

# Project Component 1

Himanshu, MDS202327

2023-09-20

## Introduction

TBA

## Data Description

The data contains six air pollution parameters like PM2.5, PM10, NO2, NH3, SO2, Ozone for 12 stations in New Delhi, collected from CPCB website from 08-02-2018 to 02-01-2021 on daily basis.

```
# Importing libraries
```

```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(TSstudio)
library(plotly)
```

```
# Reading the data into data frame
```

```
df <- read.csv("delhi.csv", header = TRUE)
set.seed(5)
df[sample(nrow(df), 5), ]
```

```
##           siteName siteCode      Date  PM2.5  PM10  NO2  NH3  SO2 Ozone
## 2255   Jahangirpuri    1423 2018-06-22 109.07 354.45 53.85 40.98 20.49 66.19
## 6859      Rohini      1430 2019-06-21  71.48 179.16 27.07 55.90  8.03 42.51
## 1833 Dwarka-Sector 8    1422 2020-03-21  86.52 241.53 57.14 34.50 22.58 39.24
## 3797   Najafgarh      1427 2019-10-17 112.84 215.94 42.99 29.68 10.12 49.32
## 7239      Rohini      1430 2020-07-05  27.91  58.40  1.34 99.40  5.08  5.00
```

```
# Variables in the data
```

```
names(df)
```

```
## [1] "siteName" "siteCode" "Date"      "PM2.5"    "PM10"     "NO2"      "NH3"
## [8] "SO2"      "Ozone"
```

```
# Dimension of the data
```

```
dim(df)
```

```
## [1] 10600      9
```

```
# Variable types
# Note the Date column has type chr which must be converted to date type.
str(df)
```

```
## 'data.frame': 10600 obs. of 9 variables:
## $ siteName: chr "Ashok Vihar" "Ashok Vihar" "Ashok Vihar" "Ashok Vihar" ...
## $ siteCode: int 1420 1420 1420 1420 1420 1420 1420 1420 1420 1420 ...
## $ Date : chr "2018-02-08" "2018-02-09" "2018-02-10" "2018-02-11" ...
## $ PM2.5 : num 237 250.5 269.7 146.4 82.1 ...
## $ PM10 : num 406 423 499 315 200 ...
## $ NO2 : num 110 79.4 183.9 41.8 23.2 ...
## $ NH3 : num 31.4 33.5 22.7 36.7 34.8 ...
## $ SO2 : num 11.2 13.24 7.16 8.38 4.43 ...
## $ Ozone : num 33.4 39.3 44.5 43 37.9 ...
```

```
df$Date <- as.Date(df$Date)
df[sample(nrow(df), 5), ]
```

```
##      siteName siteCode      Date PM2.5 PM10 NO2 NH3 SO2 Ozone
## 1527 Dwarka-Sector 8    1422 2019-05-20 117.01 350.75 52.95 27.37 31.63 64.79
## 4483      Narela      1426 2018-10-08 147.95 312.79 51.08 41.12 14.66 38.33
## 3878    Najafgarh    1427 2020-01-06 139.40 189.25 34.27 43.98 6.12 14.36
## 826    Ashok Vihar    1420 2020-05-13 72.60 127.77 20.55 34.70 22.60 52.92
## 9013    Vivek Vihar    1435 2019-07-25 23.98 81.74 12.47 26.99 10.92 22.80
```

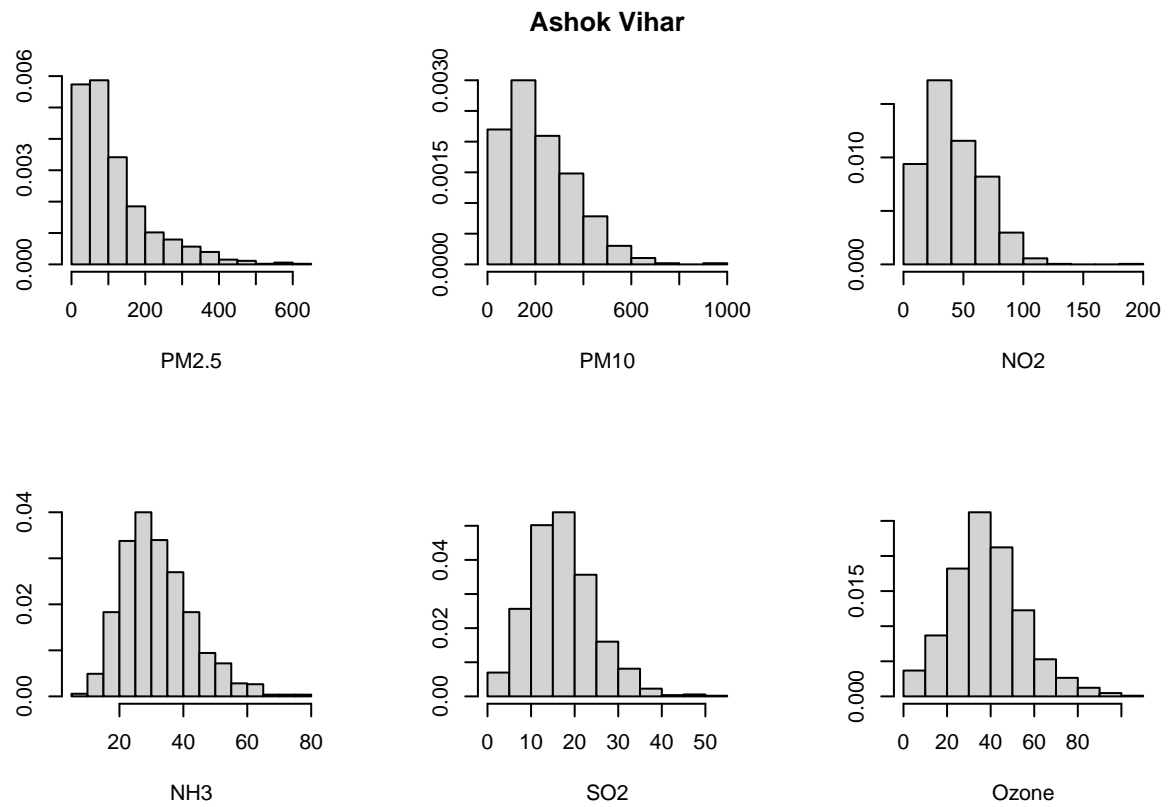
```
str(df)
```

```
## 'data.frame': 10600 obs. of 9 variables:
## $ siteName: chr "Ashok Vihar" "Ashok Vihar" "Ashok Vihar" "Ashok Vihar" ...
## $ siteCode: int 1420 1420 1420 1420 1420 1420 1420 1420 1420 1420 ...
## $ Date : Date, format: "2018-02-08" "2018-02-09" ...
## $ PM2.5 : num 237 250.5 269.7 146.4 82.1 ...
## $ PM10 : num 406 423 499 315 200 ...
## $ NO2 : num 110 79.4 183.9 41.8 23.2 ...
## $ NH3 : num 31.4 33.5 22.7 36.7 34.8 ...
## $ SO2 : num 11.2 13.24 7.16 8.38 4.43 ...
## $ Ozone : num 33.4 39.3 44.5 43 37.9 ...
```

```
unique(df$siteName)
```

```
## [1] "Ashok Vihar"      "Dwarka-Sector 8" "Jahangirpuri"    "Najafgarh"
## [5] "Narela"          "Patparganj"      "Rohini"          "Sonia Vihar"
## [9] "Vivek Vihar"      "Wazirpur"
```

```
par(mfrow = c(2,3))
hist(df[df$siteCode==1420,]$PM2.5, probability = TRUE, main = "", xlab = "PM2.5", ylab = "")
hist(df[df$siteCode==1420,]$PM10, probability = TRUE, main = "Ashok Vihar", xlab = "PM10", ylab = "")
hist(df[df$siteCode==1420,]$NO2, probability = TRUE, main = "", xlab = "NO2", ylab = "")
hist(df[df$siteCode==1420,]$NH3, probability = TRUE, main = "", xlab = "NH3", ylab = "")
hist(df[df$siteCode==1420,]$SO2, probability = TRUE, main = "", xlab = "SO2", ylab = "")
hist(df[df$siteCode==1420,]$Ozone, probability = TRUE, main = "", xlab = "Ozone", ylab = "")
```



```

par(mar=c(2,2,2,2))
par(mfrow=c(5,2))
for (i in unique(df$siteName)) {
  plot(df[df$siteName==i,]$Date, df[df$siteName==i,]$PM2.5, type = "l",
       main = i, xlab = "", ylab = "")
}

```

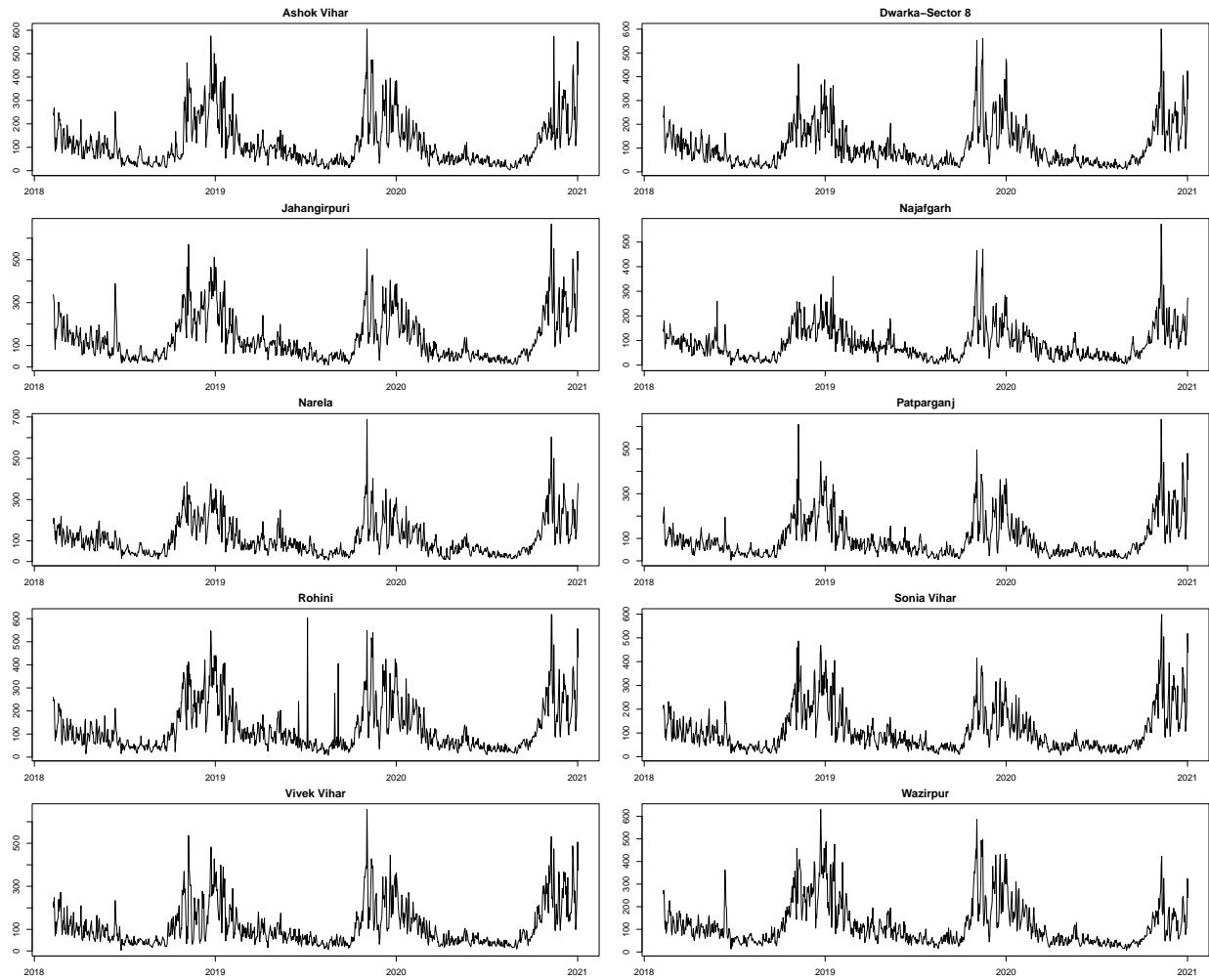


Figure: The above graph shows the time series plot of PM2.5 parameter for all 10 stations in the data.

## Exploratory Data Analysis

### Results

### Conclusion