## DATA SCIENCE DASHBOARD

# **TOPIC: AQI INDEX DASHBOARD**

**Submitted By: TANISHA PARKASH** 

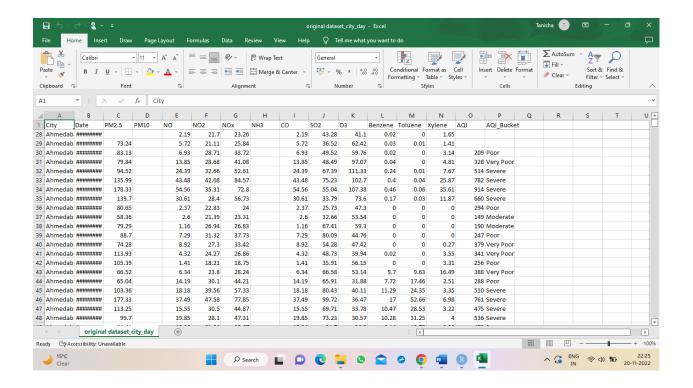
3CO18

102003447

3CO18

1. **Description:** My project aims to depict the Air Quality Index Variation in various cities of India recorded over a period of years. It represents the quantity of various Particulate matter (PM) and gases present in the air in those cities. To achieve the goal, the "City hour" dataset, which is a part of the AQI India dataset present on Kaggle has been imported in R. I have performed preprocessing in various parts, starting from removing the NULL values, replacing NULL with respective means, normalizing values, calculating missing values to splitting dataset and finally writing it into separate CSV files. The preprocessed datasets are then loaded into Tableau. I have established a relationship between the various datasets according to city names and AQI. Separate sheets present views regarding PM and gas values in some of the major cities of India like Delhi, Gurugram, Hyderabad, Shillong, Chennai, etc. One sheet represents an India map view of AQI values of the cities in a broader prospect. Finally, combining all sheets draws a comparison between all Graphs and the AQI values of cities. This is a descriptive dashboard prepared in Tableau.

2. Previous Dataset



#### 3. Preprocessing Code:

## //reading data

aqidata1<- read.csv(file.choose())
library(dplyr)</pre>

aqidata1 %>%
 distinct(City)

colSums(is.na(aqidata1))

## //removing na values

nona=aqidata1 %>% filter(!is.na(CO)) nona

nona %>% distinct(City)

cities=nona %>%
 distinct(City)

cities

colSums(is.na(nona))

```
nona %>%
filter(City=="Ahmedabad")
//replacing na values
nona$NO[is.na(nona$NO[1:62])]<- mean(nona$NO[1:62], na.rm=TRUE)
nona %>%
 filter(City=="Amritsar")
nona$NO[is.na(nona$NO[63:124])]<- mean(nona$NO[63:124], na.rm=TRUE)
nona %>%
filter(is.na(NO))
nona %>%
filter(City==cities[3,])
nona$NO[is.na(nona$NO[125:186])]<- mean(nona$NO[125:186], na.rm=TRUE)
nona$NO[is.na(nona$NO[187:248])]<- mean(nona$NO[187:248], na.rm=TRUE)
nona %>%
filter(City==cities[4,])
nona$NO[is.na(nona$NO[nona$City=="Delhi"])]<- mean(nona$NO[nona$City=="Delhi"],
na.rm=TRUE)
nona %>%
 filter(City=="Delhi")
for (i in cities[,1]){
nona$NO[is.na(nona$NO[nona$City==i])]<-mean(nona$NO[nona$City==i], na.rm=TRUE)
}
for (i in cities[,1]){
nona$PM2.5[is.na(nona$PM2.5[nona$City==i])]<-mean(nona$PM2.5[nona$City==i], na.rm=TRUE)
}
for (i in cities[,1]){
 nona$PM10[is.na(nona$PM10[nona$City==i])]<-mean(nona$PM10[nona$City==i], na.rm=TRUE)
}
for (i in cities[,1]){
 nona$NO2[is.na(nona$NO2[nona$City==i])]<-mean(nona$NO2[nona$City==i], na.rm=TRUE)
}
for (i in cities[,1]){
nona$NOx[is.na(nona$NOx[nona$City==i])]<-mean(nona$NOx[nona$City==i], na.rm=TRUE)
}
```

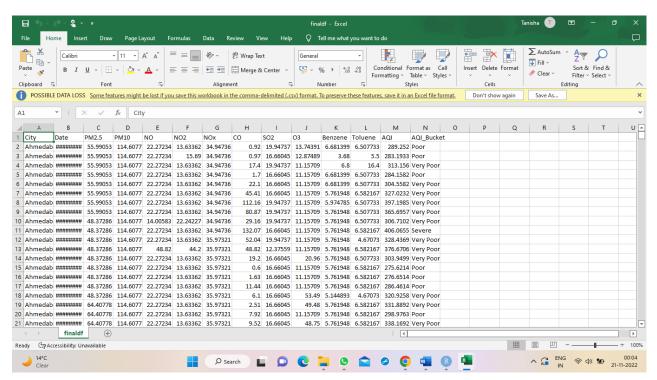
```
for (i in cities[,1]){
nona$NH3[is.na(nona$NH3[nona$City==i])]<-mean(nona$NH3[nona$City==i], na.rm=TRUE)
}
for (i in cities[,1]){
nona$SO2[is.na(nona$SO2[nona$City==i])]<-mean(nona$SO2[nona$City==i], na.rm=TRUE)
}
for (i in cities[,1]){
nona$NH3[is.na(nona$NH3[nona$City==i])]<-mean(nona$NH3[nona$City==i], na.rm=TRUE)
nona["NH3"]=aqidata1["NH3"]
library(tidyr)
nona %>%
drop_na()
//removing na columns
df = subset(nona, select = -c(8))
df
colSums(is.na(df))
for (i in cities[,1]){
df$O3[is.na(df$O3[df$City==i])]<-mean(df$O3[df$City==i], na.rm=TRUE)
}
for (i in cities[,1]){
df$Benzene[is.na(df$Benzene[df$City==i])]<-mean(df$Benzene[df$City==i], na.rm=TRUE)
}
for (i in cities[,1]){
df$Toluene[is.na(df$Toluene[df$City==i])]<-mean(df$Toluene[df$City==i], na.rm=TRUE)
}
for (i in cities[,1]){
df$Xylene[is.na(df$Xylene[df$City==i])]<-mean(df$Xylene[df$City==i], na.rm=TRUE)
df2 = subset(df, select = -c(13))
colSums(is.na(df2))
df3=df2
df3
```

```
//calculating na values of column
for(i in 1:nrow(df4)){
 if(df4[i,13]==df4[i,3]){
  df4[i,13] = df4[i,3] + df4[i,4] + df4[i,5] + df4[i,6] + df4[i,7] + df4[i,8] + df4[i,9] + df4[i,10] + df4[i,11] + df4[i,12]
}
}
aqi=df3[,13]
aqi
df4<-df3 %>%
 na.omit()
colSums(is.na(df4))
df5=df4
//determining categories
for(i in 1:nrow(df5))
{
 if(df5[i,14]==""){
  if(df5[i,13]>=400) df5[i,14]="Severe"
  else if(df5[i,13]>=300) df5[i,14]="Very Poor"
  else if(df5[i,13]>=200) df5[i,14]="Poor"
  else if (df5[i,13]>=100) df5[i,14]="Moderate"
  else if (df5[i,13]>=50) df5[i,14]="satisfactory"
  else df5[i,14]="Good"
 }
}
colSums(is.na(df5))
//writing back to csv files
write.table(df5, "D:\\AI\\data science\\dashboard project\\finaldf.csv",
       append = TRUE,
       sep = ",",
       col.names = TRUE,
       row.names = FALSE,
       quote = FALSE)
colSums(is.na(df5))
cities
delhi<-subset(df5,City=="Delhi")
bglr<-subset(df5,City=="Bengaluru")
bglr
chni<-subset(df5,City=="Chennai")</pre>
```

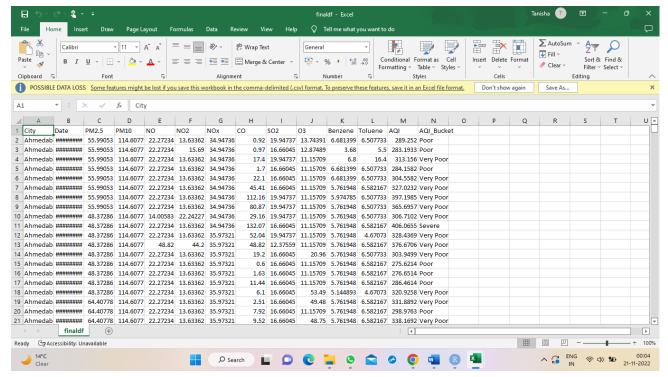
```
mbai<-subset(df5,City=="Mumbai")
kkta<-subset(df5,City=="Kolkata")
shlg<-subset(df5,City=="Shillong")
ggrm<-subset(df5,City=="Gurugram")
hybd<-subset(df5,City=="Hyderabad")
chdh<-subset(df5,City=="Chandigarh")
write.table(delhi, "D:\\AI\\data science\\dashboard project\\delhidf.csv",
      append = TRUE,
      sep = ",",
      col.names = TRUE,
      row.names = FALSE,
      quote = FALSE)
write.table(bglr, "D:\\AI\\data science\\dashboard project\\bengaloredf.csv",
      append = TRUE,
      sep = ",",
      col.names = TRUE,
      row.names = FALSE,
      quote = FALSE)
write.table(chni, "D:\\AI\\data science\\dashboard project\\chennaidf.csv",
      append = TRUE,
      sep = ",",
      col.names = TRUE,
      row.names = FALSE,
      quote = FALSE)
write.table(mbai, "D:\\AI\\data science\\dashboard project\\mumbai.csv",
      append = TRUE,
      sep = ",",
      col.names = TRUE,
      row.names = FALSE,
      quote = FALSE)
write.table(kkta, "D:\\AI\\data science\\dashboard project\\kolkatadf.csv",
      append = TRUE,
      sep = ",",
      col.names = TRUE,
      row.names = FALSE,
      quote = FALSE)
write.table(shlg, "D:\\AI\\data science\\dashboard project\\shillongdf.csv",
      append = TRUE,
      sep = ",",
      col.names = TRUE,
      row.names = FALSE,
```

```
quote = FALSE)
write.table(ggrm, "D:\\AI\\data science\\dashboard project\\gurugramdf.csv",
      append = TRUE,
      sep = ",",
      col.names = TRUE,
      row.names = FALSE,
      quote = FALSE)
write.table(hybd, "D:\\AI\\data science\\dashboard project\\hyderabaddf.csv",
      append = TRUE,
      sep = ",",
      col.names = TRUE,
      row.names = FALSE,
      quote = FALSE)
write.table(chdh, "D:\\AI\\data science\\dashboard project\\chandigarhdf.csv",
      append = TRUE,
      sep = ",",
      col.names = TRUE,
      row.names = FALSE,
      quote = FALSE)
```

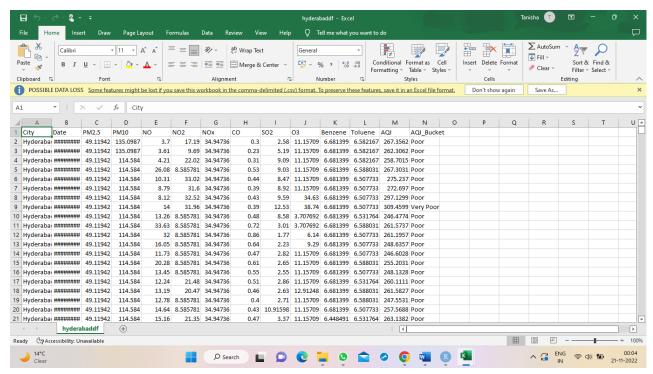
#### 4. PreProcessed Dataset



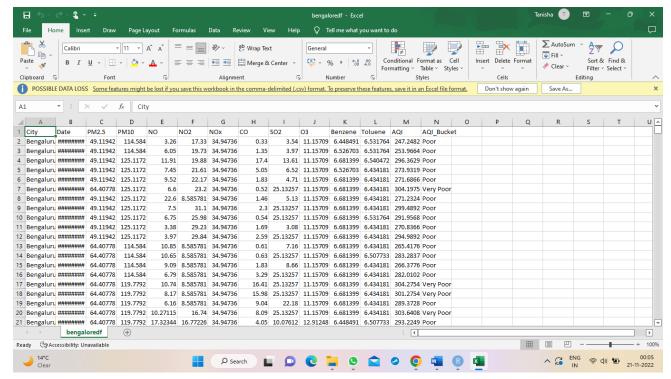
Finaldf.csv



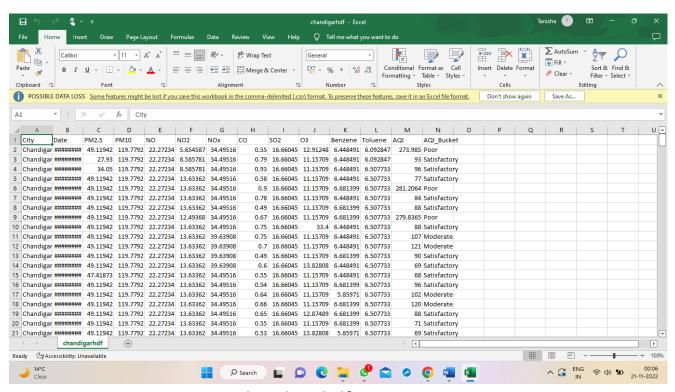
Gurugramdf.csv



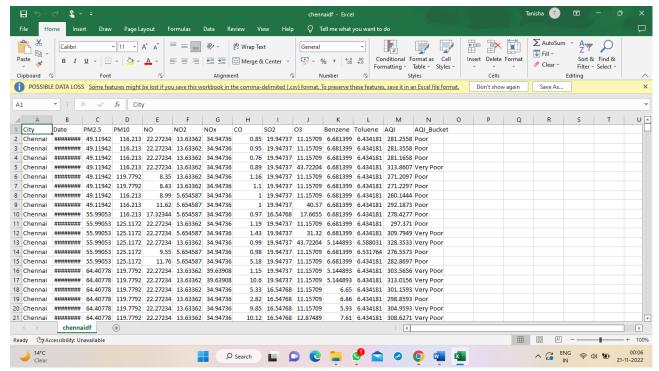
Hyderabaddf.csv



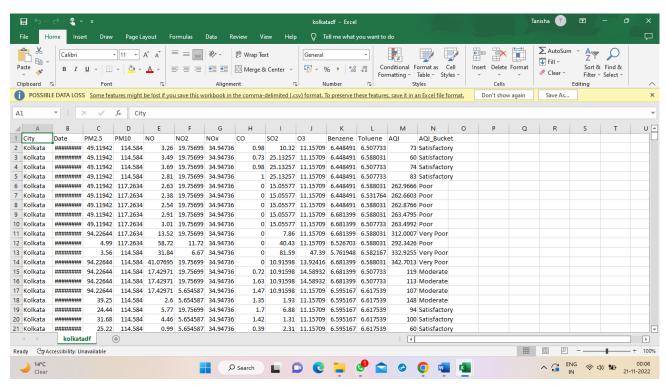
bengaloredf.csv



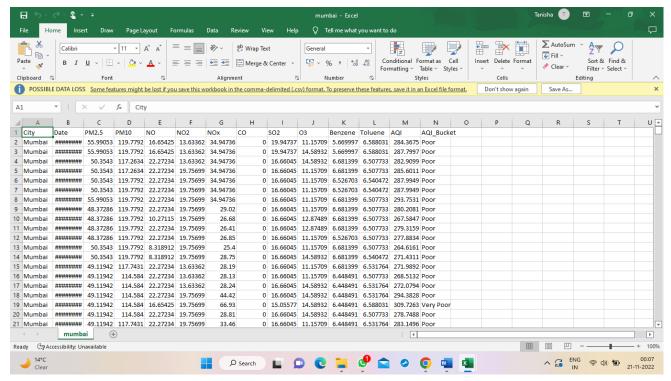
chandigarhdf.csv



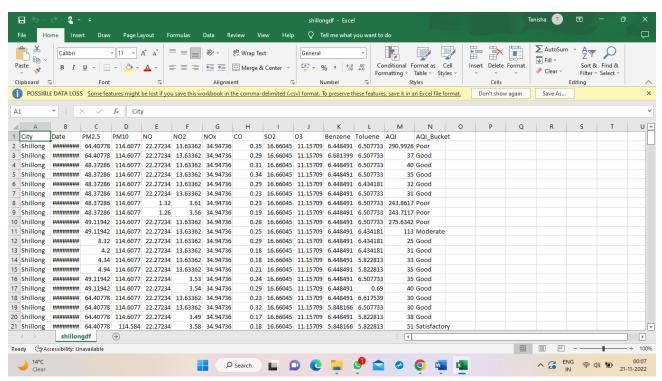
Chennaidf.csv



Kolkatadf.csv

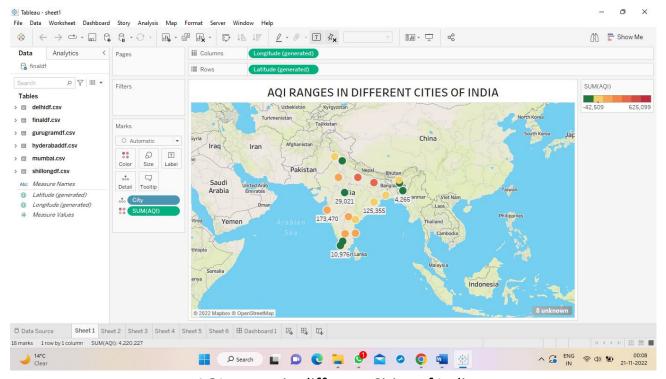


mumbaidf.csv

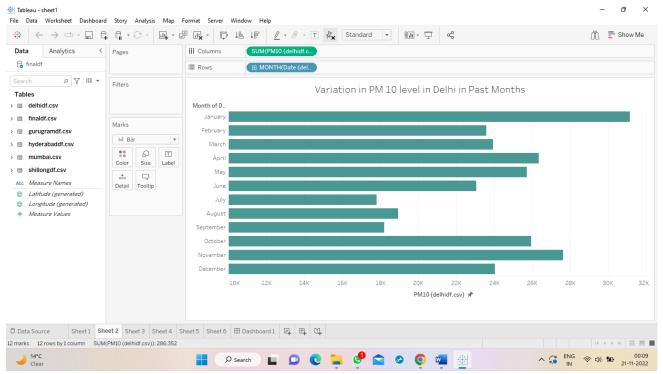


Shillongdf.csv

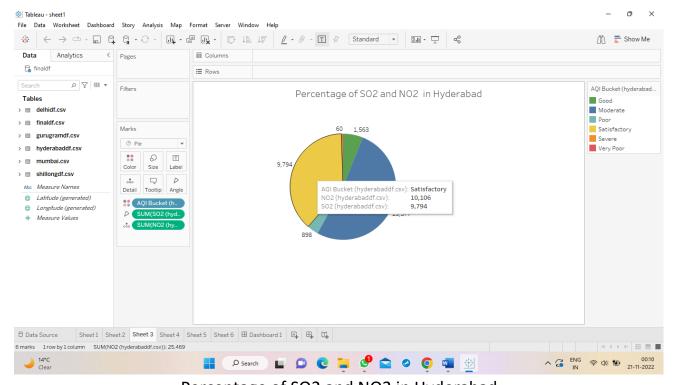
#### 5. Tableau Sheet Views



AQI ranges in different Cities of India
We have plotted the cities against the AQI measure on Graph to obtain
the desired view.

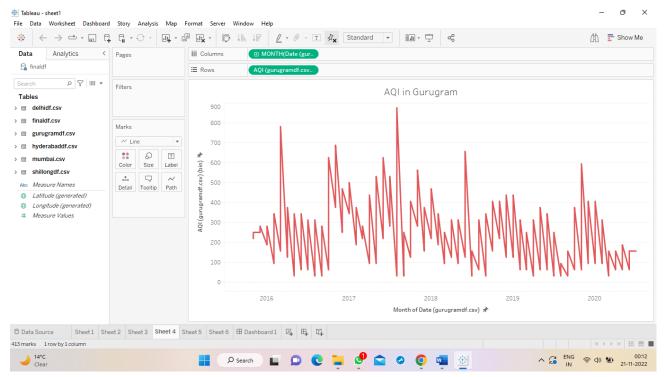


Variation in PM10 level in Delhi in Past Months
We have plotted the PM10 level against the Date field on the basis of
Months of Delhi dataset to obtain the desired view.



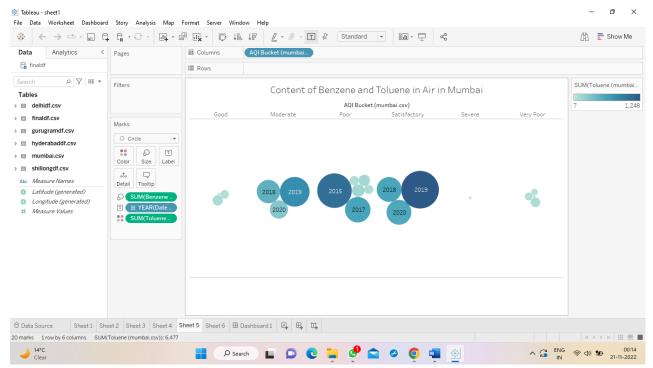
Percentage of SO2 and NO2 in Hyderabad

This pie chart is plotted using SO2 and NO2 details of Hyderabad dataset to depict the percentage of the two gases in AQI level of the city.

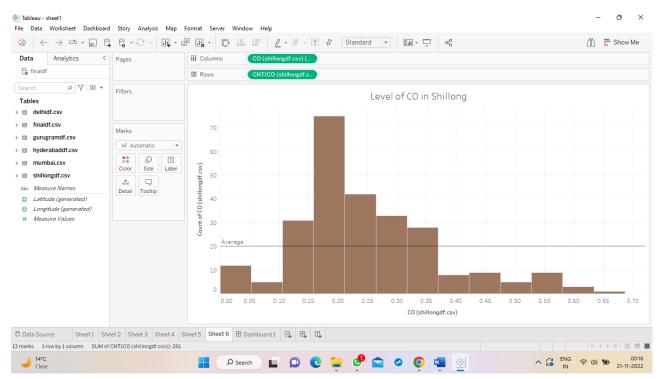


AQI in Gurugram

This line chart is an analysis of the amount of AQI pollutants present in air of Gurugram in the past five years.

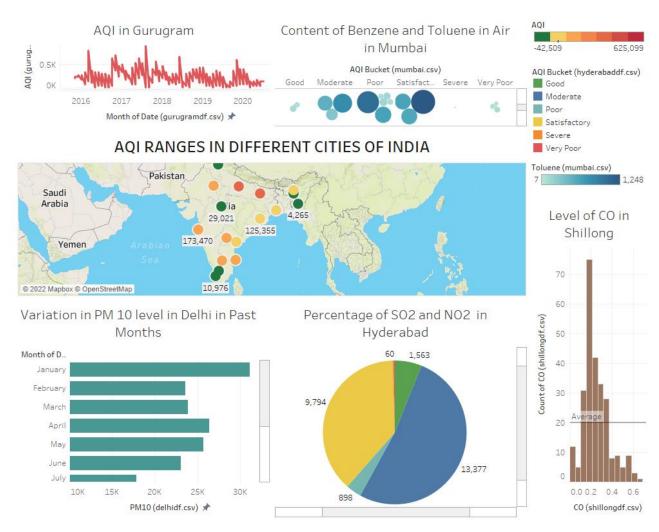


Content of Benzene and Toluene in Air in Mumbai
This is another graph describing the content of the two harmful
chemicals in air of Mumbai, and which year they have been the most in
quantity.



Level of CO in Shillong

The above histogram is a depiction of CO levels in Shillong according to the dataset. We can infer that lesser quantity of CO is present in more frequency in the city, ie Shillong has comparatively cleaner air!



This is our final AQI Dashboard.

SUBMITTED TO: Mrs. Geeta Kasana