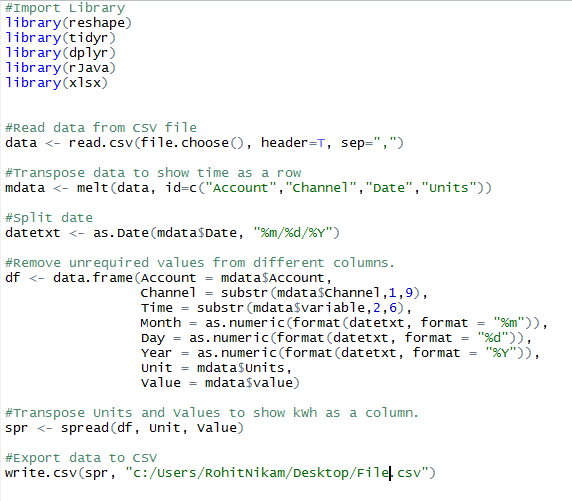
**Data Preparation**

The provided dataset includes incorrect values (outliers), missing and null values. Some of the data belonging to different channels were duplicates and some were inappropriate.

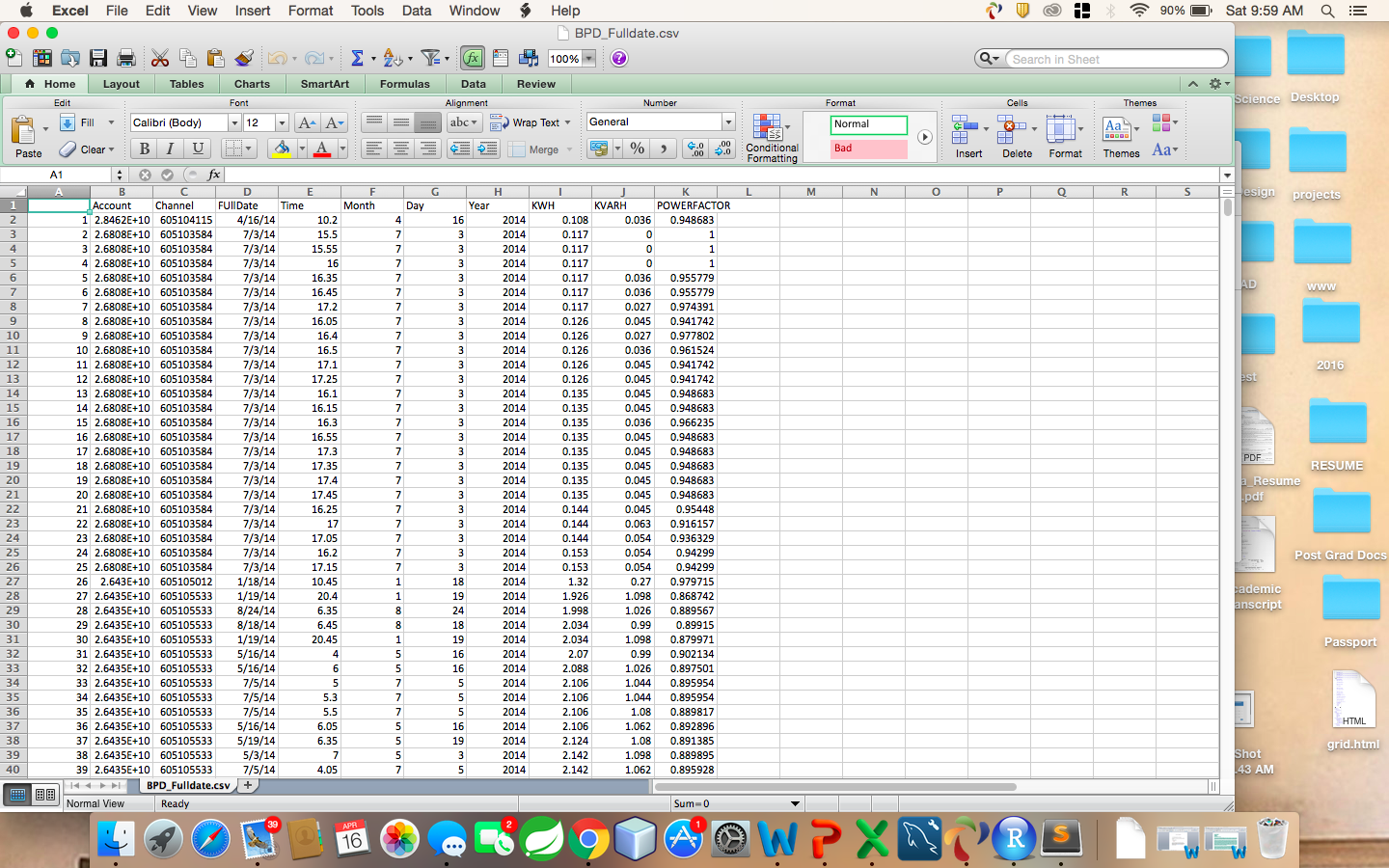
So in order to clean the dataset we executed the following steps in R:

* 1. Split date to get daily and monthly predictions
  2. Transpose Time to get it as one of the attribute
  3. Transpose Unit and Value to get kWh, kVARh and PowerFactor.
  4. Convert the format of date (string) as a date time for visualization
  5. One file per department

**R Script for Data Cleaning**

****

**Processed Dataset**

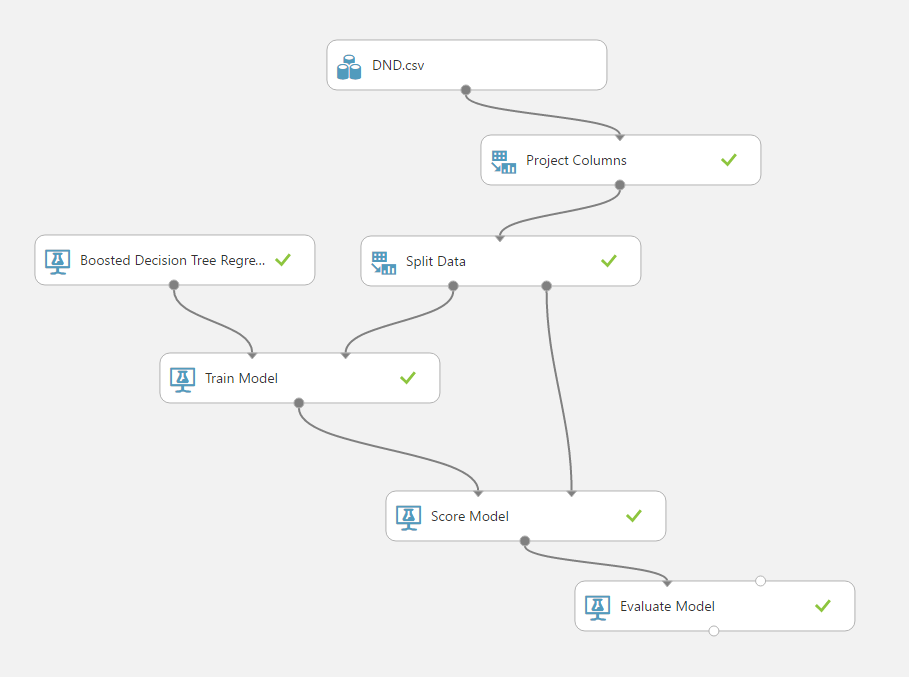
****

**Prediction Model**

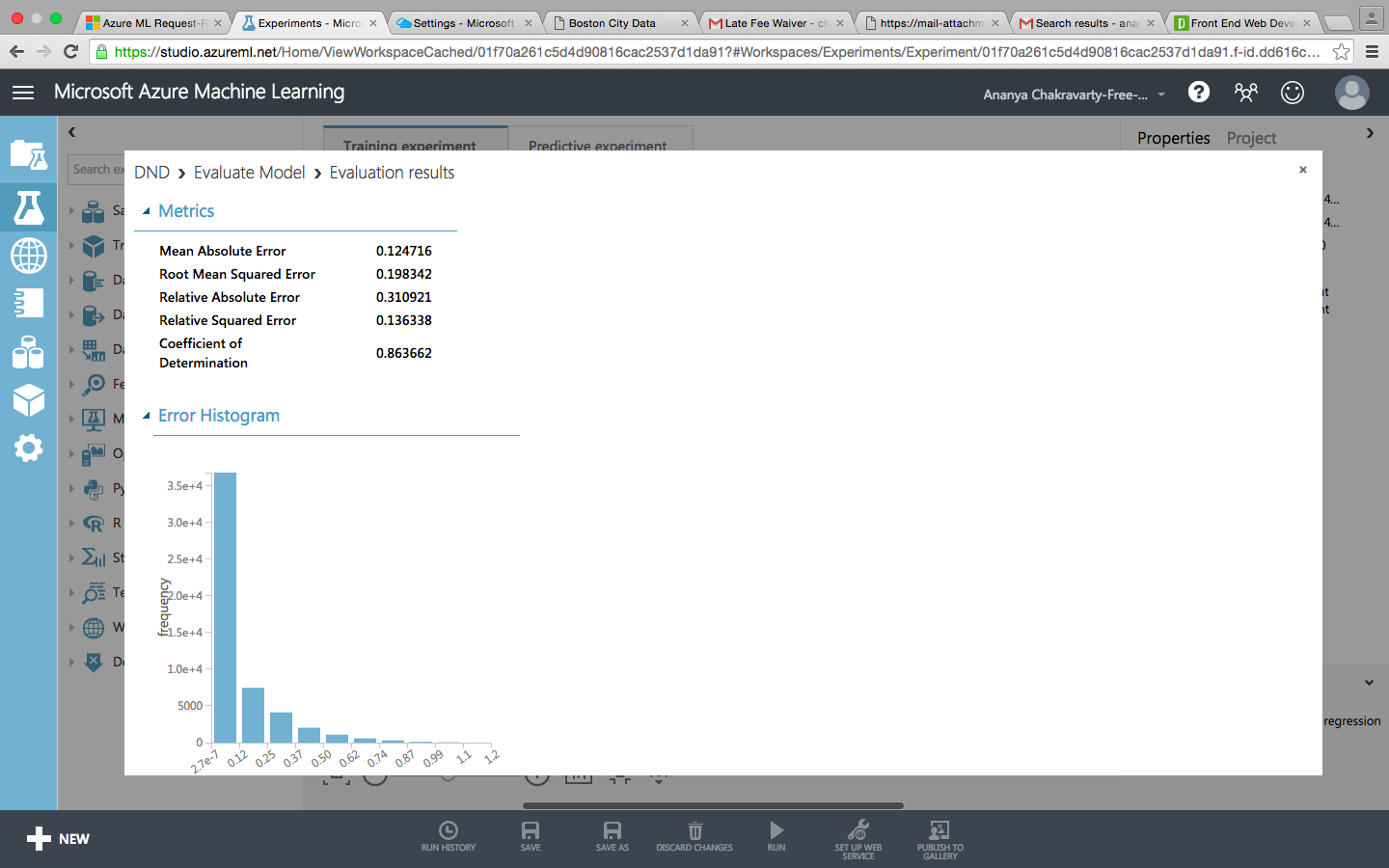
For predicting the power consumption (KWH), we used **Boosted Decision Tree Regression mode**l in azure machine learning studio. It is a supervised learning method to create ensemble of regression trees.

Properties:

* Used Projected Column to filter out the columns, which did not contribute to our prediction.
* Split the data as Train and Score (70% traning)
* Used Boosted Decision Tree Regression to predict kWh values
* Consider kWh as our output/predicted value



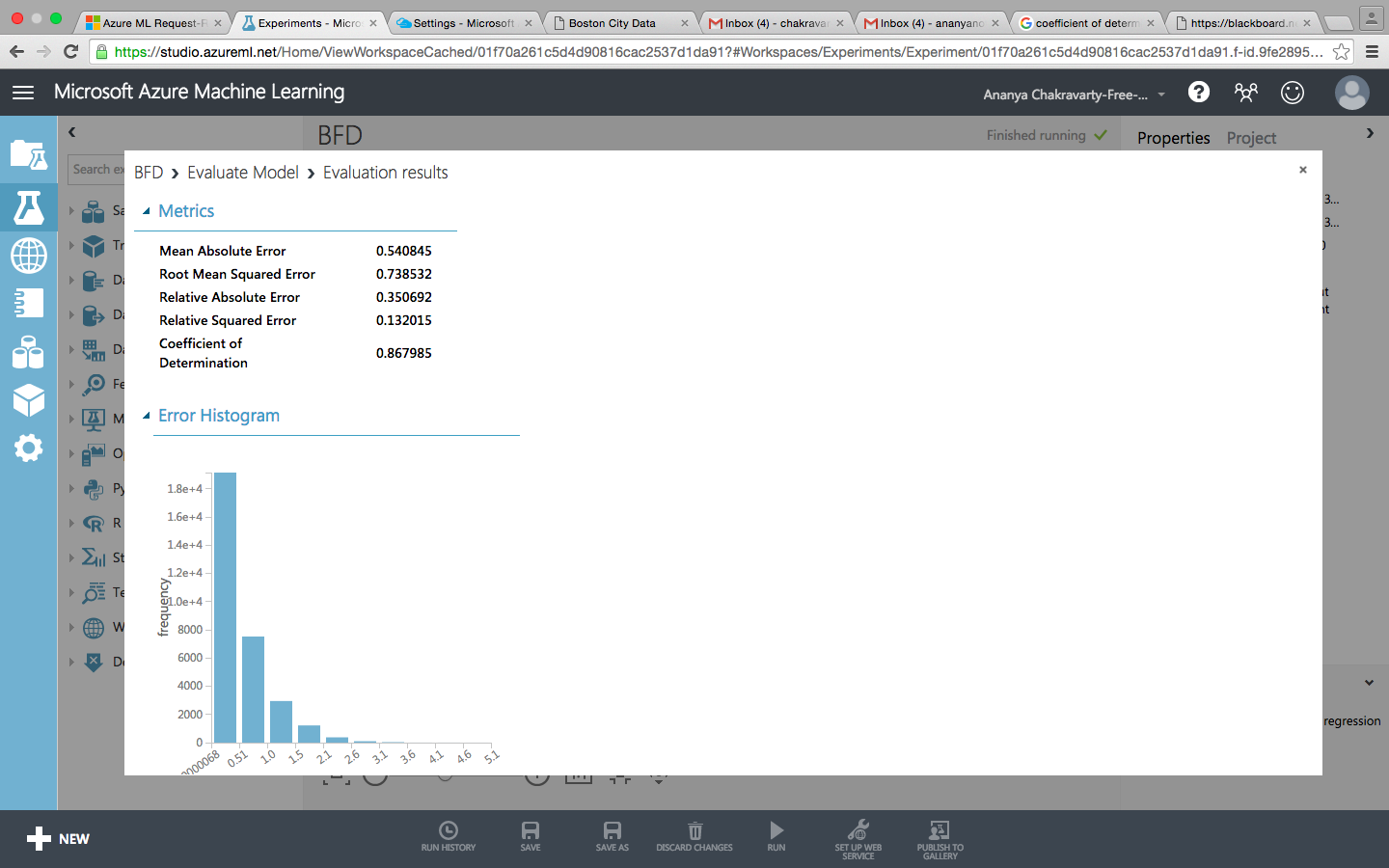
The above figure shows the Boosted Decision Tree Model for DND Department.

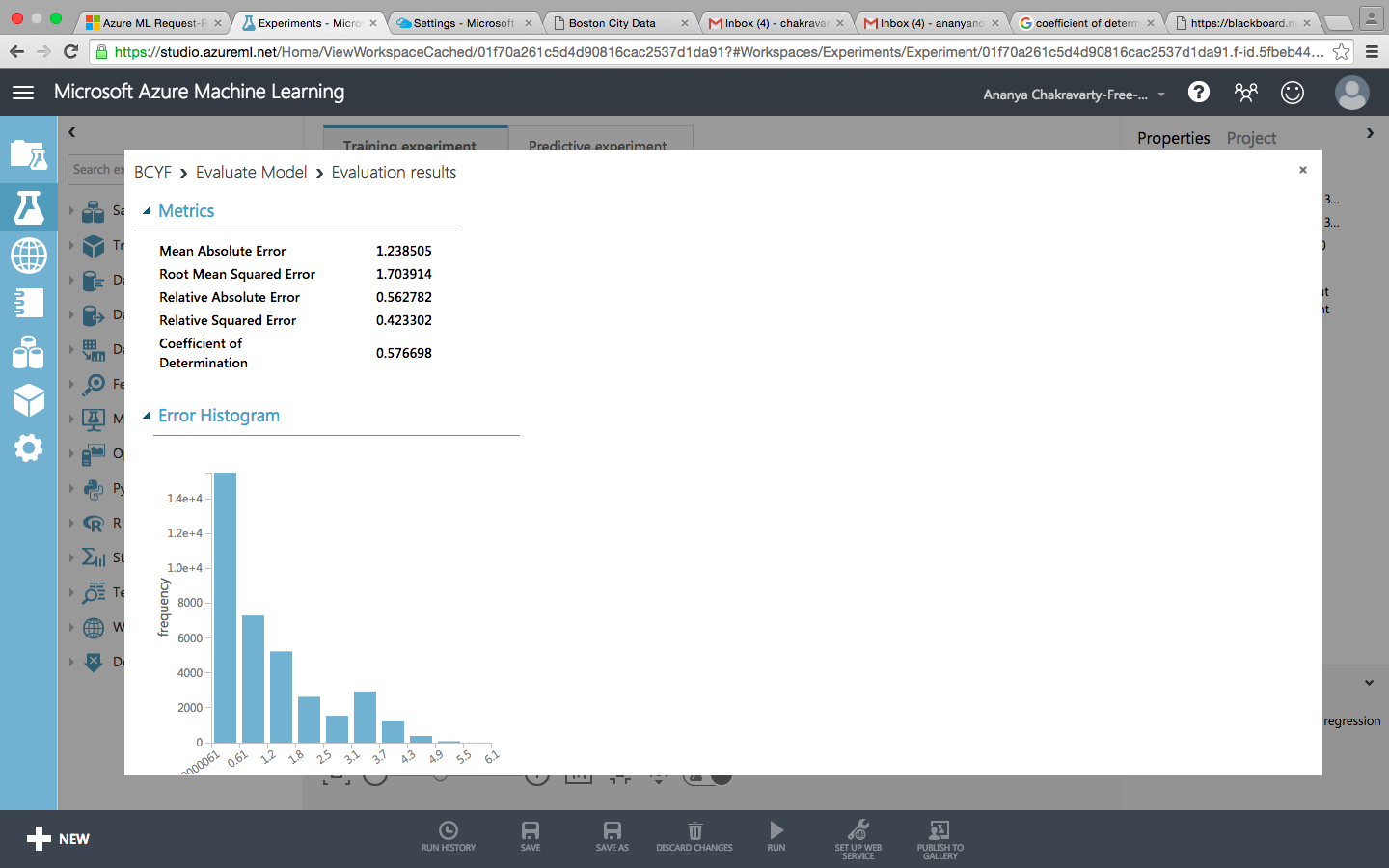


The figure on the right shows the metrics of the prediction model for Development of Neighborhood Department (DND). The coefficient of Determination for the model is evaluated as 0.86, which shows that the dependent variable can be predicted from the independent variable with less error.

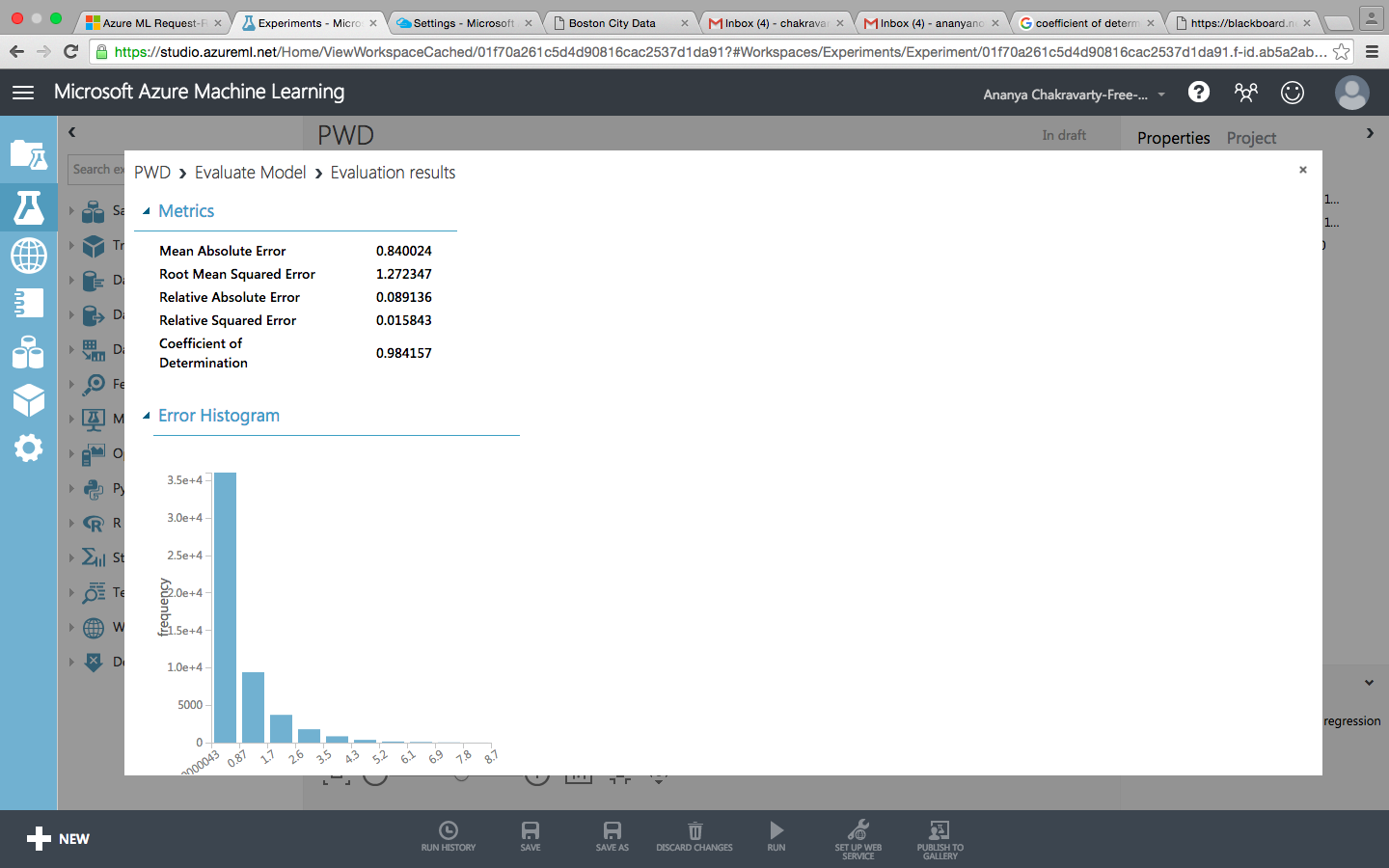
After evaluating the model we used the generated API and URL in Microsoft Azure Web App to generate dynamic platform for predicting the power consumption.

Similarly, we executed the Boosted Decision Regression Tree Model for all the other departments and the evaluation metrics of all the departments is shown below:

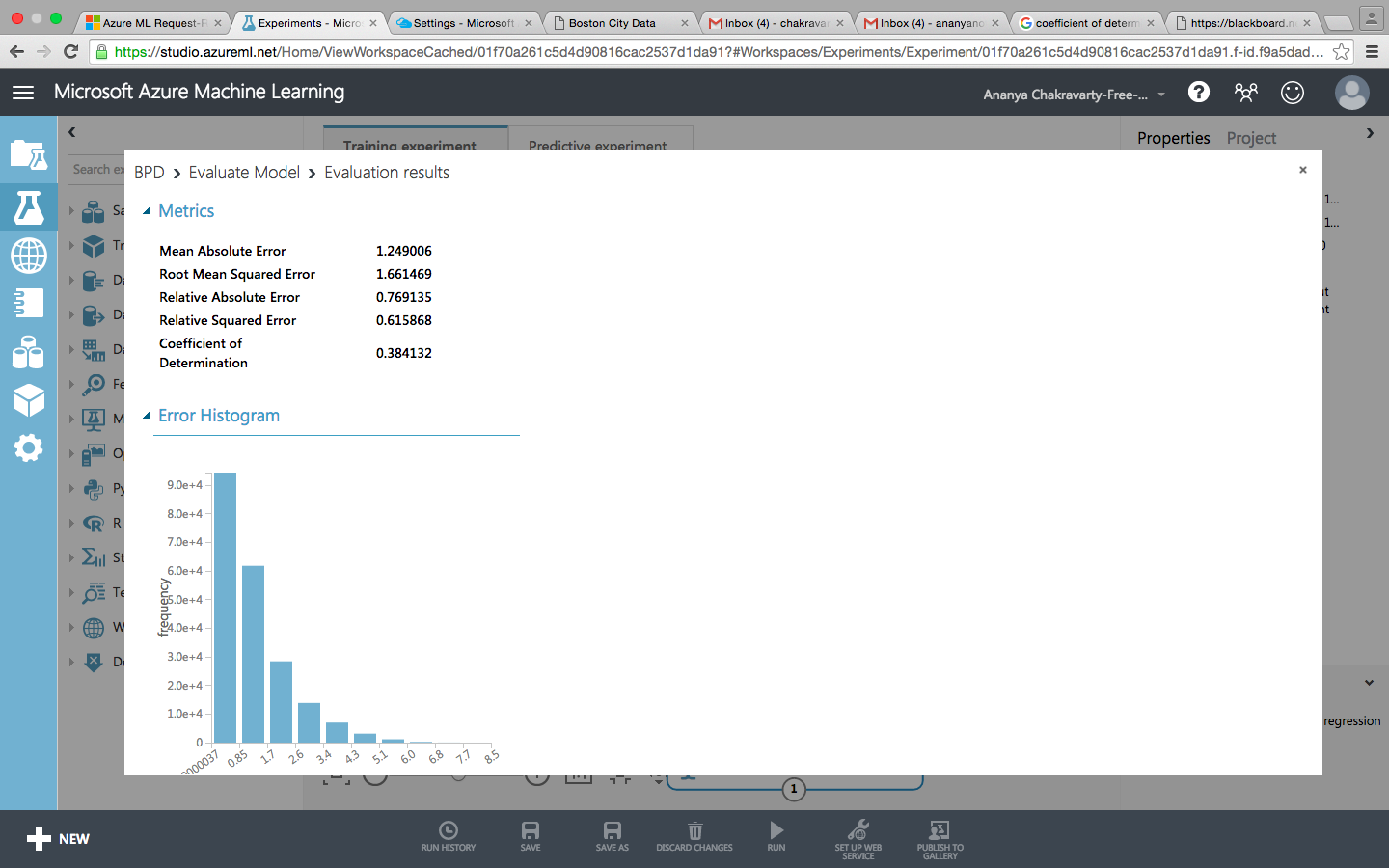
Boston Fire Department 🡪



🡨 Boston Center for Youth and Families



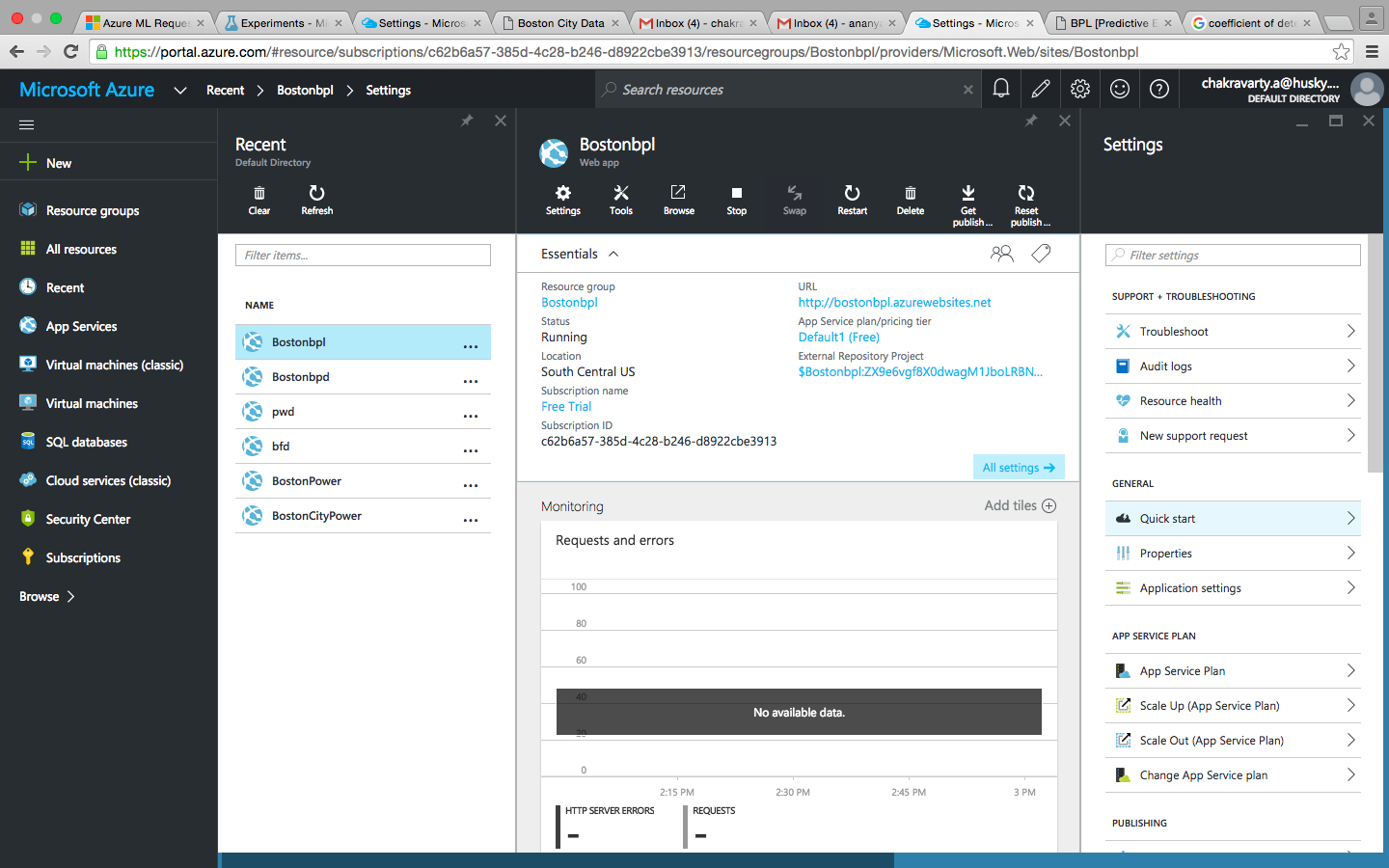
Public Works Department 🡪



🡨 Boston Police Department

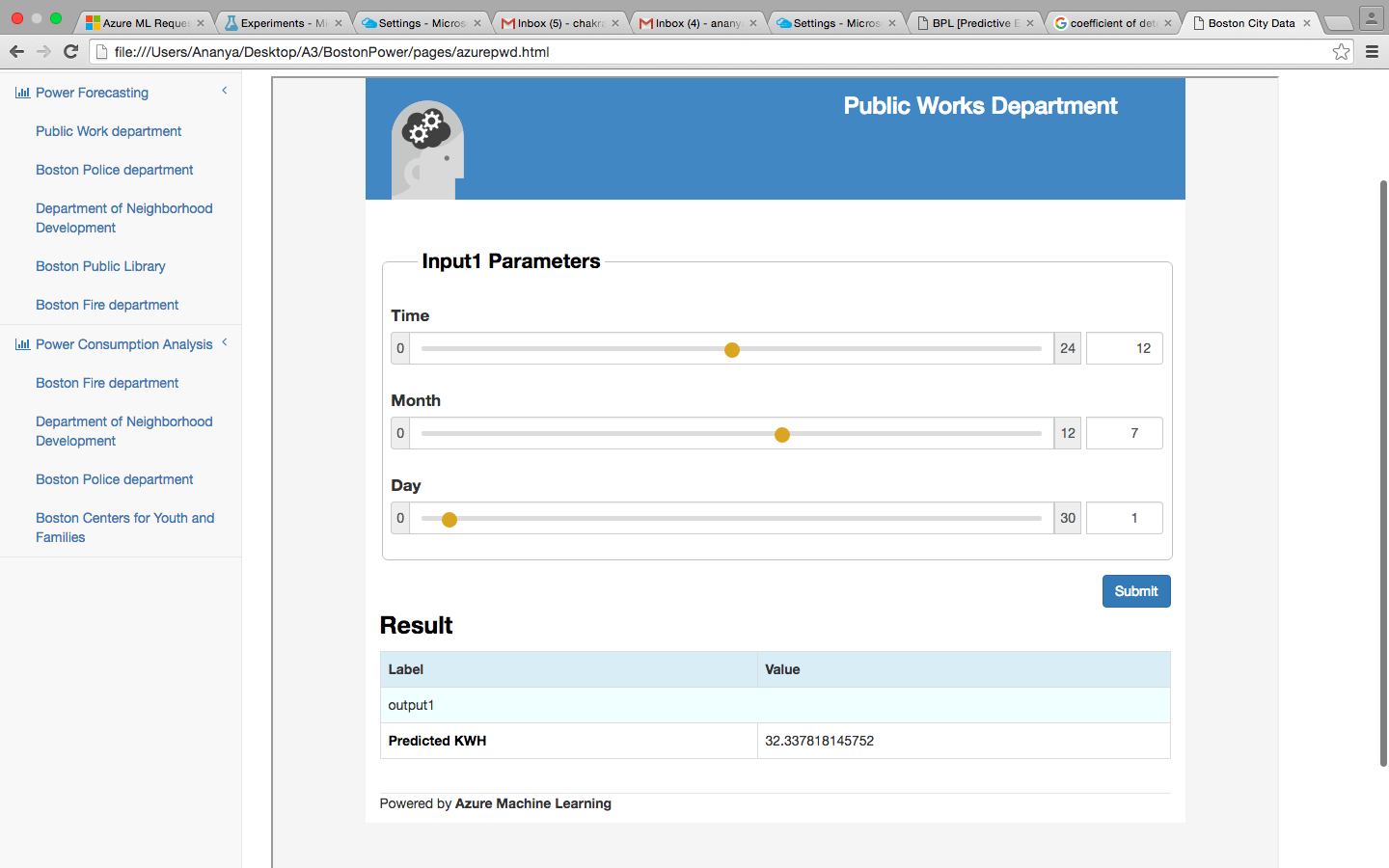
**Microsoft Azure Marketplace**

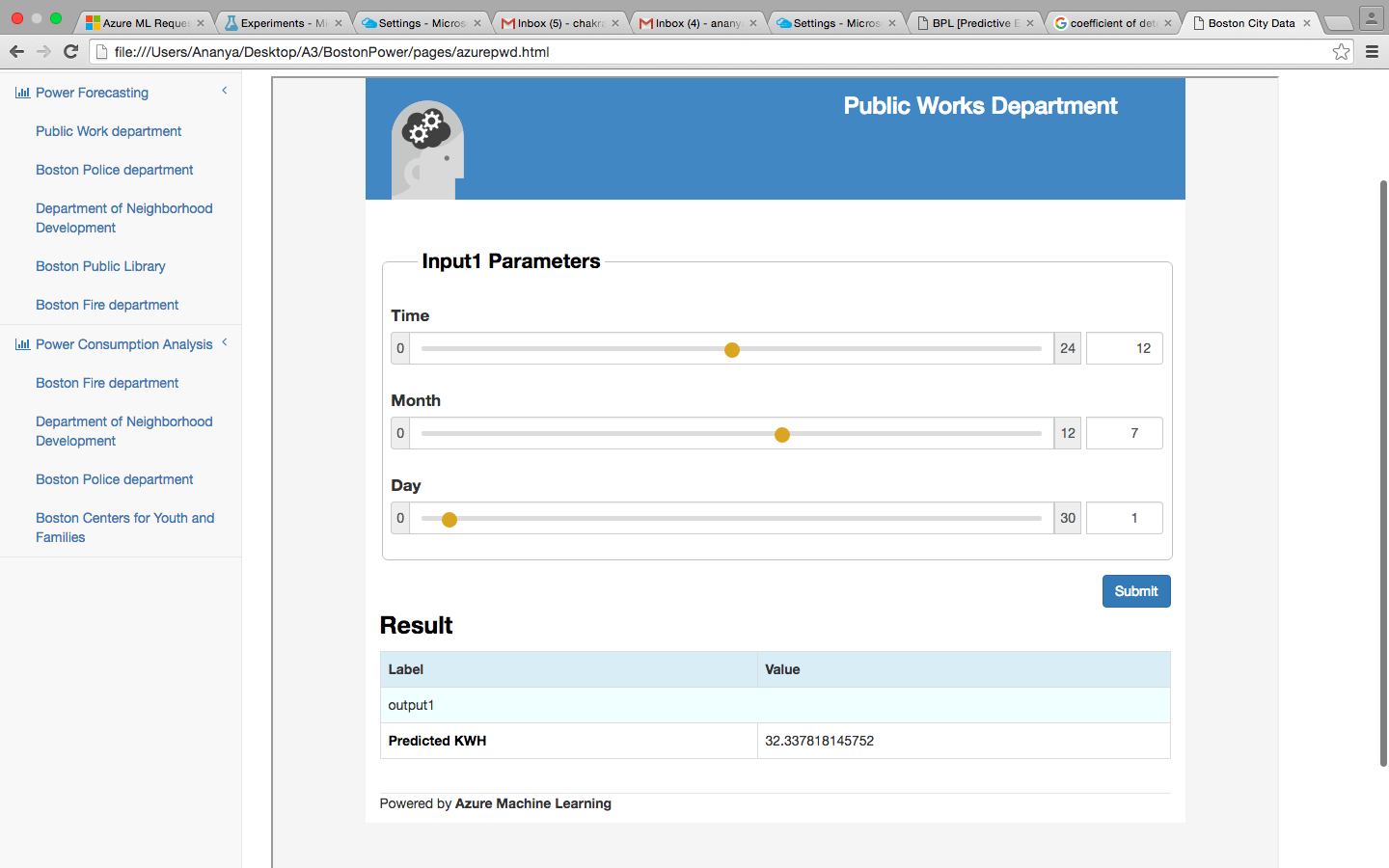
After generating the API and web URL for the prediction model of all the departments we used Microsoft Azure Marketplace platform so that any user can input valued like Date, Time and Month to get the predicted KWH

.

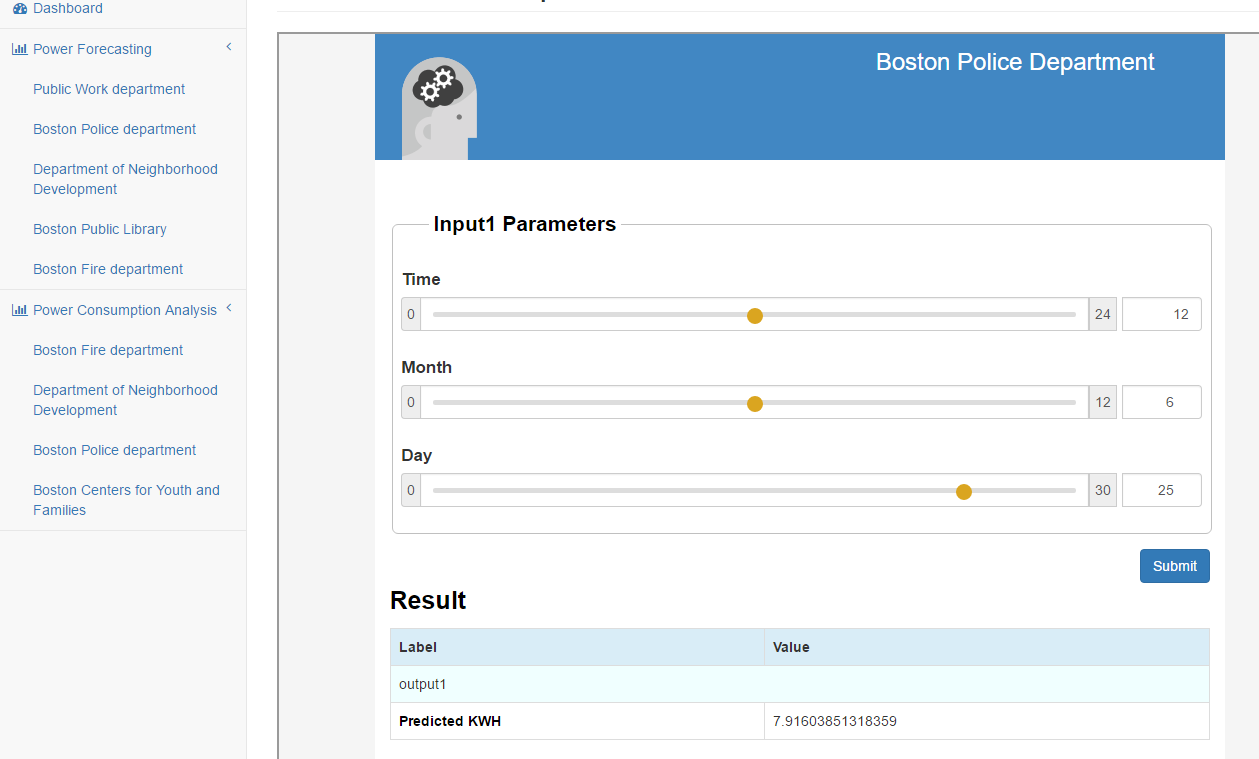
The figure on the left shows the generated web URL of the Boston Public Department using the web API and URL of the predicting model generated in Azure Studio.

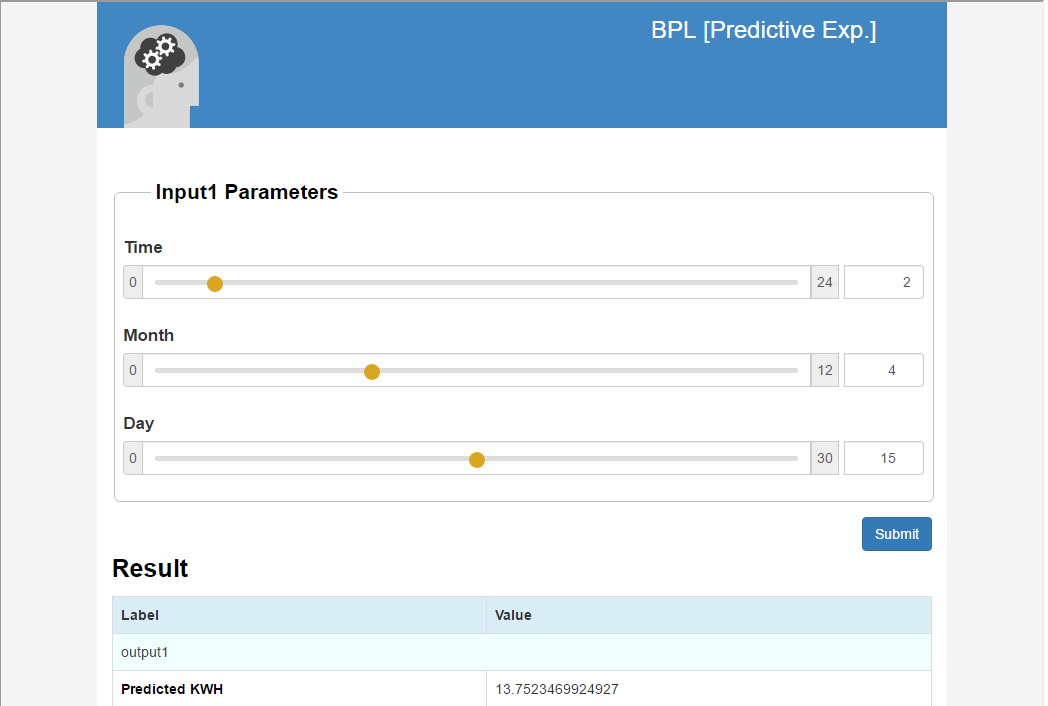
Now, the URL generated redirects to the following page:



After giving input for date, time and month; the predicted value of KWH (power consumption) is shown in the result

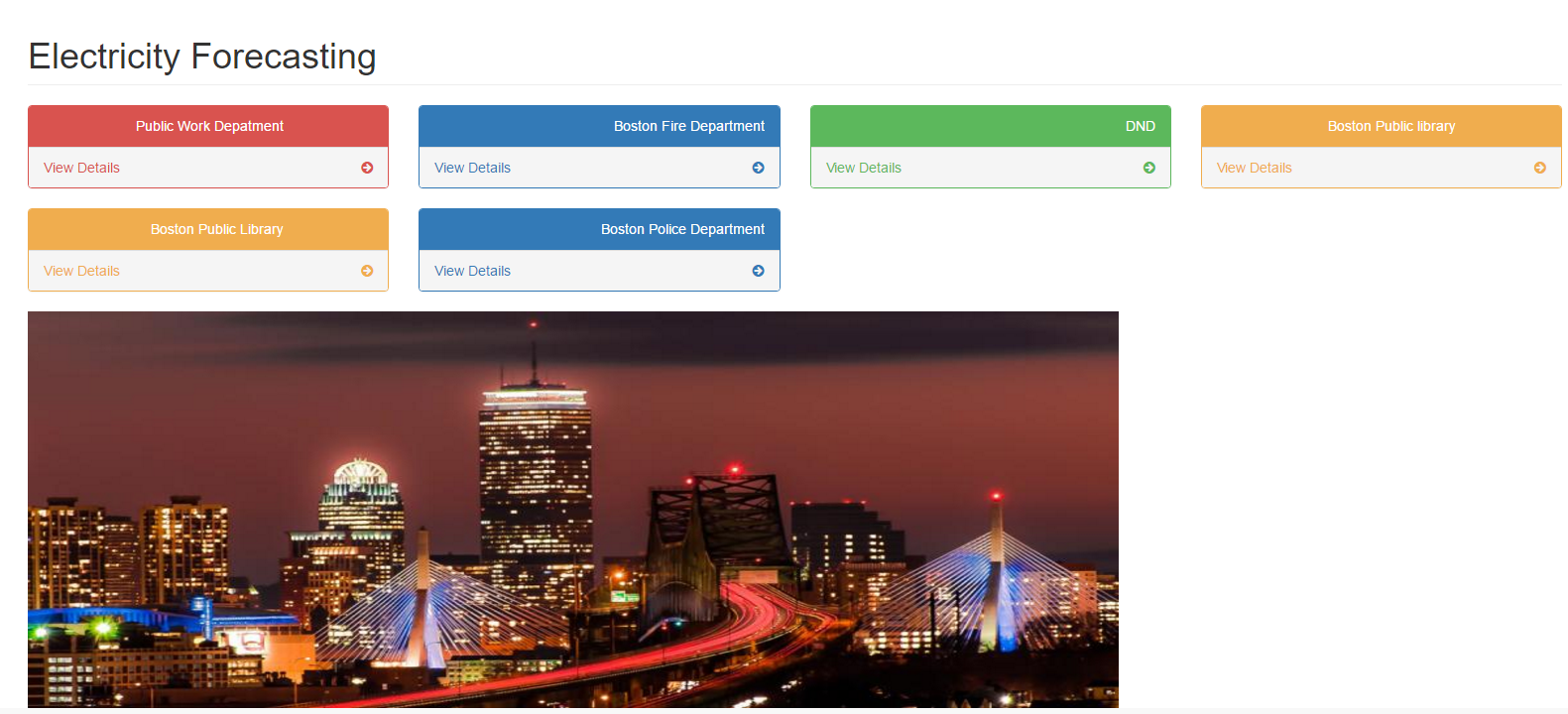
Below are the prediction given by our model for Active Power consumed by some departments -

****

****

**Web Interface –**

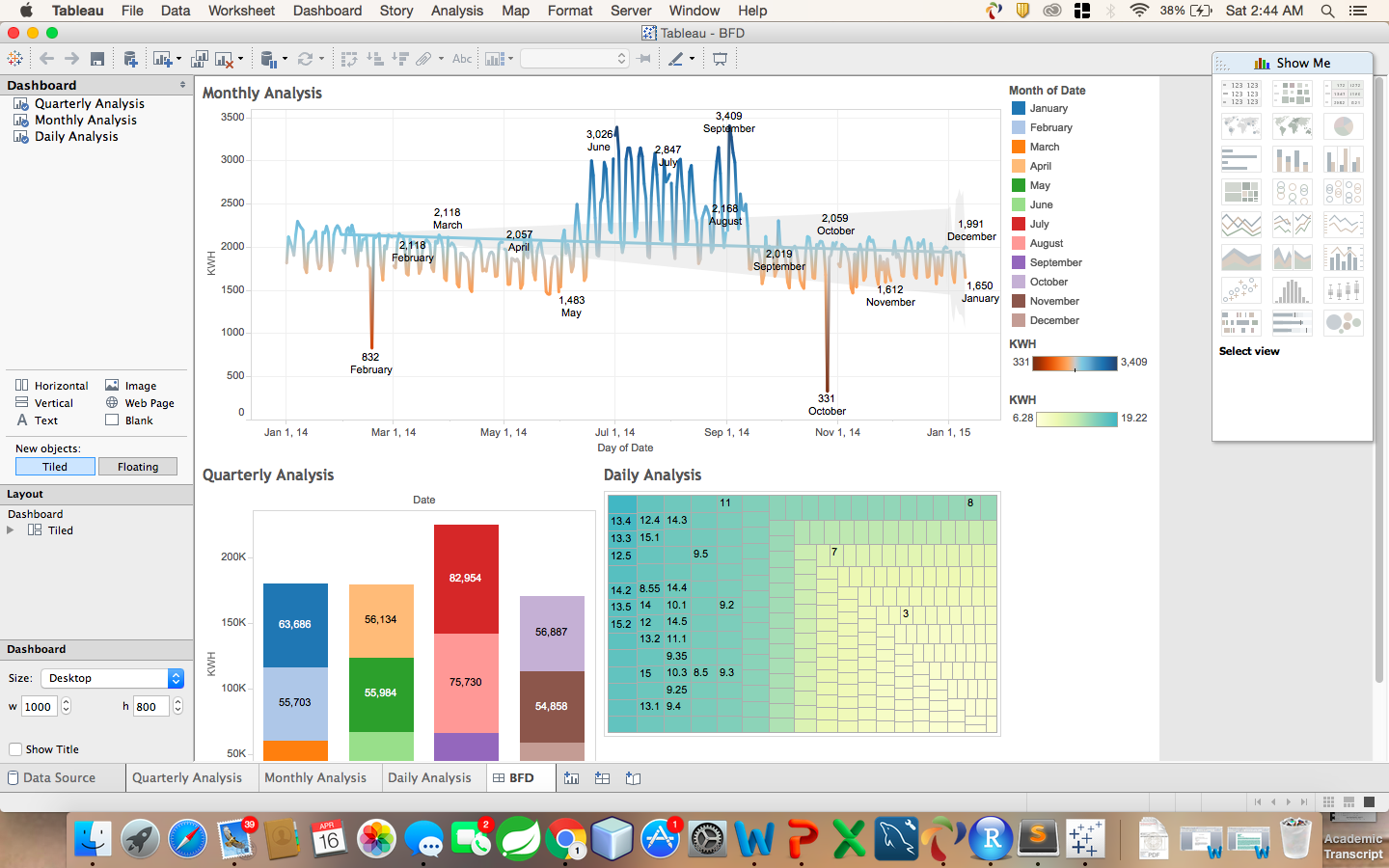
This is what our web application looks like. Users can access the visualizations and the prediction model using a simple navigation bar.



**Visualization**

To analyze the given dataset, we used Tableau to visualize the trend in power consumption of different departments. We examined the dataset based on three different criteria:

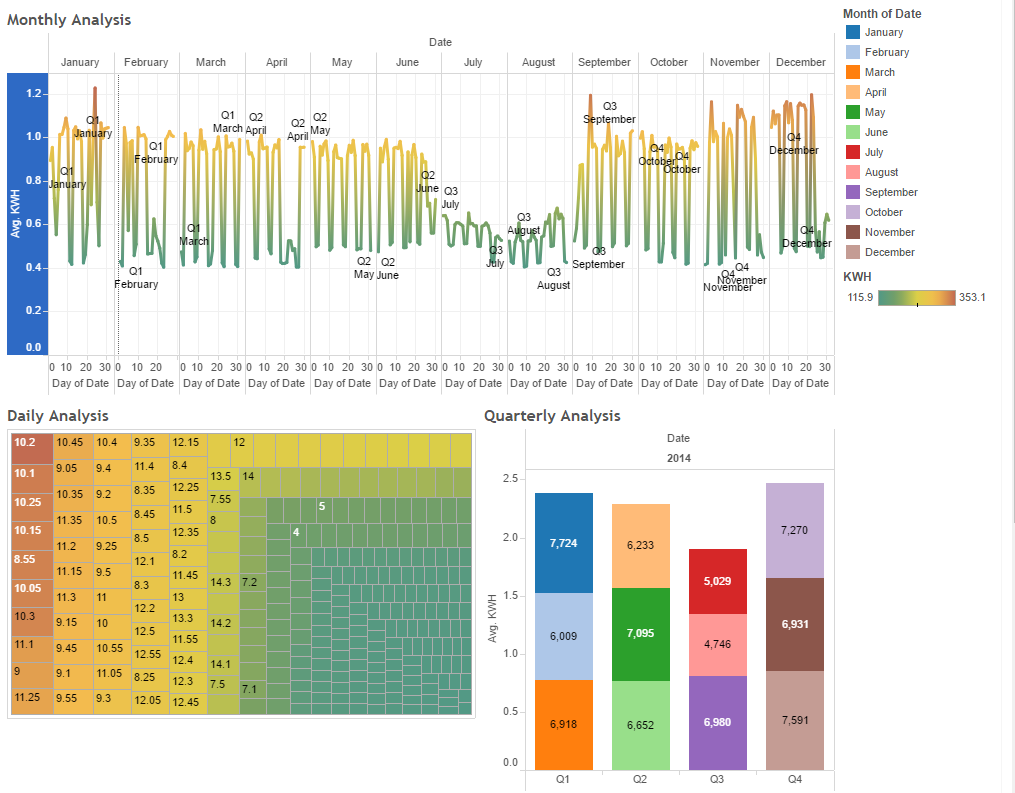
* Quarterly Analysis
* Monthly Analysis
* Daily Analysis



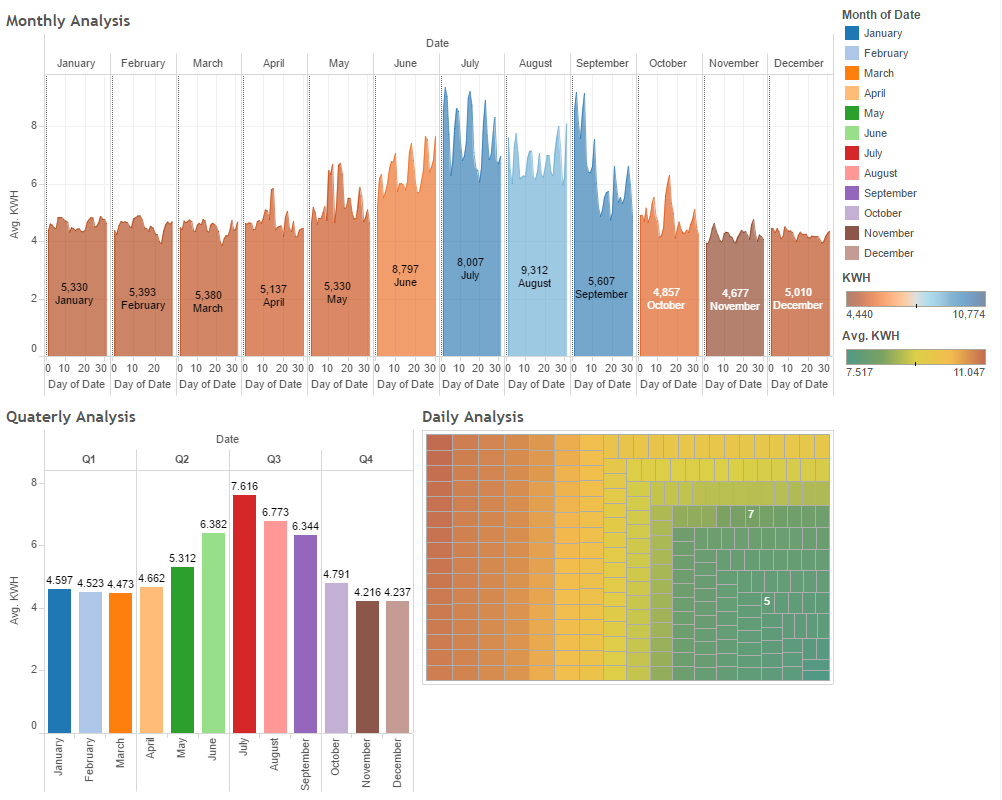
The above figure shows the dashboard for Boston Fire Department Power Consumption Analysis. Quarterly Analysis shows that the maximum energy consumption is in the third quarter and that too specifically in the month of September (showed in red). To granulize further, we analyzed the power consumption trend in each month and explored that September 2 has the highest power consumption. To further analyze, we studied the power consumption of September 2 for 24 hours.

Similarly we have created dashboards for 3 other departments which include –

**Departments of Neighborhood Development**



**Boston Police Department**

****

**Boston Centers for Youth and Families**

****

**References**

[**https://azure.microsoft.com/en-us/overview/what-is-azure/**](https://azure.microsoft.com/en-us/overview/what-is-azure/)

[**https://studio.azureml.net/**](https://studio.azureml.net/)

[**https://public.tableau.com/s/**](https://public.tableau.com/s/)