CHAITANYA BHARATHI
INSTITUTE OF TECHNOLOGY (A)
Kokapet (Village), Gandipet, Hyderabad, Telangana-500075. www.cbit.ac.in

Recognized Research Centers | Programs Accredited by NBA | Approved by | Accredited by NAAC A - Grade | All India 133rd Rank in nirf | ISO Certified 9001:2015

Analysis of Hyderabad'S AQI Data (2016-2022)

SUBMITTED TO:

Dr. M. Swamy Das,
Professor, Dept. of CSE
Smt. Ch. Vijaya Lakshmi,
Asst. Professor, Dept. of CSE

SUBMITTED BY:

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

Table of Contents

Problem statement:

- To perform an in-depth analysis of the Air Quality Index (AQI) in Hyderabad, India using the R programming language.
- The objective of the analysis is to understand the AQI trend in Hyderabad over the past few years and identify any patterns in AQI.
- This analysis will provide valuable insights into the current state of air quality in Hyderabad and help stakeholders take informed decisions to improve the air quality in the city.

Description:

- Air Quality Index (AQI) is a numerical value used to represent the overall quality of air in a specific location.
- AQI is calculated based on various air pollutants, such as nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3), among others.
- In this analysis, we will use the R programming language to explore the AQI data of Hyderabad over a period of time from 2016-2022.

Dataset Description:

- The Dataset was collected from the Official Website of Telangana Pollution Control Board.
- The initial data is in the form of 7 Excel Sheets each having the area name and the AQI in that area. The 2016 AQI Sheet is shown here.
- The dimensions of each table is 25 rows and 14 columns.

Location	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Avg
Balanagar	163	176	116	123	108	102	93	104	85	110	162	155	125
Uppal	112	102	94	92	86	91	73	67	66	104	127	144	96
Jubilee Hills	106	123	106	108	101	93	76	78	65	104	118	165	103
Paradise	126	130	123	127	119	113	92	102	90	112	133	160	119
Charminar	123	119	105	110	98	98	83	84	71	101	145	170	109
Jeedimetla	151	123	122	114	107	92	75	76	73	112	159	155	113
Abids	123	112	111	115	90	77	84	76	68	98	123	126	100
KBRN Park	80	66	78	64	45	42	36	39	30	54	85	76	58
Langahouse	113	116	107	86	70	69	65	64	54	73	79	109	84
Madhapur	94	88	88	82	68	58	55	49	42	65	98	102	74
MGBS	79	80	95	93	71	66	68	54	47	73	81	88	75
Chikkadapally	87	81	88	74	66	74	65	76	61	87	96	105	80
Kukatpally	96	82	78	83	79	79	67	65	63	91	125	126	86
Nacharam	117	124	120	117	106	100	94	67	59	100	119	126	104
Rajedranagar	63	75	67	71	65	64	71	60	57	66	75	75	67
Sainikpuri	85	87	85	71	78	79	66	72	62	88	80	110	80
BPPA	77	76	70	52	66	53	42	47	39	57	79	103	63
Shameerpet	82	64	83	78	62	77	65	75	71	74	82	60	73
HCU	101	90	101	91	62	48	43	42	52	100	159	159	87
Panjagutta	115	104	113	105	109	105	106	107					108
Sanathnagar	152	96	110	93	93	57	40	35	41	93	158	193	97
Zoopark	170	133	125	109	84			42	40	122	241	242	131
Pashamylaram	111	93	107	96	96	83	82	94	86	125	230	212	118
Bollaram													
ICRISAT													

AQI Colour Index & Health Effects:	
GOOD (0 – 50)	Minimal Impact
SATISFACTORY (51– 100)	Minor Breathing Discomfort to Sensitive People
MODERATE (101 – 200)	Breathing discomfort to People with Lung & Heart Disease, Children and Old adults
POOR (201 – 300)	Breathing discomfort to People on Prolonged Exposure
VERYPOOR (301 – 400)	Respiratory illness to People on Prolonged Exposure
SEVERE > 400	Respiratory Effects on Healthy people

Tools used:

R

R is used for analyzing the data as it has:

- Static graphics that produce good-quality data visualizations.
- Comprehensive library that provides interactive graphics and makes data visualization and representation easy to analyze.

Useful Libraries:

- Plotly for Interactive Charts
- Readxl for loading the excel sheets
- Tidyverse, a collection all R packages for data science
- GGPlot for drawing the plots
- Plotrix, for building complex Visualisations
- Nbclust, for determining the number of clusters

Methodology

1. Working Directory was set.
2. Data was loaded into the RStudio environment.
3. Packages were installed and libraries were loaded.
4. Data was refined, Missing Values were handled and data frame was created.
5. Optimum number of Clusters was evaluated and Clusters were created.
6. Validity of each Area being divided into the clusters was checked by comparing the plots.
7. The Distribution of Mean AQI of Areas in each Cluster was visualized using 3D Pie Plots.
8. From each Cluster, one location was selected to represent the whole cluster and linear model was created and future values were predicted.

CODE

```r
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$Mar,aqi2016
df$Apr,aqi2016df$May,aqi2016df$Jun,aqi2016df$Jul,aqi
2016df$Aug,aqi2016df$Sep,aqi2016df$Oct,aqi2016df$N
ov,aqi2016df$Dec)
  aqi2017df<-as.data.frame(aqi2017[i,2:13])
  aqi2017vec<-
c(aqi2017df$Jan,aqi2017df$Feb,aqi2017df$Mar,aqi2017
df$Apr,aqi2017df$May,aqi2017df$Jun,aqi2017df$Jul,aqi
2017df$Aug,aqi2017df$Sep,aqi2017df$Oct,aqi2017df$No
v,aqi2017df$Dec)
aqi2018df<-as.data.frame(aqi2018[i,2:13])
aqi2018vec<-
c(aqi2018df$Jan,aqi2018df$Feb,aqi2018df$Mar,aqi2018df$Apr,aqi2018d
f$May,aqi2018df$Jun,aqi2018df$Jul,aqi2018df$Aug,aqi2018df$Sep,aqi2
018df$Oct,aqi2018df$Nov,aqi2018df$Dec)
  aqi2019df<-as.data.frame(aqi2019[i,2:13])
aqi2019vec<-
c(aqi2019df$Jan,aqi2019df$Feb,aqi2019df$Mar,aqi2019df$Apr,aqi2019d
f$May,aqi2019df$Jun,aqi2019df$Jul,aqi2019df$Aug,aqi2019df$Sep,aqi2
019df$Oct,aqi2019df$Nov,aqi2019df$Dec)
  aqi2020df<-as.data.frame(aqi2020[i,2:13])
 aqi2020vec<-
c(aqi2020df$Jan,aqi2020df$Feb,aqi2020df$Mar,aqi2020df$Apr,aqi2020df$May,
aqi2020df$Jun,aqi2020df$Jul,aqi2020df$Aug,aqi2020df$Sep,aqi2020df$Oct,aqi
2020df$Nov,aqi2020df$Dec)
  aqi2021df<-as.data.frame(aqi2021[i,2:13])
```

```
 aqi2021vec<-
c(aqi2021df$Jan,aqi2021df$Feb,aqi2021df$Mar,aqi2021df$Apr,aqi20
21df$May,aqi2021df$Jun,aqi2021df$Jul,aqi2021df$Aug,aqi2021df$Se
p,aqi2021df$Oct,aqi2021df$Nov,aqi2021df$Dec)
  aqi2022df<-as.data.frame(aqi2022[i,2:13])
  aqi2022vec<-
c(aqi2022df$Jan,aqi2022df$Feb,aqi2022df$Mar,aqi2022df$Apr,aqi
2022df$May,aqi2022df$Jun,aqi2022df$Jul,aqi2022df$Aug,aqi2022d
f$Sep,aqi2022df$Oct,aqi2022df$Nov,aqi2022df$Dec)
  suppressWarnings(area_aqi<-
as.integer(na.omit(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),ar
ea_vector(4),area_vector(5),area_vector(6),area_vector(7),area
_vector(8),area_vector(9),area_vector(10),area_vector(11),area_v
ector(12),area_vector(13),area_vector(14),area_vector(15),area_ve
ctor(16),area_vector(17),area_vector(18),area_vector(19),area_vec
tor(20),area_vector(21),area_vector(22),area_vector(23),area_ve
ctor(24),area_vector(25))
areanames<-
c("Balanagar","Uppal","JubileeHills","Paradise","Charminar","Jeedimetl
a","Abids","KBRNPark","LangarHouse","Madhapur","MGBS","Chikkadapal
ly","Kukatpally","Nacharam","Rajendranagar","Sainikpuri","BPPA","Sham
eerpet","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)
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)
#fviz_cluster(KM,data=mydatat)
autoplot(KM,mydatat,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"),linewid
th=0.5)+geom_line(mapping=aes(y=MGBS,color="MGBS"),linewidt
h=0.5)+
geom_line(mapping=aes(y=Chikkadapally,color="Chikkadapally"),
linewidth=0.5)+
geom_line(mapping=aes(y=Rajendranagar,color="Rajendranaga
r"),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"),lin
ewidth=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"),linewid

th=0.5)+

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

geom_line(mapping=aes(y=JubileeHills,color="JubileeHills"),linewi

dth=0.5)+

geom_line(mapping=aes(y=Paradise,color="Paradise"),linewidth=

0.5)+

geom_line(mapping=aes(y=Charminar,color="Charminar"),linewi

dth=0.5)+

geom_line(mapping=aes(y=Jeedimetla,color="Jeedimetla"),linewi

dth=0.5)+

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

geom_line(mapping=aes(y=LangarHouse,color="LangarHouse"),li

newidth=0.5)+

geom_line(mapping=aes(y=Kukatpally,color="Kukatpally"),linewidt

h=0.5)+

geom_line(mapping=aes(y=Nacharam,color="Nacharam"),linewi

dth=0.5)+

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

geom_line(mapping=aes(y=Panjagutta,color="Panjagutta"),linewi

dth=0.5)+ theme_bw()+  labs(title=paste("Cluster 1

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$MGBS),mean(df$Chikkadapally),mean(df$Rajendranagar),mean(df$Sainikpuri),mean(df$BPPA),mean(df$Shameerpet))

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$JubileeHills),me
an(df$Paradise),mean(df$Charminar),mean(df$Jeedimetla),mea
n(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$Pashamyl
aram),mean(df$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
```

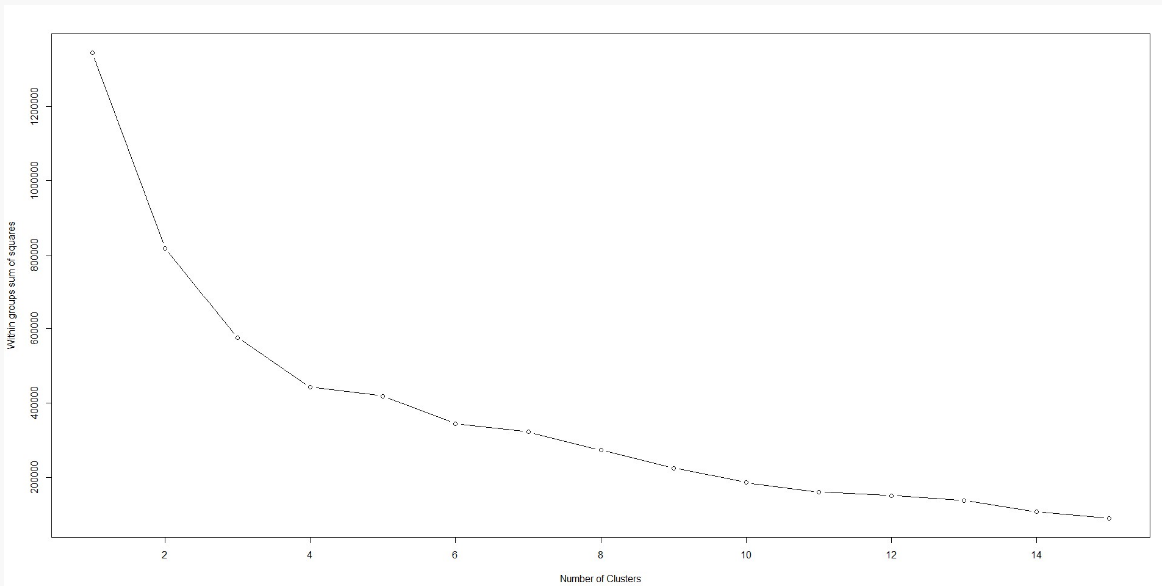
```
noRows<-25
for(i in 1:noRows){
 if(sum(is.na(aqi[i,]))){
  aqi[i,][is.na(aqi[i,])] <- mean(aqi[i,], na.rm = TRUE)
 }
}
aqi<-rowMeans(aqi)
aqi
place <-c(aqi2016[,1])
df2<-data.frame(
 Location<-place,
 aqi_overall_avg <-aqi
)
df2
```

```
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()
```

Output Plots & Explanation

WSS Plot

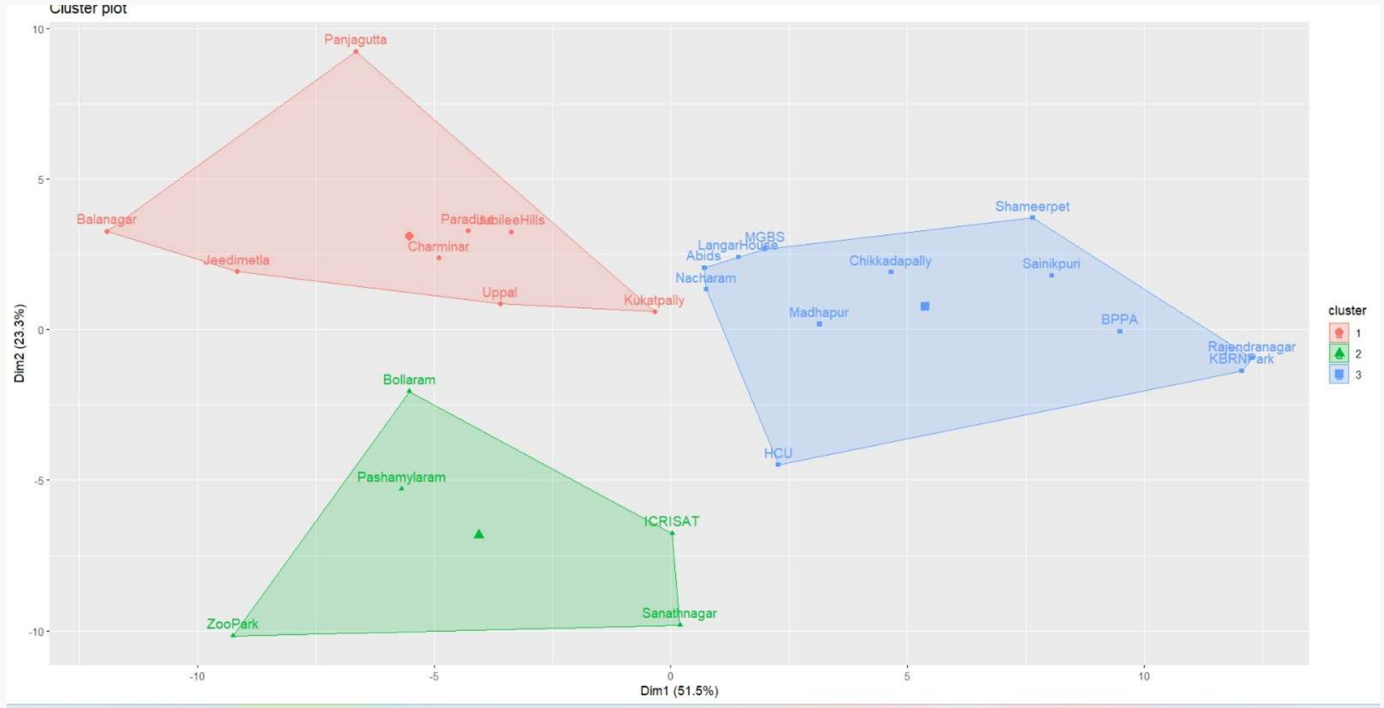


- A plot of the total within-groups sums of squares against the number of clusters in a K-means solution can be helpful. A bend in the graph can suggest the appropriate number of clusters.

Clustering Plot

- Package:
  FACTOEXTRA
- Function used:
  fviz_cluster
- Syntax:
  fviz_cluster(KM,data=mydatat)

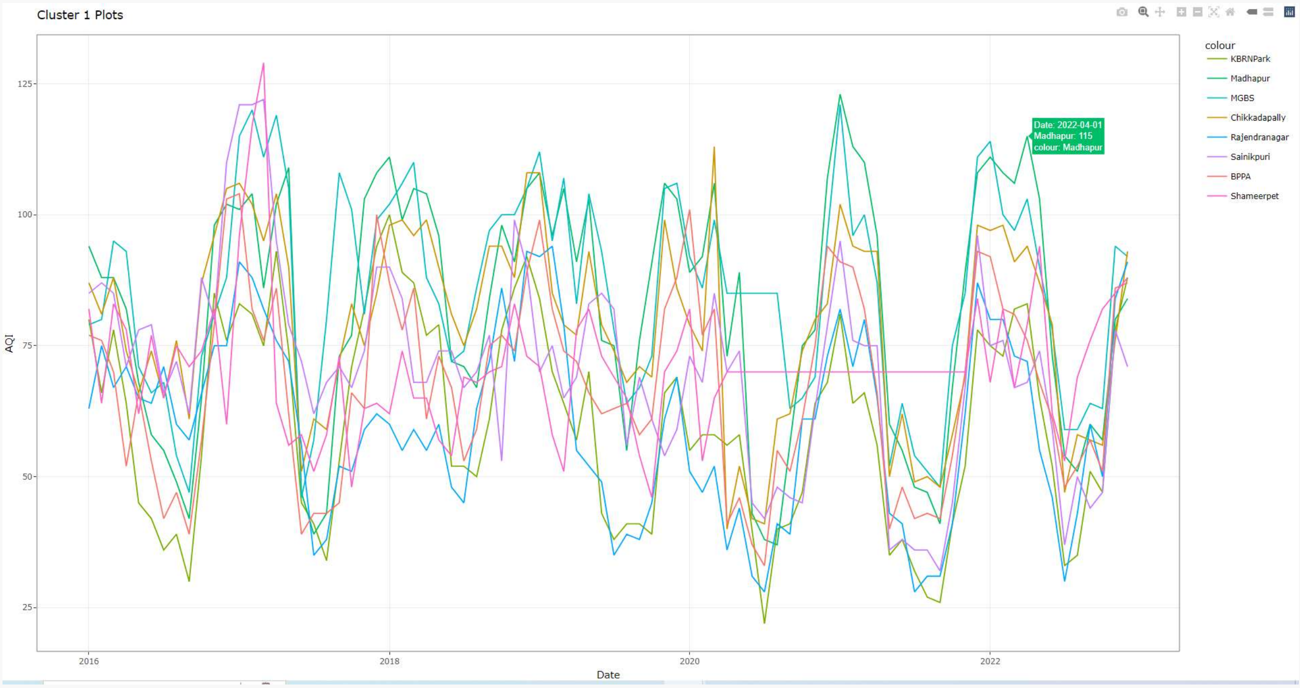
- Fviz_cluster() is a PCA Plot



- A PCA (Principal Component Analysis) plot is a type of plot that allows you to visualize the relationships between different variables in a dataset.
- It is based on a mathematical technique that reduces the dimensionality of a dataset by transforming the original variables into a smaller set of uncorrelated variables, called principal components.

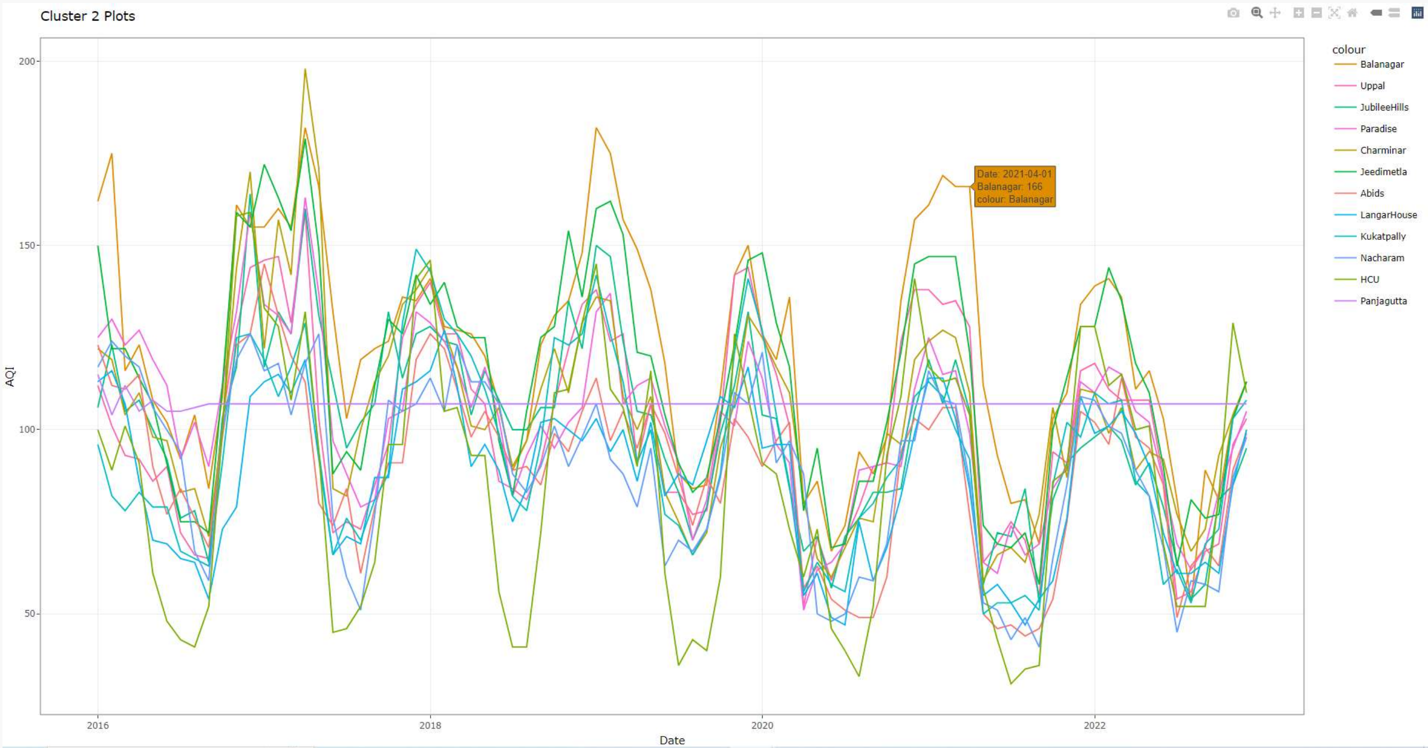
Interactive Plots of Areas in each Cluster

- Packages:
  - GGPLOT
  - PLOTLY
- Functions used:
  - geom_line()
  - ggplot()
  - plotly()

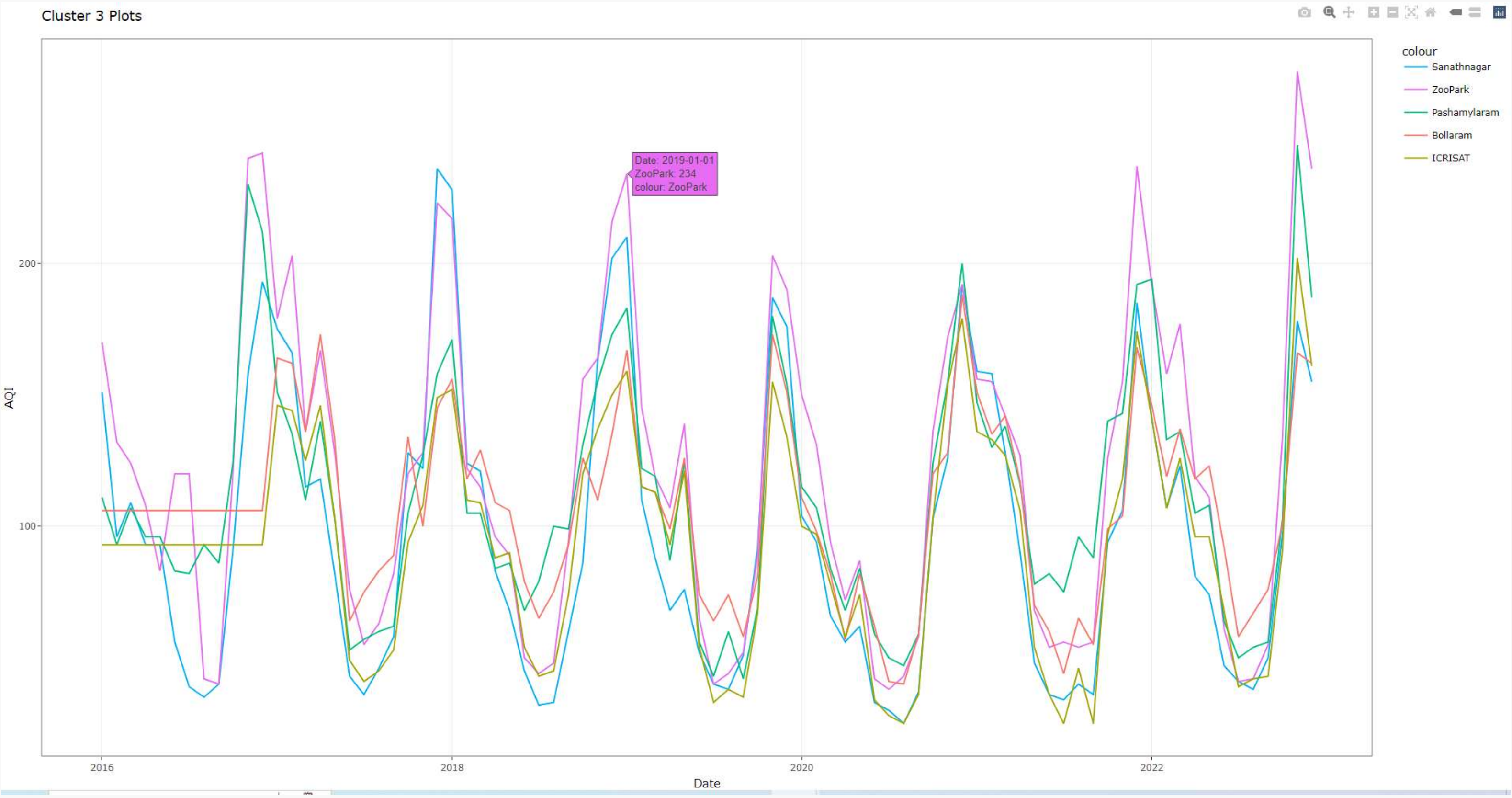


- All the areas have similar trends of the AQI and hence are categorized into the same cluster.

Interactive Plots of Areas in each Cluster

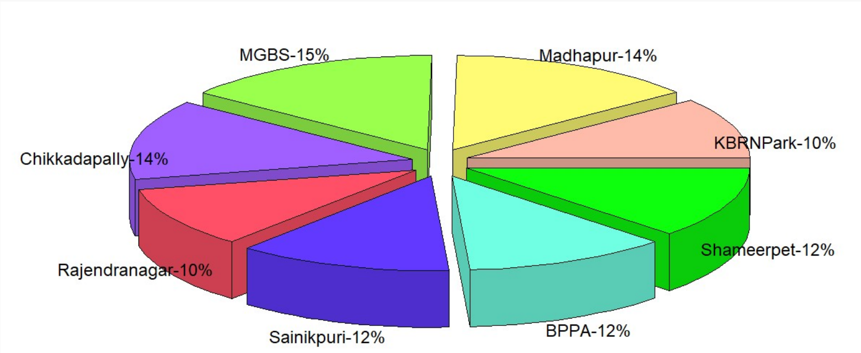


Interactive Plots of Areas in each Cluster

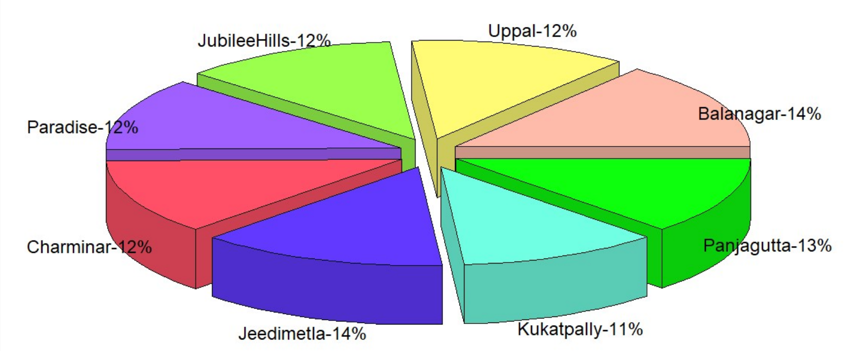


Cluster wise Mean AQI Distribution

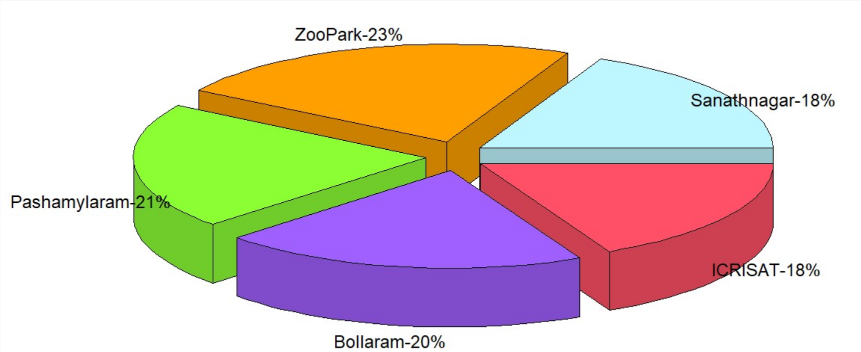
Cluster 1



Cluster 2



Cluster 3

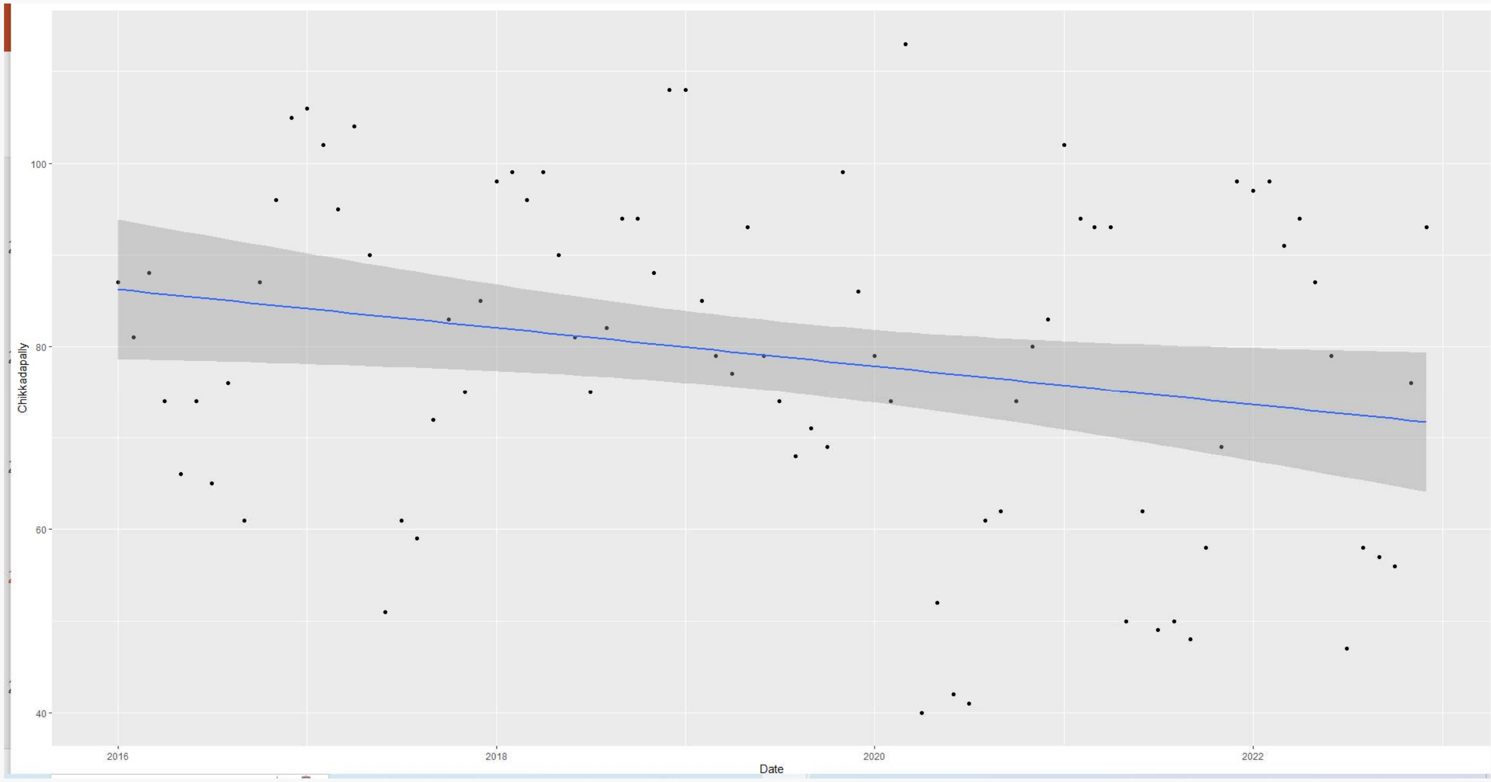


- Packages:
  PLOTRIX
- Functions used:

pie3D(x, labels, main,col,explode=0.1)

Linear Model Plot

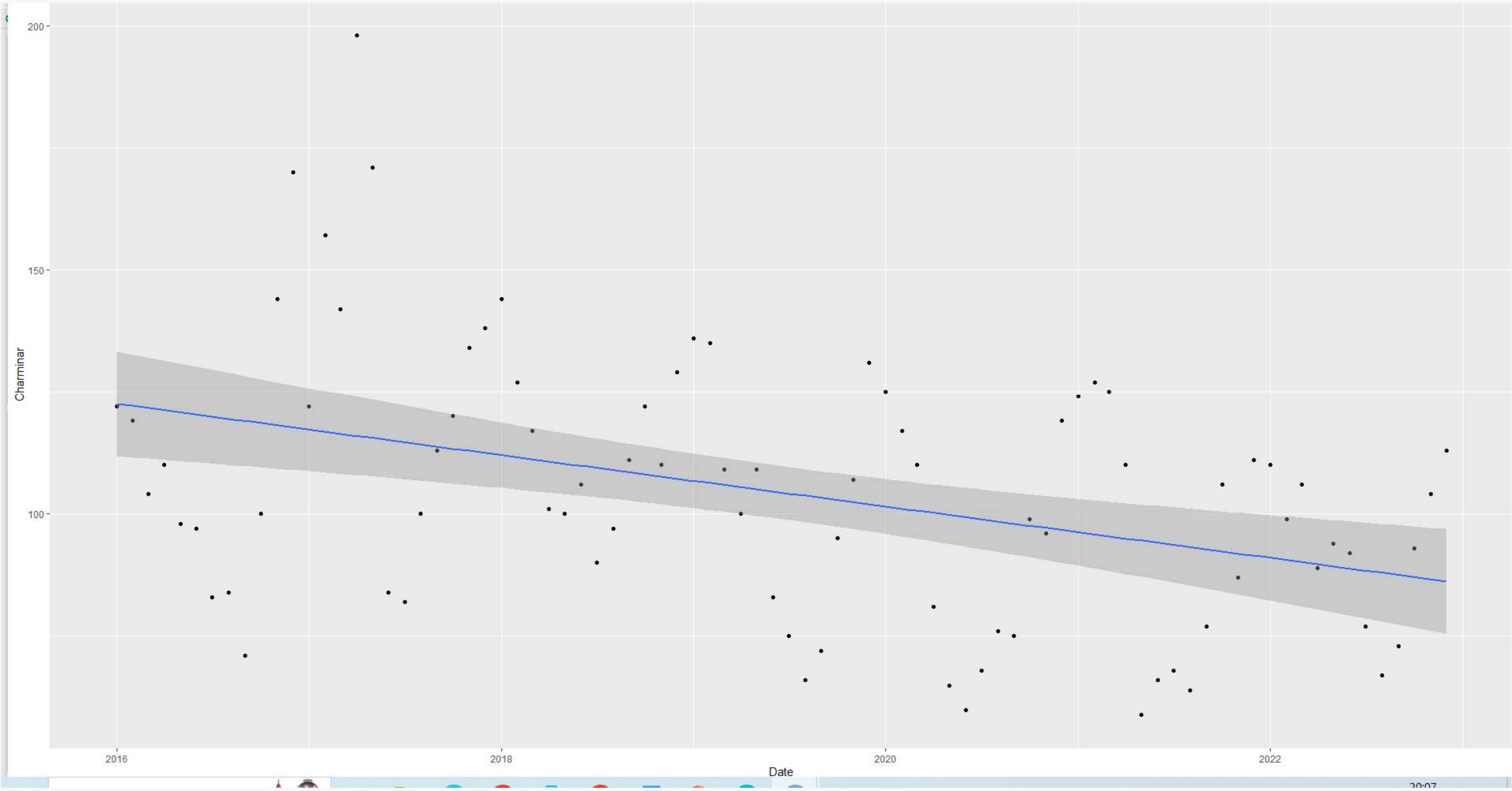
- Packages:
  - GGPLOT
- Functions used:
  - ggplot()
  - geom_line
  - geom_smooth



Chikkadapally representing the areas under Cluster 1

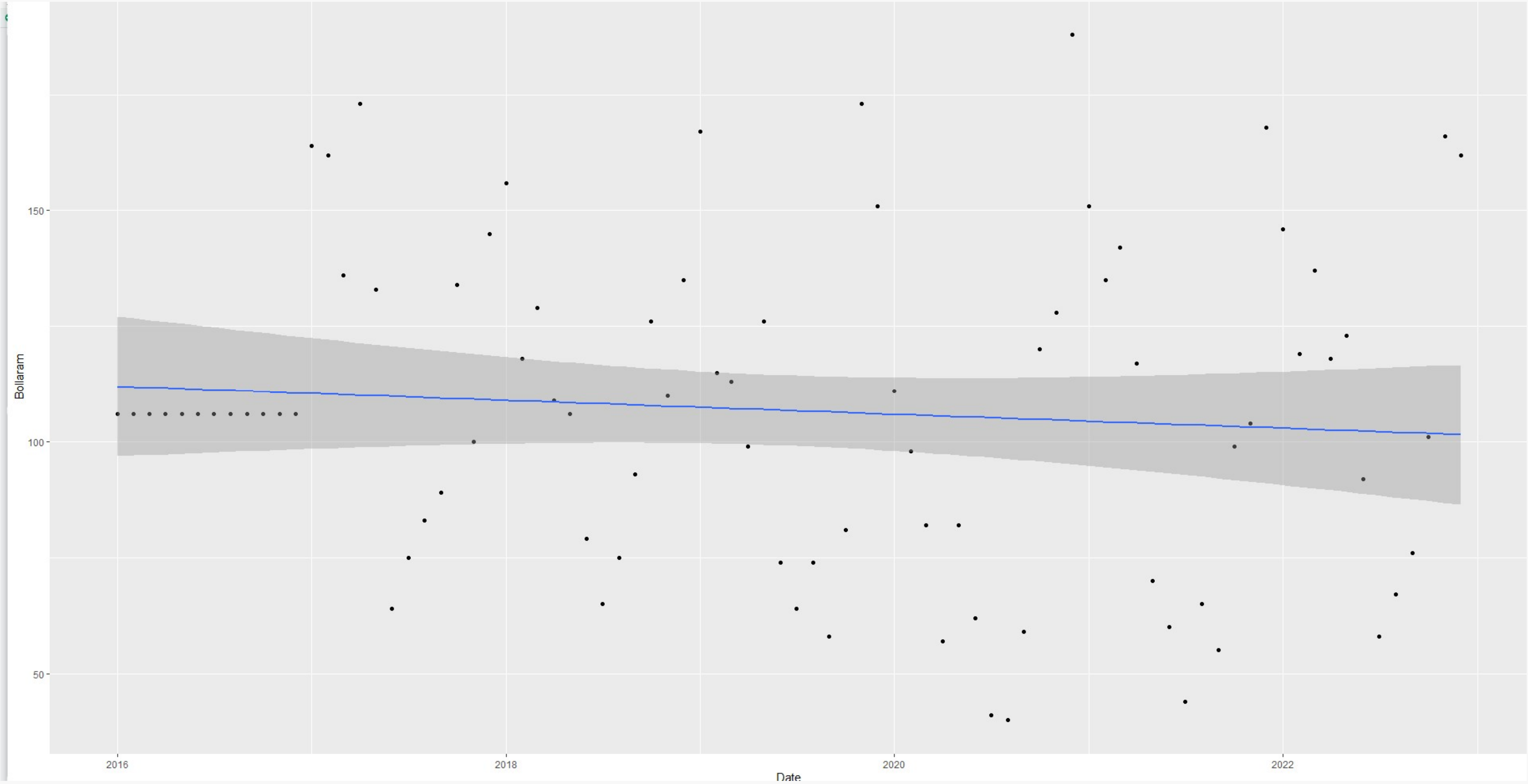
- Linear Model was used as it was beginner-friendly and easy to implement.

Linear Model Plot



Charminar representing the areas under Cluster 1

Linear Model Plot



Bollaram representing the areas under Cluster 3

Overall Mean AQI for Different Places in Hyderabad

Conclusion:

- In conclusion, R programming language is a powerful tool for analyzing AQI data to identify trends, correlations, and other patterns.
- This presentation has discussed the process of analyzing the AQI of Hyderabad, India, using R programming language.
- The results of the analysis can be used to inform decisions and to raise awareness about the importance of air quality.
- The areas under the first cluster i.e., KBRN Park, Madhapur, MGBS, Chikkadapally, Rajendranagar, Sainikpuri, BPPA and Shameerpet are good for living in the city compared to the other regions as the AQI values in these areas are lower which represents the pollution in these areas is low and the air quality is high

Conclusion:

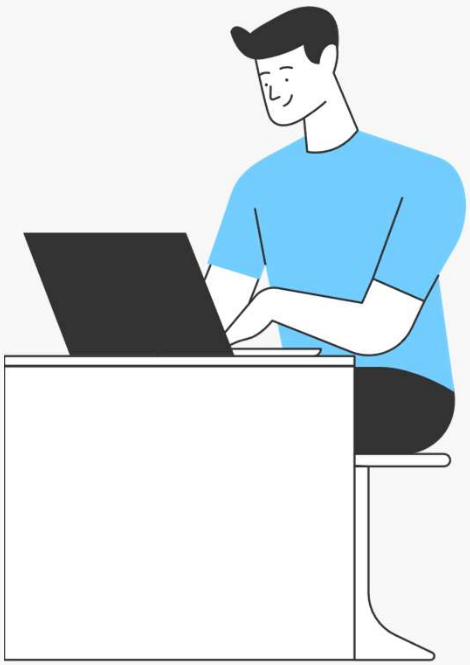
- From the individual graphs, the values of all the areas during the year 2020 has shown a great fall in the Air Quality Index Values.
- This might be due to the prevalence Covid-19 and the lockdown imposed in the city.
- After 2020, during 2021 and 2022, the values are increasing as the lockdowns have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

Conclusion:

- 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.

Conclusion:

- 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.

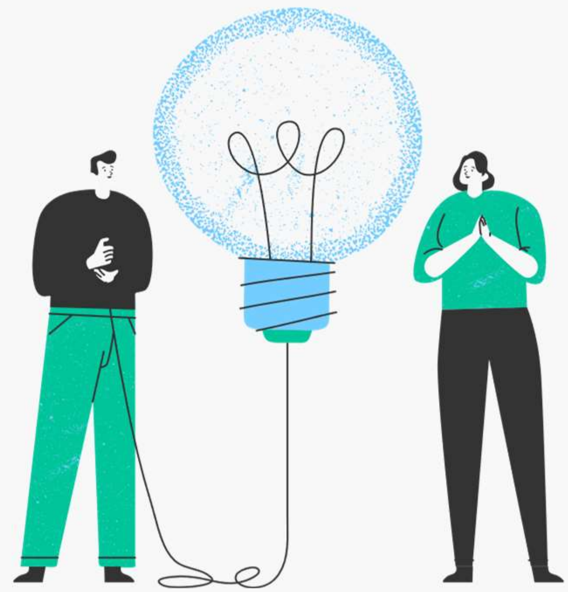


Recommendations

- Based on the analysis, it is recommended that the government must take action to reduce air pollution in the city.
- This could include implementing regulations to reduce emissions from vehicles and industries, as well as implementing public awareness campaigns to educate the public about the importance of air quality.
- Usage of Electric Vehicles must be encouraged.
- Public Transportation must be updated and the emissions must be reduced.

Future Work

- Using efficient Regression Models for the comparison of Time Series Data.
- In the case of missing data fields, the linear model can be generated by just using the available values.
- Research could be conducted to explore the relationship between the AQI and other factors such as temperature, humidity, and wind speed.
- Finally, the results of the analysis could be used to develop strategies to reduce air pollution in Hyderabad.



References:

- Dataset Source : Telangana State Pollution Control Board

Thank you!