Introduction to Inference and Interpretation

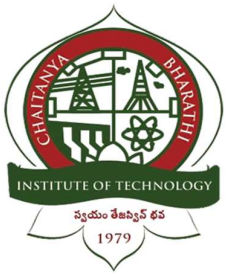
Project Report

On

Analysis of Hyderabad's Air Quality Index Data (2016-2022)

By

K.V.H. KASHYAP	-	160121729038
J. SAI RITHWIK	-	160121729037
PRANAV RAJ B.	-	160121729050
V. BHANU SAI	-	160121729061
M. SHYAM KUMAR	-	160121729066



Branch: **Artificial Intelligence and Machine Learning (J)**

BE (AI&ML) III Semester

Submitted to

Department of Computer Science and Engineering

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)

(Affiliated to Osmania University)

Gandipet, Hyderabad- 500075

2022 – 2023

Under the guidance of:

Dr. M. Swamy Das, Professor, Dept. of CSE

Smt. Ch. Vijaya Lakshmi, Asst. Professor, Dept. of CSE

INTRODUCTION

1.1 Motivation and Background

An air quality index (AQI) is used by government agencies to communicate to the public how polluted the air currently is or how polluted it is forecast to become. AQI information is obtained by averaging readings from an air quality sensor, which can increase due to vehicle traffic, forest fires, or anything that can increase air pollution. Pollutants tested include ozone, nitrogen dioxide, sulphur dioxide, among others.

With the increasing population in the country, there has been an increase in the number of migrations to the metropolitan cities, the commutators in these cities have increased which leads us to review the pollution in these cities.

The situation of Delhi has turned out to be alarming as the Air Quality Index of many places in Delhi have been marked “Unhealthy.” This might also be the plight of Hyderabad soon if no preventive measures are taken by the government immediately.

Hence, we need to analyse the trends of the Air Quality Index and also predict the future Air Quality. This would help us find the reasons for the higher observations and to control the pollution in those areas.

IMPLEMENTATION OF ANALYSIS OF HYDERABAD'S AIR QUALITY INDEX DATA (2016-2022) USING R LANGUAGE

2.1 Introduction to the Dataset[1]

The dataset has been collected from the official website of Telangana State Pollution Control Board. The yearly data was in the form of an Excel File. Each Excel File contained two sheets, one regarding the AQI of localities of Hyderabad and the other being the AQI of Localities in Telangana. Since the problem statement was restricted to Hyderabad, the data was filtered manually and the unwanted rows and columns were discarded. All the empty data fields were considered as NA.

2.2 R Code

```
setwd("C:/Users/WoW/Downloads/MonthlyAQI")


#install.packages("readxl")

library("readxl")

#install.packages('tidyverse')

library(tidyverse)

#install.packages('plotly')

library(plotly)

#install.packages("ggfortify")

library(ggfortify)

#install.packages("NbClust")

library(NbClust)

#install.packages("cluster")

library(cluster)

#install.packages("plotrix")
```

```
library(plotrix)

#install.packages("factoextra")

library(factoextra)


aqi2016<-read_excel("AQI2016.xlsx")

aqi2017<-read_excel("aqi2017.xlsx")

aqi2018<-read_excel("aqi2018.xlsx")

aqi2019<-read_excel("aqi2019.xlsx")

aqi2020<-read_excel("aqi2020.xlsx")

aqi2021<-read_excel("aqi2021.xlsx")

aqi2022<-read_excel("aqi2022.xlsx")


#Getting Area Name

area_name<- function(i) {

  a<-as.data.frame(aqi2016[i,1])

  area<-a$Location

  return(area)

}


#Creating a Vector of AQI of the Area

area_vector<- function(i) {

  aqi2016df<-as.data.frame(aqi2016[i,2:13])

  aqi2016vec<-
c(aqi2016df$Jan,aqi2016df$Feb,aqi2016df$Mar,aqi2016df$Apr,aqi2016df$May,aqi2
```

```
016df$Jun,aqi2016df$Jul,aqi2016df$Aug,aqi2016df$Sep,aqi2016df$Oct,aqi2016df$
Nov,aqi2016df$Dec)

  aqi2017df<-as.data.frame(aqi2017[i,2:13])

  aqi2017vec<-
c(aqi2017df$Jan,aqi2017df$Feb,aqi2017df$Mar,aqi2017df$Apr,aqi2017df$May,aqi2
017df$Jun,aqi2017df$Jul,aqi2017df$Aug,aqi2017df$Sep,aqi2017df$Oct,aqi2017df$
Nov,aqi2017df$Dec)

  aqi2018df<-as.data.frame(aqi2018[i,2:13])

  aqi2018vec<-
c(aqi2018df$Jan,aqi2018df$Feb,aqi2018df$Mar,aqi2018df$Apr,aqi2018df$May,aqi2
018df$Jun,aqi2018df$Jul,aqi2018df$Aug,aqi2018df$Sep,aqi2018df$Oct,aqi2018df$
Nov,aqi2018df$Dec)

  aqi2019df<-as.data.frame(aqi2019[i,2:13])

  aqi2019vec<-
c(aqi2019df$Jan,aqi2019df$Feb,aqi2019df$Mar,aqi2019df$Apr,aqi2019df$May,aqi2
019df$Jun,aqi2019df$Jul,aqi2019df$Aug,aqi2019df$Sep,aqi2019df$Oct,aqi2019df$
Nov,aqi2019df$Dec)

  aqi2020df<-as.data.frame(aqi2020[i,2:13])

  aqi2020vec<-
c(aqi2020df$Jan,aqi2020df$Feb,aqi2020df$Mar,aqi2020df$Apr,aqi2020df$May,aqi2
020df$Jun,aqi2020df$Jul,aqi2020df$Aug,aqi2020df$Sep,aqi2020df$Oct,aqi2020df$
Nov,aqi2020df$Dec)

  aqi2021df<-as.data.frame(aqi2021[i,2:13])

  aqi2021vec<-
c(aqi2021df$Jan,aqi2021df$Feb,aqi2021df$Mar,aqi2021df$Apr,aqi2021df$May,aqi2
021df$Jun,aqi2021df$Jul,aqi2021df$Aug,aqi2021df$Sep,aqi2021df$Oct,aqi2021df$
Nov,aqi2021df$Dec)

  aqi2022df<-as.data.frame(aqi2022[i,2:13])
```

```
 aqi2022vec<-
c(aqi2022df$Jan,aqi2022df$Feb,aqi2022df$Mar,aqi2022df$Apr,aqi2022df$May,aqi2
022df$Jun,aqi2022df$Jul,aqi2022df$Aug,aqi2022df$Sep,aqi2022df$Oct,aqi2022df$
Nov,aqi2022df$Dec)

 suppressWarnings(area_aqi<-
as.integer((c(aqi2016vec,aqi2017vec,aqi2018vec,aqi2019vec,aqi2020vec,aqi2021vec,
aqi2022vec))))

 return(area_aqi)

}


date<-seq(as.Date("2016-01-01"),as.Date("2022-12-01"),by="month")


df<-
data.frame(date,area_vector(1),area_vector(2),area_vector(3),area_vector(4),area_vec
tor(5),area_vector(6),area_vector(7),area_vector(8),area_vector(9),area_vector(10),ar
ea_vector(11),area_vector(12),area_vector(13),area_vector(14),area_vector(15),area_
vector(16),area_vector(17),area_vector(18),area_vector(19),area_vector(20),area_vect
or(21),area_vector(22),area_vector(23),area_vector(24),area_vector(25))

areanames<-
c("Balanagar","Uppal","JubileeHills","Paradise","Charminar","Jeedimetla","Abids","
KBRNPark","LangarHouse","Madhapur","MGBS","Chikkadapally","Kukatpally","N
acharam","Rajendranagar","Sainikpuri","BPPA","Shameerpet","HCU","Panjagutta","
Sanathnagar","ZooPark","Pashamylaram","Bollaram","ICRISAT")

colnames(df)<-c("Date",areanames)

summary(df)


df$MGBS[which(is.na(df$MGBS))]<-as.integer(mean(df$MGBS,na.rm=TRUE))

df$Nacharam[which(is.na(df$Nacharam))]<-
as.integer(mean(df$Nacharam,na.rm=TRUE))
```

```
df$Sainikpuri[which(is.na(df$Sainikpuri))]<-
as.integer(mean(df$Sainikpuri,na.rm=TRUE))

df$Shameerpet[which(is.na(df$Shameerpet))]<-
as.integer(mean(df$Shameerpet,na.rm=TRUE))

df$Panjagutta[which(is.na(df$Panjagutta))]<-
as.integer(mean(df$Panjagutta,na.rm=TRUE))

df$ZooPark[which(is.na(df$ZooPark))]<-
as.integer(mean(df$ZooPark,na.rm=TRUE))

df$Bollaram[which(is.na(df$Bollaram))]<-
as.integer(mean(df$Bollaram,na.rm=TRUE))

df$ICRISAT[which(is.na(df$ICRISAT))]<-
as.integer(mean(df$ICRISAT,na.rm=TRUE))

is.na(df)


#For Clustering, we need unlabelled data

mydata=select(df,c(2:26))


#Getting the transpose of the data for forming clusters

mydatat<-as.data.frame(t(mydata))

colnames(mydatat)<-c(date)


#Finding Optimum m=number of Clusters

#Method 1

#WSS Plot Function

wssplot <- function(data, nc=15, seed=1234){

  wss <- (nrow(data)-1)*sum(apply(data,2,var))
```

```
 for (i in 2:nc){
   set.seed(seed)
   wss[i] <- sum(kmeans(data, centers=i)$withinss)}
 plot(1:nc, wss, type="b", xlab="Number of Clusters",
      ylab="Within groups sum of squares")
 wss
}
wssplot(mydatat)


#Method 2
set.seed(1234)
nc <- NbClust(mydata, min.nc=2, max.nc=15, method="kmeans")



KM = kmeans(mydatat,3)
#KM1= kmeans(mydata,3)
fviz_cluster(KM,data=mydatat)
#autoplot(KM1,mydata,frame=TRUE)
KM$centers
KM$cluster


#Plots of Cluster 1


c1plot  <-  ggplot(data=df,mapping=aes(x=Date))+
```

```
        geom_line(mapping=aes(y=KBRNPark,color="KBRNPark"),linewidth=0.5)+
        geom_line(mapping=aes(y=Madhapur,color="Madhapur"),linewidth=0.5)+
        geom_line(mapping=aes(y=MGBS,color="MGBS"),linewidth=0.5)+

geom_line(mapping=aes(y=Chikkadapally,color="Chikkadapally"),linewidth=0.5)+

geom_line(mapping=aes(y=Rajendranagar,color="Rajendranagar"),linewidth=0.5)+
        geom_line(mapping=aes(y=Sainikpuri,color="Sainikpuri"),linewidth=0.5)+
        geom_line(mapping=aes(y=BPPA,color="BPPA"),linewidth=0.5)+
        geom_line(mapping=aes(y=Shameerpet,color="Shameerpet"),linewidth=0.5)+
        theme_bw()+
        labs(title=paste("Cluster 1 Plots"))+
        xlab("Date")+
        ylab('AQI')
ggplotly(c1plot)

#Plots of Cluster 2

c2plot  <-  ggplot(data=df,mapping=aes(x=Date))+
        geom_line(mapping=aes(y=Balanagar,color="Balanagar"),linewidth=0.5)+
        geom_line(mapping=aes(y=Uppal,color="Uppal"),linewidth=0.5)+
        geom_line(mapping=aes(y=JubileeHills,color="JubileeHills"),linewidth=0.5)+
        geom_line(mapping=aes(y=Paradise,color="Paradise"),linewidth=0.5)+
        geom_line(mapping=aes(y=Charminar,color="Charminar"),linewidth=0.5)+
        geom_line(mapping=aes(y=Jeedimetla,color="Jeedimetla"),linewidth=0.5)+
```

```
        geom_line(mapping=aes(y=Abids,color="Abids"),linewidth=0.5)+

geom_line(mapping=aes(y=LangarHouse,color="LangarHouse"),linewidth=0.5)+
        geom_line(mapping=aes(y=Kukatpally,color="Kukatpally"),linewidth=0.5)+
        geom_line(mapping=aes(y=Nacharam,color="Nacharam"),linewidth=0.5)+
        geom_line(mapping=aes(y=HCU,color="HCU"),linewidth=0.5)+
        geom_line(mapping=aes(y=Panjagutta,color="Panjagutta"),linewidth=0.5)+
        theme_bw()+
        labs(title=paste("Cluster 2 Plots"))+
        xlab("Date")+
        ylab('AQI')
ggplotly(c2plot)

#Plots of Cluster 3

c3plot  <-  ggplot(data=df,mapping=aes(x=Date))+

geom_line(mapping=aes(y=Sanathnagar,color="Sanathnagar"),linewidth=0.5)+
        geom_line(mapping=aes(y=ZooPark,color="ZooPark"),linewidth=0.5)+

geom_line(mapping=aes(y=Pashamylaram,color="Pashamylaram"),linewidth=0.5)+
        geom_line(mapping=aes(y=Bollaram,color="Bollaram"),linewidth=0.5)+
        geom_line(mapping=aes(y=ICRISAT,color="ICRISAT"),linewidth=0.5)+
        theme_bw()+
        labs(title=paste("Cluster 3 Plots"))+
```

```
        xlab("Date")+
        ylab('AQI')
ggplotly(c3plot)



#Pie Chart of Cluster 1
x1<-
c(mean(df$KBRNPark),mean(df$Madhapur),mean(df$MGBS),mean(df$Chikkadapal
ly),mean(df$Rajendranagar),mean(df$Sainikpuri),mean(df$BPPA),mean(df$Shameer
pet))
pct=round(x1/sum(x1)*100)
names1<-c("KBRNPark", "Madhapur", "MGBS",
"Chikkadapally","Rajendranagar","Sainikpuri","BPPA","Shameerpet")
label1<-paste(names1,"-",pct,"%",sep="")
col1<-c("#FFBBA9", "#FFFB74", "#9AFF4D",
"#A05FFF","#FF5067","#6139FF","#71FFE3","#0EFF0E")
pie3D(x1, labels = label1, main = "Cluster 1 AQI Average
Distribution",col=col1,explode=0.1)
legend("bottom", label1, fill = col1)


#Pie Chart of Cluster 2
x2<-
c(mean(df$Balanagar),mean(df$Uppal),mean(df$JubileeHills),mean(df$Paradise),me
an(df$Charminar),mean(df$Jeedimetla),mean(df$Kukatpally),mean(df$Panjagutta))
pct2=round(x2/sum(x2)*100)
names2<-c("Balanagar", "Uppal", "JubileeHills",
"Paradise","Charminar","Jeedimetla","Kukatpally","Panjagutta")
```

```
label2<-paste(names2,"-",pct2,"%",sep="")

col2<-c("#FF8BE4", "#FF567D", "#68FF8B", "##F7FF2A","#FF7437","#BBFF04","#A0FFA0","#5DC9FF")

pie3D(x2, labels = label2, main = "Cluster 2 AQI Average Distribution",col=col1,explode=0.1)

legend("bottom", label2, fill = col2)


#Pie Chart of Cluster 3

x3<-
c(mean(df$Sanathnagar),mean(df$ZooPark),mean(df$Pashamylaram),mean(df$Bollaram),mean(df$ICRISAT))

pct3=round(x3/sum(x3)*100)

names3<-c("Sanathnagar", "ZooPark", "Pashamylaram", "Bollaram","ICRISAT")

label3<-paste(names3,"-",pct3,"%",sep="")

col3<-c("#BEF7FF", "#FFA000", "#8BFF34", "#A05FFF","#FF5067")

pie3D(x3, labels = label3, main = "Cluster 3 AQI Average Distribution",col=col3,explode=0.1 )

legend("bottom", label3, fill = col3)


rowMeans(KM$centers)

mean(df$Chikkadapally)

mean(df$Charminar)

mean(df$Bollaram)


cl1_model<-lm(df$Chikkadapally~date)
```

```
predict(cl1_model,newdata=data.frame(date=seq(as.Date("2023-01-
01"),as.Date("2023-12-01"),by="month")),interval="confidence")

ggplot(df,aes(x=Date,y=Chikkadapally))+geom_point()+geom_smooth(method="lm"
)

cl2_model<-lm(df$Charminar~date)

predict(cl2_model,newdata=data.frame(date=seq(as.Date("2023-01-
01"),as.Date("2023-12-01"),by="month")),interval="confidence")

ggplot(df,aes(x=Date,y=Charminar))+geom_point()+geom_smooth(method="lm")

cl3_model<-lm(df$Bollaram~date)

predict(cl3_model,newdata=data.frame(date=seq(as.Date("2023-01-
01"),as.Date("2023-12-01"),by="month")),interval="confidence")

ggplot(df,aes(x=Date,y=Bollaram))+geom_point()+geom_smooth(method="lm")

aqi<-cbind(aqi2016$Avg,aqi2017$Avg)

aqi<-cbind(aqi,aqi2018$Avg)

aqi<-cbind(aqi,aqi2019$Avg)

aqi<-cbind(aqi,aqi2020$Avg)

aqi<-cbind(aqi,aqi2021$Avg)

aqi<-cbind(aqi,aqi2022$Avg)

aqi



noRows<-25

for(i in 1:noRows){

  if(sum(is.na(aqi[i,]))){

    aqi[i,][is.na(aqi[i,])] <- mean(aqi[i,], na.rm = TRUE)
```

```
  }
}
aqi<-rowMeans(aqi)
aqi
place <-c(aqi2016[,1])
df2<-data.frame(
  Location<-place,
  aqi_overall_avg <-aqi
)

df2

aqi_range<-c(ifelse(aqi_overall_avg>400,"Severe(>400)",
                    ifelse(aqi_overall_avg>300,"Very poor(301-400)",
                           ifelse(aqi_overall_avg>200,"Poor(201-300)",
                                  ifelse(aqi_overall_avg>100,"Moderate(101-200)",
                                         ifelse(aqi_overall_avg>50,"Satisfactory(51-100)","Good(0-
50)")))))

ggplot(df2,aes(x=Location,y=aqi_overall_avg,fill=aqi_range))+
  labs(title = "Overall Mean AQI for Different Places in Hyderabad",face="bold")+
  geom_bar(stat = "identity")+
  scale_fill_manual(values=c('#F9FF50','#25D84A'))+
  theme_bw()+
  coord_flip()
```

2.3 Demonstration

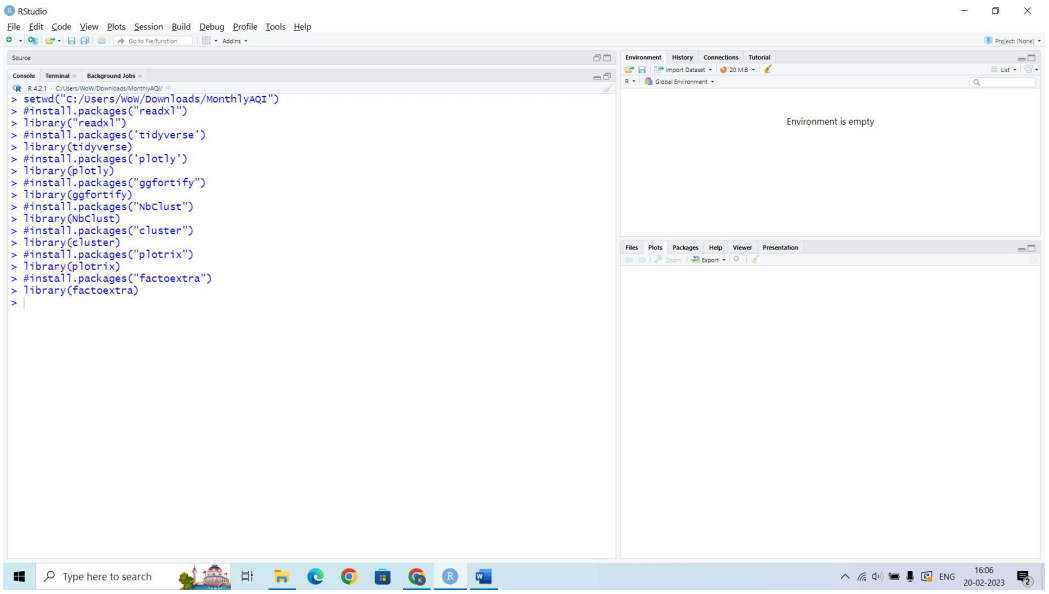


Fig 2.1: Screenshot of Setting the Directory and Installing, Loading Packages

First, the Working Directory was set to the required location. Then, the packages READXL, TIDYVERSE, PLOTLY, GGFORTIFY, NBCLUST, CLUSTER, PLOTRIX, FACTOEXTRA were loaded to the RStudio Environment. Plotly for Interactive Charts, Readxl for loading the excel sheets; Tidyverse, a collection all R packages for data science, GGPlot for drawing the plots. Nbclust was used for finding the optimum number of clusters, Plotrix for 3d Pie Plots and Factoextra for the plotting the clusterplot.

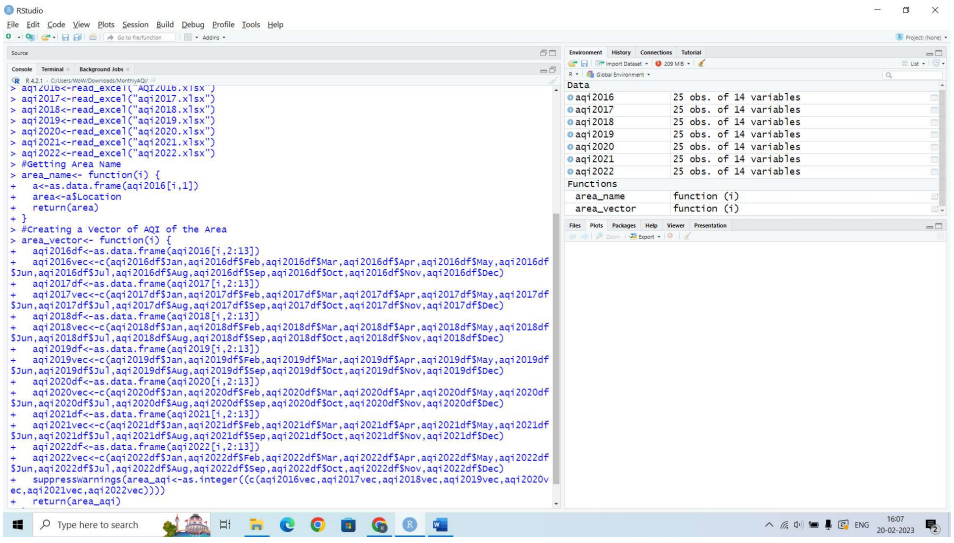


Fig 2.2: Reading the Excel Files and extracting required values to Vectors

On loading all the excel files are in the form of tibbles. Tibbles are converted to dataframe by using as.data.frame(). All the monthly AQI of each area can be converted into a vector by using the '$.' Instead of loading all the area datafields separately, only the required Area AQI values are converted by creation of a Function called area_vector(). The area vector accepts an integer value as input and provides the data of the area which is present in that row of the Excel Files. The area_name() function returns the name of the area present in that row of the Excel File.

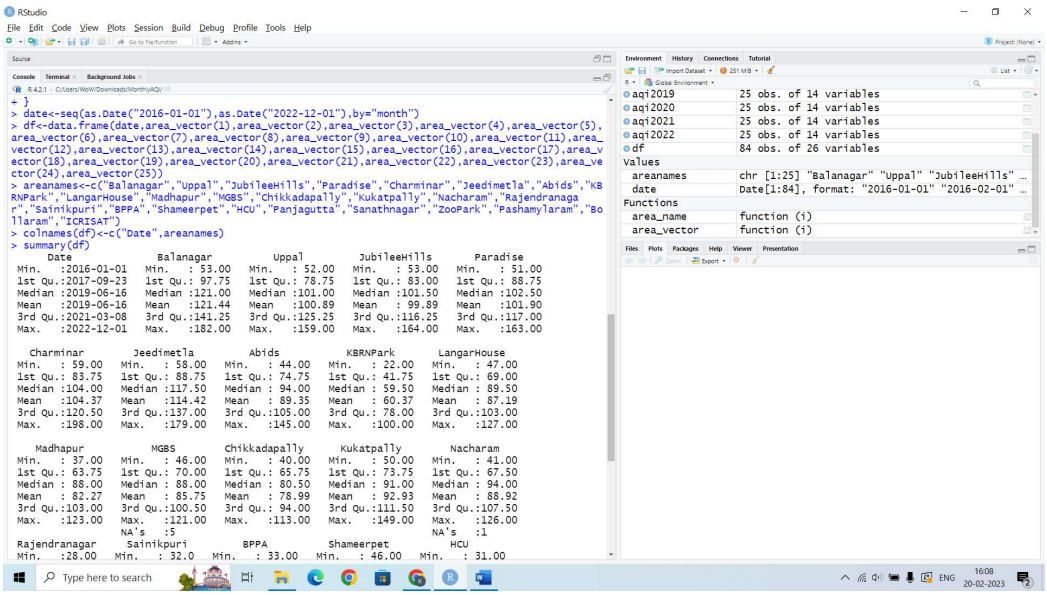


Fig 2.3: Screenshot of Function for creating a dataframe with all the areas AQI Values

All the data related to the area wise AQI are loaded into a dataframe named ‘df’. The column names of the dataframe are set using thecolnames() function. The summary function returns the summary of the dataframe about the number of missing values, first quartile, median, mean and third quartile of the data and the minimum and maximum values.

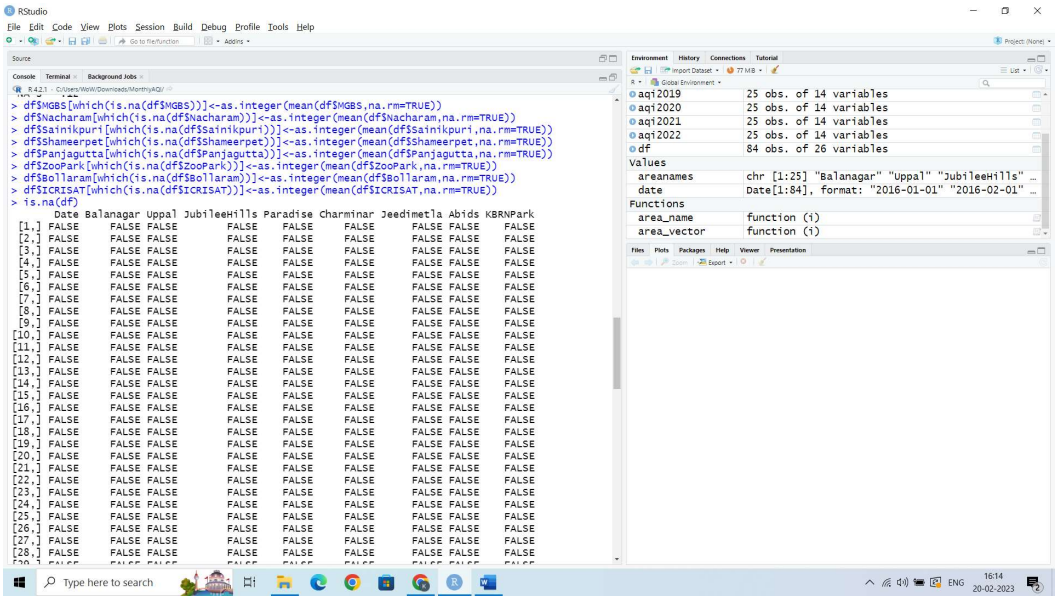


Fig 2.4: Screenshot of the console during the handling of Missing Values

The missing values in each area are replaced with the mean of AQI Values of each area. After replacing, it is checked if there are still any missing values using the is.na() function.

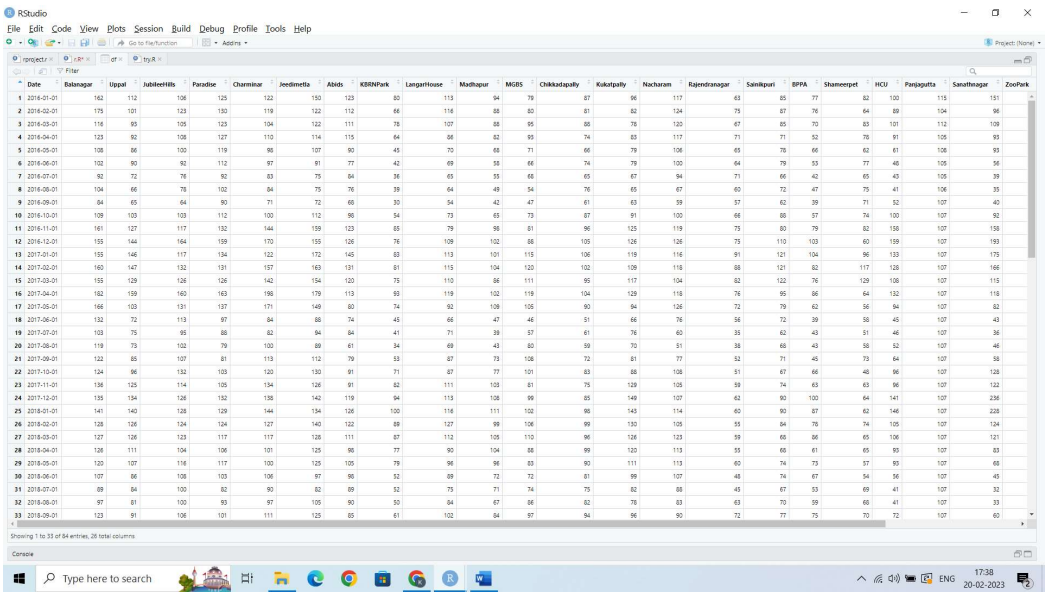


Fig 2.5: Screenshot of the Dataframe created

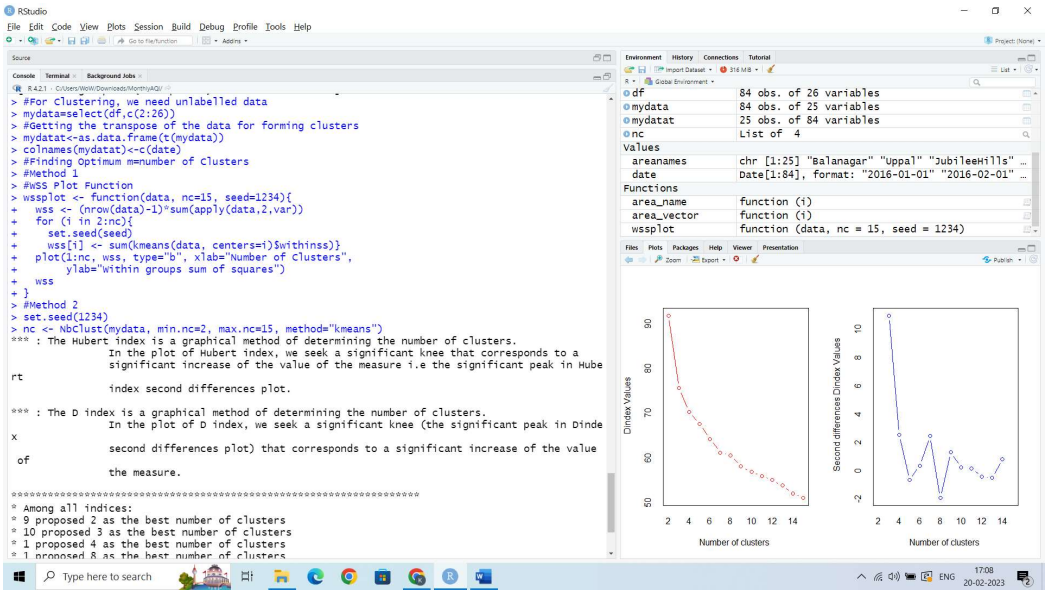


Fig 2.6: Screenshot of Function for finding the Optimum number of Clusters

The WSS Plot is used for checking the optimum number of clusters. The Nbclust function is also used for the same reason. The output was found to be 3 as a kink or an elbow shape appeared when the number of clusters was 3.

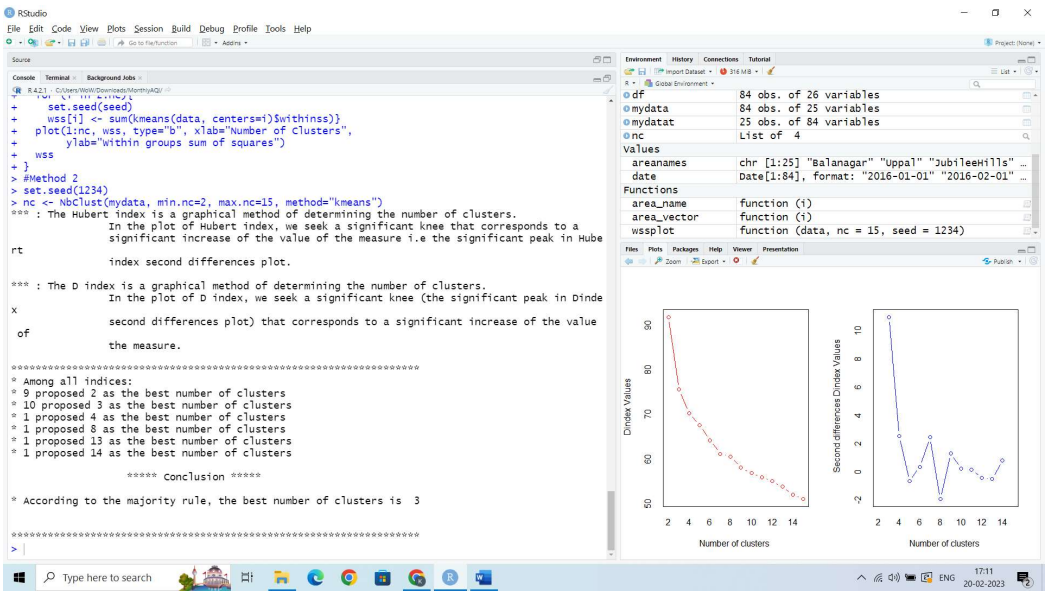


Fig 2.7: Screenshot of Output of NBCLUST Function

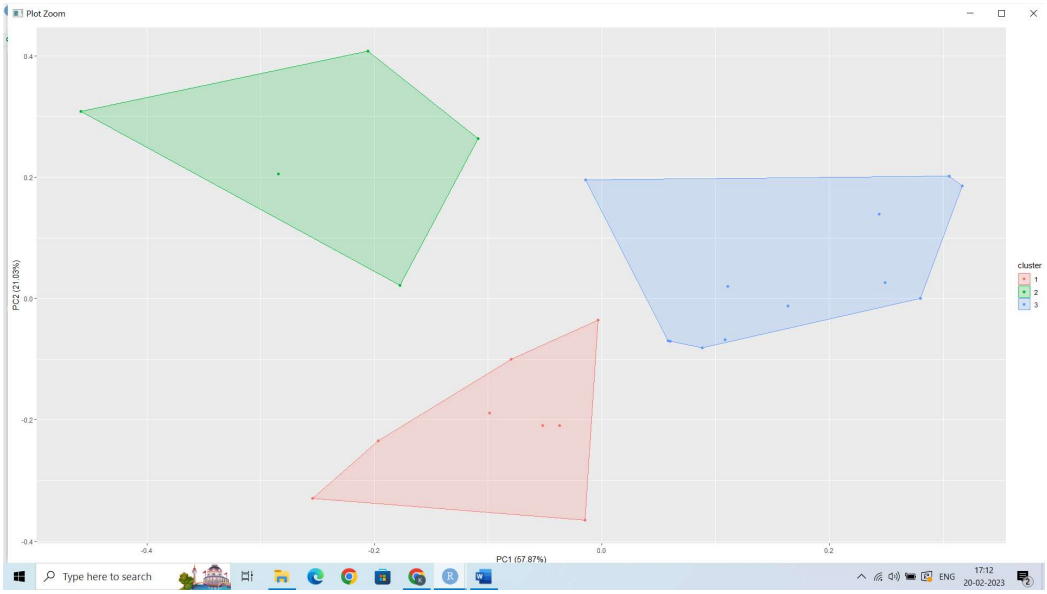


Fig 2.8: Screenshot of Representation of Three Clusters

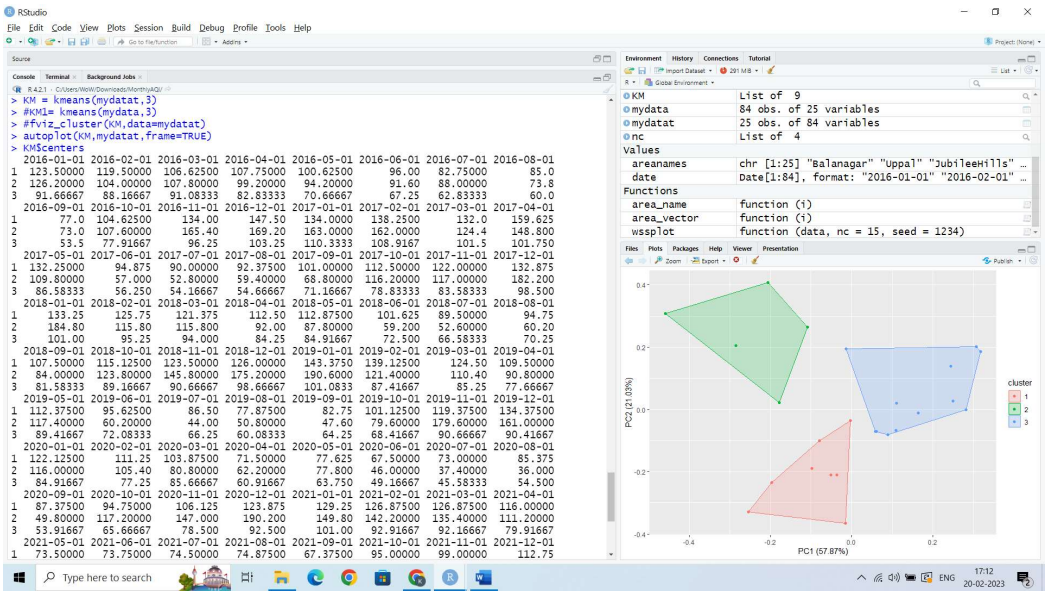


Fig 2.9: Screenshot of K Mean Cluster Centers

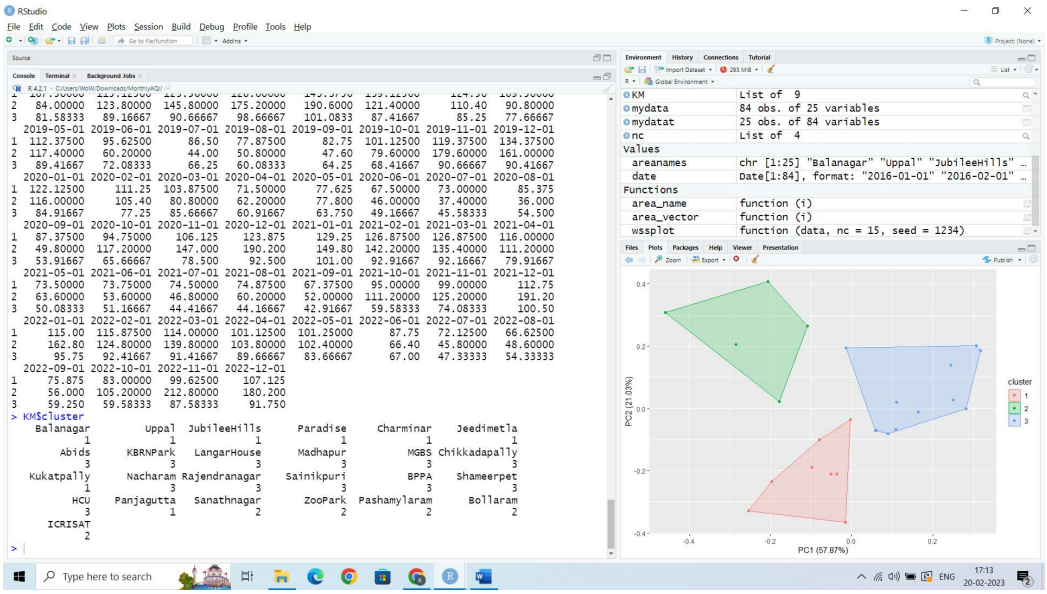


Fig 2.10: Screenshot of Division of areas into Clusters

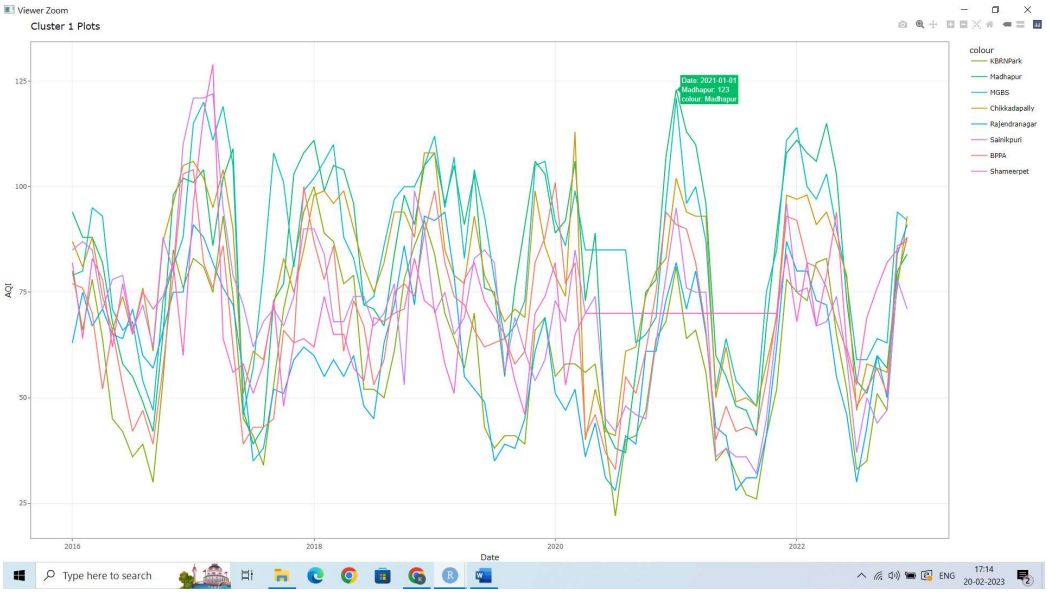


Fig 2.10: Screenshot of Interactive Plots of Areas under Cluster 1

The AQI Values v/s Date were plotted for all the areas that fall under the cluster 1. Similarly for cluster 2 and cluster 3, the plots are drawn. With these plots, it can be understood that the areas are rightly categorized into the particular clusters as the trends of these areas over years was almost similar.

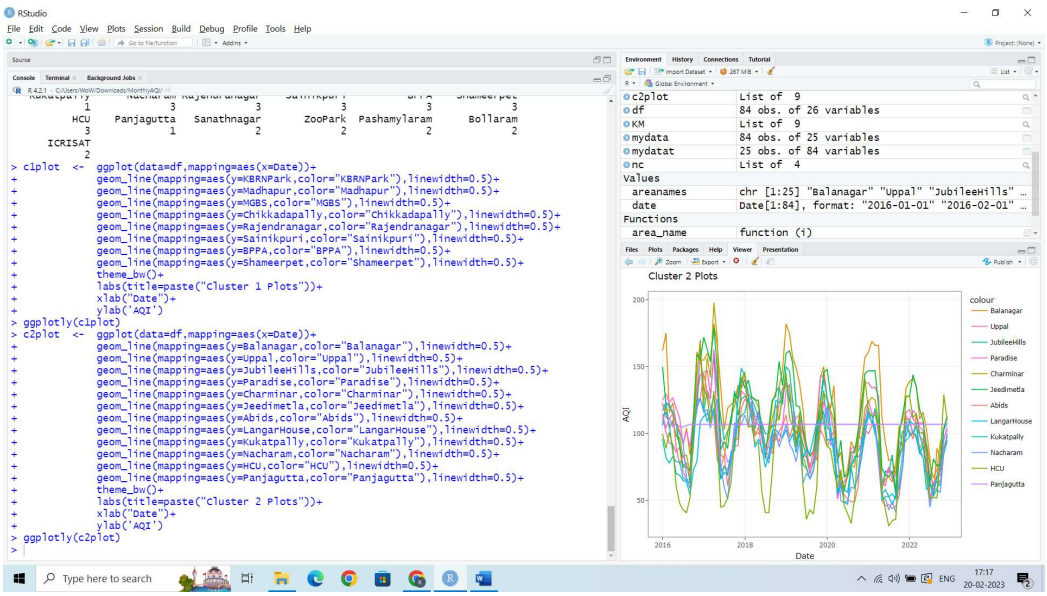


Fig 2.10: Screenshot of Console while plotting the AQI Plots of each cluster

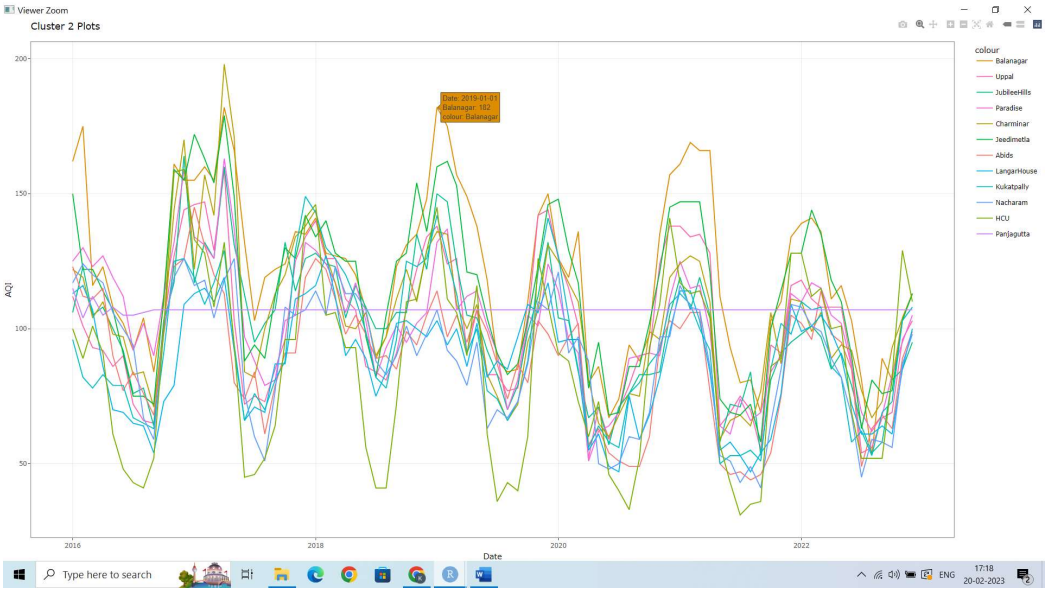


Fig 2.11: Screenshot of the Interactive Plots of Areas under Cluster 2

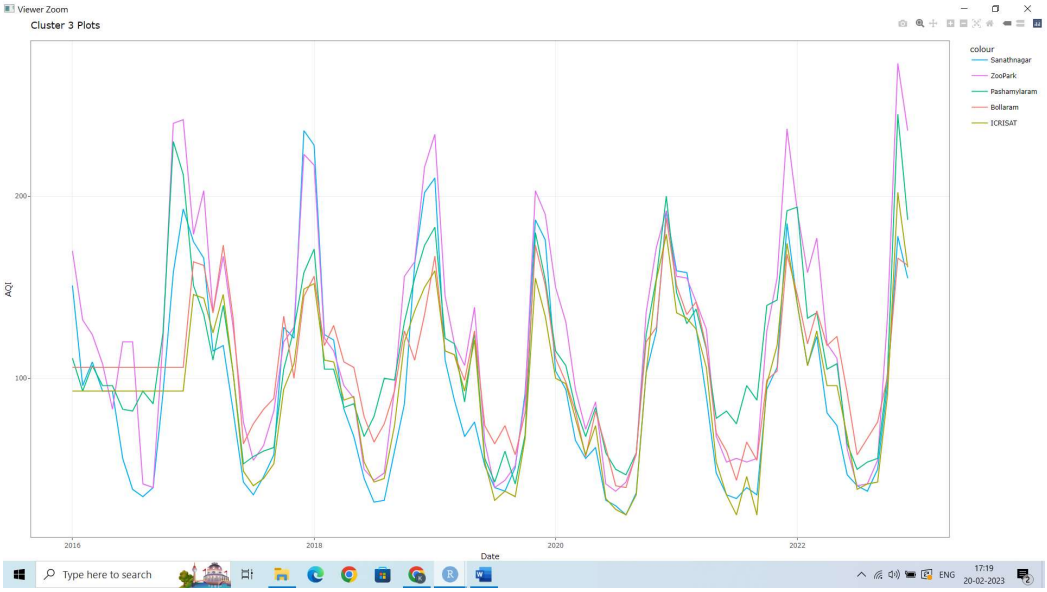


Fig 2.12: Screenshot of the Interactive Plots of Areas under Cluster 3

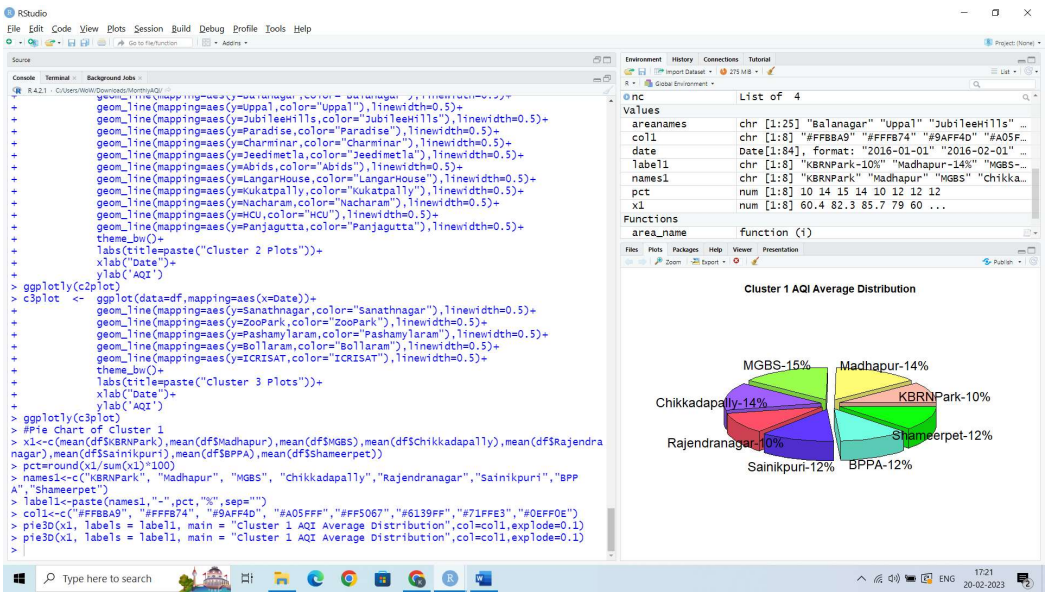
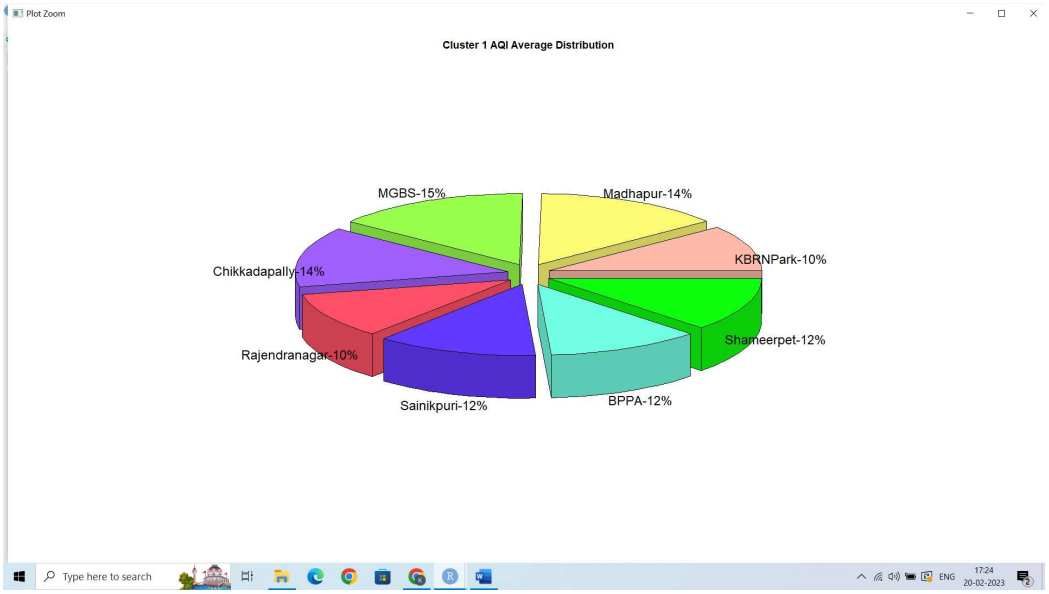


Fig 2.13: Screenshot of the Console while plotting the 3D Pie graphs

After plotting the Interactive AQI Values of the clusters, the distribution of each cluster were checked by plotting 3D Pie Plots.



. **Fig 2.14: Screenshot of the Cluster 1 AQI Average Distribution in 3D Pie graphs**

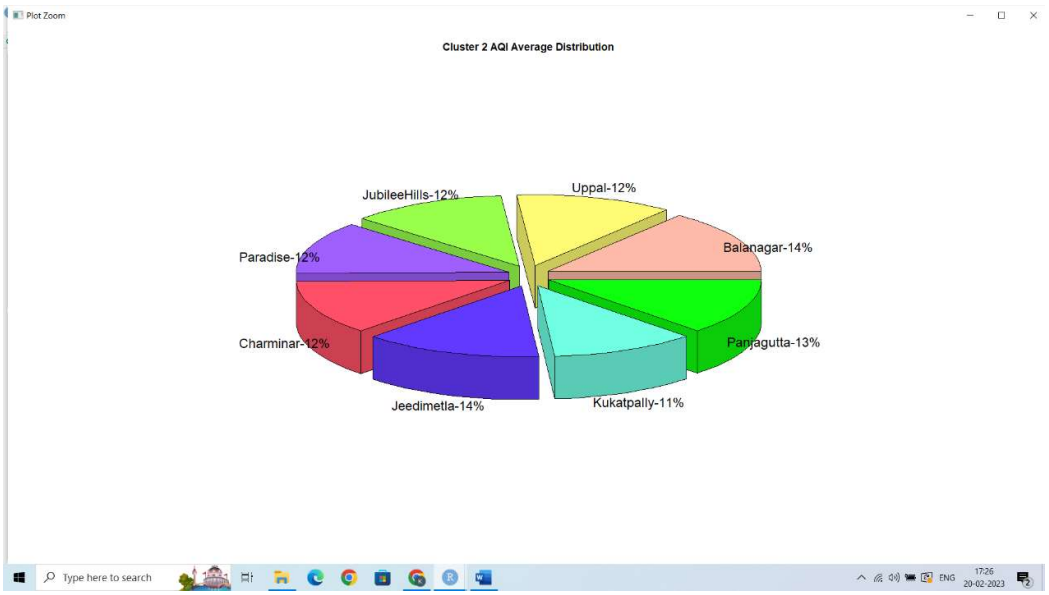


Fig 2.15: Screenshot of the Cluster 2 AQI Average Distribution in 3D Pie graphs

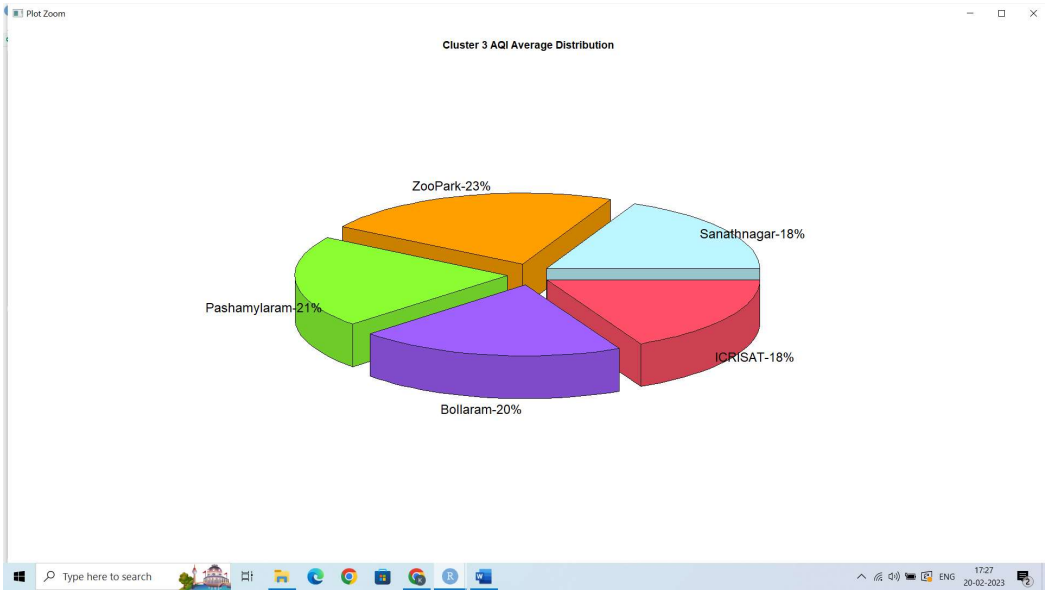


Fig 2.16: Screenshot of the Cluster 3 AQI Average Distribution in 3D Pie graphs

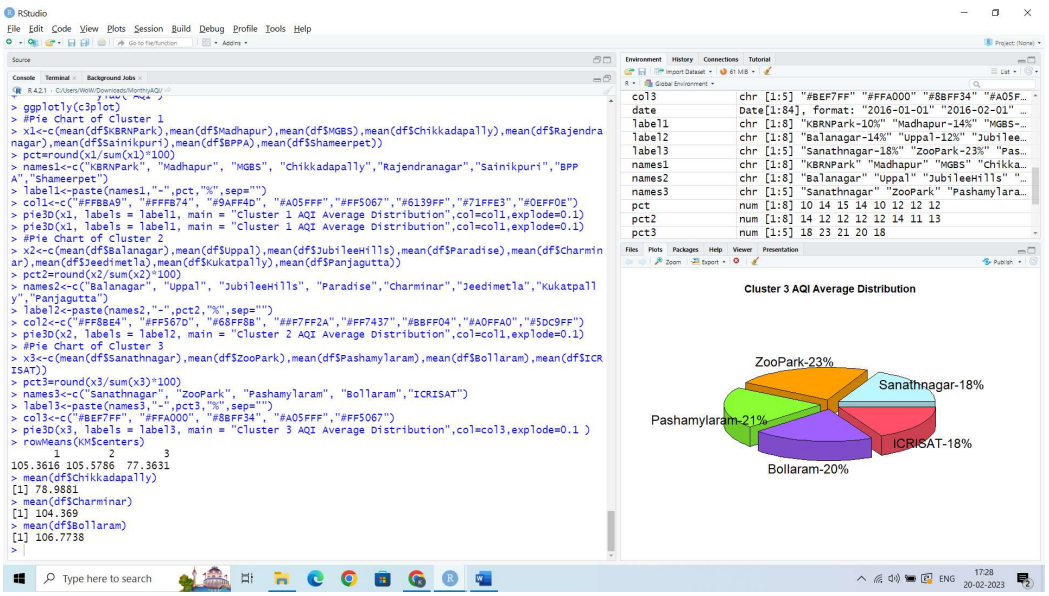


Fig 2.17: Screenshot of the after plotting the Cluster wise AQI Average Distribution in 3D Pie graphs

After plotting the 3D Pie Plots, the mean of each row of the cluster was calculated and on checking the nearest value of the areawise mean of AQI, the nearest value was chosen and it represents all the areas in that category.

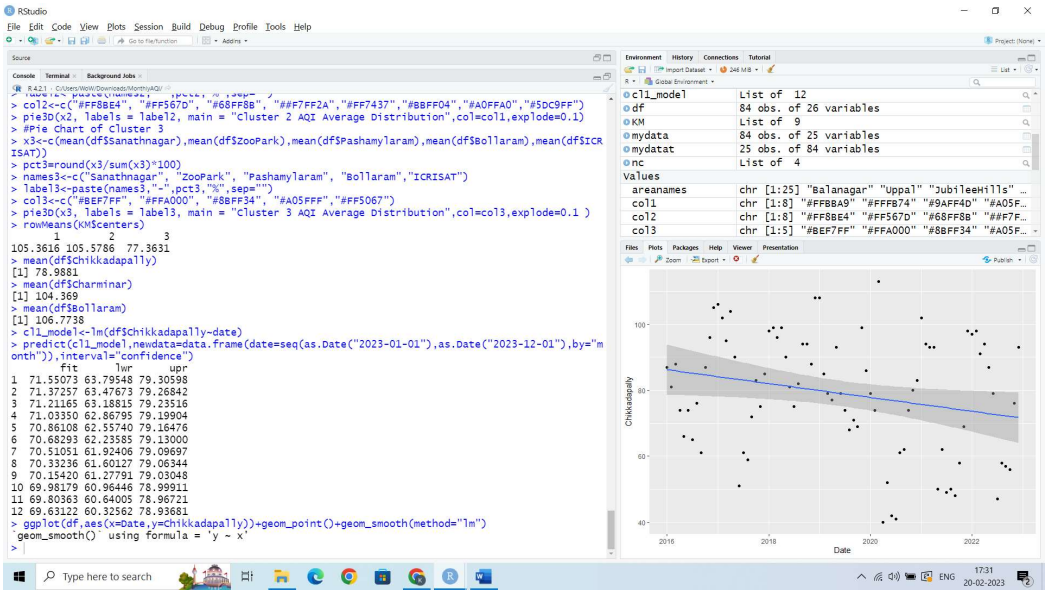


Fig 2.18: Screenshot of the Linear Model and Predicted Values and the Linear Model Plot

In the first Cluster, Chikkadapally was used to represent all the areas under the first cluster category. The linear model was created using the AQI Values and corresponding dates. Since, Linear Regression was beginner friendly and simple to use, a linear model was chosen to predict the future AQI values of the places in this area.

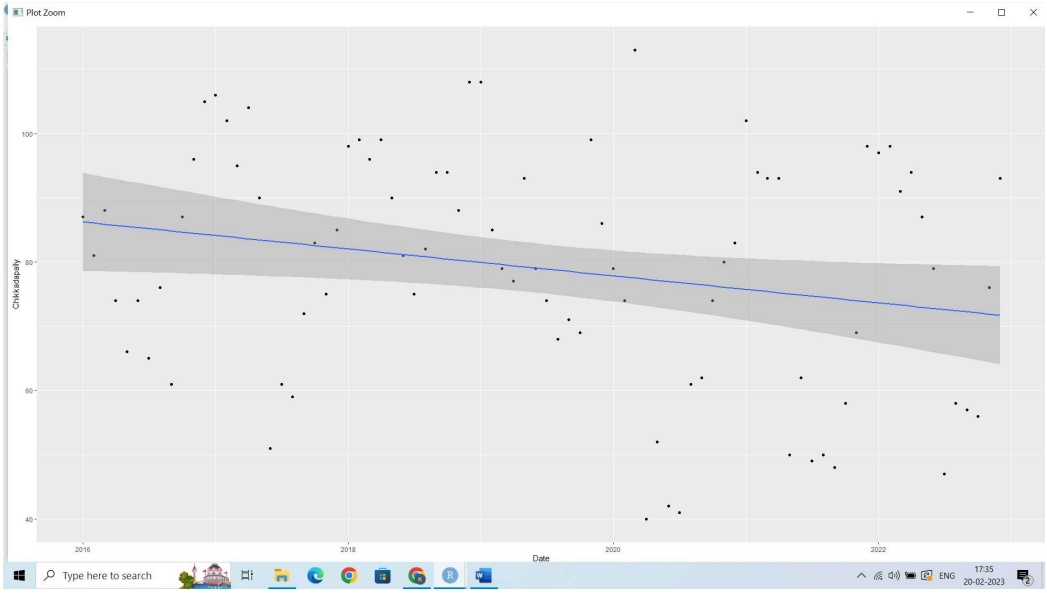


Fig 2.19: Screenshot of the Linear Model Plot of Chikkadapally Area Representing the areas in Cluster 1

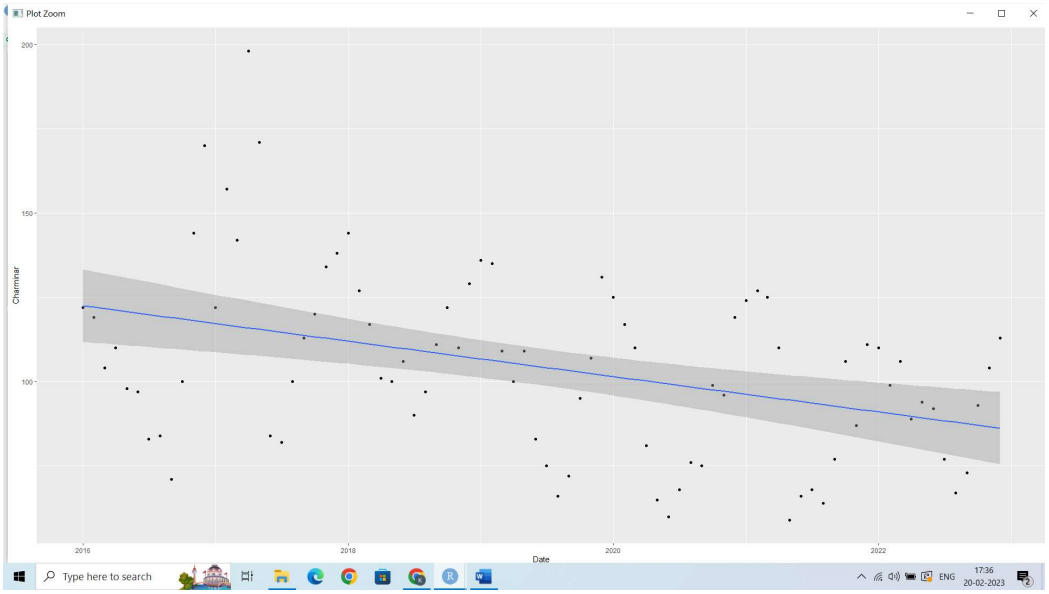


Fig 2.20: Screenshot of the Linear Model Plot of Charminar Area Representing the areas in Cluster 2

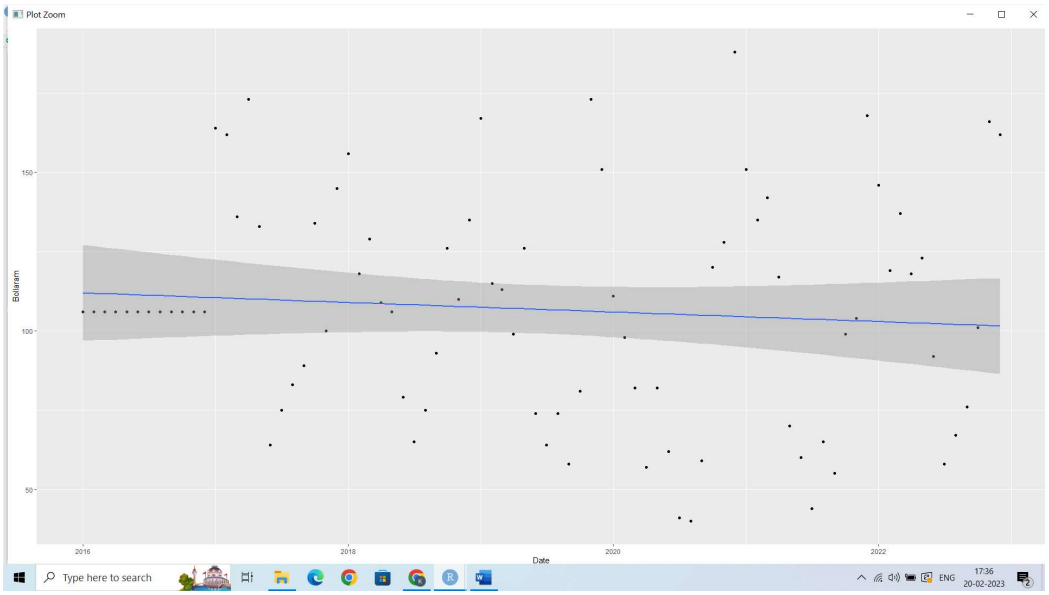


Fig 2.21: Screenshot of the Linear Model Plot of Bollaram Area Representing Cluster 3

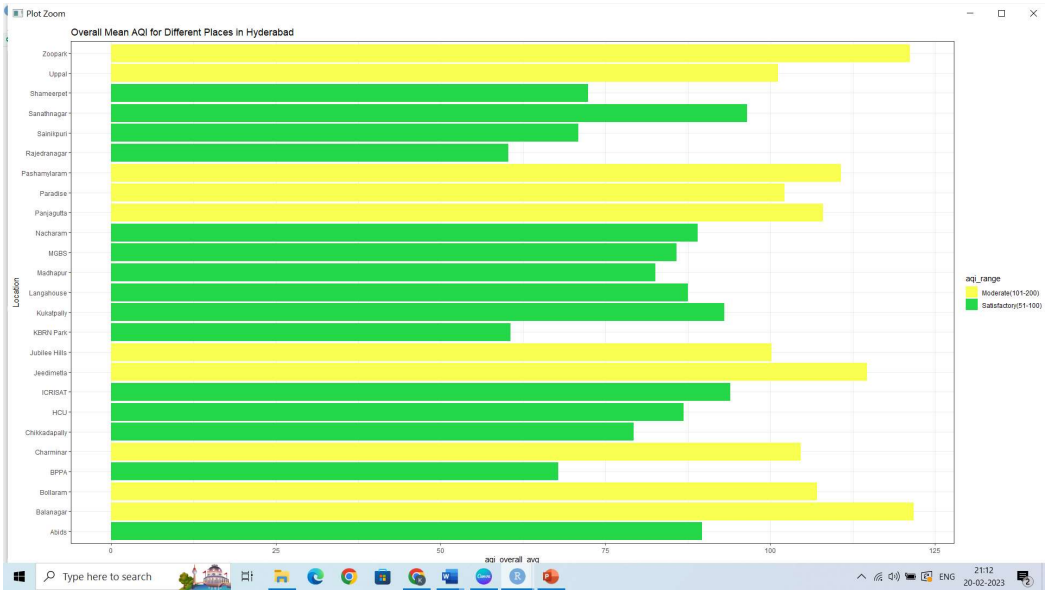


Fig 2.22: Screenshot of the Overall Mean AQI for Different Places in Hyderabad

CONCLUSION

The entire project has been developed to Analyse the Trends in the Air Quality Index of Hyderabad from 2016-2022. From the individual graphs, the values of all the areas during the year 2020 has shown a great fall in the Air Quality Index Values. This might be due to the prevalence Covid-19 and the lockdown imposed in the city. After 2020, during 2021 and 2022, the values are increasing as the lockdowns have been removed. But the values are seemingly less when compared to the data in the years 2016-2019.

Individual plots are drawn for the Linear Models which are generated for each area. By comparing the graphs, the areas Pashamylaram and Shameerpet have an increasing Linear Model which means the Air Quality Index Values would increase and the Air Quality in these areas would be decreasing. These areas are majorly industrial areas where the core reason for the pollution would be the emissions from the industries. Shameerpet has many pharmaceutical industries and the same was the case with Pashamylaram. Government must take actions regarding this, such as controlling the emissions from the industries, checking the Air Quality near these industries. Like the Effluent Treatment Plants, Industrial Air Treatment Plants must be setup in these areas.

Also in the remaining areas, measures must be taken to continue the Lower AQI readings. This could be done by providing awareness among people on the public transportation systems. If the commutators in the city travel by the means of Public Transportation instead of their using their own vehicles, the number of vehicles on road would reduce. Also, the Public Transportation System is to be able to capacitate the needs of all the civilians. Awareness must be provided to people that during the traffic signals, continuous ignition would release smokes from the vehicles which affects the Air Quality. Usage of Electric Vehicles must be encouraged by the government so that there would be an improvement in the Air Quality in the public areas.

Since, many missing values were found in the data, building efficient Linear Models was not possible. The Pollution Control Board must try to get all the readings effectively, so that Linear Models could be generated in all the areas and the graphs can be compared.

FUTURE SCOPE

In future, we can update this project by using efficient Regression Models for the comparison of Time Series Data. Furthermore, in the case of missing data fields, the linear model can be generated by just using the available values. These would also help in the comparison of area-wise Air Quality Data.

REFERENCES

1. Telangana State Pollution Control Board for the Data