Environmental Science & Technology

Project Report

AQI Meets AI

Submitted By

Aamir Farhan (20161078)

Oishika Pradhan (20161224)

Under the guidance of

Dr. Rama Chandra Prasad P.

Contents

Introduction

a) Literature Survey

Attention to every day levels of air contamination is imperative to the citizens, particularly for the individuals who experience the ill effects of sicknesses brought about by exposure to air pollution. Further, achievement of a country to improve air quality relies upon the help of its citizens who are well-educated about neighborhood and national air contamination issues. Consequently, a straightforward yet powerful correspondence of air quality is important. The idea of an air quality index(AQI) that changes weighted estimations of individual air contamination related parameters (for example PM2.5, SO2, CO,etc.) into a solitary number or set of numbers is generally utilized for air quality correspondence and decision making in numerous nations.

Definition

The Air Quality Index (AQI) is a measurement that has been developed by the Environmental Protection Agency (EPA) to provide the regular citizen with accurate, up-to-date, and easily understandable information about daily levels of air pollution. AQI can be calculated using different formulae, like weighted average, geometric mean and so on. The most commonly used formula is

$$I = \frac{I_{\text{high}} - I_{\text{low}}}{C_{\text{high}} - C_{\text{low}}}(C - C_{\text{low}}) + I_{\text{low}},$$

To make it easier to understand, the AQI is divided into six levels of health concern as shown below in Figure 1

Air Quality Index (AQI) Values	Levels of Health Concern	Colors
When the AQI is in this range:	*...air quality conditions are:*	*...as symbolized by this color:*
0 to 50	Good	Green
51 to 100	Moderate	Yellow
101 to 150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon

Figure 1

b) Objectives Of The Study

i) Creating a structured, clean and colour-coded dataset of historical AQI and PM2.5 concentration of major Indian cities.

ii) Developing a script to find the AQI value from the PM2.5 concentration and analysing the trend of Air Quality Index over a period starting from January 2016 to November 2019.

iii) Predicting AQI in the near future given past AQI measurements using regression model. More specifically, for a given location and for a given near future date, predicting the AQI value.

c) Study area

The area of study is concerned with air pollution. This area of study was particularly relevant to us because a few Indian cities got completely enveloped with smoke/smog during Diwali. This made the AQI rise to alarming levels. We hope to see if we can come up with a model that can learn previous trends of AQI and predict the near future of AQI.

Materials and Methods

The data collected and the study procedure is comprehensively described below

a) Data Source

We decided to do our analysis and prediction of AQI using the PM2.5 concentration because it is a dominant and major air pollutant and the historical data of PM2.5 is readily available at various sources.

We obtained the historical data of the PM2.5 from the website of Central Control Room for Air Quality Management - All India (https://app.cpcbccr.com/ccr/). Every city has different locations at which the Air Quality is monitored. (for eg: Delhi: JNU, New Delhi etc.). We chose the locations which had the maximum historical data available.

Cities chosen for the study : Bengaluru, Delhi, Jaipur, Chennai and Lucknow.

b) Procedure

i) Cleaning the Dataset

Raw data of historical PM2.5 concentration of the major cities is gathered in form of spreadsheets. The data was then structured date-wise where each date is assigned an integer value from 1 to n to serialize the days. There were many days which did not have data entries. Those entries as well as redundant entries were removed and a cleaned dataset is formed.

ii) Calculating the AQI and comparing average values of different cities.

Once the dataset is cleaned, we developed a python script to find the AQI value from the PM2.5 concentration. The script can be found on this link **.**

After the AQI calculations, all the values are colour-coded as per Figure 1.

Now we calculated an average AQI of each city and plotted them to compare as shown in Figure in the Results section

iii) Building the Prediction Model for near future

Now that we have a clean and structured dataset, we can feed it to a regression model to predict the near future AQI values. The dataset contains around 900 entries which are divided among two parts for training and testing. Most of the given entries are used for training while the last few and newly added entries are used for testing.

The model works on two variables, which are, the Day number and the AQI value where the AQI value is the goal variable. Now, the Regression model of *Tada Predictive Solutions* is used to generate statistical model which predicts the AQI values of near future days. The process pipeline is as shown in Figure 2.

Figure 2

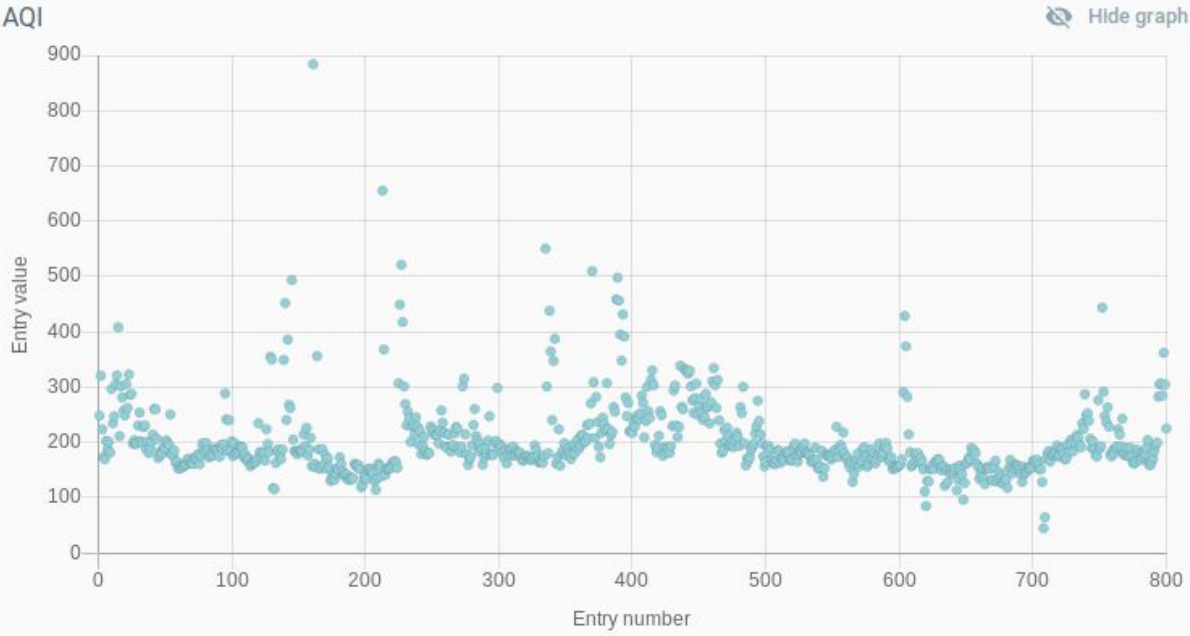
Results and Discussion

1) Data Analysis

The historical AQI data of different cities is plotted and trends are shown. The minimum, maximum, mean and standard deviation of AQI values of each city is also calculated as shown below.

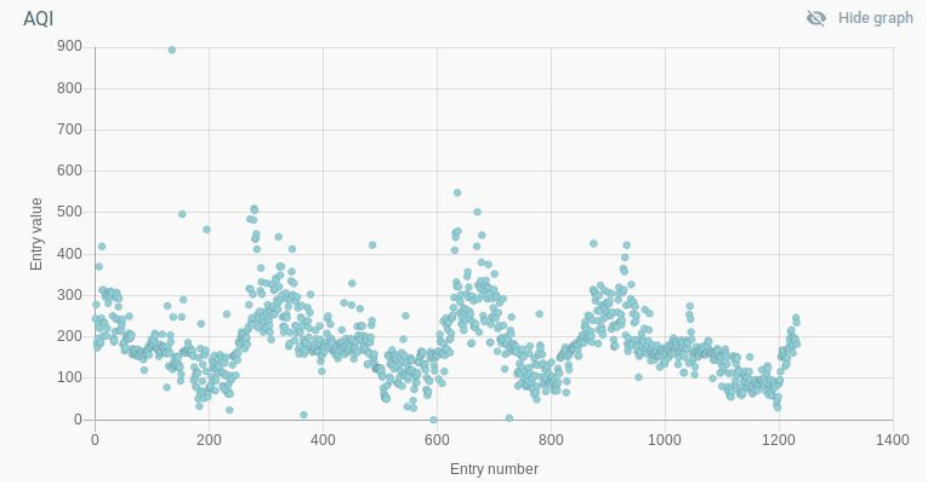
The x-axis represents the Day number starting from January 2016. The y-axis represents the corresponding AQI value on that day.

Delhi AQI Analysis



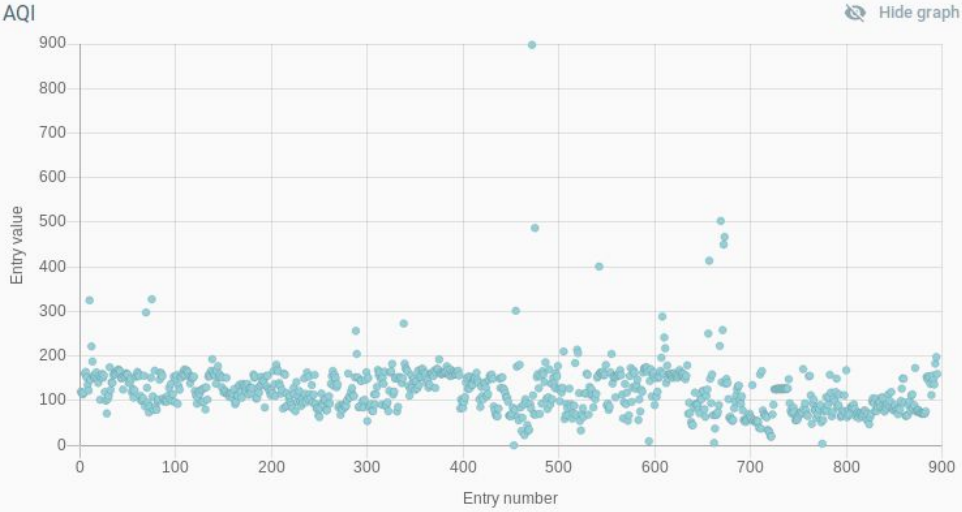
Type	Min	Max	Mean	Standard deviation
DOUBLE	44.875	884.05646	201.9961	67.9612

Lucknow AQI Analysis



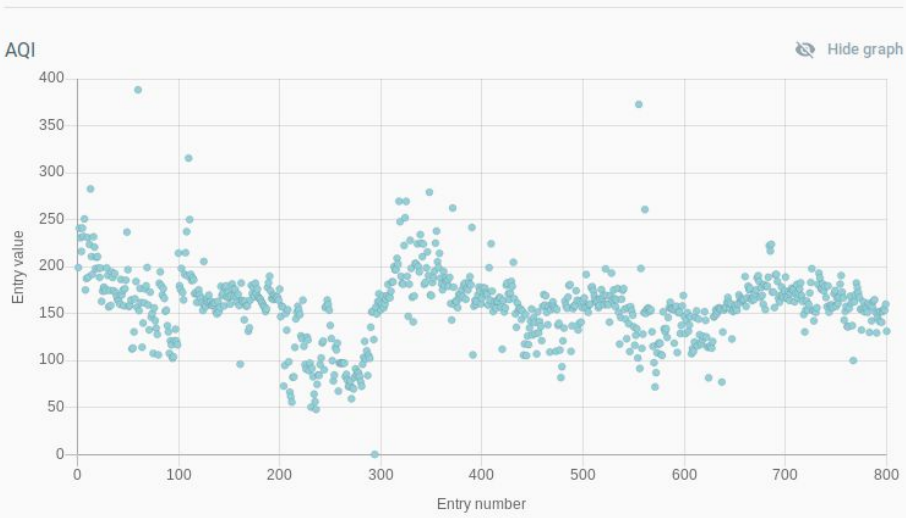
Type	Min	Max	Mean	Standard deviation
DOUBLE	0	893.1504	182.7554	80.2668

Chennai AQI Analysis



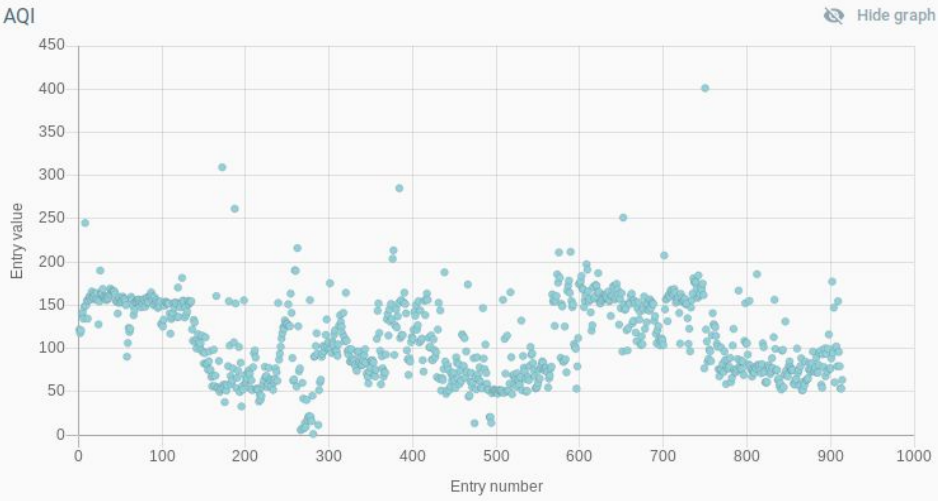
Type	Min	Max	Mean	Standard deviation
DOUBLE	0.08333	897.8328	122.6976	56.6214

Jaipur AQI Analysis



Type	Min	Max	Mean	Standard deviation
DOUBLE	0	388.2048	157.3549	37.2925

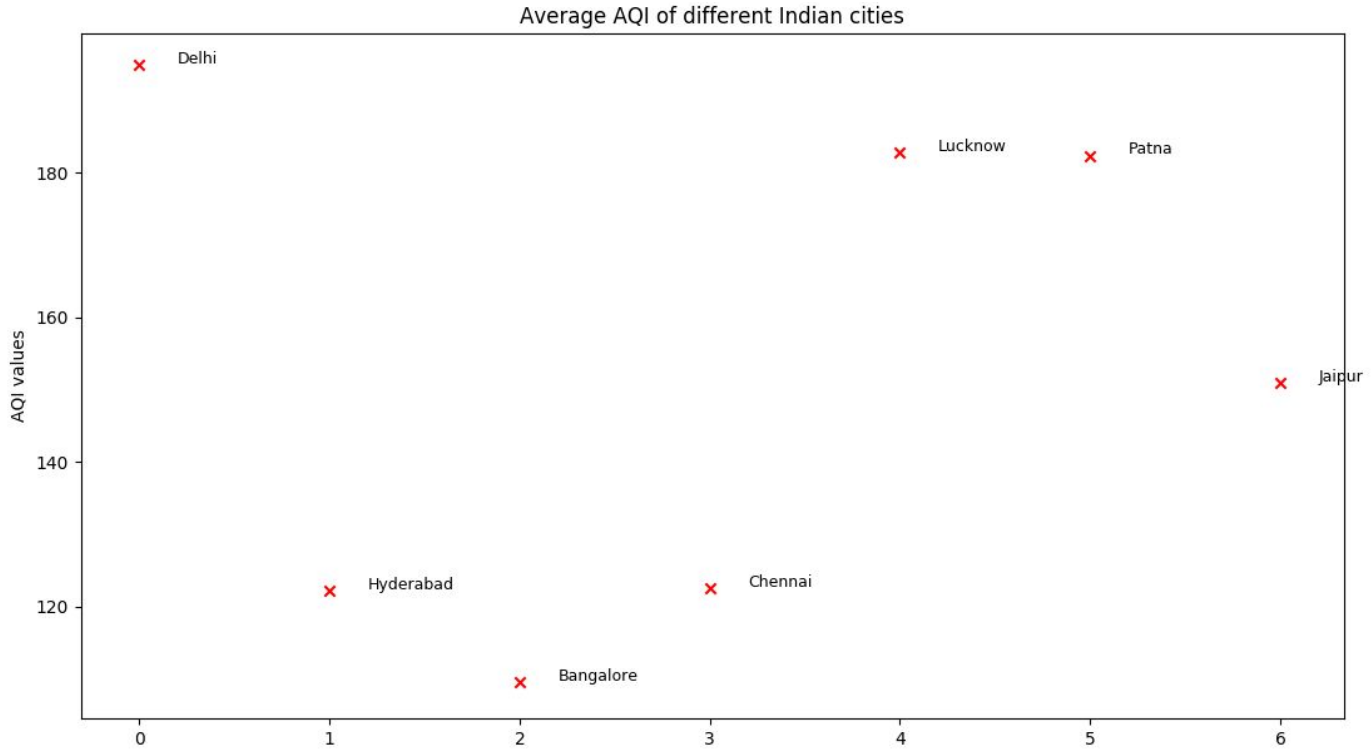
Bangalore AQI Analysis



Type	Min	Max	Mean	Standard deviation
DOUBLE	1.2917	401.06539	109.5605	45.07458

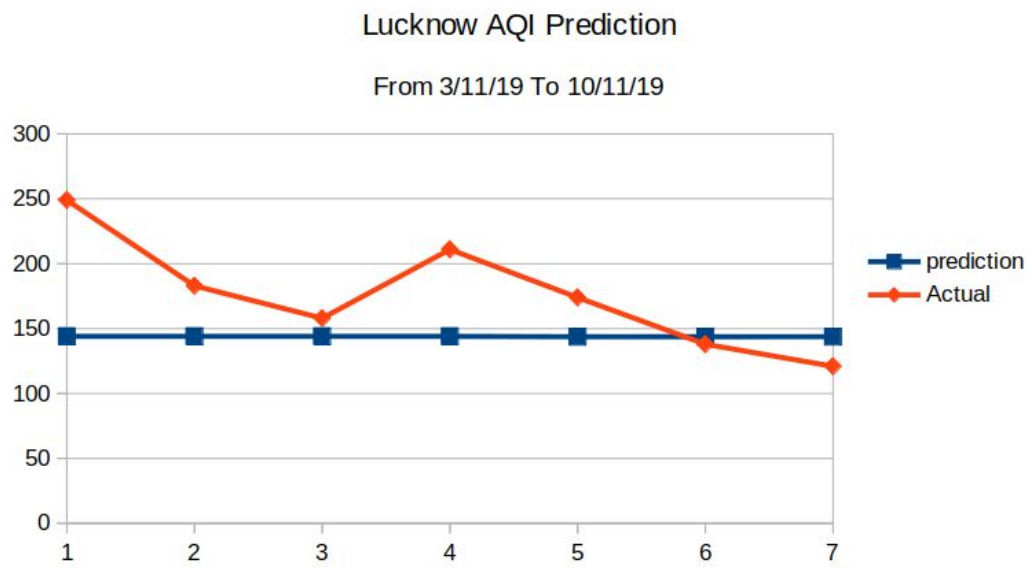
2) Comparison between different cities

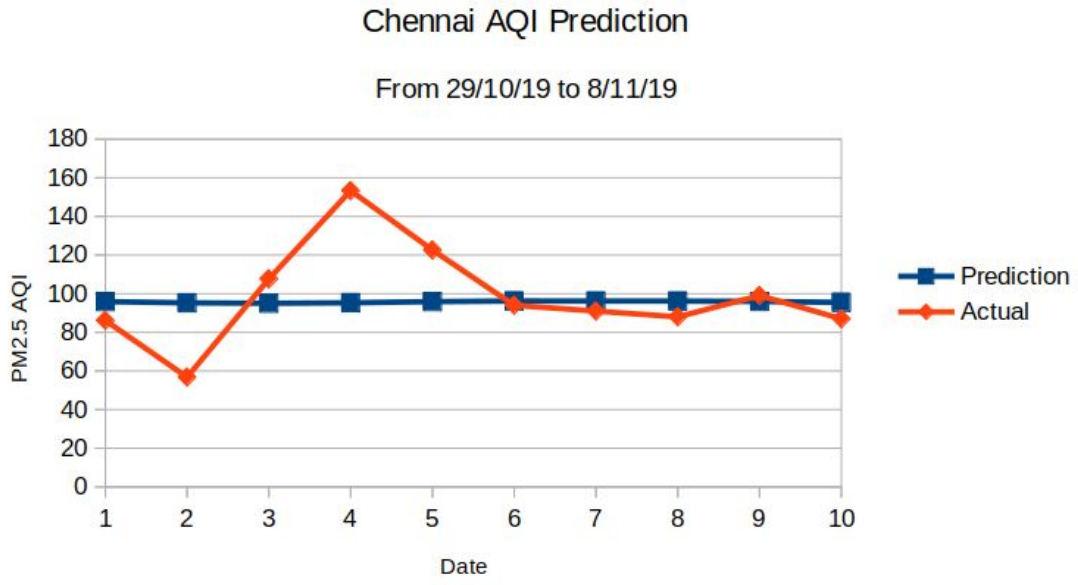
Comparisons are drawn between different major Indian cities, taking the average AQI value over the last 4 years. It's evident from the graph shown below that **Delhi has the highest average AQI (~200)**, which is hazardous for human health. **Bangalore on the other hand has the lowest average AQI(~110)** among the cities compared which is unhealthy but much better than other places.

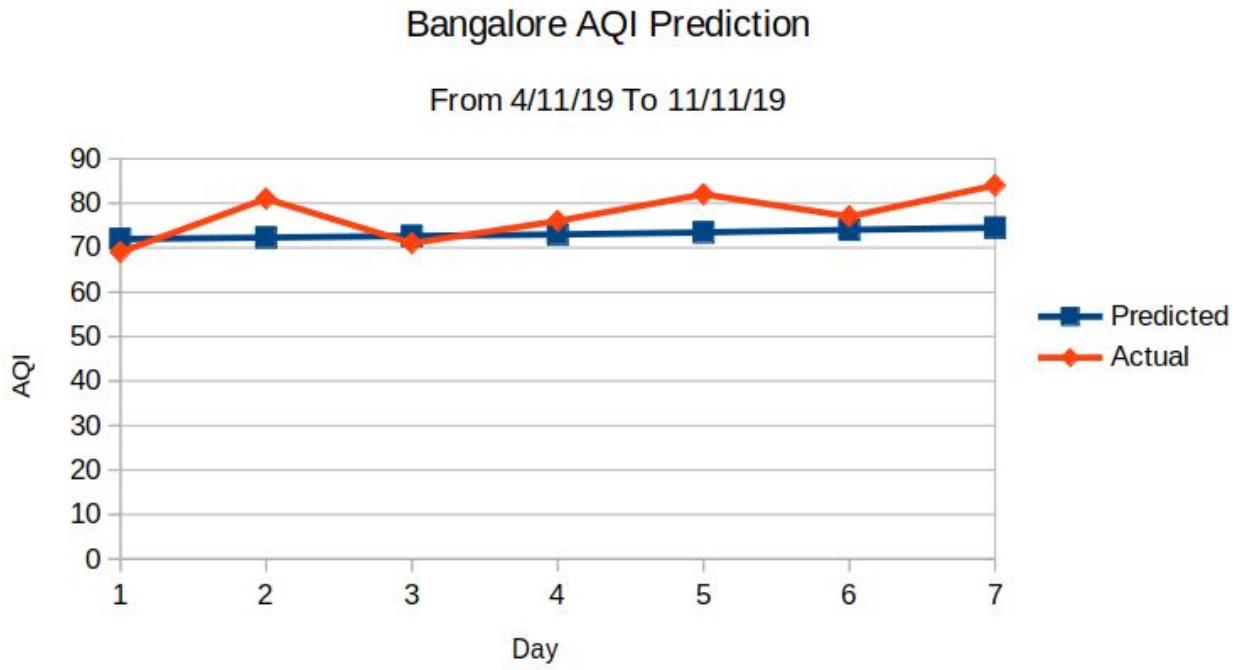


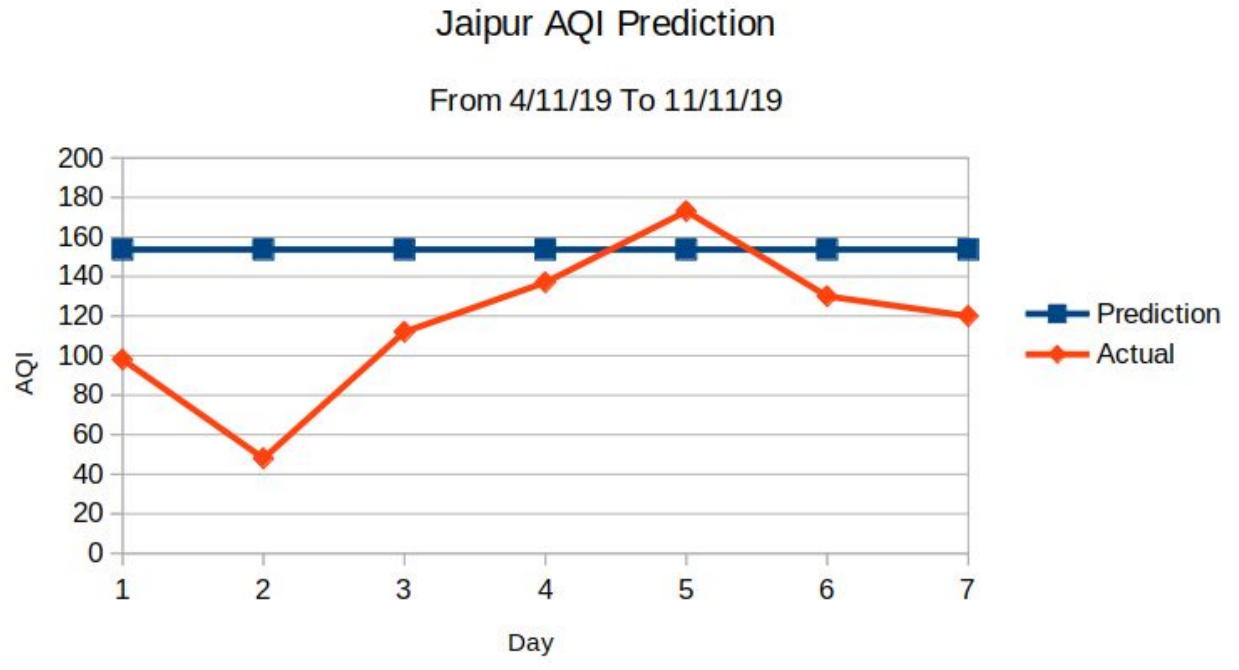
3) AQI Prediction

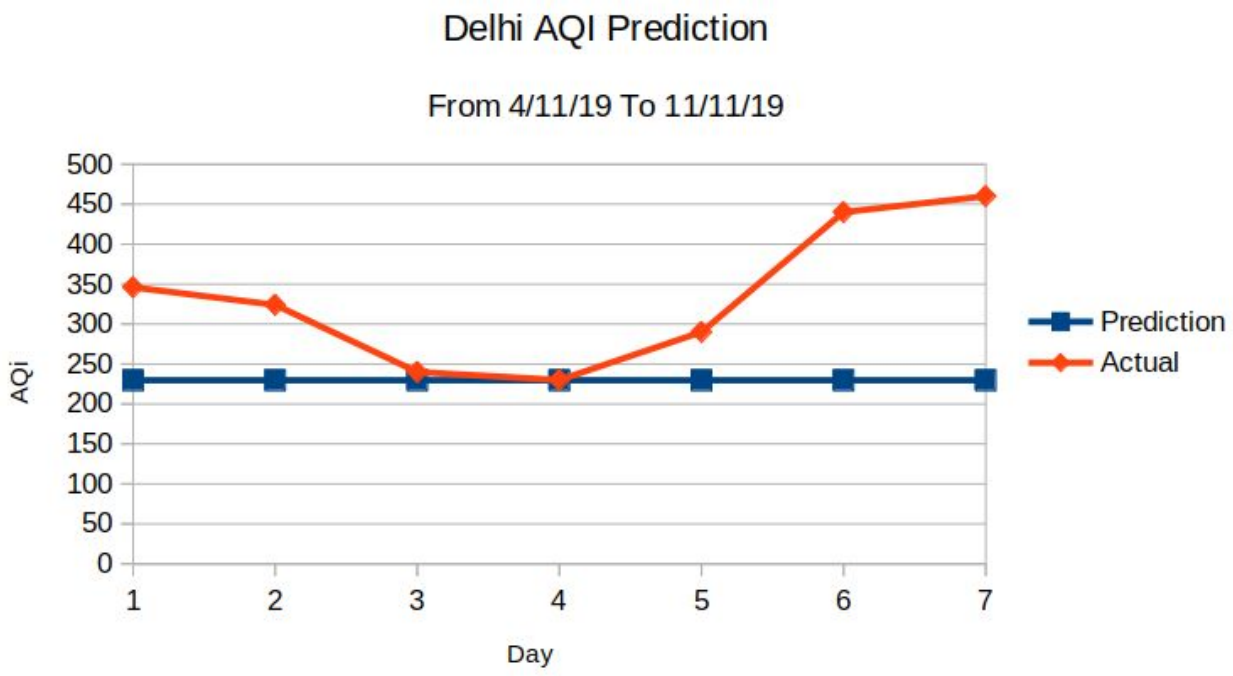
As discussed in the procedure, a regression model was run on the dataset with AQI value as goal variable and Day number as input variable. For each city, the model was run to predict the AQI values for near future dates to the training dates and the following results were obtained






Bangalore AQI Prediction
From 4/11/19 To 11/11/19
AQI
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
Day
Predicted
Actual


Jaipur AQI Prediction
From 4/11/19 To 11/11/19
AQI
200
180
160
140
120
100
80
60
40
20
0
1
2
3
4
5
6
7
Day
Prediction
Actual



As seen in the above graphs, the prediction curve's rate of change is very small because of the fact that the model is trained using just Day number as the input variable. The sharp rise and falls in the actual curve adhere to the fact that AQI depends on various factors like temperature, human intervention, disasters, humidity, festivals etc.

The best prediction results are seen in case of Bangalore as there is the least error/ difference in the predicted and the actual AQI curve.

The final data predictions including training and testing datasets can be seen on this link.

Acknowledgements

We acknowledge the Central Control room for Air Quality management for maintaining their website and for providing the data for this project. We also acknowledge Tada Predictive Solutions for giving us a platform to run our regression model and predict the AQI values. Furthermore we acknowledge Dr. Rama Chandra Prasad P. for giving us this opportunity to work on this project and guiding us throughout the semester.

References

1. ANALYSIS OF AMBIENT AIR QUALITY USING AIR QUALITY INDEX – A CASE STUDY: International Journal of Advanced Engineering Technology, Prakash Mamta, Basin JK
2. Air Quality assessment and its relation to potential health impacts in Delhi, India: Sanjoy Maji, Sirajuddin Ahmed
3. http://www.njaqinow.net/App_AQI/AQI.en-US.pdf
4. https://app.cpcbccr.com/ccr/