IDENTIFYING PATTERNS IN NEW DELHI'S AIR POLLUTION

CAPSTONE PROJECT- FINAL REPORT

Submitted towards partial fulfillment of the criteria for award of PGP-BABI by GLIM

SUBMITTED BY

Karthikeyan Gnanasekaran
Shrinivasabharathi Balasubramanian
Sankaranarayanan Mahadevan
Nagesh Shenoy M

PGP-BABI 8 Bangalore 2015-16

Project Mentor & Research Supervisor: Mr. Jatinder Bedi, New Delhi

Great Lakes Institute of Management

www.greatlakes.edu.in/Bangalore

Acknowledgements

We wish to place on record our deep appreciation for the guidance and help provided to us by our Mentor Mr. Mr. Jatinder Bedi, New Delhi. Mr. Jatinder Bedi helped us narrow down on the choice of the Project as well as the scope and focus area of the Project. He gave us valuable feedback at every stage to enhance the process and the outputs.

We would also like to place on record our appreciation for the guidance provided by Dr. P.K. Viswanathan for giving us valuable feedback and being a source of inspiration in helping us to work on this project.

We certify that the work done by us for conceptualizing and completing this project is original and authentic.

Date: August 28, 2016
Place: Bangalore

Karthikeyan Gnanasekaran
Shrinivasabharathi Balasubramanian
Sankaranarayanan Mahadevan
Nagesh Shenoy M

Certificate of Completion

I hereby certify that the project titled "**IDENTIFYING PATTERNS IN NEW DELHI'S AIR POLLUTION**" was undertaken and completed under my supervision by Karthikeyan Gnanasekaran, Shrinivasabharathi Balasubramanian, Nagesh Shenoy & Sankaranarayanan Mahadevan, students of the Postgraduate Program in Business Analytics & Business Intelligence (PGPBABI-SEPTEMBER-2016).

Date: August 28, 2016 (Jatinder Bedi)

Place: Gurgaon Mentor

TABLE OF CONTENTS

PAGE

ACKNOWLEDGEMENTS 2
CERTIFICATE OF COMPLETION 3
TABLE OF CONTENTS 4
EXECUTIVE SUMMARY……………………………………………………6

CHAPTER 1. INTRODUCTION..……………….8
1.1 Problem Statement……………………………………………………9
1.2 Objective and Scope of the Project ……………………………............... 10
1.3 Data Sources………………………………………………………… 11
1.4 Tools and Techniques……………………………………………...12
1.5 Limitations…………………………………………………………… 14

CHAPTER 2. DATA DESCRIPTION AND PREPARATION............................ 14
2.1 Data Management……………………………………………………… 14
2.2 Data Table…………………………………………………………… 15
2.3 Data Quality ………………………………………………………… 16
2.4 Data Preparations……………………………………………………..16

CHAPTER 3. EXPLORATORY DATA ANALYSIS……………………………17
3.1 Impact of Vehicle Density & Vehicle Population ……………………...17
3.2 Histogram & Box Plots...……………………………………………..19
3.3 Seasonality Analysis ………………………………………………….22
3.4 Correlation Matrix and Analysis ……………………………………..24

CHAPTER 4. PREDICTIVE MODEL DEVELOPMENT……………………….27
4.1 Multiple Linear Regression & Neural Network Models & Results…………..27
4.2 Validation...……………………………………………………………….33
4.3 Conclusion & Recommendations …………………………………………….34

CHAPTER 5. ODD-EVEN CAMPAIGN……………………………………35
4.1 Average Pollution Analysis ………………………………………………...35
4.2 Pollution Level Trend Analysis …………………………………………….36
4.3 PM 2.5 & PM 10 – Before & After Campaign ……………………….…….37
4.4 Odd-Even Impact on Traffic ………………………………………………..38
4.5 Bio-Mass Residual Burning impact on Pollution Levels …………………...38
4.6 Quantifying the Bio-Mass burning Campaign ……………………...............39
4.7 Sentiment Analysis on Odd-Even Campaign ……………………………….40
4.8 Conclusions …………………. ……………………………………………..45
4.9 Recommendations……………… …………………………………………..46

LIST OF TABLES & CHARTS..47
LIST OF EXHIBITS IN APPENDIX ……………………..............................49

APPENDIX……………………………………………………………........50

LIST OF REFERENCES…………………………………………………......63

EXECUTIVE SUMMARY

The rate at which urban air pollution has grown across India is alarming. A vast majority of cities are caught in the toxic web as air quality fails to meet health-based standards. Almost all cities are reeling under severe particulate pollution while newer pollutants like oxides of nitrogen and air toxics have begun to add to the public health challenge. New Delhi is among the most polluted cities in the world today.

In the above context, we felt, if we closely study the Air Quality Data for New Delhi, we should be able to identify patterns (spike in air pollution levels), identify correlating factors on key levels of Air Pollution across key locations of New Delhi. Also as part of the exercise, we wanted to study the impact of Government sponsored Initiative like 'Odd-Even' Pilot Project Phase II.

There were conflicting reports on media on the actual cause of air pollution in New Delhi. Through this study we hope to develop some insights that can help organizations (State/Central Pollution Control Boards and NGOs) to advocate more stringent policy frame work to control air pollution.

The Primary objectives of the study are:

- Identify patterns of spike in Air Pollution levels w.r.t to various monitored parameters
- Identify the Metrological factors that correlate with the air pollution levels
- Develop a Predictive Model (for each location) for predicting the level for key pollutants
- Study the Odd-Even Pilot Project (Phase II) and its impact on air pollution levels in Delhi.

The data for the Project was downloaded from Central Pollution Control Board (CPCB) website. Currently, CPCB track the Air Pollution levels across 26 dimension (variables). Day wise, hour wise (for some variables) data are available on-line across the following dimensions:

Data used for the Project includes nearly 13 months' data starting 1st April'15 to 30th April'15. The locations include Anand Vihar, Punjabi Bagh, R.K. Puram & Shadipur. Shadipur data was used only for analysis of Odd-Even Campaign impact. One location data each for Bangalore and Chennai was considered for Vehicle population & density impact on air pollution. The data covers 15 days prior to the pilot and the 15 days of the pilot.

THE KEY HIGH LEVEL FINDINGS:

Patters in New Delhi air pollution

- Vehicle density (measured as vehicles/km of road) does not have any impact on the air pollution. New Delhi has the least vehicle density but significantly higher levels of PM 2.5 as compared to Bangalore & Chennai. Chennai has the highest density of vehicles, has a lower pollution level (PM 2.5)
- If you consider the absolute vehicle population, then there seem to be a positive correlation between the number of vehicles and the Air Pollution levels of PM 2.5 and to a lesser extent on NO2.

Seasonality Analysis:

- Concentration of Particulate matter known as PM2.5 and PM10 are lower during Monsoon (July-August)
- PM2.5 and PM10 averages are exceeding its permissible values of 60 µg/m3 and 100 µg/m3 during WINTER (November-January) followed by AUTUMN (September-October), SUMMER (April-June) and to a lesser extend during SPRING (February-March)
- Some kind of association between PM 2.5/PM 10 levels and Wind Speed as well as Temp can be seen in the graph

Predictive Model Performance conclusion:

- Multiple Linear Regression Model is able to explain almost 76% of variations in PM 2.5.
- Neural Network overall is able to provide slightly lower RMSE values for PM 2.5 & PM 10 across locations except for Punjabi Bagh (PM 2.5) where MLR gives a slightly lower RMSE value.
- Wind Speed seem to be the most important independent variable followed by Previous day's level for the pollutant and Temperature.
- Model Fit seem to be significant for PM 2.5 for both the models across locations.

PART II: ODD-EVEN CAMPAIGN:

- No apparent impact of 'Odd-Even' on the air pollution levels both during Phase I & Phase II as key pollutants showed increased levels during the Campaign periods as compared to the preceding 15 days.
- The Bio Mass (Crop Residual) burning in the neighbourhood states like Punjab, Haryana & Rajasthan contributed to the increased levels of air pollutants post 19/20th April'16.
- The average levels of Wind Speed went down during the Odd-Even Campaign Phase I & II contributing marginally to the increase in pollution Levels.
- Actual reduction in vehicle was only 13% during the campaign as compared to the normal period.

Key Recommendation: Use the Predictive Model to Predict the following day's Pollutant levels and put in place Trigger based Strick Norms like 'ALARM SYSTEM' FOR Specific Decisive Interventions for those days where the pollution levels are expected to be exceed levels.

1. INTRODUCTION:

The rate at which urban air pollution has grown across India is alarming. A vast majority of cities are caught in the toxic web as air quality fails to meet health-based standards. Almost all cities are reeling under severe particulate pollution while newer pollutants like oxides of nitrogen and air toxics have begun to add to the public health challenge.

WHO says India ranks among the world's worst for its polluted air. Out of the 20 most polluted cities in the world, 13 are in India. Delhi is among the most polluted cities in the world today.

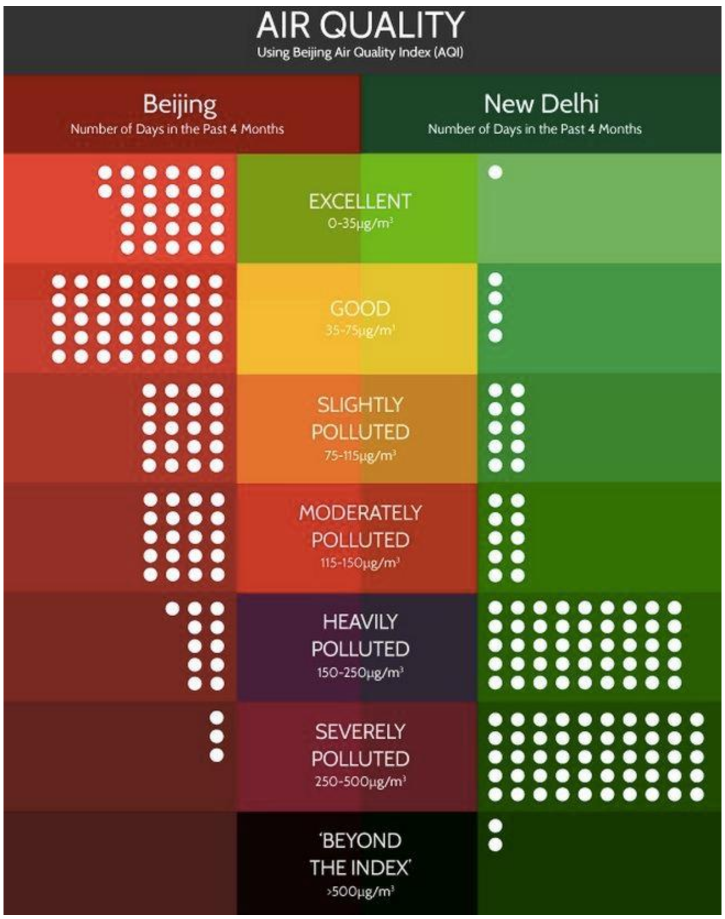
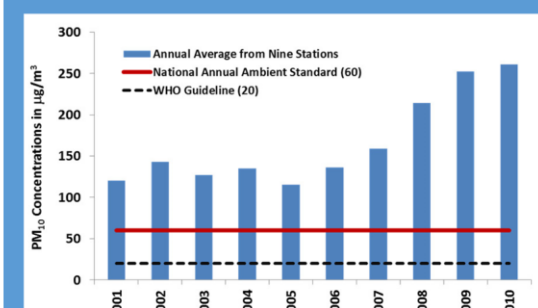


Figure 1: Chart showing the Air Quality Index for Beijing and New Delhi for a 4 Month period

AIR QUALITY INDEX

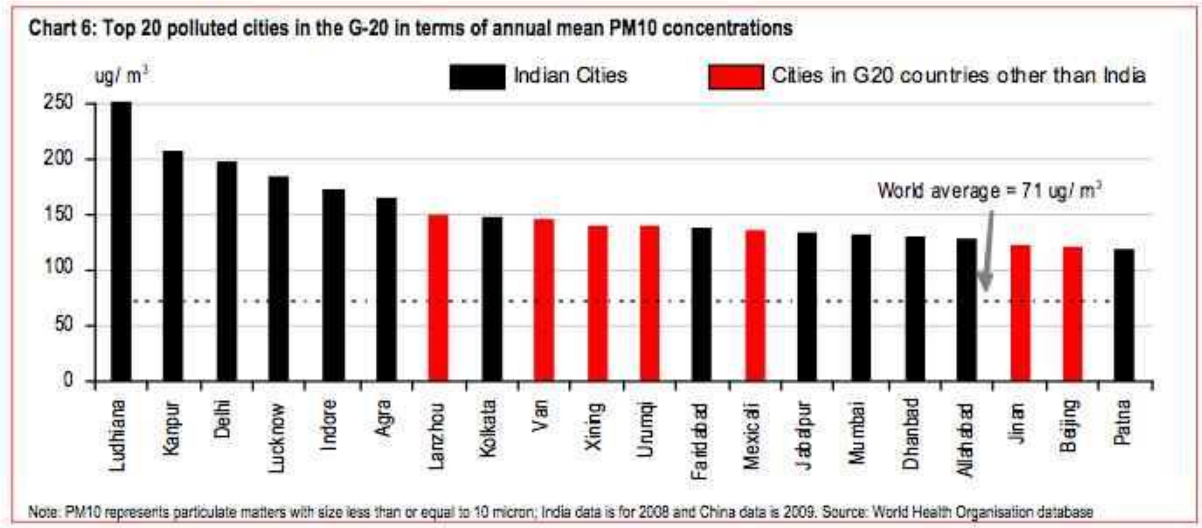
WHO says India ranks among the world's worst for its polluted air. Delhi is among the most polluted cities in the world today.

Figure 2: Chart showing New Delhi's PM10 Levels over a 10-year period against Indian Standard & WHO Standard



Exposure to particulate matter for a long time can lead to respiratory and cardiovascular diseases such as asthma, bronchitis, lung cancer and heart attacks. Last year, the Global Burden of Disease study pinned outdoor air pollution as the fifth largest killer in India after high blood pressure, indoor air pollution, tobacco smoking, and poor nutrition; about 620,000 early deaths occurred in India from air pollution-related diseases in 2010." The Central Pollution Control Board (CPCB) sponsored the study that links the pollutant, pm 10 (particulate matter smaller than 10 microns), to these illnesses. The central regulatory authority recently prescribed stricter norms for a number of air toxins and pollutants but omitted revision of the standard for pm 10.

Figure 3: Chart showing Top 20 polluted cities in the G-20 Countries in terms of annual mean PM10



Sunita Narain, director general, Centre for Science and Environment (CSE) says, "This data confirms our worst fears about how hazardous air pollution is in our region". In addition to this, Narain points out, 18 million years of healthy lives are lost due to illness burden that enhances the economic cost of pollution. Half of these deaths have been caused by ischemic heart disease triggered by exposure to air pollution and the rest due to stroke, chronic obstructive pulmonary disease, lower respiratory track infection and lung cancer.

1.1 PROBLEM STATEMENT:

In the above context, we feel, if we closely study the Air Quality Data, we should be able to identify patterns (spike in air pollution levels), identify correlating factors on key levels of Air Pollution across key locations of New Delhi. Also as part of the exercise, we wanted to study the impact of Government sponsored Initiative like 'Odd-Even' Pilot Project Phase II. The Phase I of the 'Odd-Even' experiment was a huge success in terms of people compliance and reduction of traffic congestion, it had very little impact on the Air Pollution levels during the Campaign period.

It is also important to understand the behaviour of meteorological parameters in the planetary boundary layer because, atmosphere is the medium in which air pollutants are transported away from the source, which is governed by the meteorological parameters such as atmospheric wind speed, wind direction, and temperature.

Air pollutants are being let out into the atmosphere from a variety of sources, and the concentration of pollutants in the ambient air depends not only on the quantities that are emitted but also the ability of the atmosphere, either to absorb or disperse these pollutants.

There were conflicting reports on media on the actual cause of air pollution in New Delhi. Some section said it is Vehicular population was the major cause and others saying the road dust and construction debris/dust and Industrial pollution were the actual root cause. Through this study we hope to develop some insights that can help organizations (State/Central Pollution Control Boards and NGOs) to advocate more stringent policy frame work to control air pollution.

1.2 OBJECTIVE AND SCOPE OF THE PROJECT:

1.2.1. Objective:

The Primary objectives of the study are:

- Study the Air Pollution Data for various locations in New Delhi to identify patterns of spike in Air Pollution levels w.r.t to various monitored parameters
- Identify the Metrological factors that correlate with the air pollution levels for the respective locations
- Explore the possibility of developing a Predictive Model for predicting the level for key pollutants like PM 2.5
- Study the Odd-Even Pilot Project (Phase II) and its impact on air pollution levels in New Delhi. As part of this, also study the people's response to this by studying the social conversation around 'Odd-Even'.

1.2.2 Scope:

- The scope of the study covers 3 major polluting centers in New Delhi
- The study covers one-year Data starting 1st April'15. This is done to ensure seasonality factors are covered
- The Study's focus is on factors for which authentic secondary data are available that can be used for Statistical Analysis

1.2.3 Out of Scope:

- Experimental measures like developing first-hand data are not considered I.e. factors like Vehicle density during the given period at each location, measuring & monitoring level of road dust, Industrial pollution etc.
- The scope of the study will cover 3 to 4 major cities in India and will include 2-3 key monitoring stations per city (depending on the data availability)
- The study will cover up to one year data starting 1st April'15 to 31st March'16. This is done to ensure seasonality factors are covered

1.3. DATA SOURCE:

The data for the Project was obtained from Central Pollution Control Board (CPCB) website. Currently, CPCB track the Air Pollution levels across 23 dimension (variables). Day wise, hour wise (for some variables) data are available on-line across the following dimensions:

1. Nitric Oxide (NO)
2. Carbon Monoxide(CO)
3. Suspended Particulate Matter/RPM/PM10/
4. Nitrogen Dioxide (No2)
5. Ozone
6. Sulphur Dioxide (SO2)
7. PM 2.5 (DUST PM2.5)
8. Toluene
9. Ethyl Benzene (Ethylben)
10. M & P Xylene
11. Oxylene
12. Oxides of Nitrogen (Nox)
13. PM10 DUST
14. PM10 RSPM
15. Ammonia NM3
16. Non Methane Hydro Carbon (NMHC)
17. Total Hydro carbon (THC)
18. Relative Humidity (RH)
19. Temperature
20. Wind Speed (Wind speed S)
21. Vertical Wind speed (Wind speed V)
22. Wind Direction
23. Solar Radiation

Not all monitoring stations track Air Pollution on all the above mentioned parameters and for all days.

India's Central Pollution Control Board now routinely monitors four air pollutants namely Sulphur dioxide (SO2), oxides of nitrogen (NOx), suspended particulate matter (SPM) and respirable particulate matter (PM10) & (PM 2.5). These are target air pollutants for regular monitoring at 308 operating stations in 115 cities/towns in 25 states and 4 Union Territories of India.

The monitoring of meteorological parameters such as wind speed and direction, relative humidity and temperature has also been integrated with the monitoring of air quality. The monitoring of these pollutants is carried out for 24 hours (4-hourly sampling for gaseous pollutants and 8-hourly sampling for particulate matter) with a frequency of twice a week, to yield 104 observations in a year.

- Data includes odd-even pilot project (phase I & II) for 4 locations.
- The data covers 15 days prior to the pilot and the 15 days of the pilot.
- Data on social conversation that took place around the odd-even experiment (phase II). Primarily twitter.

1.4. TOOLS & TECHNIQUES:

We have used the following Analytical techniques/Methodology for analyzing the Data

1. Summary Statistics for each variable
2. Identification of frequency of standard violation for each of the factors
3. Using Graphs and Box Plots to visually represent them
4. Identification of significant Metrological factors through correlation and regression methodology
5. Using Multiple Linear Regression & Neural Network for Model Development
6. Tools used: R, Tableau & Excel
7. Techniques: Box Plot, Histogram, Bar Chart, Line Chart, Infographics, Visual Clues, Correlation Matrix, Multiple Linear Regression, Artificial Neural Network
8. We have used R Programming environment and Microsoft Excel for our analysis and Tableau for data visualization.

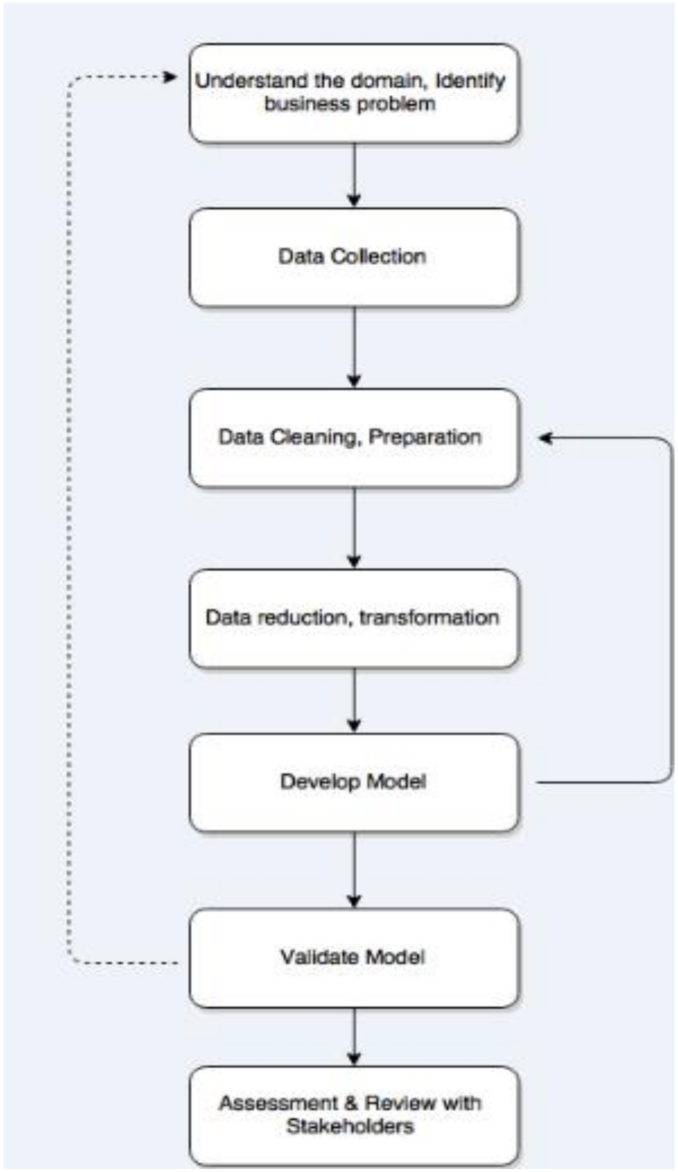
ANALYTICAL APPROACH:

The Analytical Approach will involve the following (not necessarily in the order) activities:

- Data extraction from Primary Data source as well as secondary data sources
- Data quality check
- Data cleaning and data preparation
- Study each of the variables by exploring the data
- Study the variables for its relevance for the study
- Identifying Y variable(s).
- Performing Univariate analysis for all variables
- Division of data into train and test
- Model Development
- Final Model
- Model Validation & Model Validation on Test
- Intervention Strategies and recommendations

We plan to use the following Seven Step Analytical Approach to the Project:

Figure 4: High Level Process Flow



1.5. LIMITATIONS

There are few limitations that this study has w.r.t data and the methodology that can be used.

- Due to time and cost constraints we could not deploy a primary source for data collection. We were not in a position to deploy primary pollution data collection by deploying near ground level monitoring system that are typically used in advanced countries for such Air

Pollution studies. They help accurately capture the road level air pollution contributed maximum by the automobiles.

- Due to a very short window of 15 days for the Odd-Even Campaign, we had to live with a very small data size rendering the data unusable for any kind of rigorous statistical analysis.
- Since the Analysis & Models were built specifically for a particular location, the insights and the Models cannot be used for other locations in New Delhi or for other locations outside New Delhi.
- Since the Models were built on rather small data size (about a year), the models need to be strengthened with at least another year or two data. Till such time the Models are likely to work in a larger range of values. i.e. The variance is likely to be higher.

CHAPTER 2. DATA DESCRIPTION AND PREPARATION

2.1 DATA MANAGEMENT:

Based on the scope, we have extracted data for a year across 23 variables. This was collected for about 4 centres in New Delhi, One Centre in Bangalore and one in Chennai. Data was extracted from CPCB's real Time Air Quality data monitoring application that is available on-line. We have also extracted Data for Odd-Even Pilot project (Phase I & II). This data covers 4/5 major pollutant parameters like SO2, NO2, CO, PM2.5 & PM 10. The data covers 15 days prior to The Pilot and the 15 days of Pilot.

As part of exercise we have also collected data on social conversation that took place around the Odd-Even experiment (Phase II). We were able to collect nearly 1000 social mentions/conversation around this theme.

2.2. DATA TABLE – LIST OF VARIABLES

Table: List of Variables and Their Type				
Variable Abbreviation	Variable	Variable type	Unit of Measurement	Data Type
NO	Nitric Oxide	Pollutant	µg/m3	Continous
CO	Carbon Monoxide	Pollutant	mg/m3	Continous
NO2	Nitrogen Dioxide	Pollutant	µg/m3	Continous
OZONE	Ozone	Pollutant	µg/m^3	Continous
SO2	Sulphur Dioxide	Pollutant	µg/m3	Continous
NOx	Oxides of Nitrogen	Pollutant	µg/m^3	Continous
RSPM	Respiratory Susupended Particulate Matter	Pollutant	µg/m^3	Continous
PM2.5	Particulate Matter less than 2.5 Micrometer	Pollutant	µg/m^3	Continous
PM10	Particulate Matter less than 10 Micrometer	Pollutant	µg/m^3	Continous
Benzene	Benzene	Pollutant	µg/m^3	Continous
Toulene	Toulene	Pollutant	µg/m^3	Continous
Ethylben	Ethyl Benzene	Pollutant	µg/m^3	Continous
M_P_Xylene	M & P Xylene	Pollutant	µg/m^3	Continous
O_Xylene	O Xylene	Pollutant	µg/m^3	Continous
P_Xylene	P Xylene	Pollutant	µg/m^3	Continous
NH3	Ammonia	Pollutant	µg/m^3	Continous
CH4	Methane	Pollutant	µg/m^3	Continous
NMHC	Non Methane Hydro Carbon	Pollutant	µg/m^3	Continous
THC	Total Hydro Carbon	Pollutant	µg/m^3	Continous
RH	Relative Hydrocarbon	Meterological	%	Continous
Temp	Temperature	Meterological	°C	Continous
WS	Wind Speed	Meterological	m/s	Continous
VWS	Vertical Wind Speed	Meterological	m/s	Continous
WD	Wind Direction	Meterological	°	Continous
SR	Solar Radiation	Meterological	W/m2	Continous
Bar Pressure	Bar Pressure	Meterological	mmHg	Continous

Table 1: Table showing List of Variables

2.3. DATA QUALITY:

- Pollutant level Data for certain days were missing. Some days had data for only few of the variables. Data for those days where there were no data for key variables like PM 2.5, PM 10, NO2, SO2, CO were removed. There were no data available for few of the days on the source system itself.

- Specially for Odd-Even Campaign, data was not reported for few days (already on a short window of 15 days pre campaign and 15 days post campaign) on the source system. After plummeting all such variables and observations, the data was merged.
- There were 26 variables with 284 records for Anand Vihar; 289 records for Punjabi Bagh & 345 records for R.K. Puram location.

2.4.DATA PREPARATION

2.4.1. Variables Transformation

- For building the Multiple Linear Regression Model, all the variables were transformed using logarithm function.
- For Neural Network, no data transformation was used.

2.4.2. Missing values and Outliers

- No specific missing value treatment was used.
- Days for which no data was available for the key variables, then that day's record was removed from analysis.
- Only days where observations were recorded for key variables were included for the analysis
- Days when Outliers were present, the day's record was removed from the data.

CHAPTER 3. EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS:

The Exploratory Data Analysis is divided in to three parts. They are:

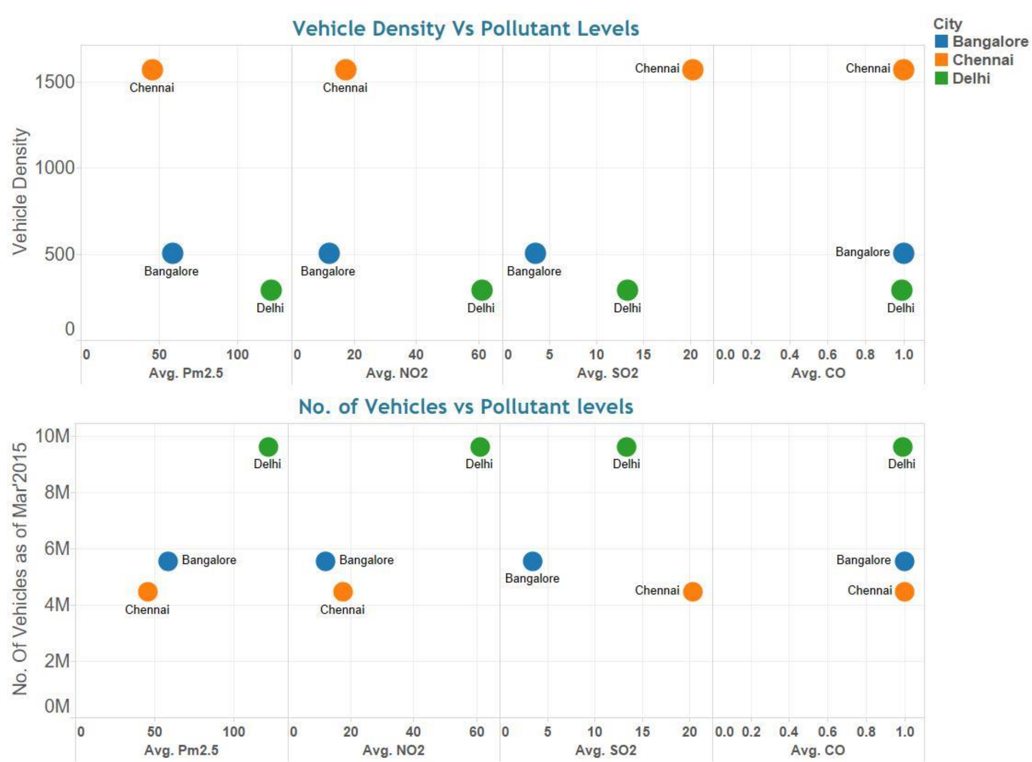
- Analyzing three City Air Pollution Data and check whether the number of vehicle and vehicle density have any impact on the Air pollution levels
- Analyzing the New Delhi's three location data across various factors and find out any correlation exists between the factors
- Analyzing the New Delhi Data to find out the impact of 'Odd-Even' experiment on the pollution levels (i.e. measured across4/5 key parameters). Also explore the social data and do a sentimental analysis for gauging people's reaction to the experiment.

.1. Analyzing the impact of Vehicle Density & Vehicle Population

Analyzing three City Air Pollution Data and check whether the number of vehicle and vehicle density have any impact on the Air pollution levels:

We used simple Graph to plot the Pollutant levels for PM2.5, SO2, NO2 & CO across New Delhi, Bangalore & Chennai. The Average Pollution levels of the Pollutants were mapped on X – axis and the Vehicle Density and the Number of vehicles were plotted on the Y-axis.

Figure 5: Graph showing 3 City Pollution Level Vs Vehicle Density & Vehicle Population



INSIGHTS:

- Vehicle density (measured as vehicles/km of road) does not have any impact on the air pollution. New Delhi has the least vehicle density amongst the three cities we have considered for the study, but the PM 2.5levels are significantly higher in New Delhi as compared to Bangalore and Chennai. Though Chennai has the highest density of vehicles, has a lower pollution levels for (PM 2.5)

- If you consider the absolute vehicle population, then there seem to be a positive correlation between the number of vehicles and the Air Pollution levels of PM 2.5 and to a lesser extent on NO2.
- CO levels does not seem to have any correlation with either vehicle density or with vehicle population as the levels of CO are almost at same levels across the 3 cities.
- The results probably indicates factors other than vehicular pollution are also contributing to the overall air pollution in the three cities in equal measure if not more.
- New Delhi has vast stretch of roads, so the vehicle density tends to get averaged out to a lower number.
- But there is a high probability that the vehicle density in many of the observatory locations are high and contributing to higher air pollution levels

Identifying Patterns in New Delhi Area Air Pollution

Our secondary research identified the three most polluted areas of New Delhi. They are Anand Vihar, R.K. Puram & Punjabi Bagh.

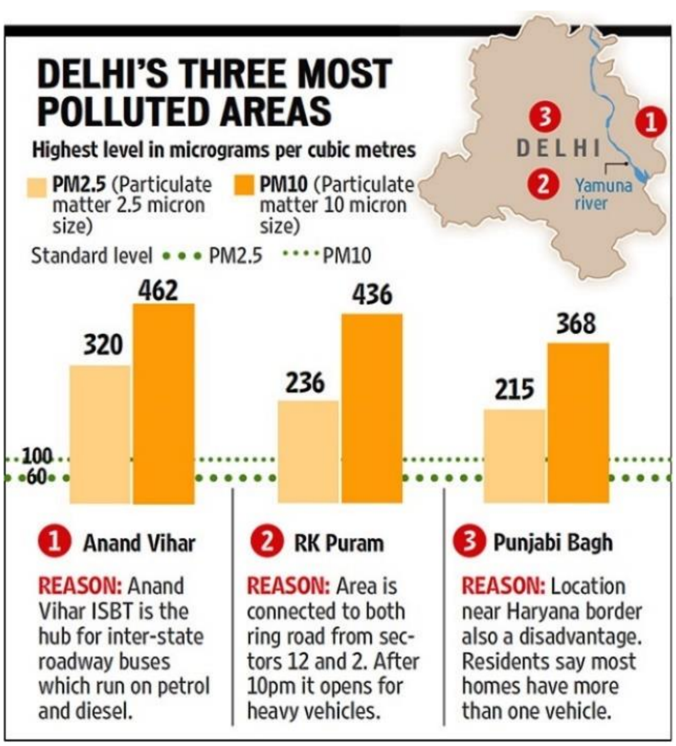
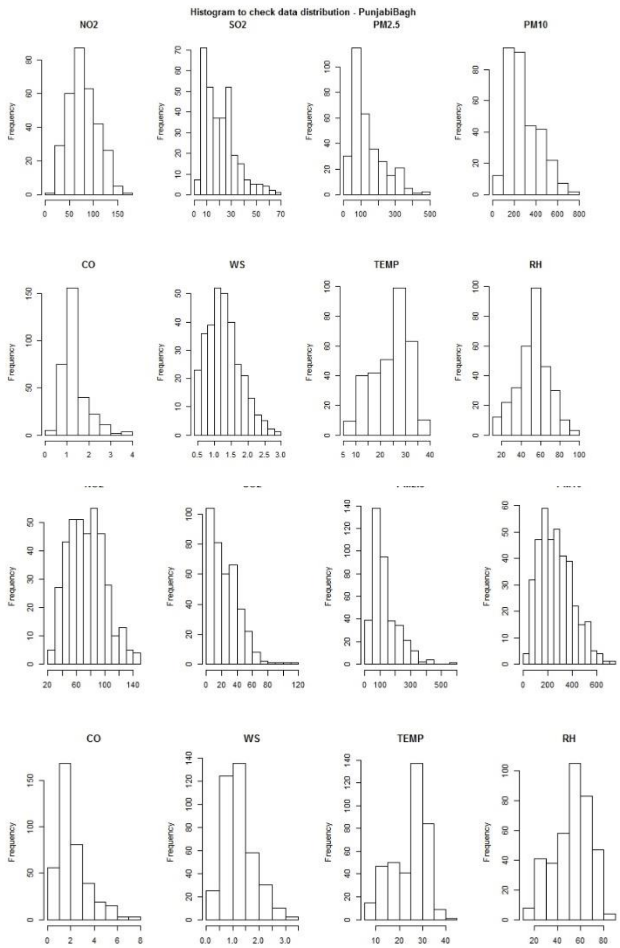


Figure 6: Chart showing the three most polluted areas of New Delhi.

.2. EXPLORATORY DATA ANALYSIS - Histogram for Various Pollutants:



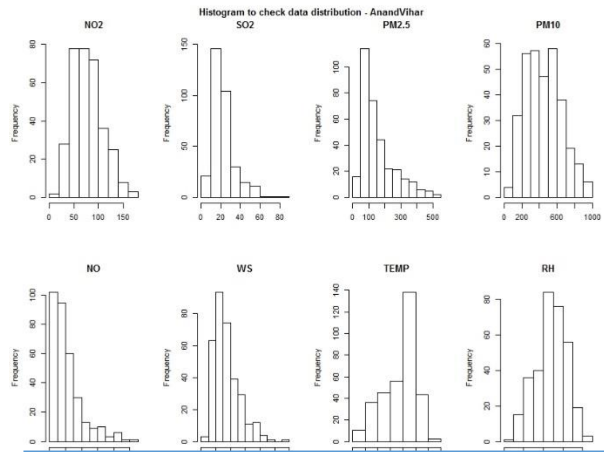


Figure 7:

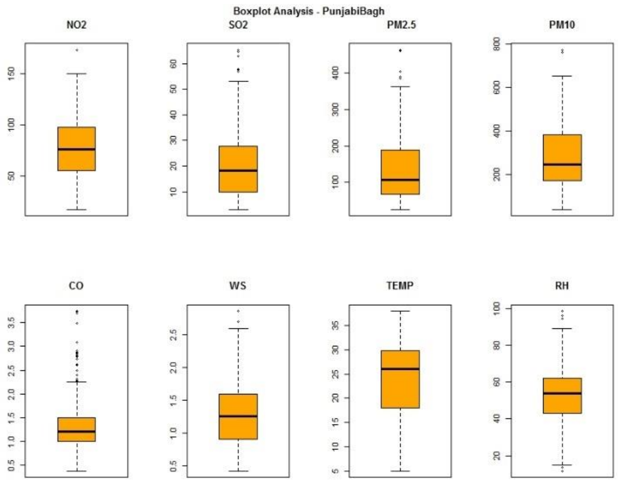
HISTOGRAM CHART SHOWING VARIOUS POLLUTANT LEVELS FOR EACH OF THE THREE LOCATIONS – ANAND VIHAR, PUNJABI BHAG & R.K. PURAM

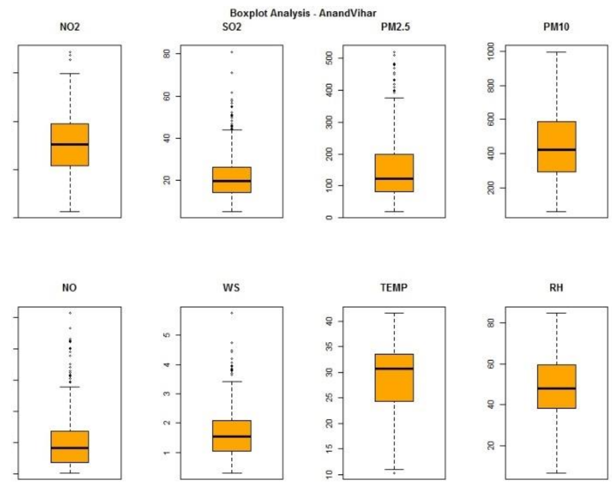
The histogram shows a few key attributes about the distribution of the different pollutants.

- Distribution is asymmetric – Left or right skewed
- Distribution is Unimodal in most pollutant data

There are some Outliers near the low and high ends

Box Plot for Various Pollutants – All Locations:





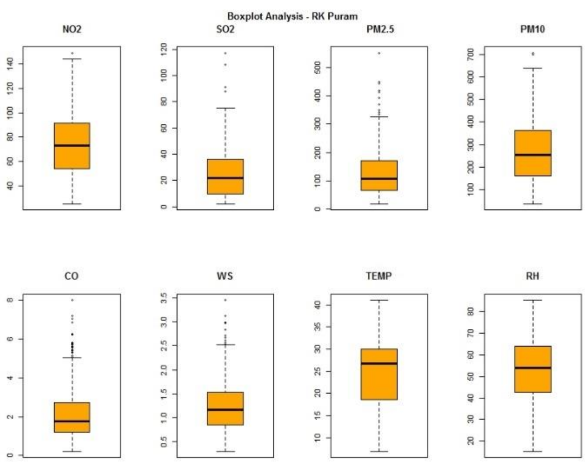


Figure 8: Box Plots

BOX PLOT SHOWING KEY POLLUTANT LEVELS IN THE THREE LOCATIONS

- All the pollutants are almost at the same level in the 3 areas (Centres and spreads are equally likely for all 3 areas).
- Indicating the area between Anand Vihar and Punjabi Bagh including RK Puram are equally polluted.
- The data has outliers caused by external factors and that needs to be investigated.

SUMMARY DATA ON THE KEY VARIABLES FOR EACH LOCATION

Figure 9: AnandVihar

```
       WS              TEMP              WD                RH                SR
 Min.   :0.300   Min.   :10.30   Min.   : 63.74   Min.   : 6.52   Min.   : 12.29
 1st Qu.:1.040   1st Qu.:22.89   1st Qu.:133.81   1st Qu.:39.13   1st Qu.:176.89
 Median :1.520   Median :30.41   Median :194.77   Median :49.20   Median :204.90
 Mean   :1.699   Mean   :28.21   Mean   :189.52   Mean   :48.43   Mean   :201.25
 3rd Qu.:2.060   3rd Qu.:33.37   3rd Qu.:247.50   3rd Qu.:60.23   3rd Qu.:221.54
 Max.   :5.760   Max.   :41.54   Max.   :287.03   Max.   :84.86   Max.   :429.69
  Bar.Pressure        NO2               SO2              PM2.5             PM10
 Min.   :739.0   Min.   :  6.55   Min.   :  5.33   Min.   : 19.51   Min.   : 60.65
 1st Qu.:740.0   1st Qu.: 54.88   1st Qu.: 14.51   1st Qu.: 83.11   1st Qu.:297.94
 Median :740.0   Median : 76.29   Median : 19.56   Median :128.58   Median :429.78
 Mean   :739.9   Mean   : 78.65   Mean   : 22.54   Mean   :161.93   Mean   :450.64
 3rd Qu.:740.0   3rd Qu.: 97.14   3rd Qu.: 25.91   3rd Qu.:213.56   3rd Qu.:586.75
 Max.   :740.0   Max.   :279.51   Max.   :101.10   Max.   :519.68   Max.   :996.62
```

Figure 10: R.K. Puram

```
       WS              TEMP              WD              RH              SR           Bar.Pressure
 Min.   :0.300   Min.   : 6.92   Min.   :126.7   Min.   :15.19   Min.   :  3.31   Min.   :721.7
 1st Qu.:0.850   1st Qu.:18.98   1st Qu.:180.7   1st Qu.:41.46   1st Qu.: 85.32   1st Qu.:731.7
 Median :1.160   Median :26.77   Median :209.8   Median :53.83   Median :122.18   Median :733.4
 Mean   :1.238   Mean   :24.50   Mean   :203.3   Mean   :52.09   Mean   :113.34   Mean   :732.6
 3rd Qu.:1.520   3rd Qu.:29.91   3rd Qu.:229.0   3rd Qu.:63.86   3rd Qu.:141.51   3rd Qu.:733.8
 Max.   :3.450   Max.   :41.05   Max.   :258.1   Max.   :86.82   Max.   :345.68   Max.   :736.1
      VWS               NO2              SO2             PM2.5             PM10              CO
 Min.   :-2.8700   Min.   : 25.27   Min.   :  0.00   Min.   : 18.75   Min.   : 36.63   Min.   : 0.220
 1st Qu.:-0.1700   1st Qu.: 54.14   1st Qu.:  9.55   1st Qu.: 68.72   1st Qu.:163.66   1st Qu.: 1.210
 Median :-0.0600   Median : 73.52   Median : 21.68   Median :108.41   Median :252.83   Median : 1.780
 Mean   : 0.1873   Mean   : 74.29   Mean   : 26.07   Mean   :131.03   Mean   :266.97   Mean   : 2.207
 3rd Qu.: 0.3900   3rd Qu.: 91.30   3rd Qu.: 37.25   3rd Qu.:171.20   3rd Qu.:358.75   3rd Qu.: 2.730
 Max.   : 3.5600   Max.   :149.01   Max.   :371.75   Max.   :550.23   Max.   :705.70   Max.   :19.900
```

Figure 11: Punjabi Bagh

```
       WS              TEMP              WD                RH              SR            Bar.Pressure
 Min.   :0.360   Min.   : 5.12   Min.   : 38.72   Min.   :11.60   Min.   : 15.90   Min.   :553.5
 1st Qu.:0.950   1st Qu.:19.35   1st Qu.: 83.48   1st Qu.:39.03   1st Qu.: 60.13   1st Qu.:553.5
 Median :1.270   Median :27.01   Median :100.32   Median :51.62   Median :101.33   Median :553.5
 Mean   :1.289   Mean   :24.88   Mean   : 99.54   Mean   :50.16   Mean   : 89.83   Mean   :553.8
 3rd Qu.:1.570   3rd Qu.:30.41   3rd Qu.:117.77   3rd Qu.:60.33   3rd Qu.:110.97   3rd Qu.:553.7
 Max.   :2.860   Max.   :39.09   Max.   :170.25   Max.   :98.45   Max.   :165.90   Max.   :561.8
      VWS                NO2              SO2             PM2.5             PM10             CO
 Min.   :-0.13000   Min.   : 17.68   Min.   :  3.02   Min.   : 25.39   Min.   : 39.97   Min.   :0.38
 1st Qu.:-0.04000   1st Qu.: 57.49   1st Qu.: 10.60   1st Qu.: 68.30   1st Qu.:177.09   1st Qu.:0.96
 Median : 0.01000   Median : 76.77   Median : 18.29   Median :107.70   Median :249.73   Median :1.20
 Mean   : 0.01946   Mean   : 79.71   Mean   : 21.00   Mean   :135.62   Mean   :286.54   Mean   :1.30
 3rd Qu.: 0.07000   3rd Qu.: 98.05   3rd Qu.: 27.50   3rd Qu.:179.28   3rd Qu.:370.50   3rd Qu.:1.45
 Max.   : 0.70000   Max.   :196.74   Max.   :150.59   Max.   :462.91   Max.   :772.23   Max.   :3.74
```

.3. Seasonality Analysis:

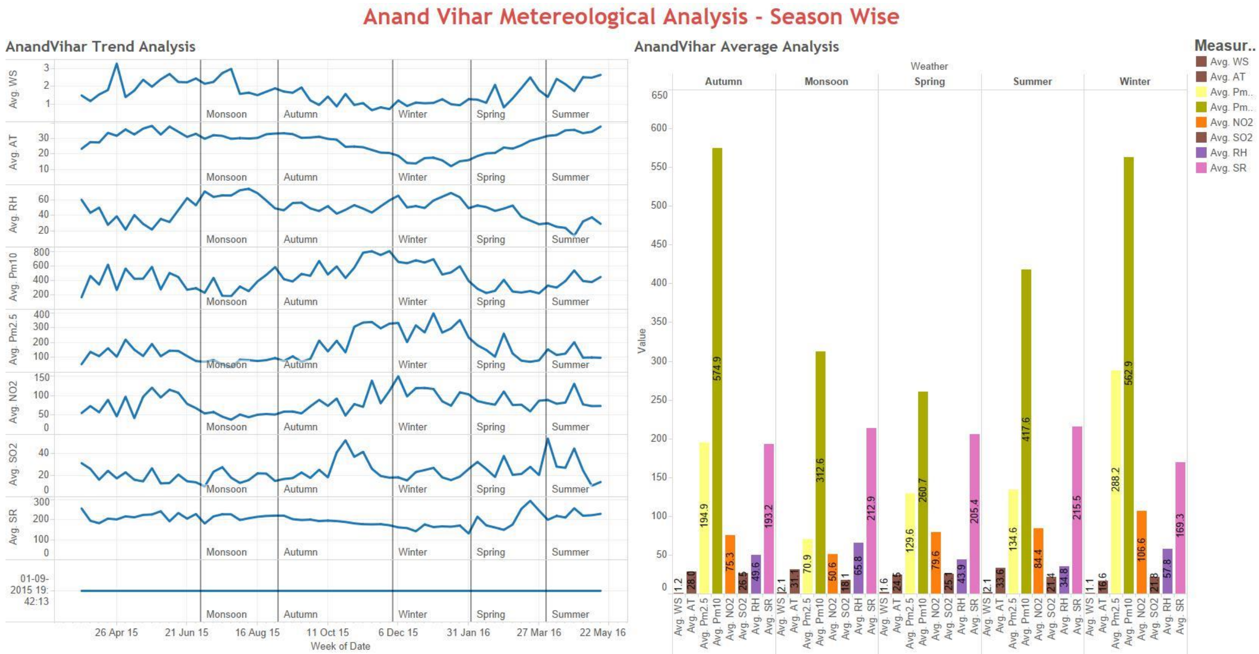


Figure 12: Anand Vihar - Graph & Chart showing pollutant levels across seasons

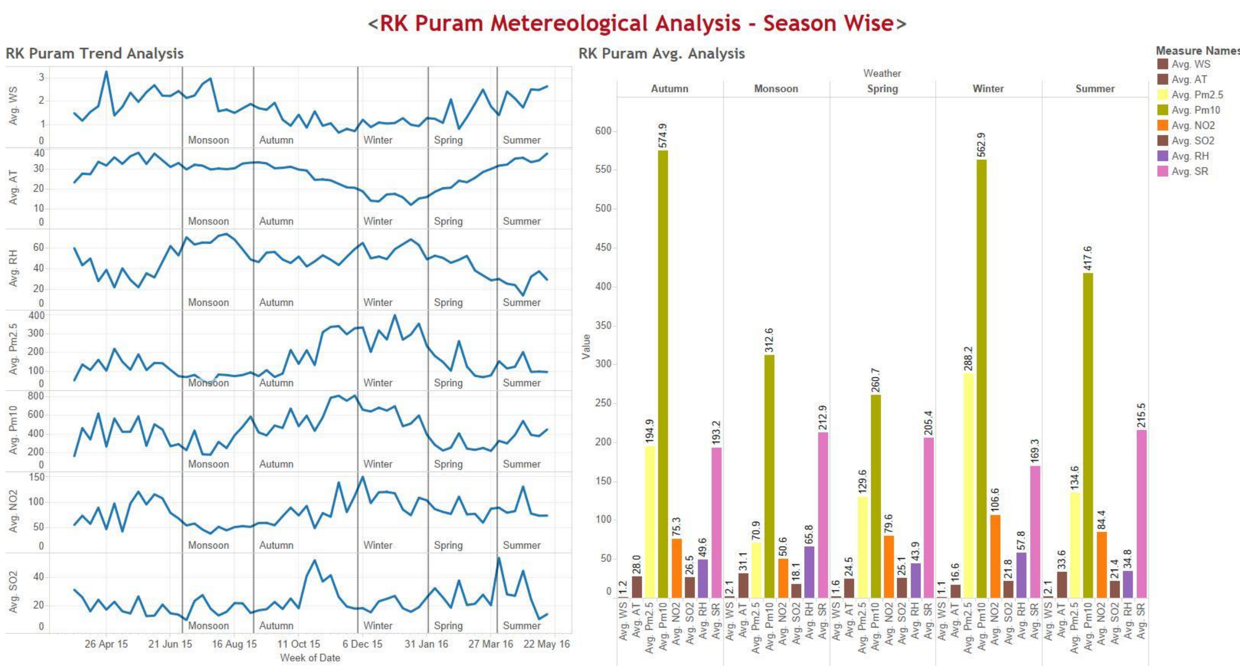


Figure 13: R.K. Puram - Graph & Chart showing pollutant levels across seasons

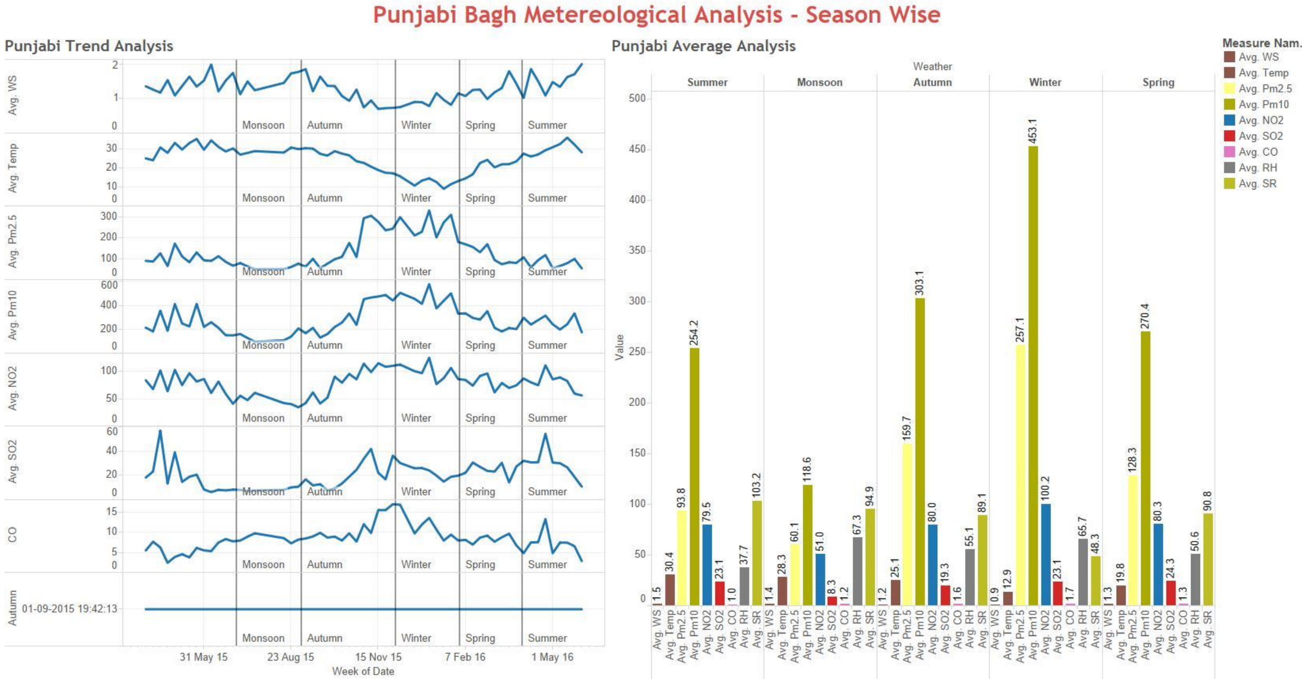


Figure 14: Punjabi Bagh - Graph & Chart showing pollutant levels across seasons

Seasonality Analysis – Conclusion:

- Concentration of Particulate matter known as PM2.5 and PM10 are lower during Monsoon (July-August)
- PM2.5 and PM10 averages are exceeding its permissible values of 60 µg/m3 and 100 µg/m3 during WINTER (November-January) followed by AUTUMN (September-October), SUMMER (April-June) and to a lesser extend during SPRING (February-March)
- Some kind of association between PM 2.5/PM 10 levels and Wind Speed as well as Temp can be seen in the graph
  - Relatively lower Pollution levels seem to be associated with higher Wind Speed
  - Very low Atmospheric Temperature is associated with relatively higher Pollution levels of PM 2.5/PM 10
- Other pollutants data remains significantly same throughout the year except for NO2, peaks during winter and is at its lowest during monsoon

.4. Correlation Matrix & Analysis: Anand Vihar

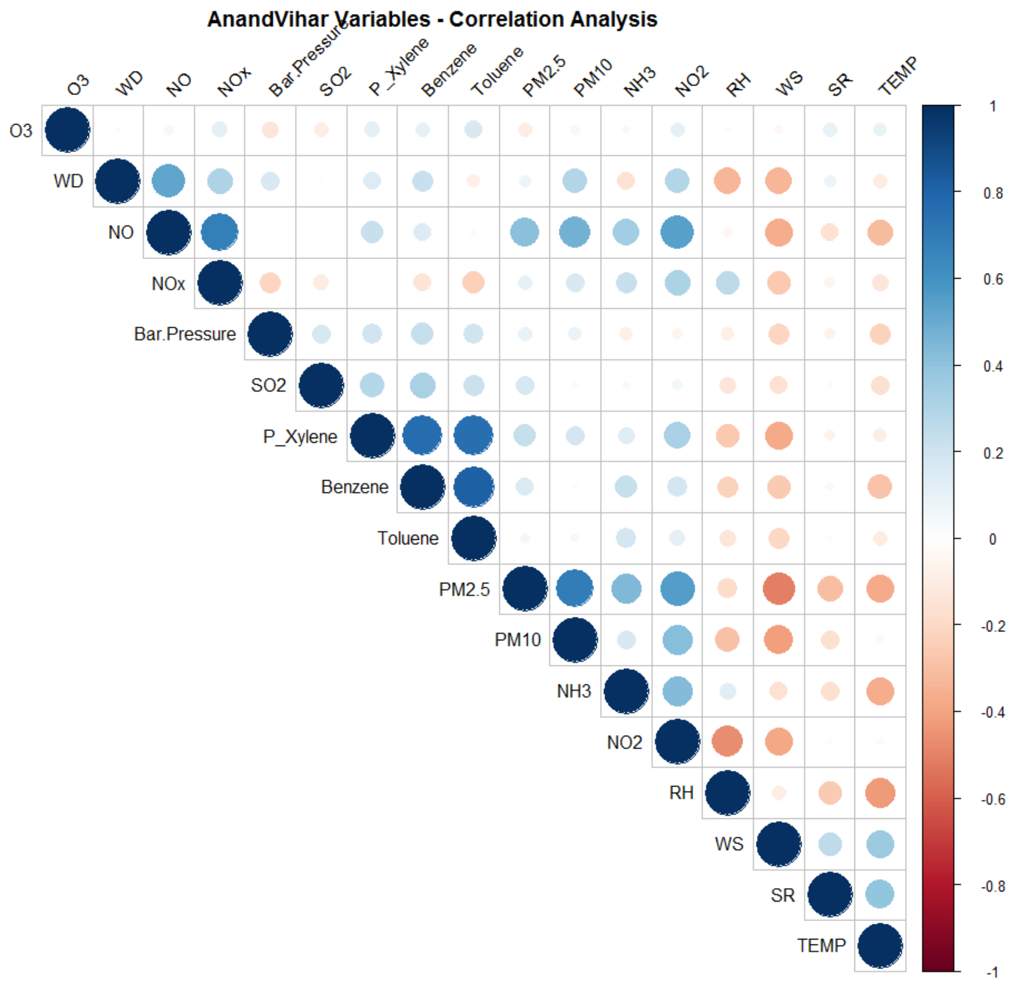


Figure 15: - Correlation Matrix for Anand Vihar

Insights:

- PM 2.5 & 10 have a strong negative correlation with Wind Speed
- Temp has a negative correlation with PM 2.5, NH3 & Relative Humidity
- PM 2.5 also has a positive correlation with NO2
- Xylene, Toluene & Benzene are positively correlated with each other

Correlation Matrix: Punjabi Bagh

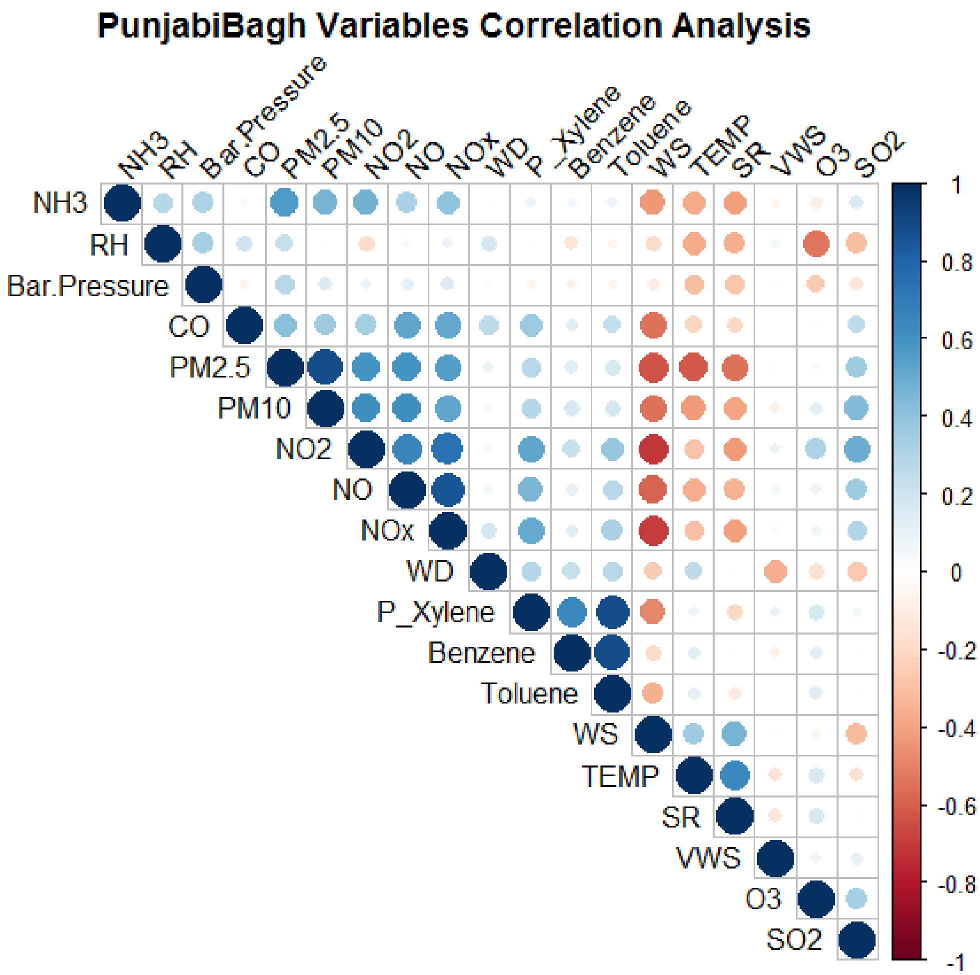


Figure 16: - Correlation Matrix for Punjabi Bagh

Insights:

- Wind Speed have a strong negative correlation with PM 2.5, 10, NO2, NO, CO, NH3 & NOx Wind Speed
- O3 has a strong negative correlation with RH
- Temp & SR also have some negative correlation with PM 2.5, PM 10, NO2, NH3
- Xylene, Toluene & Benzene are positively correlated with each other

Correlation Matrix: R. K. Puram

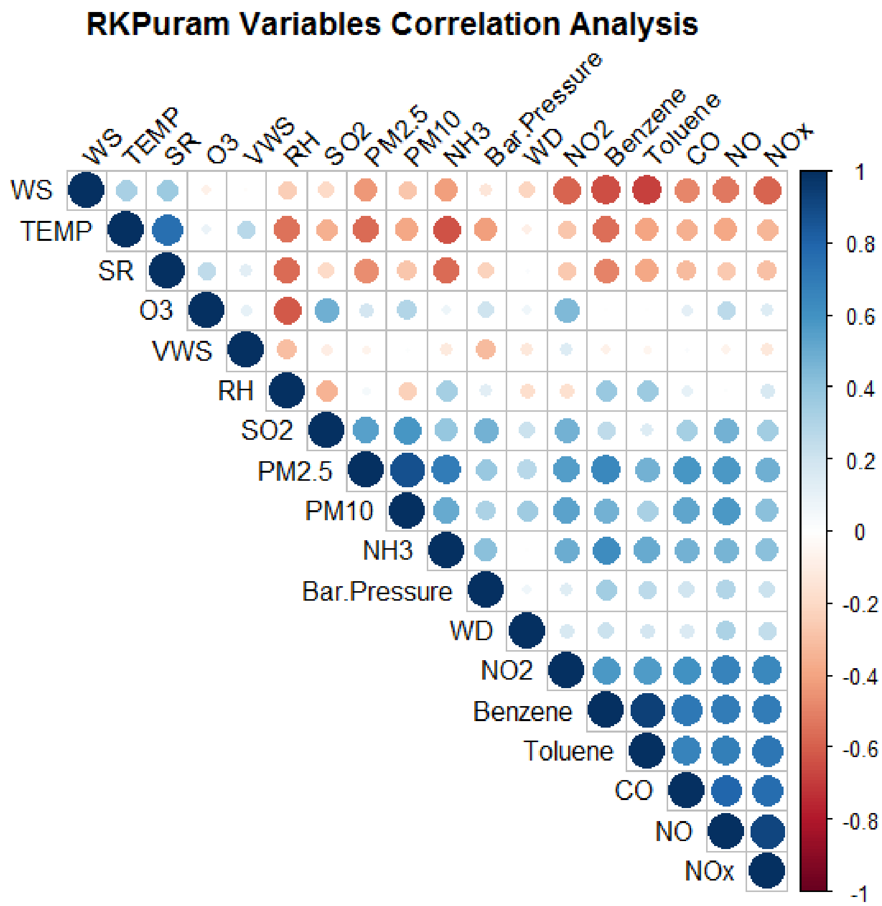


Figure 17: - Correlation Matrix for R.K. Puram

Insights:

- PM 2.5, NO2, Benzene, Toluene, CO, NO have a strong negative correlation with Wind Speed and a negative correlation with Temp & SR
- O3 has a strong negative correlation with RH
- PM 2.5 also has a positive correlation with NO2, NO, CO, Benzene, Toluene
- Xylene, Toluene & Benzene are positively correlated with each other

CHAPTER 4.0: PREDICTIVE MODEL DEVELOPMENT

4.1. MULTIPLE LINEAR REGRESSION MODEL (MLR) & Neural Network Model (NN)

The objective for the Predictive Model Development was to Develop a Model that can predict the next day's level for key pollutants like PM 2.5, PM 10, SO2, CO etc.

The Model Development was done at multiple levels to arrive at a most suitable model. At first level we developed two sets of Model using Multi Linear Regression (MLR). The first one with the actual available variables. The second Model (MLR) was developed using one additional variable i.e. Previous Day's level for that particular Pollutant (Dependent Variable).

Then at the second level we developed the Model using Neural Network (NN). Once again this was further divided in two parts. First with using all the available variables as they are. The second NN Model was developed using one additional variable i.e. Previous Day's level for that particular Pollutant (Dependent Variable).

This Model building approach helped us with 4 sets of Model for each of the predictor variables i.e. Key pollutants.

The data for the modeling was split into two parts. Training & and Test data. The Split of the data as follows:

Table 2: Location wise Modeling Data

Modeling Data Location wise				
Location	**Total Data Size**	**Data after Treatment**	**Training**	**Test**
Anand Vihar	424	284	204	80
R. K. Puram	427	345	271	74
Punjabi Bagh				

The following are the details for the Models

Multiple Linear Regression	
Sampling	Jacknife(LOOCV -Caret Package)
Method	Step-wise regression
Validation	VIF and regression Assumptions
Transformation	Dependent variable Log transformation

Neural Network Model	
Package	NNET & Neuralnet
Sampling	Jacknife(LOOCV -Caret Package)
hidden layer	1
Size and Decay	Optimised by RMSE value

Since the objective is to predict the next day's value we have included the previous day's level as Multiple Linear Regression was run on Training Data set using R package. Multi Linear Regression Model was used on Metrological variables like wind speed (WS), wind direction (WD), relative humidity (RH), solar radiation (SR) and temperature. The key pollutants like PM 2.5, PM 10, SO2, NO2, CO were kept as Dependent. Variables with low information value & high P -vale were dropped. The resulting significant predictors, their p-values and the estimated signs for numeric predictors are shown in Tables 3.1 to 3.4; 4.1 to 4.4 & 5.1 to 5.4.

Table 3.1: Table showing Anand Vihar Air Pollution Predictive Model Results

AnandVihar Air Pollution Level Data Analysis												
Multiple Linear Regression on Metrological and other variables												
										Jacknife		
MLR Exp No.	Dependent Variable	Independent Variables	Intercept Value	R- Squared	Adjusted R-Squared	F-Value	P-Value	RMSE	Relative Error	R-sq	RMSE	RE
1	log(PM 2.5)	WS,RH,WD, TEMP	8.424	0.641	0.634	88.98	<2.2e-16	49.4	27.53	0.62	54.65	31.52
	log(PM 2.5)	WS,RH,WD,log(PD_PM2.5)	2.96	0.772	0.767	168.5	<2.2e-16	38.58	20.54	0.763	45.19	24.15
2	log(PM10)	WS, RH	7.14	0.374	0.368	60.27	<2.2e-16	158.42	32.43	0.345	161.02	35.15
	log(PM10)	WS,RH,log(PD_PM10)	3.4	0.624	0.619	111	<2.2e-16	113.19	21.02	0.618	116.84	24.26
3	log(NO2)	WS,SR, RH	330.8	0.389	0.376	31.69	<2.2e-16	19.9	17.97	0.419	19.5	26.39
	log(NO2)	WS, RH,log(PD_NO2)	2.2	0.574	0.568	90.1	<2.2e-16	14.55	14.89	0.586	15.4	19.46
4	log(SO2)	WS,RH	3.55	0.184	0.176	22.73	<2.2e-16	6.634	32.55	0.09	7.093	31.71
	log(SO2)	WS, RH,log(PD_SO2)	1.679	0.538	0.531	77.86	<2.2e-16	5.34	24.42	0.462	5.54	22.84

Multiple Linear Regression Model - Beta Coefficient Table													
			Beta Coefficients										
MLR Exp No.	Dependent Variable	Independent Variables	WS	TEMP	BP	RH	WD	SR	log (PD_PM 2.5)	log (PD_PM 10)	log (PD_NO 2)	log (PD_SO 2)	log (PD_CO)
1	log(PM 2.5)	WS,RH,WD, TEMP	-0.414	-0.044	0	-0.022	-0.002	0	0	0	0	0	0
	log(PM 2.5)	WS,RH,WD,log(PD_PM2.5)	-0.321	0	0	-0.006	-0.001	0	0.608	0	0	0	0
2	log(PM10)	WS, RH	-0.312	0	0	-0.013	0	0	0	0	0	0	0
	log(PM10)	WS,RH,log(PD_PM10)	-0.219	0	0	-0.008	0	0	0	0.559	0	0	0
3	log(NO2)	WS,SR, RH	-0.215	-0.016	-0.439	-0.014	0	0	0	0	0	0	0
	log(NO2)	WS, RH,log(PD_NO2)	-0.165	0	0	-0.005	0	0	0	0	2.207	0	0
4	log(SO2)	WS,RH	-0.179	0	0	-0.006	0	0	0	0	0	0	0
	log(SO2)	WS, RH,log(PD_SO2)	-0.131	0	0	-0.003	0	0	0	0	0	0.63	0

Table 3.2: Multiple Linear Regression Model Beta Coefficient Table

Neural Network model on Metrological and other variables(threshold = 0.1)												
NN Exp No	Dependent Variable	Independent Variables	Hidden Layer Optimum	Train RMSE	Test RMSE	%Variation	Converted Test RMSE	Relative Error	Jacknife R-sq	Jacknife RMSE	Jacknife RE	
1	PM 2.5	WS,TEMP, BP, RH,SR,VWS,V	1	0.577	0.493	-14.56	43.69	29.06	0.739	45.12	28.83	(5,0.5)
2	PM 10	WS,TEMP, BP, RH,SR,VWS,V	1	0.795	0.798	0.38	151.09	33.63	0.476	136.97	31.03	(6,0.5)
3	NO2	WS,TEMP, BP, RH,SR,VWS,V	1	0.686	0.679	-1.02	18.16	17.37	0.55	18.32	26.35	(6,0.5)
4	SO2	WS,TEMP, BP, RH,SR,VWS,V	1	0.889	0.813	-8.55	5.96	31.62	0.28	6.21	29.4	(5,0.5)

Neural Network model on Metrological and other variables(threshold = 0.1)												
NN Exp No	Dependent Variable	Independent Variables	Hidden Layer Optimum	Train RMSE	Test RMSE	%Variation	Converted Test RMSE	Relative Error	Jacknife R-sq	Jacknife RMSE	Jacknife RE	
1	PM 2.5	WS,TEMP, BP, RH,SR,,WDPD	1	0.48	0.397	-17.29	35.21	22.29%	0.818	37.67	21.66	(7,0.5)
2	PM 10	WS,TEMP, BP, RH,SR,WD,PD	1	0.592	0.605	2.20	114.56	22.56	0.697	103.99	22.47	(5,0.5)
3	NO2	WS,TEMP, BP, RH,SR,WD,PD	1	0.54	0.49	-9.26	13.44	13.65	0.721	14.43	19.51	(5,0.5)
4	SO2	WS,TEMP, BP, RH,SR,WD,PD	1	0.721	0.711	-1.39	5.2	24.53	0.478	5.31	23.95	(7,0.5)

Table 3.3 & 3.4: Neural Network Model Results for w/o Previous Day's and with PD's

Inference:

- Almost 76.7% of the Variations in PM 2.5 seem to be explained by the MLR Model & 73.9% by the Neural Network Model.
- NN gives a shade better RMSE value as compared to MLR. Model Fit seem to be significant for PM 2.5.

Table 4.1: Table showing Punjabi Bagh Air Pollution Predictive Model Results

Punjabi Bagh Air Pollution Level Data Analysis												
Multiple Linear Regression on Metrological and other variables												
MLR Exp No.	Dependent Variable	Independent Variables	Intercept Value	R-Squared	Adjusted R-Squared	F-Value	P-Value	RMSE	RE	Jacknife R-sq	Jacknife RMSE	Jacknife RE
1	log(PM 2.5)	WS,TEMP	6.29	0.573	0.569	159.39	<2.2e-16	44.32	30.9	0.523	46.4	32.42
	log(PM 2.5)	WS,TEMP,log(PD_PM2.5)	2.82	0.769	0.766	257	<2.2e-16	35.39	24.42	0.734	33.27	22.74
2	log(PM10)	WS,TEMP, RH,WD	7.08	0.397	0.39	51.11	<2.2e-16	94.8	28.85	0.35	91.48	30.5
	log(PM10)	WS,log(PD_PM10)	2.35	0.677	0.674	244.35	<2.2e-16	74.96	22.27	0.628	70.17	22.97
3	log(NO2)	WS,SR, RH	5.59	0.605	0.6	118.7	<2.2e-16	11.87	13.94	0.703	13.94	15.13
	log(NO2)	WS,SR, RH,log(PD_NO2)	3.204	0.735	0.731	160.894	<2.2e-16	11.47	12.83	0.737	12.97	14.2
4	log(SO2)	WS,SR, RH,BP	5.54	0.445	0.426	23.32	<2.2e-16	8.1	42.27	0.33	8.63	43.78
	log(SO2)	WS,SR, RH,log(PD_SO2)	2.25	0.776	0.766	74.88	<2.2e-16	6.36	27.15	0.646	6.807	29.11
5	log(CO)	WS,TEMP,WD	0.697	0.351	0.343	41.96	<2.2e-16	0.286	22.33	0.247	28.65	22.46
	log(CO)	WS,RH, log(PD_CO)	-0.07	0.505	0.499	79.15	<2.2e-16	0.279	19.76	0.44	0.27	18.4

MLR Exp No.	Dependent Variable	Independent Variables	Intercept Value	Beta Coefficients WS	TEMP	BP	RH	WD	SR	log (PD_PM2.5)	log (PD_PM10)	log (PD_NO2)	log (PD_SO2)	log (PD_CO)
1	log(PM 2.5)	WS,TEMP	6.29	-0.648	-0.03	0	0	0	0	0	0	0	0	0
	log(PM 2.5)	WS,TEMP,log(PD_PM2.5)	2.82	-0.407	-0.009	0	0	0	0	0.564	0	0	0	
2	log(PM10)	WS,TEMP, RH,WD	7.08	-0.456	-0.023		-0.007	0	0	0	0	0	0	0
	log(PM10)	WS,log(PD_PM10)	2.35	-0.289	0	0	0	0	0	0	0.639	0	0	0
3	log(NO2)	WS,SR, RH	5.59	-0.508	18.1	0	-0.007	0	-0.003	0	0	0	0	0
	log(NO2)	WS,SR, RH,log(PD_NO2)	3.204	-0.34	0	0	-0.004	0	-0.001	0	0	0.444	0	0
4	log(SO2)	WS,SR, RH,BP	5.54	-0.763	0	0	-0.012	-0.01	0	0	0	0	0	0
	log(SO2)	WS,SR, RH,log(PD_SO2)	2.25	-0.326	0	0	-0.007	-0.004		0	0	0	0.654	0
5	log(CO)	WS,TEMP,WD	0.697	-0.328	-0.013	0	0	-0.002	0	0	0	0	0	0
	log(CO)	WS,RH, log(PD_CO)	-0.07	-0.265	0		-0.003	0	0	0	0	0	0	0.323

Table 4.2: Multiple Linear Regression Model with Beta coefficients

Table 4.3: Neural Network Model without Previous Day's value

Neural Network model on Metrological and other variables(threshold = 0.1) w/o Previous Day's Level											
NN Exp No	Dependent Variable	Independent Variables	Hidden Layer Optimum	Train RMSE	Test RMSE	%Variation	Converted Test RMSE	Relative error	Jacknife R-sq	Jacknife RMSE	Jacknife RE
1	PM 2.5	WS,TEMP, BP, RH,SR,VWS,WD	1	0.56	0.55	-1.79	40.3	31.81	0.698	39.72	31.43
2	PM 10	WS,TEMP, BP, RH,SR,VWS,WD	1	0.686	0.827	20.55	95.66	32.68	0.618	71.29	27.21
3	NO2	WS,TEMP, BP, RH,SR,VWS,WD	1	0.558	0.557	-0.18	13.72	18.63	0.69	13.22	15.66
4	SO2	WS,TEMP, BP, RH,SR,VWS,WD	1	0.75	0.767	2.27	8.06	49.47	0.44	7.9	42.41
5	CO	WS,TEMP, BP, RH,SR,VWS,WD	1	0.745	0.96	28.86	0.321	24.32	0.483	0.24	18.44

NN model on Metrological and other variables(threshold = 0.1) with Previous day Pollutant Level											
NN Exp No	Dependent Variable	Independent Variables	Hidden Layer Optimum	Train RMSE	Test RMSE	%Variation	Converted Test RMSE	Relative error	Jacknife R-sq	Jacknife RMSE	Jacknife RE
1	PM 2.5	WS,TEMP, BP, RH,SR,VWS,WD,PD_PM2.5	1	0.41	0.46	12.20	33.27	25.33	0.818	30.78	23.46
2	PM 10	WS,TEMP, BP, RH,SR,VWS,WD,PD_PM10	1	0.54	0.634	17.41	73.39	23.13	0.7	62.87	22.07
3	NO2	WS,TEMP, BP, RH,SR,VWS,WD,PD_NO2	1	0.56	0.479	-14.46	11.78	14.57	0.757	12.08	13.81
4	SO2	WS,TEMP, BP, RH,SR,VWS,WD,PD_SO2	1	0.6	0.633	5.50	6.65	34.7	0.588	6.77	31.58
5	CO	WS,TEMP, BP, RH,SR,VWS,WD,PD_CO	1	0.698	0.696	-0.29	0.233	17.81	0.477	0.244	17.8

Table 4.4: Neural Network Model with Previous Day's value

Inference:

- 76.6% of the Variations in PM 2.5 seem to be explained by the MLR Model as compared to it NN is able to explain 81.8%.
- NN also gives a better RMSE value as compared to MLR but with slightly higher Relative error %. Model Fit seem to be significant for PM 2.5.

Table 5.1: Table showing R.K. Puram Air Pollution Multiple Linear Regression Model Results

RKPuram Air Pollution Level Data Analysis												
Multiple Linear Regression on Metrological and other variables												
MLR Exp No.	Dependent Variable	Independent Variables	Intercept Value	R- Squared	Adjusted R-Squared	F-Value	P-Value	RMSE	RE	Jacknife R-sq	Jacknife RMSE	Jacknife RE
1	log(PM 2.5)	WS,TEMP, RH	7.34	0.525	0.52	96.02	<2.2e-16	47.88	29.32	0.547	54.7	34
	log(PM 2.5)	WS,RH,log(PD_PM2.5	1.711	0.762	0.76	278.69	<2.2e-16	32.87	18.47	0.778	39.7	21.98
2	log(PM10)	WS,TEMP, RH,WD	7.31	0.537	0.53	75.14	<2.2e-16	99.13	30.63	0.529	99.11	32.36
	log(PM10)	WS,RH,WD,log(PD_PM10)	1.094	0.777	0.773	225.65	<2.2e-16	60.6	17.57	0.782	69.35	20.98
3	log(NO2)	WS,SR, RH	5.86	0.605	0.601	133.22	<2.2e-16	17.4	18.6	0.54	18.49	20.42
	log(NO2)	WS,SR, RH,log(PD_NO2)	3.53	0.723	0.719	169.78	<2.2e-16	11.8	12.29	0.693	14.7	16.25
4	log(SO2)	WS,SR, RH,BP	-147	0.653	0.647	121.57	<2.2e-16	14.73	39.35	0.62	14.7	41.58
	log(SO2)	WS,SR, RH,BP,log(PD_SO2)	-54.59	0.827	0.824	245.45	<2.2e-16	12.29	31.84	0.857	9.9	24.23
5	log(CO)	WS,TEMP	1.4	0.297	0.291	55.19	<2.2e-16	0.88	33	0.329	1.04	42.55
	log(CO)	WS,TEMP,log(PD_CO)	0.091	0.568	0.563	114.08	<2.2e-16	0.676	25.3	0.587	0.892	30.69

Table 5.2: Multiple Linear Regression Model Results – Beta Coefficients

Multiple Linear Regression on Metrological and other variables - Table showing the Beta Coefficients															
				Beta Coefficients											
MLR Exp No.	Dependent Variable	Independent Variables	Intercept Value	WS	TEMP	BP	RH	WD	SR	log (PD_PM 2.5)	log (PD_PM10)	log (PD_NO2)	log (PD_SO2)	log (PD_CO)
1	log(PM 2.5)	WS,TEMP, RH	7.34	-0.35	-0.052	0	-0.019	0	0	0	0	0	0	0
	log(PM 2.5)	WS,RH,log(PD_PM2.5	1.711	-0.234	0	0	-0.002	0	0	0.711	0	0	0	0
2	log(PM10)	WS,TEMP, RH,WD	7.31	-0.148	-0.048	0	-0.023	0.003	0	0	0	0	0	0
	log(PM10)	WS,RH,WD,log(PD_PM10)	1.094	-0.092	0	0	-0.003	0.003	0	0	0.719	0	0	0
3	log(NO2)	WS,SR, RH	5.86	-0.413	0	0	-0.011	0	-0.004	0	0	0	0	0
	log(NO2)	WS,SR, RH,log(PD_NO2)	3.53	-0.277	0	0	-0.008		-0.003	0	0	0.424	0	0
4	log(SO2)	WS,SR, RH,BP	-147	-0.225	0	0.209	-0.035	0	-0.01	0	0	0	0	0
	log(SO2)	WS,SR, RH,BP,log(PD_SO2)	-54.59	-0.131	0	0.077	-0.014		-0.003	0	0	0	0.63	0
5	log(CO)	WS,TEMP	1.4	-0.404	-0.014	0	0	0	0	0	0	0	0	0
	log(CO)	WS,TEMP,log(PD_CO)	0.091	-0.248	0	0	0	0.002	0	0	0	0	0	0.551

Table 5.3 & 5.4: Neural Network Model Results – w/o PD's value and with PD value

Neural Network model on Metrological and other variables(threshold = 0.1)											
NN Exp No	Dependent Variable	Independent Variables	Hidden Layer Optimum	Train RMSE	Test RMSE	%Variation	Converted Test RMSE	Relative error	Jacknife		
									R-sq	RMSE	RE
1	PM 2.5	WS,TEMP, BP, RH,SR,V	1	0.59	0.49	-16.95	38.3	34.45	0.67	38.88	29.45
2	PM 10	WS,TEMP, BP, RH,SR,V	1	0.66	0.72	9.09	93.27	39	0.64	69.6	25.96
3	NO2	WS,TEMP, BP, RH,SR,V	1	0.66	0.61	-7.58	15.9	16.9		15.58	18.18
4	SO2	WS,TEMP, BP, RH,SR,V	1	0.65	0.62	-4.62	11.3	30.9	0.61	13.59	15.75
5	CO	WS,TEMP, BP, RH,SR,V	1	0.66	0.79	19.70	1	39	0.364	0.799	41.22

Neural Network model on Metrological and other variables(threshold = 0.1) with Previous day Pollutant Level											
NN Exp No	Dependent Variable	Independent Variables	Hidden Layer Optimum	Train RMSE	Test RMSE	%Variation	Converted Test RMSE	Relative error	Jacknife		
									R-sq	RMSE	RE
1	PM 2.5	WS,TEMP, BP, RH,SR,V	1	0.46	0.38	-17.39	29.95	23.75	0.82	28.52	20.67
2	PM 10	WS,TEMP, BP, RH,SR,V	1	0.5	0.505	1.00	64.9	21.04	0.789	53.63	18.85
3	NO2	WS,TEMP, BP, RH,SR,V	1	0.57	0.45	-21.05	11.8	13.2	0.689	13.7	15.89
4	SO2	WS,TEMP, BP, RH,SR,V	1	0.5	0.64	28.00	11.6	33.5	0.768	7.48	9.49
5	CO	WS,TEMP, BP, RH,SR,V	1	0.61	0.62	1.64	0.79	32.4	0.545	0.672	29.87

Inference:

- 76% of the Variations in PM 2.5 seem to be explained by the MLR Model where as NN is able to explain 82%.
- NN gives a better RMSE value as compared to MLR with lower Relative Error %.
- Model Fit seem to be significant for PM 2.5

MODEL FIT GRAPHS for ANAND VIHAR, PUNJABI BAGH & R.K. PURAM

Figure 18: Anand Vihar – Comparative Model Fit graph for PM 2.5

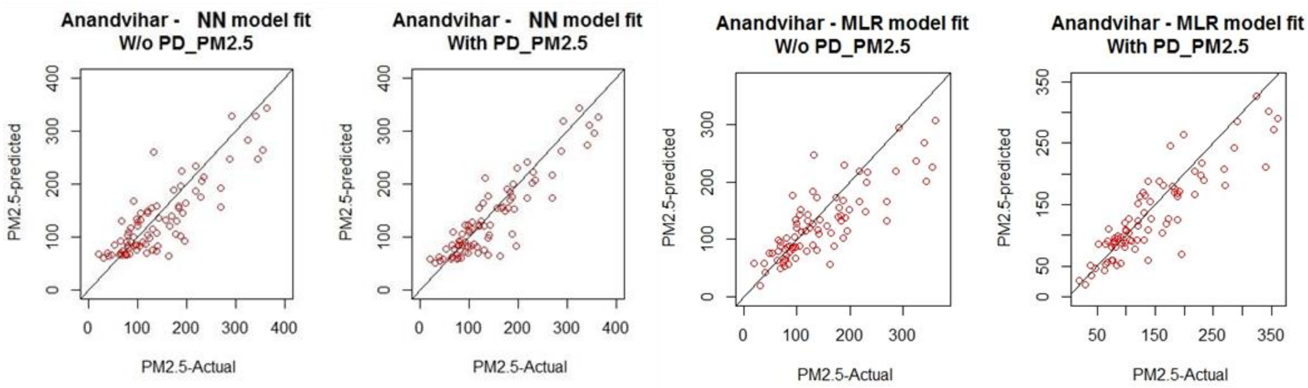


Figure 19: Punjabi Bagh – Comparative Model Fit Graph for PM 2.5

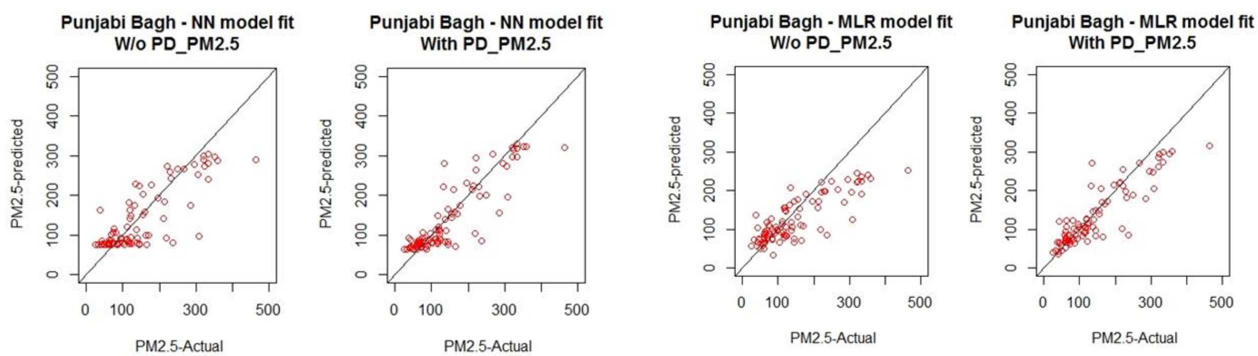


Figure 20: R.K. PURAM – Comparative Model Fit Graph for PM 2.5

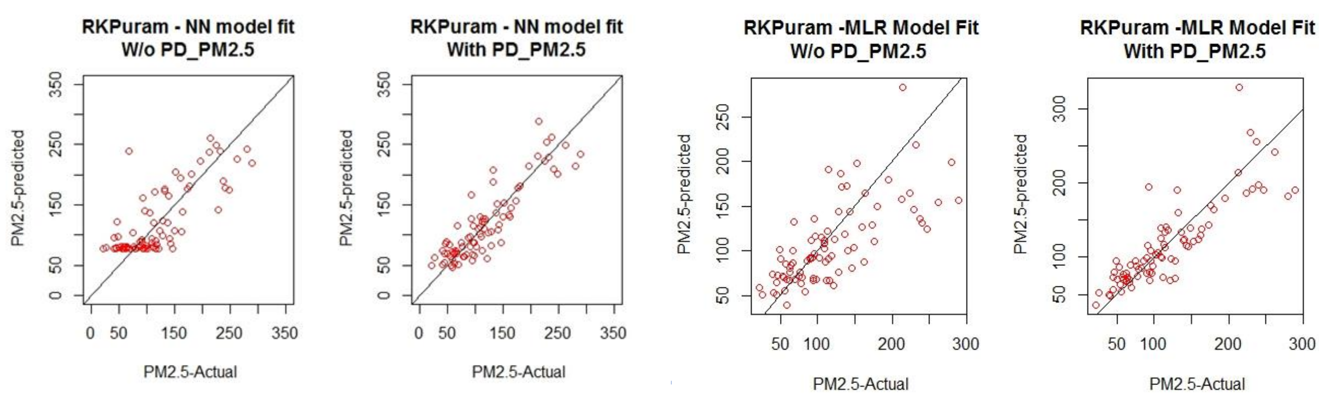
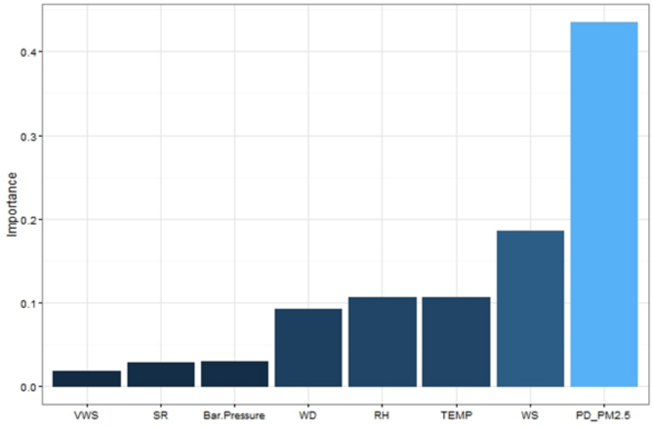
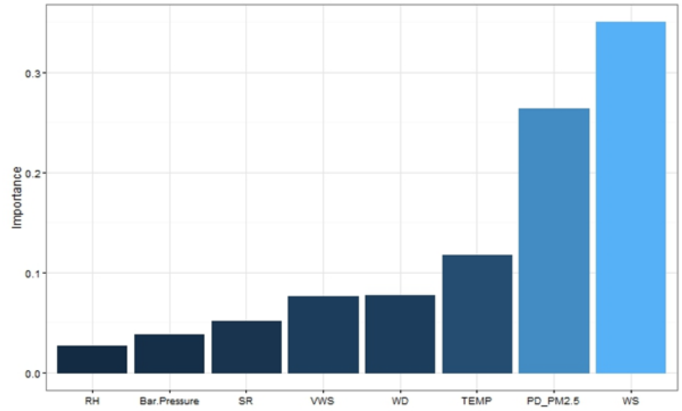


Figure 21,22 & 23: Relative Importance Variables for the Three Locations

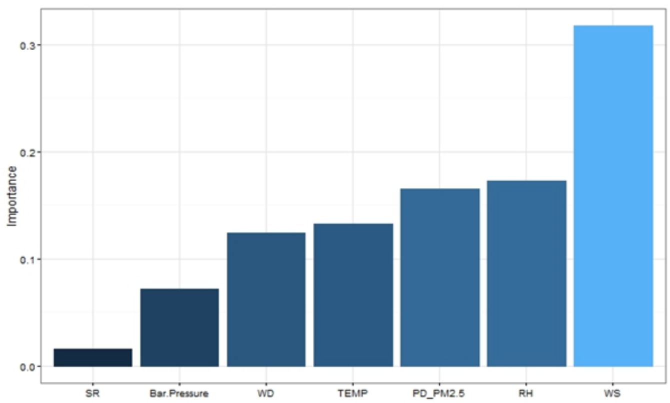


R.K. Puram



Punjabi Bagh

Anand Vihar



- Wind Speed is the most important variable for Punjabi Bag as well as Anand Vihar. It is the 2nd most important variable for R.K. Puram.
- Previous Day's level is the second most important variable for PB and the most important variable for R.K. Puram.
- Temp is the next important variable

4.2. Model Validation:

We used Jackknife Validation Method for validating the 4 Models and their relative performance

We also used Root Mean Square Error (RMSE) Value method to validate and compare the relative performance of the 4 Models that we have developed.

We also performed the relative error check to validate the model.

The results of the three validations are presented in the Tables 6.

TABLE 6: Comparative Performance of Models with Jacknife Validation										
		R.K. PURAM			ANAND VIHAR			PUNJABI BAGH		
Variables	With Previous Day value and without PD value	MLR Relative Error in %	NN Relative Error in %	% Variance between models	MLR Relative Error in %	NN Relative Error in %	% Variance between models	MLR Relative Error in %	NN Relative Error in %	% Variance between models
PM 2.5	with out PD value	34	29.45	13.38235	31.52	28.83	8.534264	32.42	31.43	3.053671
	with PD value	21.98	**20.67**	5.959964	24.15	**21.66**	10.31056	**22.74**	23.46	-3.16623
PM 10	with out PD value	32.36	25.96	19.7775	35.15	31.03	11.72119	30.5	27.21	10.78689
	with PD value	20.98	**18.85**	10.15253	24.26	**22.47**	7.378401	22.97	**22.07**	3.918154
NO2	with out PD value	20.42	18.18	10.96964	26.39	18.18	31.11027	15.13	15.66	-3.50297
	with PD value	16.25	**15.89**	2.215385	**19.46**	19.51	-0.25694	14.2	**13.81**	2.746479
SO2	with out PD value	39.9	15.75	60.52632	31.71	29.4	7.284768	43.78	42.41	3.129283
	with PD value	24.23	**9.49**	60.83368	**22.84**	23.95	-4.85989	**29.11**	31.58	-8.48506
CO	with out PD value	39.55	41.22	-4.2225	N/A	N/A	N/A	22.46	18.44	17.89849
	with PD value	30.69	**29.87**	2.67188	N/A	N/A	N/A	18.4	**17.8**	3.26087

Inference:

- For all the predictor variables Model built with Previous Day's value provides the lowest Relative error. Across most of Predictor variable, Neural Network gives the lowest Relative Error in prediction. Only for Punjabi Bagh PM 2.5; SO2 & Anand Vihar's NO2; SO2 MLR provides lower Relative error.

Predictive Model Development Conclusions:

- Multiple Linear Regression Model is able to explain almost 76% of variations in PM 2.5. across all location and in comparison, Neural Network Model is able to explain up to 82% in R.K. Puram & Punjabi Bagh and to a lower 73.9% in Anand Vihar.
- Neural Network overall is able to provide lower RMSE values for PM 2.5 & PM 10 across locations except for Punjabi Bagh (PM 2.5) where MLR gives a slightly lower RMSE value.
- Wind Speed seem to be the most important independent variable followed by the Previous Day's Value and temperature.
- Model Fit seem to be significant for PM 2.5 for both the models across locations.
- **Overall Neural Network Model was able to relatively perform better as compared to Multiple Linear Regression Model for predicting many pollutants across location.**

Next Steps:

- Further strengthen the Model by including another 12-24 months of data. This will help further increase the accuracy of the Models.
- There is some opportunity to do PCA Analysis, Factor Analysis and Discriminant Analysis to further separate the pollutant factors and identify the combinations of pollutants and its impact at each location. This could help the local administration to chart out a localized strategy for Pollution reduction.

CHAPTER 5: ODD-EVEN CAMPAIGN

Analyzing the impact of the campaign on New Delhi's air pollution levels

For the Odd-Even Campaign Analysis, we have taken 4 locations for consideration. They are:

- Anand Vihar
- Punjabi Bagh
- R.K. Puram
- Shadipur

The Key Air Pollutant levels were obtained for the 15 days prior to the Campaign and for the 15 days Campaign period. For purpose of record, these days are:

Pre Campaign Period: 1st April 2016 to 14th April 2016

Campaign Period: 15th April 2016 to 30th April 2016

5.1. Average Pollutant Level Analysis

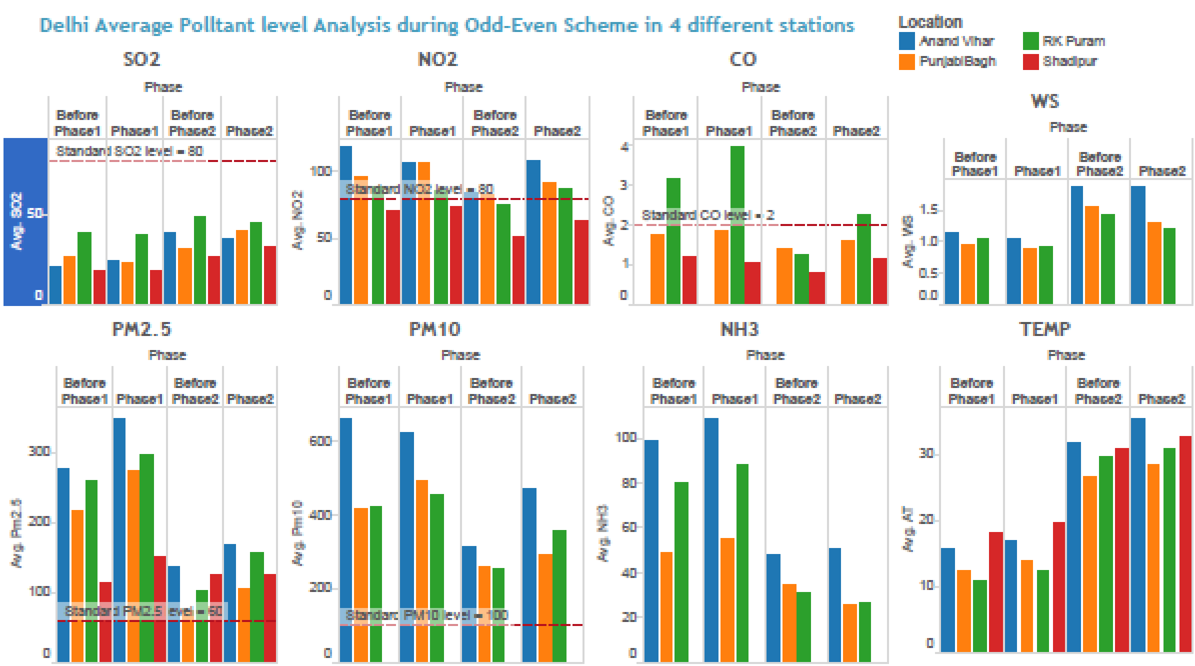


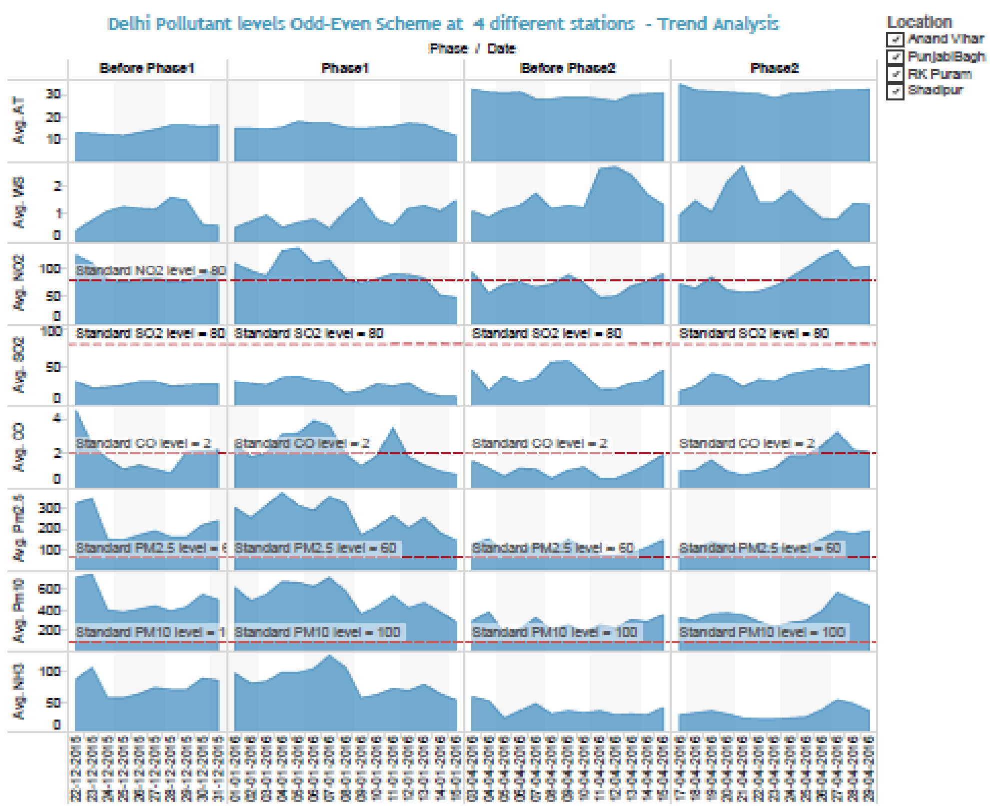
Figure 24: Average Pollutant Levels across 4 locations.

Insights:

- PM 2.5, PM 10, CO & NO2 showed significant increase in levels during Odd-Even
- SO2 & NO3 showed marginal decline during Phase II
- **All locations showed a drop in wind speed during the phase I & II of the ODD-EVEN Campaign**

5.2 Pollutant levels Trend Analysis

Figure 25: Pollution Level Trend Analysis Graph -All Locations combined



- Pollutant levels went up towards the end of Phase II accompanied by lower WS.
- Pollutant levels dropped towards the end of Phase I accompanied by higher WS.

5.3. PM 2.5 & PM 10 Levels during Phase 2

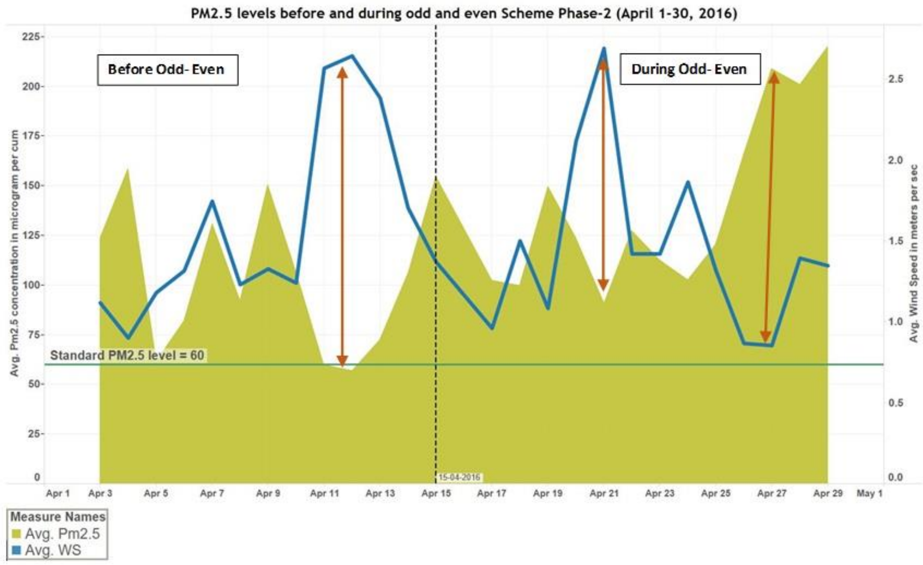
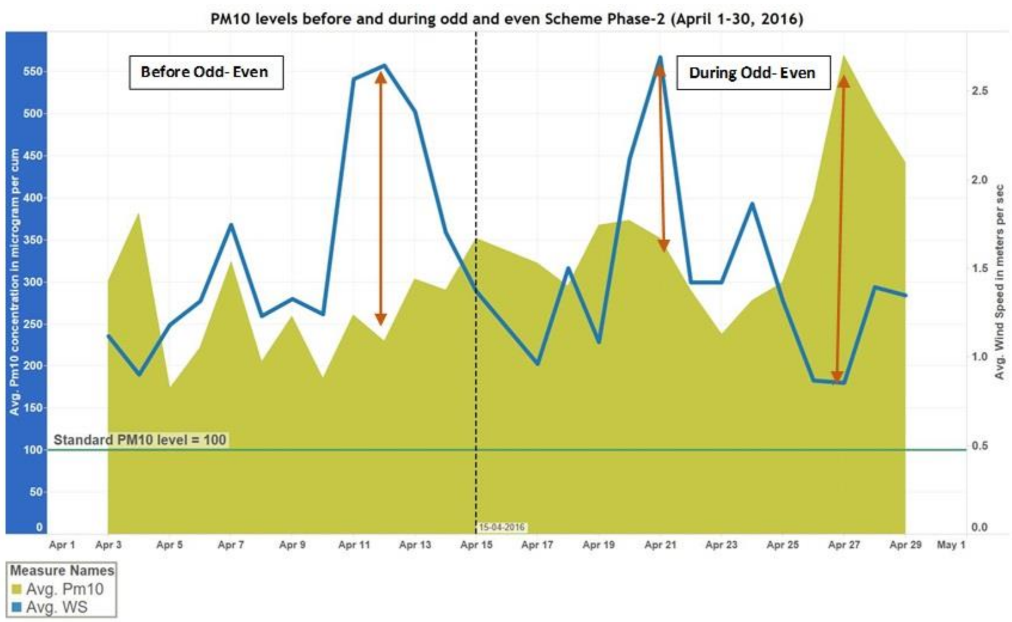


Figure 26 (up) & 27(down): Graphs showing the PM 10 & PM 2.5 levels before and during Odd-Event Campaign (II)



Insights:

- There is clear correlation between wind speed and PM 2.5 & PM 10 Levels.
- Drop in wind Speed after 24^{th} accompanied by spike in PM 2.5 levels

5.4. ODD-EVEN Impact on Traffic (Cars):

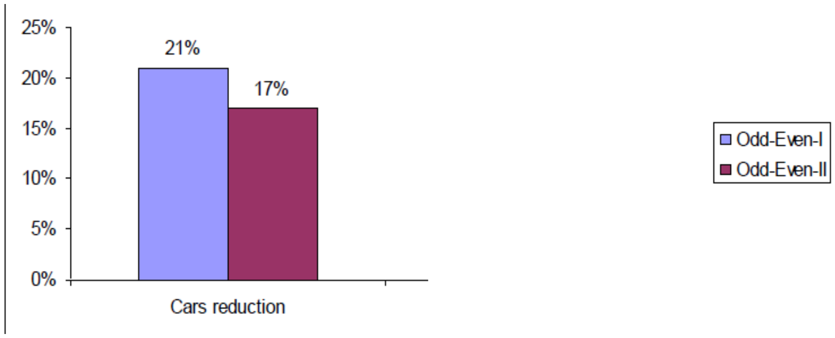


Figure 28: Impact on Number of Cars on the Road

Insights: Reduction in Cars on road between 8AM -8PM was 17% during Phase I, this dropped to 13% during phase II. Lower reduction rate attributed to: using 2nd car, taxis & CNG kit installation.

5.5. Impact of Bio Mass Residual Burning on ODD-EVEN Campaign:

- Satellite image substantiate impact of bio mass burning
- 1st April image establish a near absence of any fire
- 21st April image shows the start of the fire across Punjab, Haryana and Himalaya
- 26th & 31st image establish the widespread fire phenomenon

Figure 29: Picture showing the Bio Mass Burning across North India

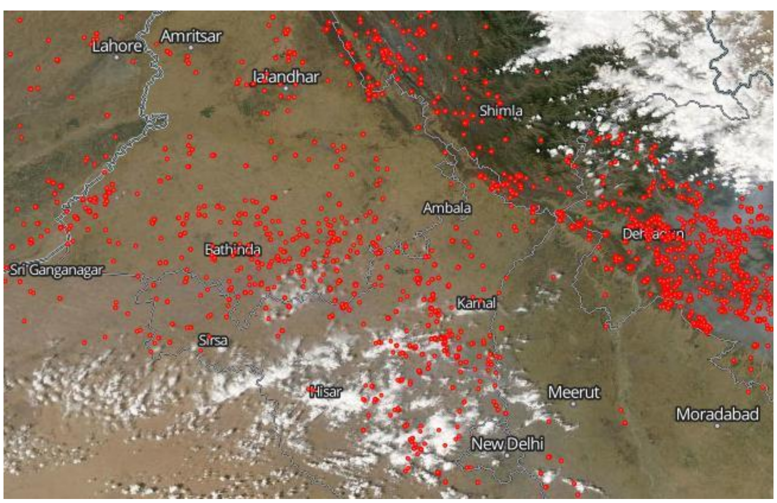
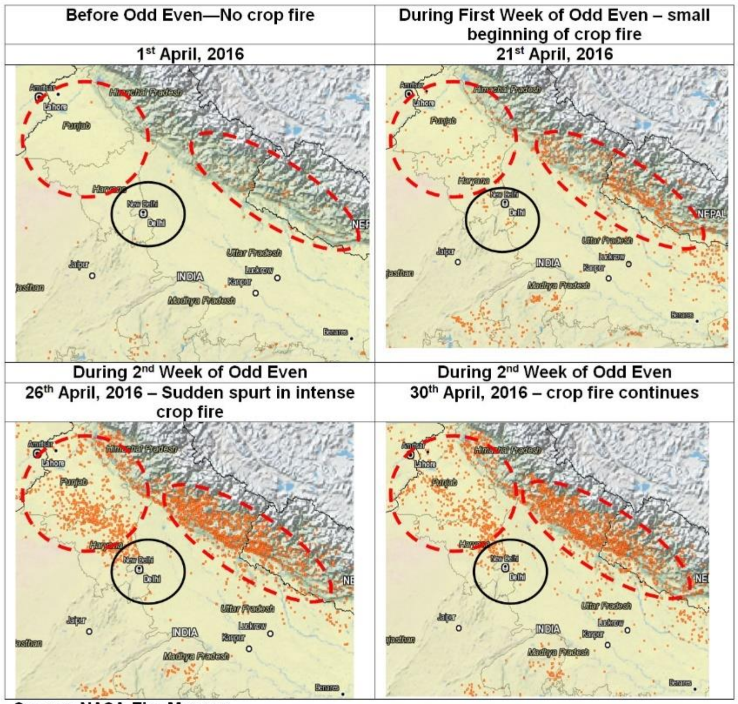


Figure 30: Picture showing the impact of Bio-Mass burning

Figures: NASA Satellite Images showing open crop burning in Punjab, Haryana (From April 1 – 30, 2016



Source: NASA Fire Mapper

- Satellite image showing the extent of Bio Mass burning immediately after the harvest.
- This year started around 19-21th April.
- Picture dated 26th April'16
- Setting of smog captured at the bottom

5.6. QUANTIFYING THE BIO MASS BURNING IN INDIA:

Bio Mass Residual Burning – 2008-09 - State wise

- 56% of PM 2.5 is contributed by the 4 neighbouring states of New Delhi. i.e. Haryana, Punjab, Rajasthan & Uttar Pradesh
- Aided by wind speed and favourable wind direction the pollutants drift to New Delhi and compounding the air Pollution levels of the capital

Table 7: Table showing the amount of Pollutant generated due to Bio-Mass burning across various States of India

States	CO_2	CO	NO_x	SO_x	NMVOC	NMHC	NH_3	HCN	PAH	TPM	$PM_{2.5}$	BC
	Gg/yr											
Andhra Pradesh	8009.96	486.41	13.22	2.11	83.01	37.01	6.87	0.79	0.13	68.73	20.62	3.65
Arunanchal Pradesh	80.78	4.91	0.13	0.02	0.84	0.37	0.07	0.01	0.00	0.69	0.21	0.04
Assam	1460.41	88.69	2.41	0.39	15.13	6.75	1.25	0.14	0.02	12.53	3.76	0.67
Bihar	5077.03	308.31	8.38	1.34	52.61	23.46	4.36	0.50	0.08	43.57	13.07	2.31
Chhattisgarh	1110.69	67.45	1.83	0.29	11.51	5.13	0.95	0.11	0.02	9.53	2.86	0.51
Goa	39.19	2.38	0.06	0.01	0.41	0.18	0.03	0.00	0.00	0.34	0.10	0.02
Gujarat	6835.92	415.12	11.28	1.80	70.84	31.59	5.87	0.68	0.11	58.66	17.60	3.11
Haryana	13907.71	844.56	22.95	3.67	144.13	64.26	11.93	1.38	0.23	119.34	35.80	6.33
Himachal Pradesh	635.45	38.59	1.05	0.17	6.59	2.94	0.55	0.06	0.01	5.45	1.64	0.29
Jammu & Kashmir	1403.12	85.21	2.32	0.37	14.54	6.48	1.20	0.14	0.02	12.04	3.61	0.64
Jharkhand	1939.61	117.78	3.20	0.51	20.10	8.96	1.66	0.19	0.03	16.64	4.99	0.88
Karnataka	8987.46	545.77	14.83	2.37	93.14	41.53	7.71	0.89	0.15	77.12	23.14	4.09
Kerala	184.66	11.21	0.30	0.05	1.91	0.85	0.16	0.02	0.00	1.58	0.48	0.08
Madhya Pradesh	3032.18	184.13	5.00	0.80	31.42	14.01	2.60	0.30	0.05	26.02	7.81	1.38
Maharashtra	10335.70	627.65	17.06	2.73	107.11	47.76	8.87	1.02	0.17	88.69	26.61	4.71
Manipur	109.00	6.62	0.18	0.03	1.13	0.50	0.09	0.01	0.00	0.94	0.28	0.05
Meghalaya	76.61	4.65	0.13	0.02	0.79	0.35	0.07	0.01	0.00	0.66	0.20	0.03
Mizoram	15.56	0.95	0.03	0.00	0.16	0.07	0.01	0.00	0.00	0.13	0.04	0.01
Nagaland	141.23	8.58	0.23	0.04	1.46	0.65	0.12	0.01	0.00	1.21	0.36	0.06
Orissa	1984.66	120.52	3.28	0.52	20.57	9.17	1.70	0.20	0.03	17.03	5.11	0.90
Punjab	32299.31	1961.41	53.30	8.53	334.72	149.24	27.72	3.20	0.53	277.16	83.15	14.71
Rajasthan	4202.19	255.18	6.93	1.11	43.55	19.42	3.61	0.42	0.07	36.06	10.82	1.91
Sikkim	18.95	1.15	0.03	0.01	0.20	0.09	0.02	0.00	0.00	0.16	0.05	0.01
Tamil Nadu	5099.67	309.68	8.42	1.35	52.85	23.56	4.38	0.50	0.08	43.76	13.13	2.32
Tripura	173.76	10.55	0.29	0.05	1.80	0.80	0.15	0.02	0.00	1.49	0.45	0.08
Uttar Pradesh	33701.42	2046.55	55.61	8.90	349.25	155.72	28.92	3.34	0.56	289.19	86.76	15.35
Uttarakhand	1146.20	69.60	1.89	0.30	11.88	5.30	0.98	0.11	0.02	9.84	2.95	0.52
West Bengal	8219.03	499.11	13.56	2.17	85.17	37.98	7.05	0.81	0.14	70.53	21.16	3.74
A & N Islands	5.66	0.34	0.01	0.00	0.06	0.03	0.00	0.00	0.00	0.05	0.01	0.00
D & N Haveli	6.81	0.41	0.01	0.00	0.07	0.03	0.01	0.00	0.00	0.06	0.02	0.00
Delhi	25.40	1.54	0.04	0.01	0.26	0.12	0.02	0.00	0.00	0.22	0.07	0.01
Daman & Diu	1.61	0.10	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.01	0.00	0.00
Pondicherry	30.07	1.83	0.05	0.01	0.31	0.14	0.03	0.00	0.00	0.26	0.08	0.01
All India	**149240.68**	**9062.80**	**6.90**	**246.27**	**39.40**	**1546.59**	**128.06**	**14.78**	**2.46**	**1280.61**	**384.18**	**67.97**

5.7. Text Mining of Tweets for Odd-Even Phase-II (April 15th 2016 – April 30th 2016) for Sentiment Analysis

5.7.1. Introduction

As part of the study "Identifying Patterns in New Delhi's Air Pollution", text mining of tweets was undertaken to identify the sentiment of people towards Odd-Even Phase-II in New Delhi.

Odd-Even rule is Delhi Government's new proposed rule to run vehicles with odd and even numbers on alternate days, and as a result is expected to reduce Air Pollution in New Delhi. The first trial period of this rule Phase-I was applied from 1st January 2016 to 15th January 2016. The second trial period of this rule Phase-II was applied from 15th April 2016 to 30th April 2016. During Phase-II of the rule the following vehicles were exempt from the rule

i. Emergency services vehicles, such as, ambulances, fire engines, and those belonging to the hospitals, prisons, hearses, and law enforcement vehicles.
ii. SPG (Special Protection Group) protectees.
iii. Vehicles with defence ministry numbers.
iv. Pilot Cars.
v. Embassy Cars.
vi. Two-wheelers.

5.7.2. Scope

The document describes the approach to mining of tweets for Odd-Even Phase-II.

5.7.3. Mining of Tweets - Obtaining Tweets

The data pipeline built for mining of tweets is as shown below

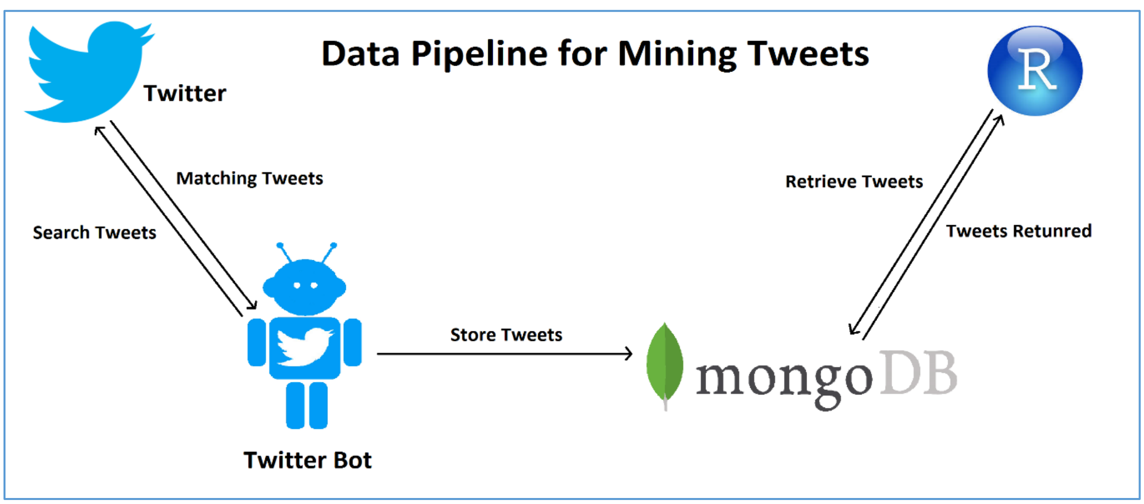


Figure 31: Data Pipeline for Mining Tweets

a. A twitter bot implemented in Node.js is used for retrieving tweets from Twitter. The bot is configured to use the OAuth credentials received from the Twitter Developer account.
b. The bot is executed every day during the period of Odd-Even Phase-II.
c. The bot uses the Twitter search API to retrieve tweets filtered on ‘Odd Even’.
d. The response of the Twitter search API is a JSON object which is then stored in MongoDB, which is a NoSQL database.
e. The twitter search APIs returns a maximum of 100 tweets for one request.
f. The response of Twitter search API contains tweets that were returned in an earlier search query thus resulting in duplication of tweets.
g. To resolve the duplication of tweets the ‘id’ (identifier) field of the tweet is used. Each tweet is identified by a unique ‘id’ which is returned in the response of the Twitter Search API. The ‘id’ is then used as a unique identifier rule set on the MongoDB collection, which ensures that only single copy of the tweet for a given ‘id’ is stored in MongoDB.
h. A total of 1172 unique tweets are collected during the Odd-Even Phase-II using this approach.

5.7.4. Analysis of tweets

The tweets collected were analyzed using R through the following steps:

a. First using R package ‘rmongodb’ tweets are imported into R and converted into a data frame.
b. The ‘text’ column in the resulting data frame contains the tweet which is to be further analyzed.
c. The tweets in the ‘text’ column is then cleaned to remove punctuation characters, URLs etc.
d. The tweets are then normalized by converting all tweet to lower case alphabets.
e. The cleansed tweets are then analyzed. The objective is to first create a word cloud and then analyze the sentiment of the cleansed tweets.
f. To create a word cloud R package ‘tm’ and ‘word cloud’ is used.
g. The tweets are first converted to ‘Corpus’ which is the data structure used for ‘tm’ package.
h. As a result, all tweets are converted to documents.
i. Then the stopwords are removed from these documents. Stopwords are common words that occur in a natural language.
j. After this the tweets in ‘Corpus’ is converted to ‘Term Document Matrix’. The ‘Term Document Matrix’ contains words as rows and documents(tweets) as columns. That is if a term (word) at the i^{th} row of the matrix appears in a document (tweet) at the j^{th} column of the matrix then the value 1 is stored at location [i][j] of the matrix else 0 is stored.
k. Then using the ‘Tern Document Matrix’ term (word) frequencies are calculated which are then stored in a data frame with its associated word. Now we have each word with its frequency stored as a data frame.

l. This is then visualized as a word cloud using the ‘wordcloud’ package.
m. The cleansed tweets available at step ‘e’ is now analyzed for sentiments.
n. Two kinds of scores are arrived at for each tweet. First scoring is based on emotional sentiments that a tweet has which can be – Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise and Trust. The second type of score is based on polarity which indicates if a tweet carries a ‘positive’ sentiment or a ‘negative’ sentiment.
o. R packages ‘syuzhet’, ‘lubridate’, ‘scales’, ‘reshape2’, ‘dplyr’ are used to arrive at sentiment scores for each tweet.
p. To analyze the sentiment over time the time stamp associated with each time frame is used.
q. Each tweet has a timestamp which is specific to Twitter service. To process this in R these are converted into POSIX timestamps.
r. Then R package ‘ggplot’ is used to visualize the sentiments over the period of Odd-Even Phase-II.

5.7.5. Analysis Results

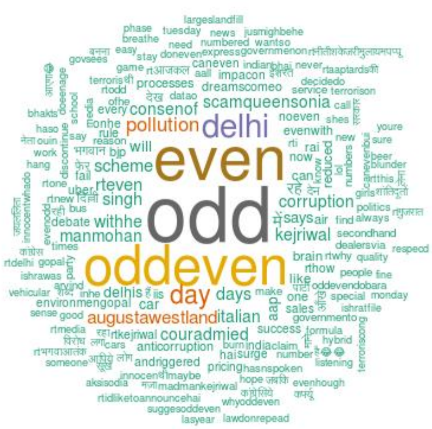


Figure32: Word Cloud for Tweets Collected

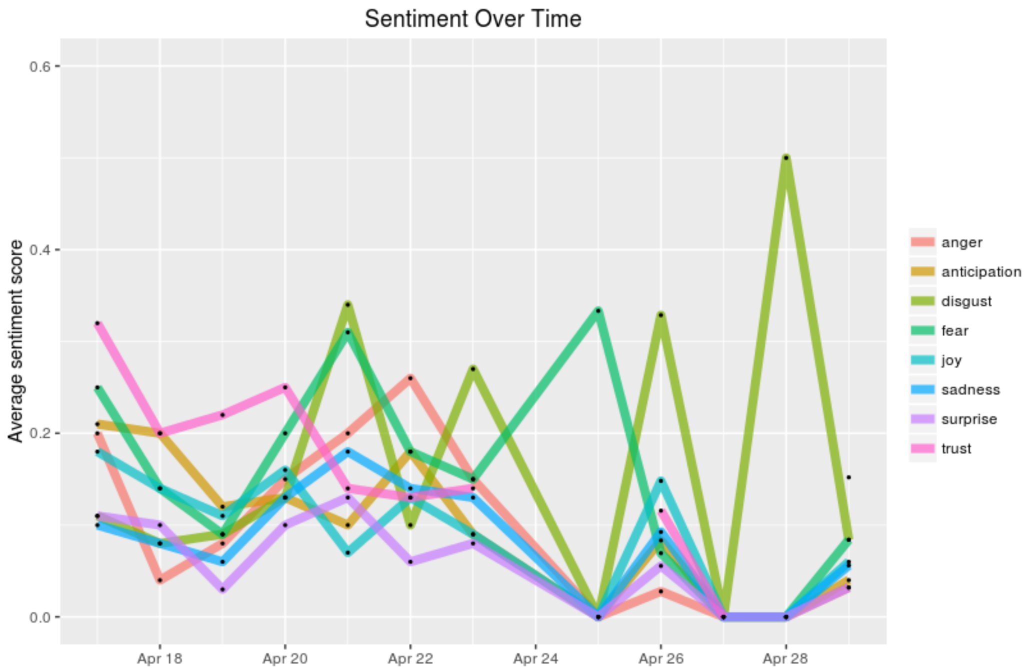


Figure33: Emotional Sentiment of Tweets over Time

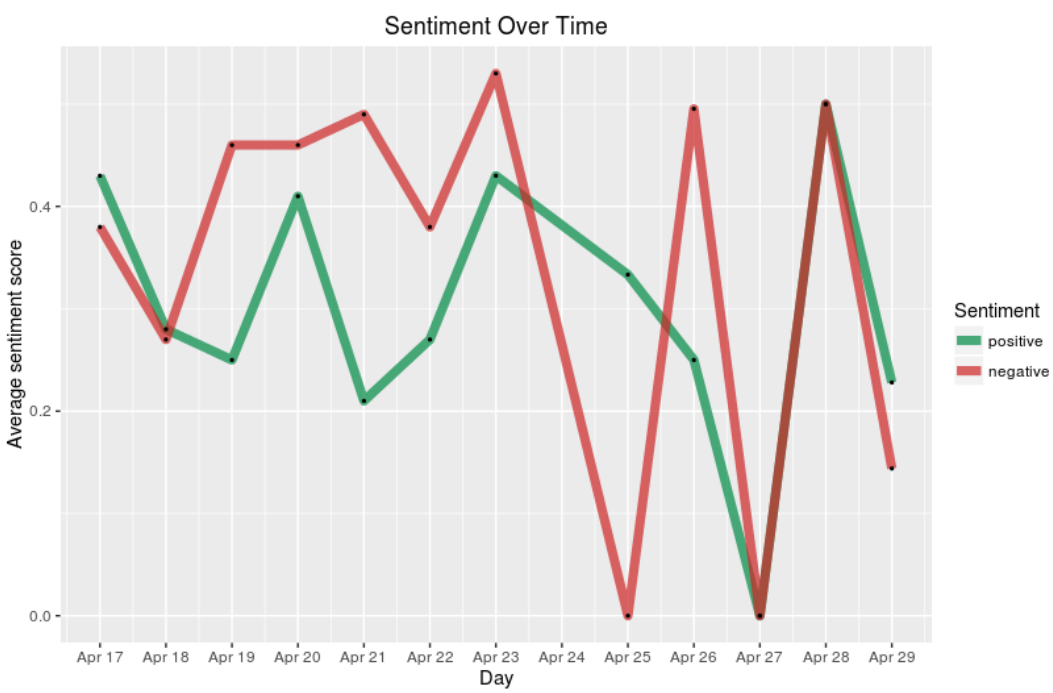


Figure34: Sentiment Polarity of Tweets over Time

5.7.6. Insights & Conclusions

- From the Sentiment analysis of the tweets collected for ‘Odd-Even’ Phase-II, it can be concluded that Twitterati largely holds negative sentiment towards this rule.
- Twitterati mostly holds negative sentiment about Odd Even Phase 2 with increase in negative sentiments towards the end of the Odd Even Phase 2 duration.
- Campaign started with good sentiments like Trust, Joy, Surprise. Unfortunately, negative sentiments like disgust took over from the second week onwards overriding the positive sentiments.

5.8. Conclusions: Odd-Even Campaign

- No apparent impact of ‘Odd-Even’ on the air pollution levels both during Phase I & Phase II
- PM 2.5, PM 10, CO, NO2 & SO2 all showed increased levels during the Campaign periods as compared to the preceding 15 days.
- The Bio Mass (Crop Residual) burning in the neighbourhood states like Punjab, Haryana & Rajasthan also contributed to the increased levels of air pollutants post 19/20th April’16.
- The average levels of Wind Speed went down during the Odd-Even Campaign Phase I & II contributing marginally to the increase in pollution Levels.
- There is a strong possibility that any gains from Odd-Even scheme in terms of air quality levels were entirely eclipsed by ***"other sources of pollution".***
- Some of the reasons for the lack of impact could be:
  - Vehicular pollution contributes only to 20% of Delhi’s air pollution.
  - Of this, only 13-14% is contributed by Cars (10% petrol and 4% diesel) a segment that was involved in the experiment.
  - Actual reduction in vehicle was only 13% during the campaign as compared to the normal period.
  - The other major contributing factors could be Road Dust -38%; domestic source-12% & Industrial pollutants-11%.
- Any spike in any of these other factors could drastically alter the air pollution levels in Delhi.
- Odd-Even Concept can work if it is not a for very long duration. It can work as an emergency short-term measure as done in Beijing for specific days when the pollution levels are expected/projected to exceed certain targeted levels.
- If it is implemented at semi-permanent measure for longer duration, the impact is likely to be diluted as citizens are expected to circumvent the rule by opting for multiple car, two-wheelers, hire taxi etc.

5.9. Recommendations:

- Introduce wet/machined vacuum sweeping of Roads
- Evolve a system for reporting of garbage/municipal solid waste burning through a mobile based application and other social media platforms directly linked with control rooms
- Set-up bio-mass based power generation units in the peripheral areas and neighbouring states
- Regulate carriage of construction materials in covered carriage
- Take stringent action against open burning of bio-mass, tyres etc.
- Control dust pollution at construction sites with appropriate covers
- Take steps for retrofitting the diesel vehicles with particulate filters
- Extend LPG/PNG coverage to 100%. Follow it with a phase-out of charcoal and kerosene cooking in New Delhi
- Engage Citizens actively and educate them on the need for participation as they are nor too happy with the Odd-Even Campaign. After the initial euphoria the sentiments about the Campaign turned negative.

Strick Norms with 'ALARM SYSTEM' FOR Specific Decisive Interventions as illustrated here

HERE'S WHY BEIJING'S AIR HAS BECOME A LOT CLEANER THAN DELHI'S

INDIA HAS LAX STANDARDS AND NO ALERTS

Alerts	Health Impact	Description	AQI
None	Minimal Impact	Good	0-50
	May cause minor breathing discomfort to sensitive people	Satisfactory	51-100
	May cause breathing discomfort to the people with lung disease; discomfort to people with heart disease, children and elders	Moderate	101-200
	May cause breathing discomfort to people on prolonged exposure; discomfort to people with heart disease	Poor	201-300
	May cause respiratory illness to the people on prolonged exposure; pronounced effect in people with lung and heart diseases	Very Poor	301-400
	May cause respiratory effects on healthy people; serious health impacts on people with lung/heart diseases	Severe	401-500

CHINA HAS STRICTER NORMS AND A FOUR-LEVEL ALARM SYSTEM

AQI	Description	Health Impact
0-50	Excellent	Air quality satisfactory, and air pollution poses little or no risk
51-100	Good	For some pollutants there may be a moderate health concern for a small number of people
101-150	Slightly Polluted	Sensitive groups may experience health effects. General public not likely to be affected
151-200	Moderately Polluted	Heart and respiratory system of everyone may be affected; serious health effects on sensitive groups
201-300	Heavily Polluted	Patients with heart and lung diseases severly affected; healthy people are commonly affected
>300	Severely Polluted	Weaker endurance in activities and significant severe health warnings among healthy; certain diseases have early appearance

Alerts:

- **When PM 2.5 levels are higher than 150 µm/m³** | Children, elderly with cardiovascular or respiratory conditions warned, dust prevention at construction sites
- **PM 2.5 levels above 150 µm/m3 for 3 days** | Vulnerable groups to stay indoors, power plants, factories asked to reduce emissions
- **PM 2.5 levels more than 150 µm/m3 for 3 days and more than 250 µm/m3 on some days** | factories close, no outdoor activity
- **PM2.5 above 250 µm/m3** | Schools close, power plants cut emissions, car use regulated as per licence number

Figure 35: Chart showing Trigger Alarm and corrective action

LIST OF TABLES & CHARTS (FIGURES)

LIST OF TABLES **PAGE**

Table 1: Table showing List of Variables

Table 2: Location wise Modeling Data

Table 3.1: Anand Vihar Air Pollution Predictive Model Results

Table 3.2: Anand Vihar MLR Model Beta Coefficient Table

Table 3.3 & 3.4: Anand Vihar NN Model Results

Table 4.1: Punjabi Bagh Air Pollution Predictive Model Results

Table 4.2: Punjabi Bagh -MLR Model Beta Coefficient Table

Table 4.3 & 4.4: Punjabi Bagh NN Model Results

Table 5.1: R.K. Puram Air Pollution Predictive Model Results

Table 5.2: R.K. Puram -MLR Model Beta Coefficient Table

Table 5.3 & 5.4: R.K. Puram NN Model Results

Table 6. Comparative Performance of Models with Jackknife Validation

Table 7: Amount of Pollutant generated due to Bio-Mass burning

LIST OF FIGURES – CHARTS, IMAGES & INFOGRAPHICS
PAGE

Figure 1: Air Quality Index for Beijing and New Delhi for a 4 Month period

Figure 2: New Delhi's PM10 Levels over a 10-year period

Figure 3: Top 20 polluted cities in the G-20 Countries (PM10)

Figure 4: High Level Process Flow

Figure 5: 3 City Pollution Level Vs Vehicle Density & Vehicle Population

Figure 6: The three most polluted areas of New Delhi (Infographics)

Figure 7: Histogram chart showing various pollutant levels

Figure 8: Box plot - Key pollutant levels in the three locations

Figure 9: Summary data-Anand Vihar

Figure 10: Summary data-R.K. Puram

Figure 11: Summary data- Punjabi Bagh

Figure 12: Anand Vihar – Graph/Chart - Pollutant levels across seasons

Figure 13: R.K. Puram – Graph/Chart- Pollutant levels across seasons

Figure 14: Punjabi Bagh – Graph/Chart-Pollutant levels across seasons

Figure 15: Correlation Matrix for Anand Vihar

Figure 16: Correlation Matrix for Punjabi Bagh

Figure 17: Correlation Matrix for R.K. Puram

Figure 18: Anand Vihar-Comparative Model Fit graph for PM 2.5

Figure 19: Punjabi Bagh-Comparative Model Fit graph for PM 2.5

Figure 20: R.K. Puram-Comparative Model Fit graph for PM 2.5

Figure 21,22 & 23: Relative Importance Variables for the Three Locations

Figure 24: Average Pollutant Levels across 4 locations.

Figure 25: Pollution Level Trend Analysis Graph -All Locations

Figure 26 (up) & 27(down): Graphs showing the PM 10 & PM 2.5 levels

Figure 28: Impact on Number of Cars on the Road

Figure 29: Picture showing the Bio Mass Burning across North India

Figure 30: Picture showing the impact of Bio-Mass burning

Figure 31: Data Pipeline for Mining Tweets

Figure32: Word Cloud for Tweets Collected

Figure33: Emotional Sentiment of Tweets over Time

Figure34: Sentiment Polarity of Tweets over Time

Figure 35: Chart showing Trigger Alarm and corrective action

TABLE OF EXHIBITS FOR APPENDIX

EXHIBIT DETAILS **PAGE**

Exhibit 1: R.K. Puram-Neural Network Model Fit Graph for PM2.5

Exhibit 2: Relative Importance of Meterological Factors

Exhibit 3 : Response & Explanatory Graph for PM2.5

Exhibit 4: Neural Network Model Fit Graph for PM10

Exhibit 5: Relative Importance of Meterological Factors for PM10

Exhibit 6 : Response & Explanatory Graph for PM 10

Exhibit 7: Neural Network Model Fit Graph for NO2

Exhibit 8: Relative Importance of Meterological Factors for NO2

Exhibit 9 : Response & Explanatory Graph for NO2

Exhibit 10: Neural Network Model Fit Graph for SO2

Exhibit 11: Relative Importance of Meterological Factors for SO2

Exhibit 12 : Response & Explanatory Graph for SO2

Exhibit 13: Neural Network Model Fit Graph for CO

Exhibit 14: Relative Importance of Meterological Factors for CO

Exhibit 15 : Response & Explanatory Graph for CO

Exhibit 16: Multiple Linear Regression Model Fit Graph for PM 2.5

Exhibit 17: Multiple Linear Regression Model Fit Graph for PM 10

Exhibit 18: Multiple Linear Regression Model Fit Graph for NO2

Exhibit 19: Multiple Linear Regression Model Fit Graph for SO2

Exhibit 20: Multiple Linear Regression Model Fit Graph for CO

APPENDIX- (One Location Sample) R.K. PURAM

EXHIBIT 1: R.K. PURAM – NEURAL NETWORK MODEL FIT GRAPH PM 2.5

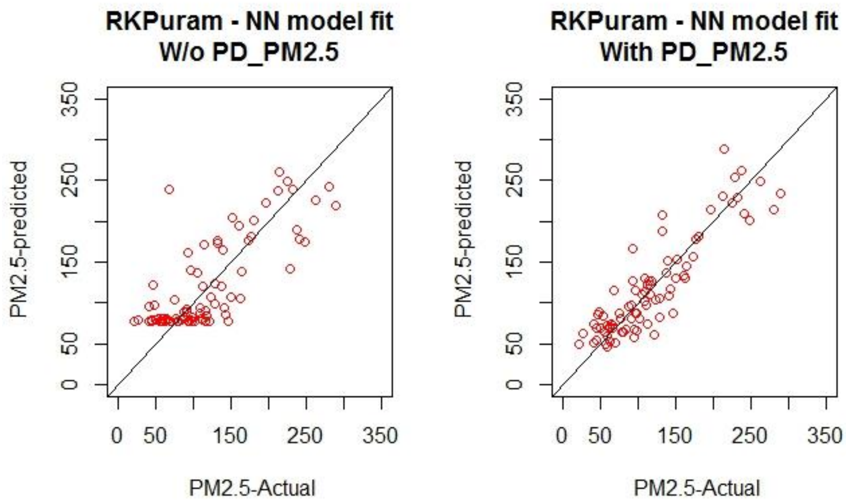


EXHIBIT 2: RELATIVE IMPORTANCE OF METEROLOGICAL FACTORS

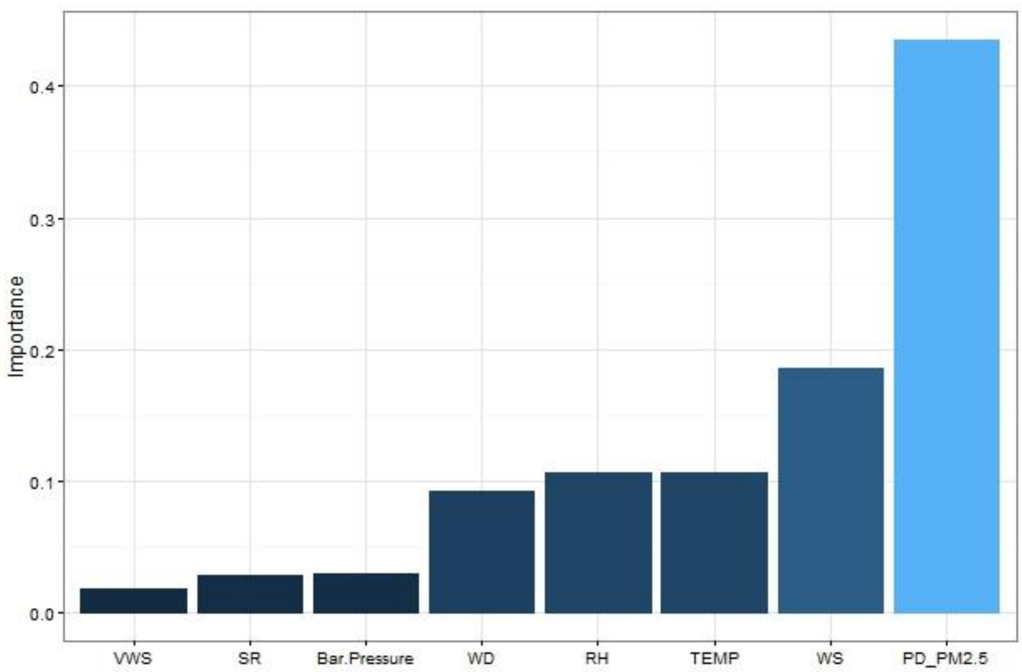


EXHIBIT 3 : RESPONSE & EXPLANATORY GRAPH FOR PM 2.5

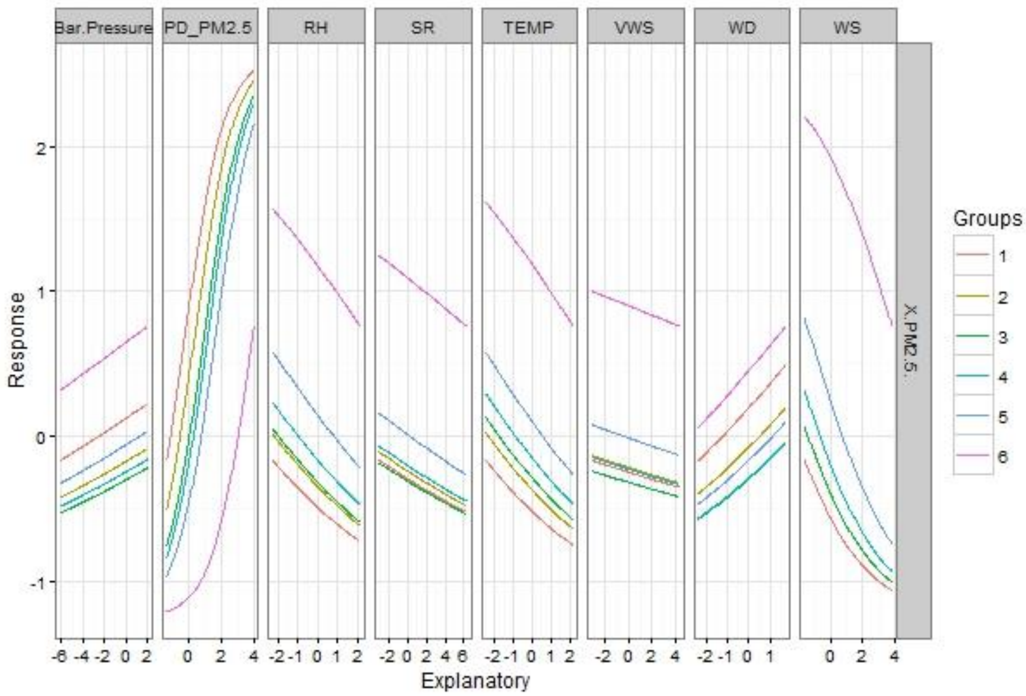


EXHIBIT 4: – NEURAL NETWORK MODEL FIT GRAPH PM 10 – WITH & w/o PD

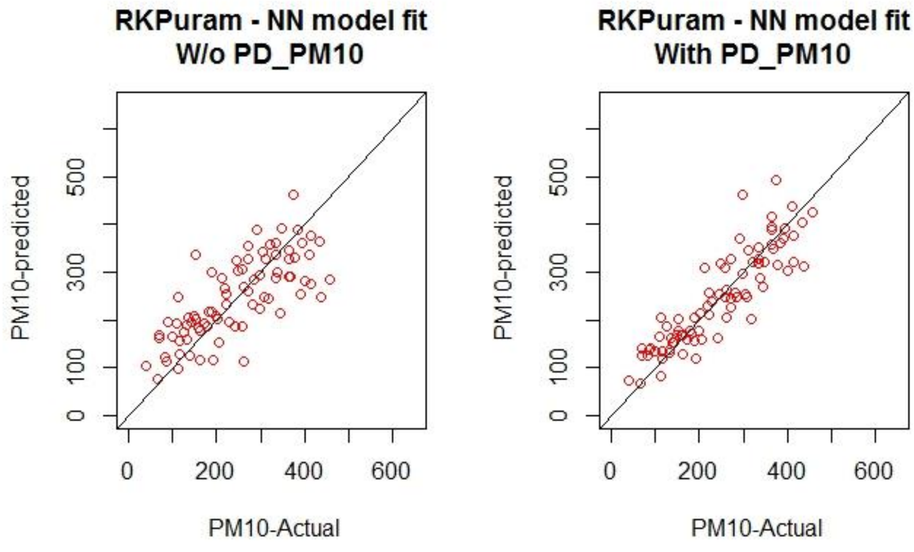


EXHIBIT 5: RELATIVE IMPORTANCE OF METEROLOGICAL FACTORS FOR PM 10

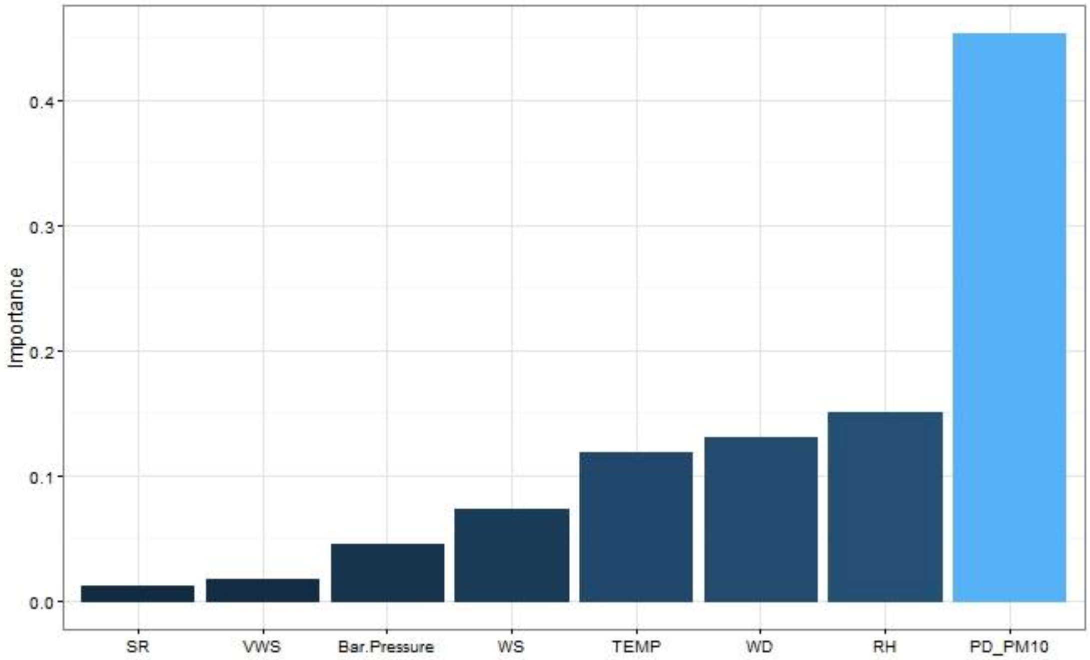


EXHIBIT 6 : RESPONSE & EXPLANATORY GRAPH FOR PM 2.5

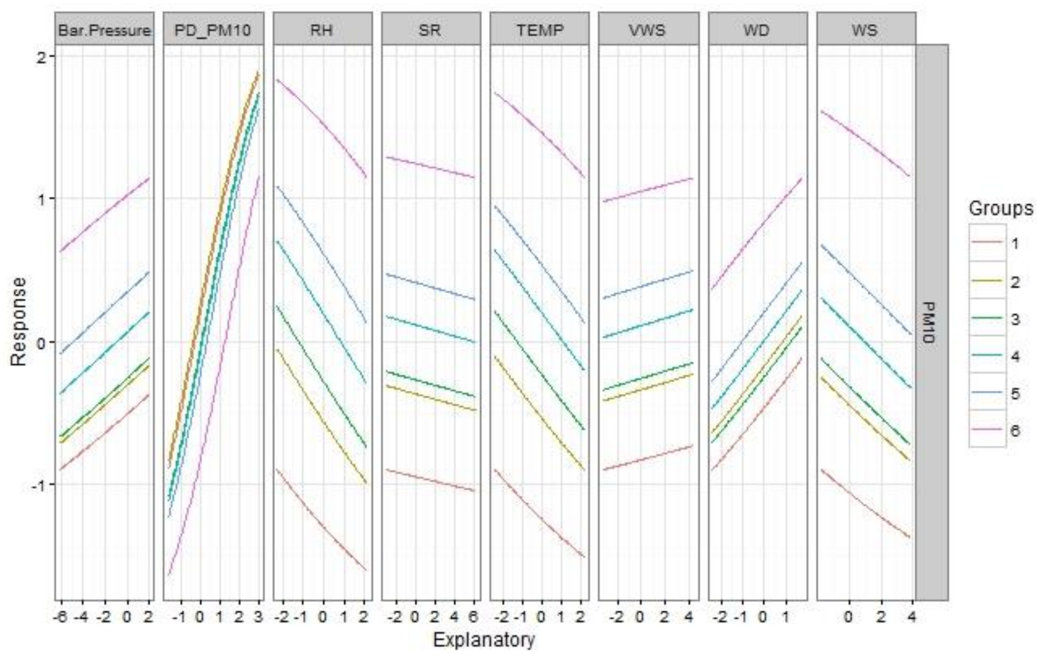


EXHIBIT 7: – NEURAL NETWORK MODEL FIT GRAPH NO2 – WITH & w/o PD

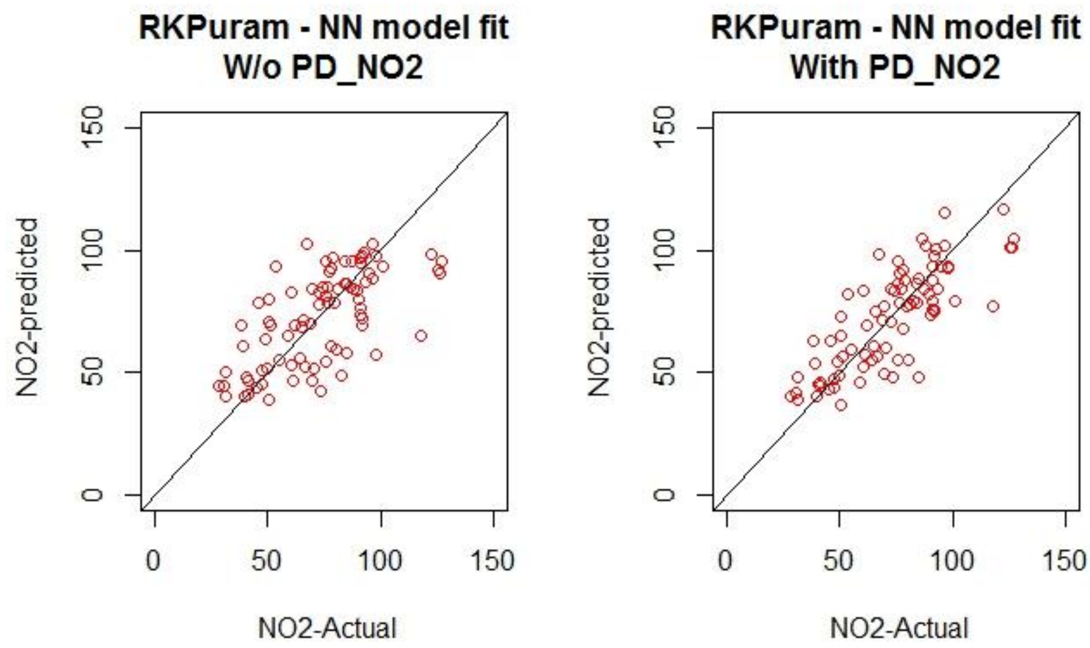


EXHIBIT 8: RELATIVE IMPORTANCE OF METEROLOGICAL FACTORS FOR NO2

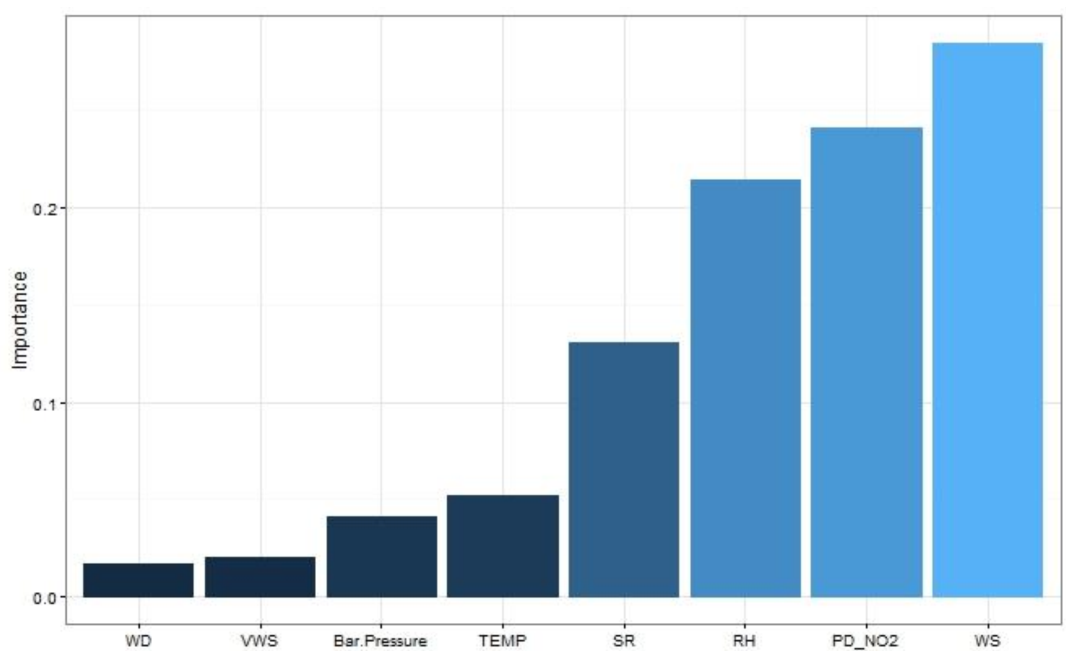


EXHIBIT 9 : RESPONSE & EXPLANATORY GRAPH FOR NO2

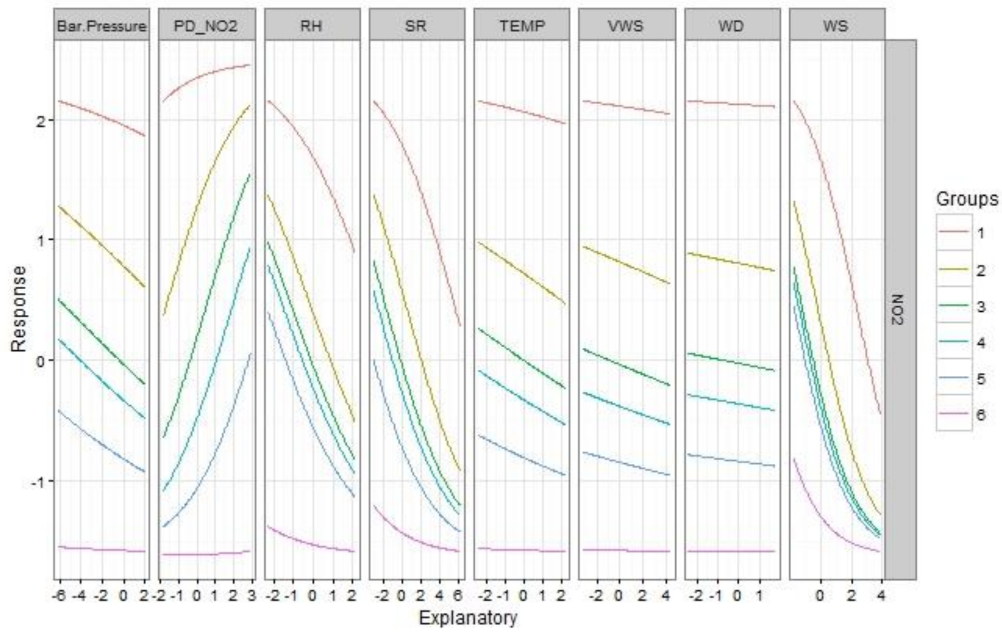


EXHIBIT 10: – NEURAL NETWORK MODEL FIT GRAPH SO2 – WITH & w/o PD

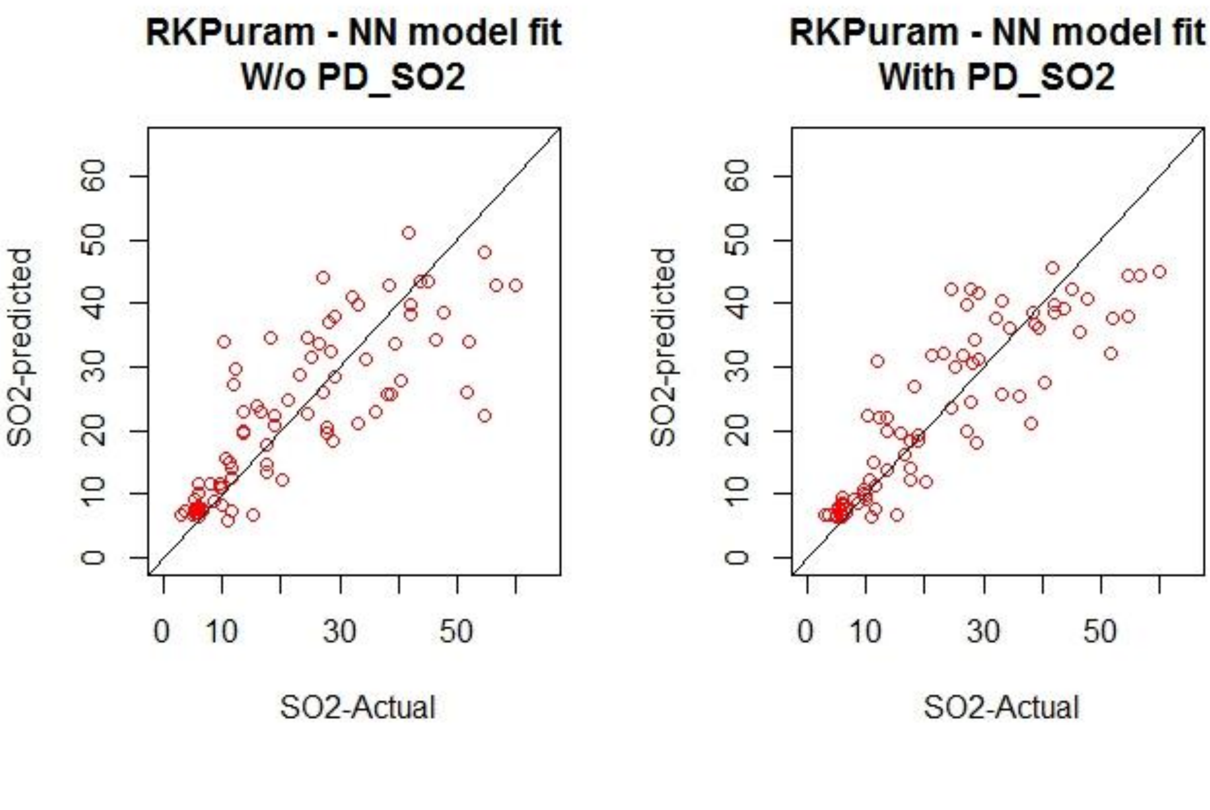


EXHIBIT 11: RELATIVE IMPORTANCE OF METEROLOGICAL FACTORS FOR SO2

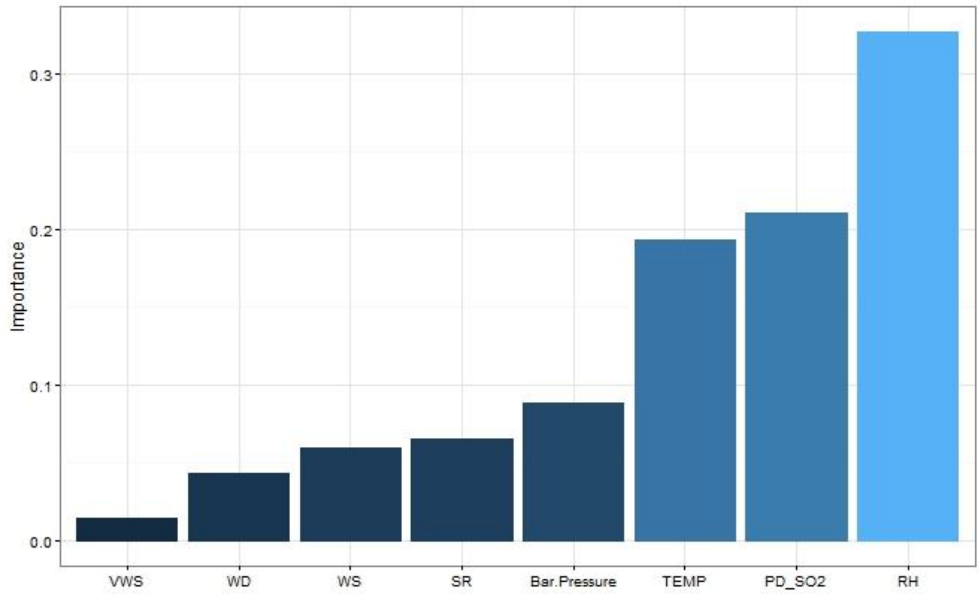


EXHIBIT 12 : RESPONSE & EXPLANATORY GRAPH FOR SO2

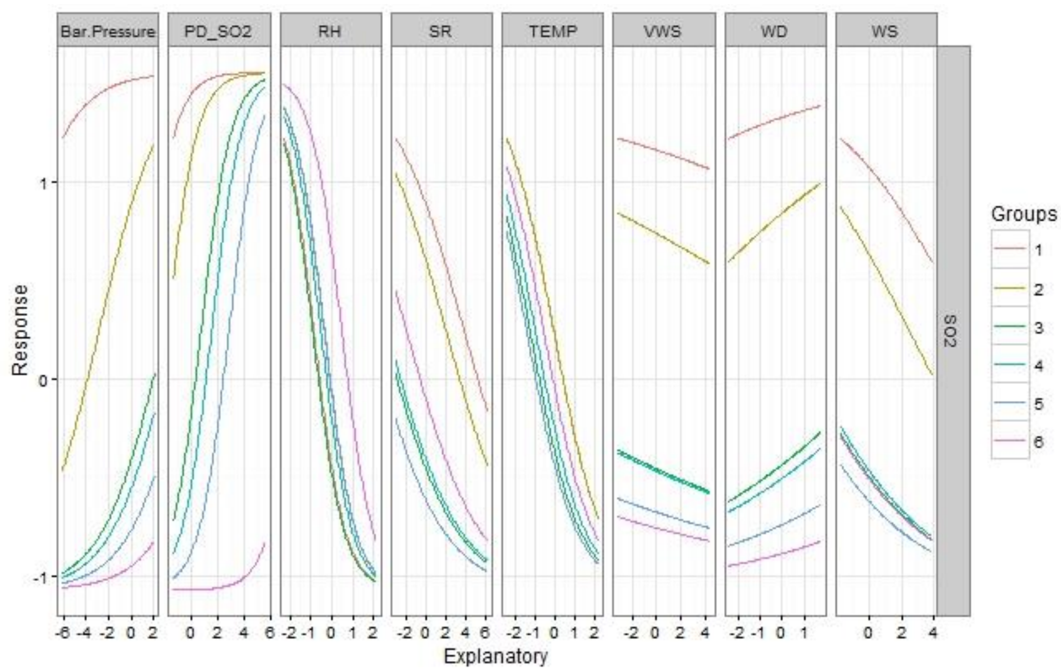


EXHIBIT 13: – NEURAL NETWORK MODEL FIT GRAPH CO – WITH & w/o PD

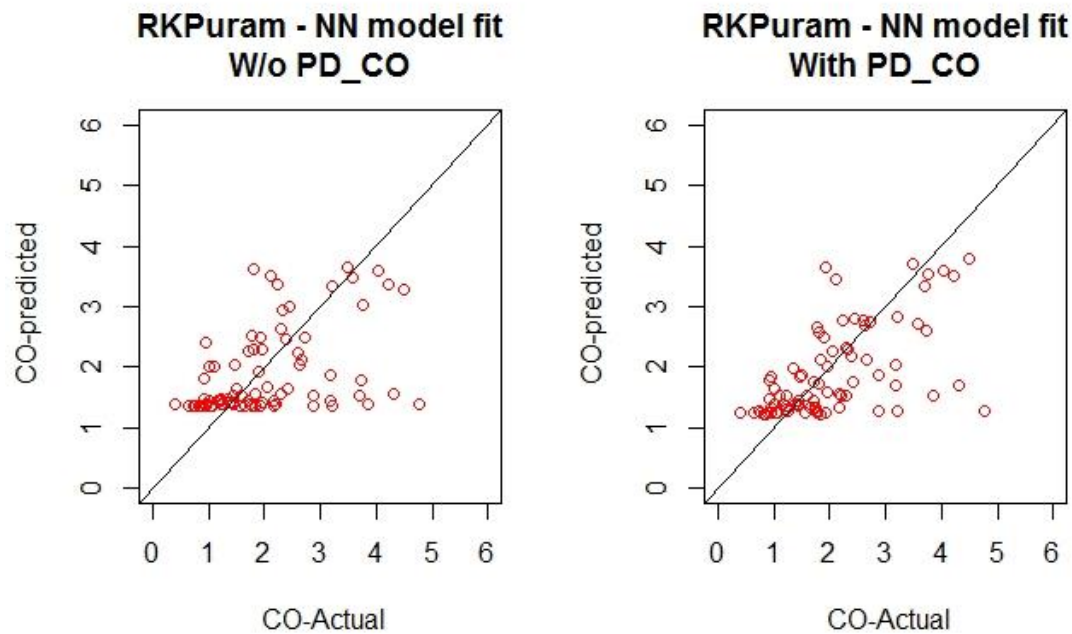


EXHIBIT 14: RELATIVE IMPORTANCE OF METEROLOGICAL FACTORS FOR CO

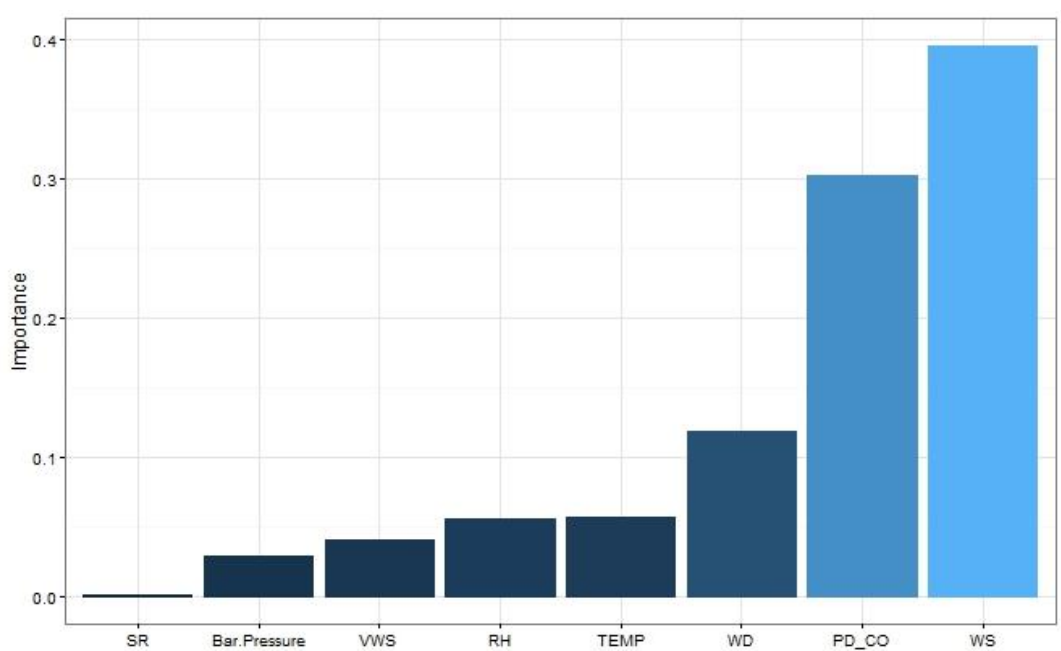


EXHIBIT 15 : RESPONSE & EXPLANATORY GRAPH FOR CO

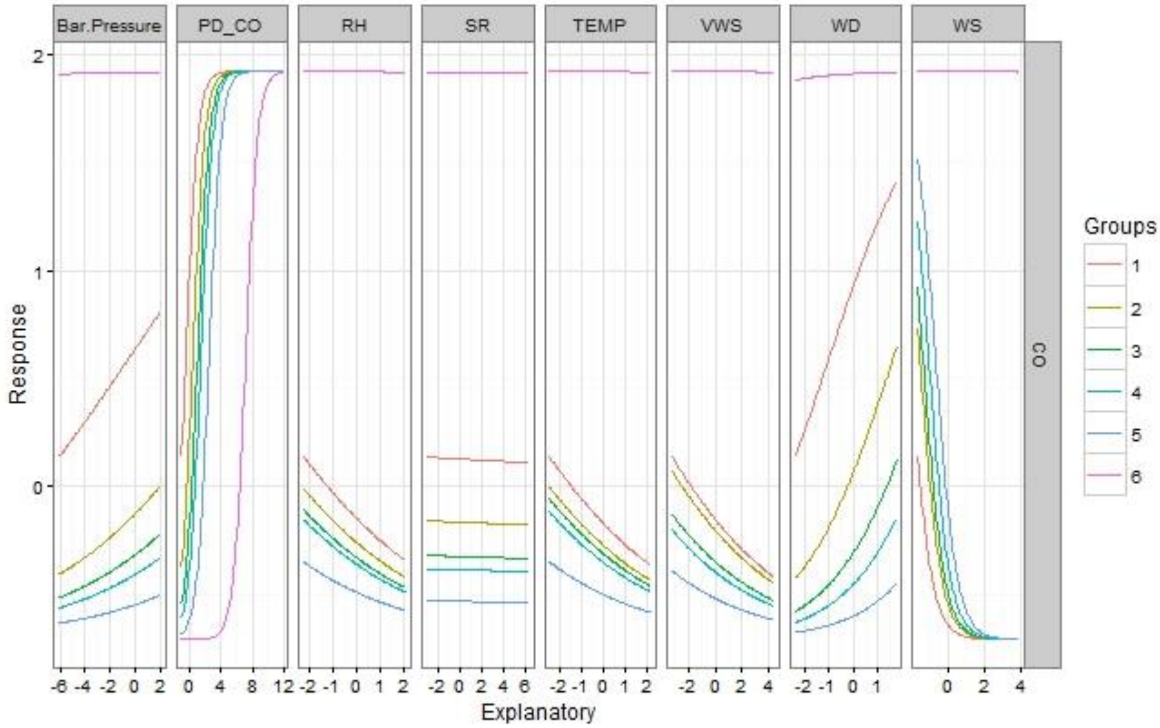


EXHIBIT 16: – MULTIPLE LINEAR REGRESION MODEL FIT GRAPH -PM 2.5

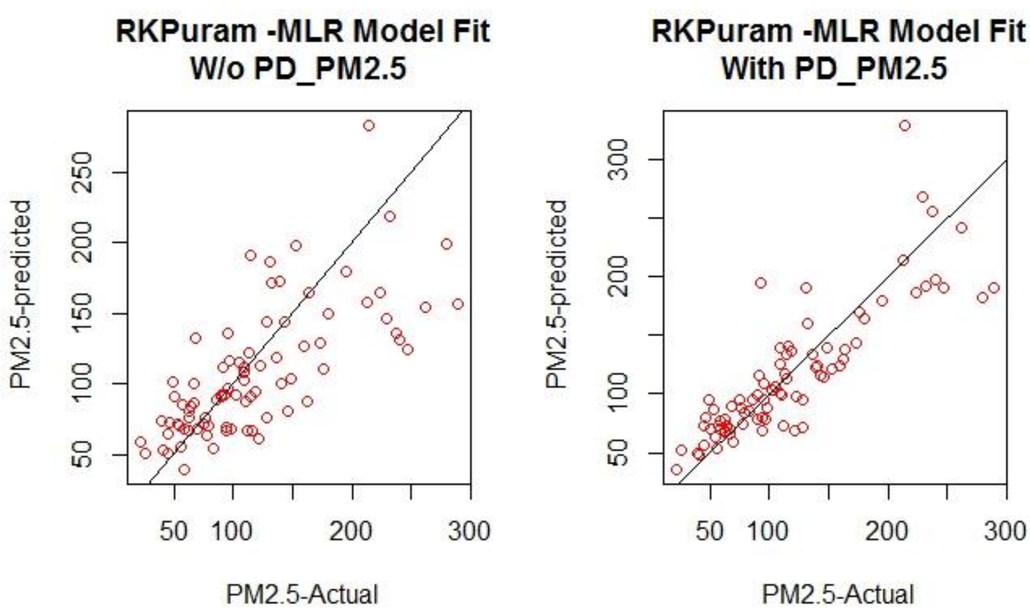
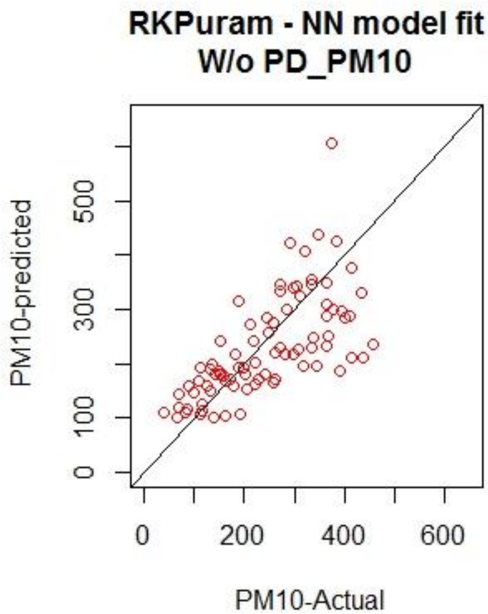


EXHIBIT 17: – MULTIPLE LINEAR REGRESSION MODEL FIT GRAPH PM 10 – WITH & w/o PD



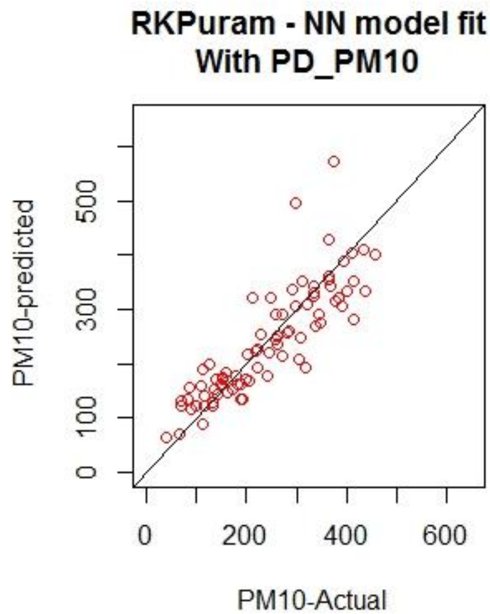
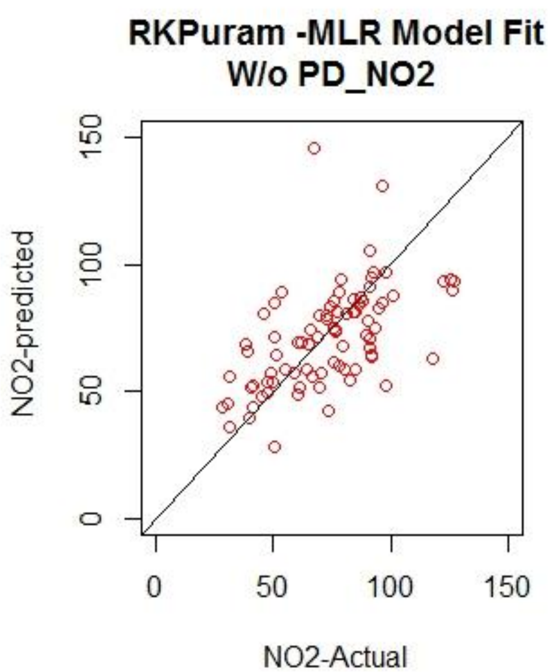


EXHIBIT 18: MULTIPLE LINEAR REGRESSION MODEL FIT GRAPH NO2



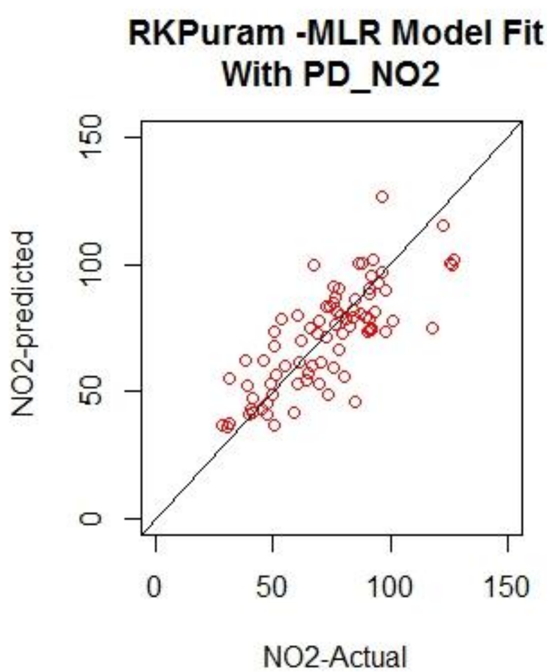


EXHIBIT 19: – MULTIPLE LINEAR REGRESSION MODEL FIT GRAPH SO2 – WITH & w/o PD

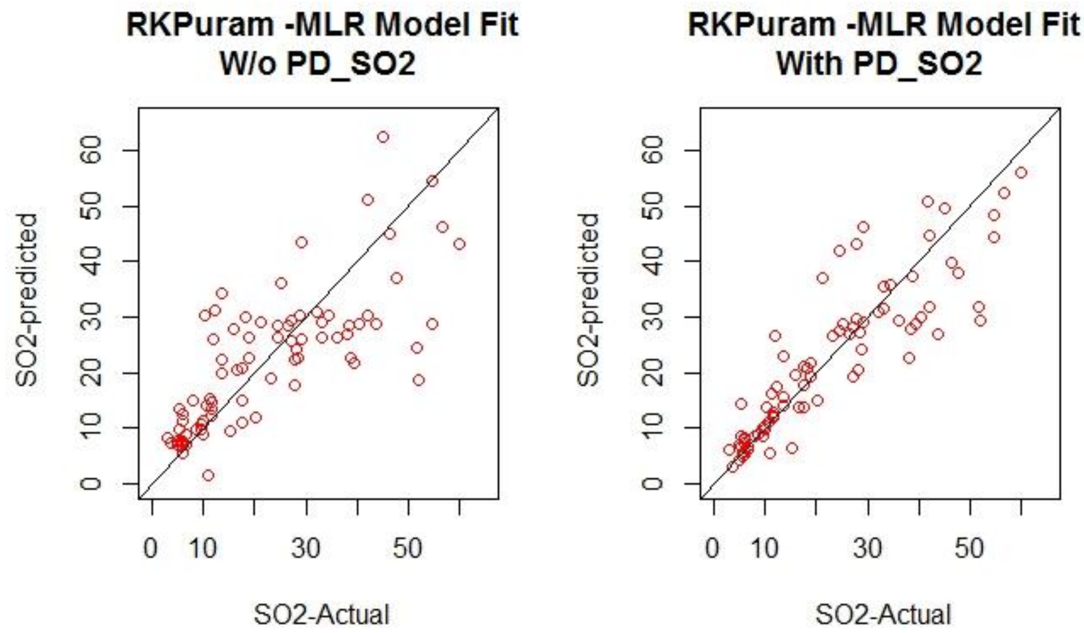
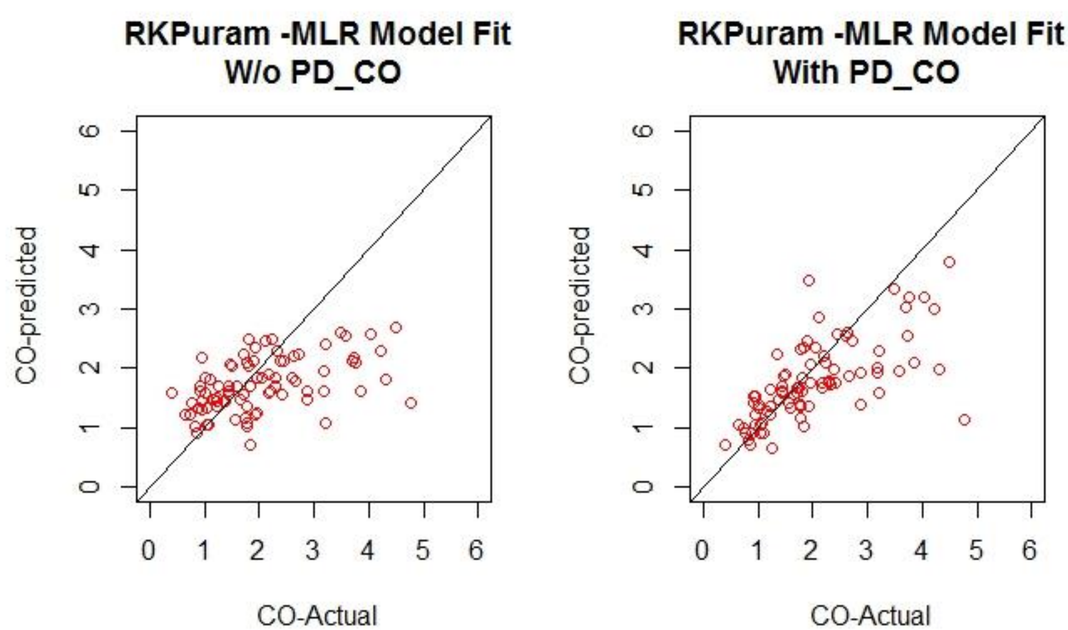


EXHIBIT 20: – MULTIPLE LINEAR REGRESSION MODEL FIT GRAPH CO



LIST OF TABLES & CHARTS (FIGURES)

LIST OF TABLES	PAGE
Table 1: Table showing List of Variables	15
Table 2: Location wise Modeling Data	27
Table 3.1: Anand Vihar Air Pollution Predictive Model Results	28
Table 3.2: Anand Vihar MLR Model Beta Coefficient Table	28
Table 3.3 & 3.4: Anand Vihar NN Model Results	29
Table 4.1: Punjabi Bagh Air Pollution Predictive Model Results	29
Table 4.2: Punjabi Bagh -MLR Model Beta Coefficient Table	29
Table 4.3 & 4.4: Punjabi Bagh NN Model Results	30
Table 5.1: R.K. Puram Air Pollution Predictive Model Results	30
Table 5.2: R.K. Puram -MLR Model Beta Coefficient Table	31
Table 5.3 & 5.4: R.K. Puram NN Model Results	31
Table 6. Comparative Performance of Models with Jackknife Validation	34
Table 7: Amount of Pollutant generated due to Bio-Mass burning	40

LIST OF FIGURES – CHARTS, IMAGES & INFOGRAPHICS	PAGE
Figure 1: Air Quality Index for Beijing and New Delhi for a 4 Month period	8
Figure 2: New Delhi's PM10 Levels over a 10-year period	8
Figure 3: Top 20 polluted cities in the G-20 Countries (PM10)	9
Figure 4: High Level Process Flow	12
Figure 5: 3 City Pollution Level Vs Vehicle Density & Vehicle Population	17
Figure 6: The three most polluted areas of New Delhi (Infographics)	18
Figure 7: Histogram chart showing various pollutant levels	19
Figure 8: Box plot - Key pollutant levels in the three locations	20
Figure 9: Summary data-Anand Vihar	21
Figure 10: Summary data-R.K. Puram	21

Figure 11: Summary data- Punjabi Bagh 21
Figure 12: Anand Vihar – Graph/Chart - Pollutant levels across seasons 22
Figure 13: R.K. Puram – Graph/Chart- Pollutant levels across seasons 22
Figure 14: Punjabi Bagh – Graph/Chart-Pollutant levels across seasons 23
Figure 15: Correlation Matrix for Anand Vihar 24
Figure 16: Correlation Matrix for Punjabi Bagh 25
Figure 17: Correlation Matrix for R.K. Puram 26
Figure 18: Anand Vihar-Comparative Model Fit graph for PM 2.5 32
Figure 19: Punjabi Bagh-Comparative Model Fit graph for PM 2.5 32
Figure 20: R.K. Puram-Comparative Model Fit graph for PM 2.5 32
Figure 21,22 & 23: Relative Importance Variables for the Three Locations 33
Figure 24: Average Pollutant Levels across 4 locations 35
Figure 25: Pollution Level Trend Analysis Graph -All Locations 36
Figure 26 (up) & 27(down): Graphs showing the PM 10 & PM 2.5 levels 37
Figure 28: Impact on Number of Cars on the Road 38
Figure 29: Picture showing the Bio Mass Burning across North India 38
Figure 30: Picture showing the impact of Bio-Mass burning 39
Figure 31: Data Pipeline for Mining Tweets 41
Figure32: Word Cloud for Tweets Collected 43
Figure33: Emotional Sentiment of Tweets over Time 44
Figure34: Sentiment Polarity of Tweets over Time 44
Figure 35: Chart showing Trigger Alarm and corrective action 46

TABLE OF EXHIBITS FOR APPENDIX

EXHIBIT DETAILS **PAGE**

EXHIBIT DETAILS	PAGE
Exhibit 1: R.K. Puram-Neural Network Model Fit Graph for PM2.5	50
Exhibit 2: Relative Importance of Meterological Factors	50
Exhibit 3 : Response & Explanatory Graph for PM2.5	51
Exhibit 4: Neural Network Model Fit Graph for PM10	51
Exhibit 5: Relative Importance of Meterological Factors for PM10	52
Exhibit 6 : Response & Explanatory Graph for PM 10	52
Exhibit 7: Neural Network Model Fit Graph for NO2	53
Exhibit 8: Relative Importance of Meterological Factors for NO2	53
Exhibit 9 : Response & Explanatory Graph for NO2	54
Exhibit 10: Neural Network Model Fit Graph for SO2	54
Exhibit 11: Relative Importance of Meterological Factors for SO2	55
Exhibit 12 : Response & Explanatory Graph for SO2	55
Exhibit 13: Neural Network Model Fit Graph for CO	56
Exhibit 14: Relative Importance of Meterological Factors for CO	56
Exhibit 15 : Response & Explanatory Graph for CO	57
Exhibit 16: Multiple Linear Regression Model Fit Graph for PM 2.5	57
Exhibit 17: Multiple Linear Regression Model Fit Graph for PM 10	58
Exhibit 18: Multiple Linear Regression Model Fit Graph for NO2	58
Exhibit 19: Multiple Linear Regression Model Fit Graph for SO2	59
Exhibit 20: Multiple Linear Regression Model Fit Graph for CO	59

LIST OF REFERENCES

1. Central Pollution Control Board Website Website: www.cpcb.nic.in
2. National Air Quality Index by CPCB
3. Centre for Science & Environment website: www.cseindia.org
4. The Energy and Resources Institute (TERI) website: www.teriin.org
5. International Research Journal of Earth Sciences, Review Paper "Emissions from Crop/Biomass Residue Burning Risk to Atmospheric Quality"
6. Atmosphere, review paper "A Study on the Use of a Statistical Analysis Model to Monitor Air Pollution Status in an Air Quality Total Quantity Control District", by Edward Ming-Yang Wu 1 and Shu-Lung Kuo 2.
7. "Emission of Air Pollutants from Crop Residue Burning in India" by Niveta Jain, Arti Bhatia, Himanshu Pathak, *Centre for Environment Science and Climate Resilient Agriculture, Indian Agricultural Research Institute, New Delhi-110012, India*
8. "Identifying pollution sources and predicting urban air quality using ensemble learning methods" by Kunwar P. Singh a,b, Shikha Gupta a,b, Premanjali Rai a,b. Atmospheric Environment Journal.