SM

Google Earth Engine

Engineering Economics

Coding Project

SO2 and Its Role in Climate Change

PRESENTED BY

GROUP 11

Sidharth Menon (SM)	B180561CS
Vishnu Sajith (VS)	B180474CS
Mathew Jose Mammoottil (MJ)	B180586CS
Emmanuel Marian Mathew (EM)	B180347CS
Shaaheen A M (SA)	B181134CS

MJ

INTRODUCTION

In this problem-based assignment, we try to investigate the impact of sulfur dioxide on the environment. Sulfur dioxide comes from both human activities and natural sources. Burning coal and other fossil fuels is the largest source of sulfur dioxide from human activities, while volcanic activity is the prime natural cause of Sulfur Dioxide.

RECENT OVERVIEW

Power plant emissions of sulfur dioxide have increased across India in recent years. India passed the United States in 2010 to become the world's second largest emitter of sulfur dioxide, after China. NASA's Aura satellite found that emissions of sulfur dioxide from Indian power plants have increased by more than 60 percent between 2005 and 2012.

OUR OBJECTIVES

To develop JavaScript code for extracting zonal statistics from major cities in India using the Google Earth Engine Code Editor.

Extract the Zonal Statistics of the attribute ‘Pollution levels’ by using Sulphur dioxide as a proxy for pollution in cities as a CSV file.

Examine the cities - Bengaluru, Delhi, Kolkata, Mumbai and Hyderabad and use them as our frame of study.

Perform Python analysis on the extracted data with the help of Jupyter notebooks, Pandas library, etc. and draw inferences.

Conduct Correlation analysis and form ML regression models with other correlated datasets.

PROGRAMMING LANGUAGES USED

JavaScript
Python

TECHNICAL TOOLS

Google Earth Engine
Jupyter Notebook
Pandas library
Seaborn library
Sklearn library
NumPy library
SciPy library
Matplot library

SM

DATA SETS

For **Data Extraction**:

Sentinel-5P NRTI SO2: Near Real-Time Sulphur Dioxide (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_NRTI_L3_SO2)

For **Analysis, Correlation and ML Regression Model**:

Sentinel-5P NRTI SO2: Near Real-Time Sulphur Dioxide (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_NRTI_L3_SO2)

ERA5 Monthly aggregates - Latest climate reanalysis produced by ECMWF / Copernicus Climate Change Service (https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_MONTHLY)

Sentinel-5P NRTI AER AI: Near Real-Time UV Aerosol Index (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_NRTI_L3_AER_AI)

MJ

EXPLORATORY DATA ANALYSIS

CITIES STUDIED

BANGALORE
DELHI
HYDERABAD
KOLKATA
MUMBAI

VS

Storing SO2 data from Bangalore into a pandas dataframe.
The same proccess is repeated for all cities.

```
data2=pd.read_csv(r"C:\Users\mjose\Desktop\India\India\SO2_Sentinel\SO21\SO2Bangalore.csv")
data2.drop(['City'],axis=1)
data2['City'] = 'Bangalore'
data2.rename(columns = {'mean':'Mean'}, inplace = True)
data2['Month'] = [month[int(x[3:5]) -1] for x in data2['date']]
data2['Year'] = [x[6:10] for x in data2['date']]

temp= pd.to_datetime(data2['date'])
temp = temp.dt.strftime('%d-%m-%Y')
data2['Day'] = pd.to_datetime(temp).dt.day_name()
```

Concatenating all the dataframes to form one single dataframe.

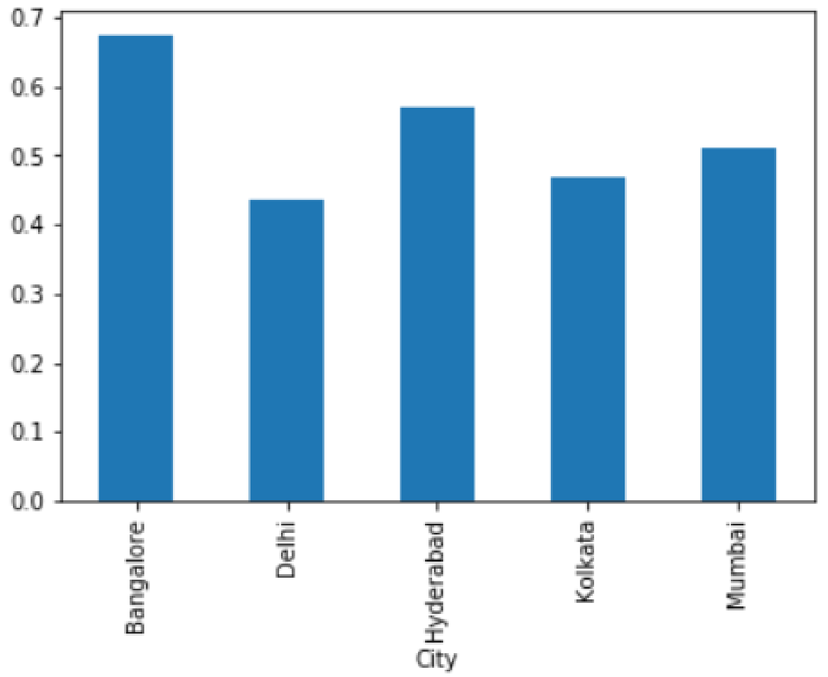
```
data7=pd.concat([data1, data2,data3,data4,data5], ignore_index=True, sort=True)
data7=data7.dropna(axis=1)
data7.rename(columns = {'mean':'Mean'}, inplace = True)
data7['Month'] = [month[int(x[3:5]) -1] for x in data7['date']]

temp= pd.to_datetime(data7['date'])
temp = temp.dt.strftime('%d-%m-%Y')
data7['Day'] = pd.to_datetime(temp).dt.day_name()
```

Grouping the rows in the concatenated dataframe by City, taking the mean of the 'Mean' value in rows of each group and displaying data as a bar plot.

```
data7.groupby('City').Mean.mean().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e284f86948>
```

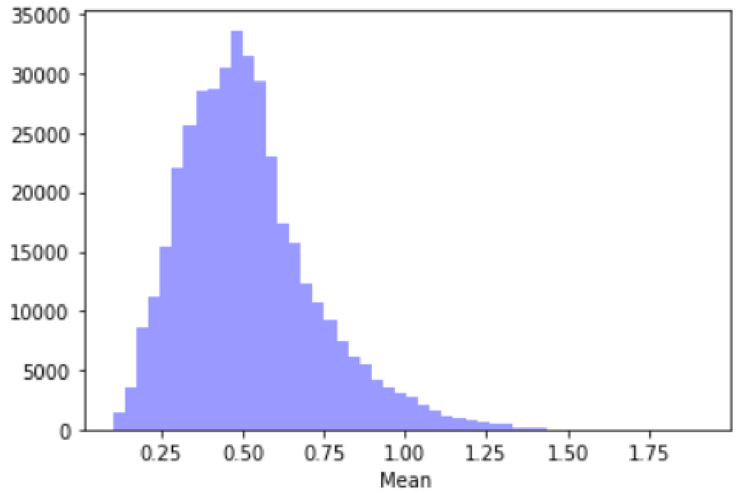


Inferences

- Banglore has the highest mean SO2 value among all the cities studies
- Delhi has the lowest mean SO2 value

```
sns.distplot(data7['Mean'],kde=False,norm_hist=False,bins=50,color='blue')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e28036b6c8>
```



Inferences

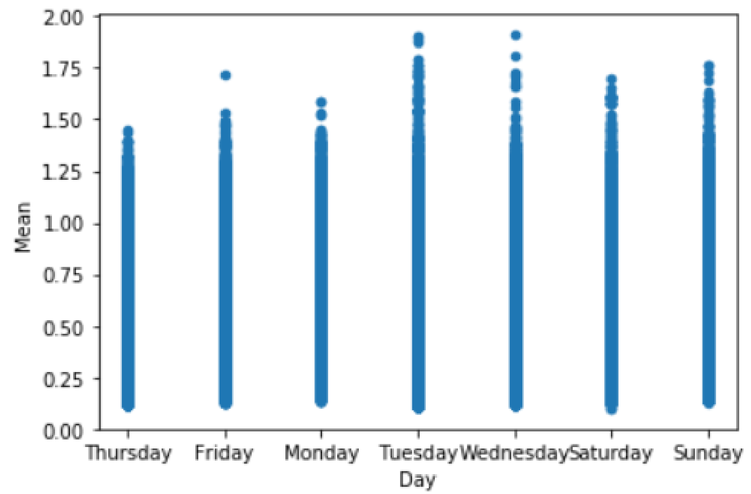
- The graph resembles bell curve of normal distibution with mean 0.50
- Distribution curve of 'Mean' peaked at about 0.50

Displaying a seaborn distplot to show the distribution of 'Mean' value in the concatenated dataframe.

Displaying a scatter plot to show the distribution of 'Mean' value in the concatenated dataframe grouped by the days of the week.

```
data7.plot.scatter(x='Day',y='Mean')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e283494d88>
```

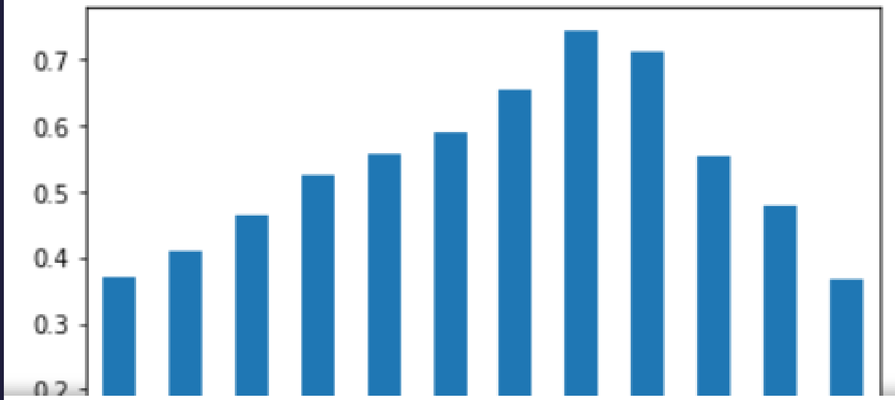


Inferences

- Maximum 'Mean' values were observed on a Tuesday and a Wednesday

```
data7.groupby('Month').Mean.mean().reindex(month).plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e2852dab08>
```



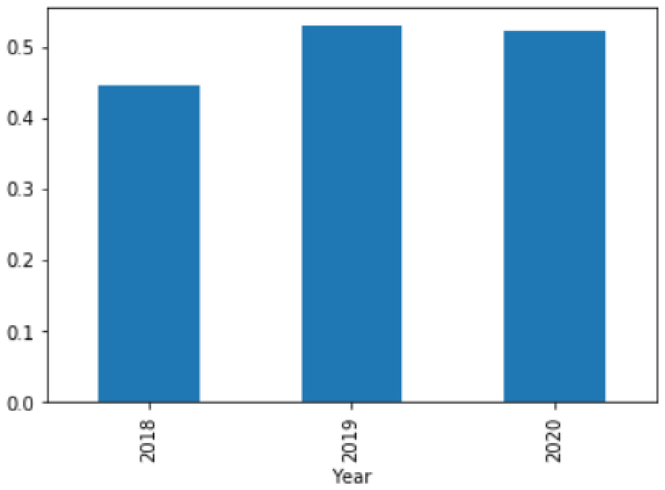
Inferences

- The SO2 level was the highest in August and September
- The SO2 level was the lowest in December and January

Showing how the mean of 'Mean' value varies as per the month of the year using a bar plot for the concatenated dataframe.

```
data7.groupby('Year').Mean.mean().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e2857ec348>
```



Inferences

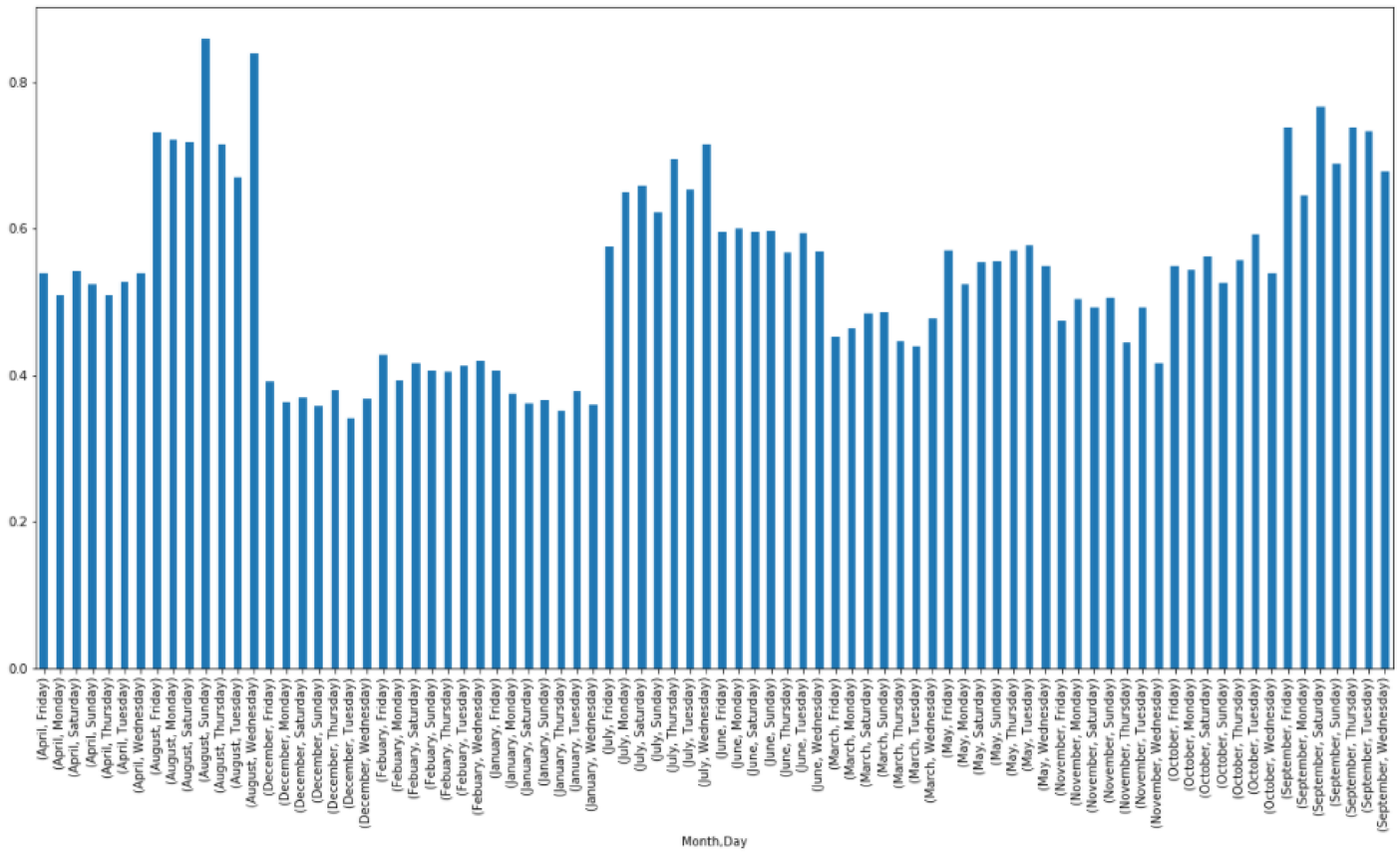
- The maximum mean SO2 values in 2019 (in the graph)
- We cannot say for sure that the actual SO2 emission in 2018 was lower than 2019 or 2020 because we have only the SO2 values from October to December in 2018 which are typically the months with low SO2 levels
- We can infer that the SO2 emission in 2020 was significantly lower than 2019 because we only have SO2 values until October 2020, so the yearly average of 2020 will be lower than the value shown in the graph since November and December are the months with lowest SO2 values historically. Thwas can be attributed to the industrial shutdown due to the COVID 19 pandemic.

Variation of 'Mean' according to the year using a bar plot for the concatenated dataframe.

Showing the variation of the 'Mean' value according to the day of the week in each month for the concatenated dataframe.

```
plt.figure(figsize=(20,10))
data7.groupby(['Month','Day']).Mean.mean().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e2868cd048>
```



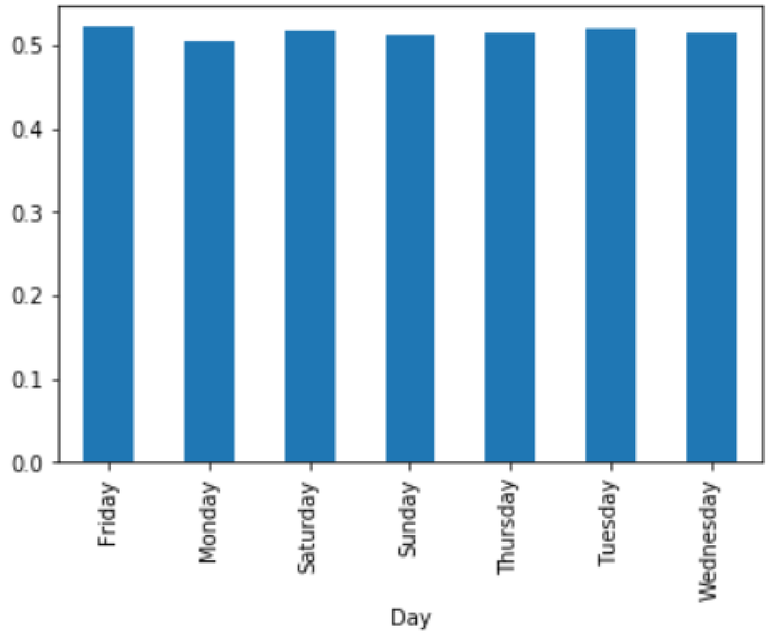
Inferences

- It was observed that the SO2 emission on all days in a week in each month was roughly uniform except for a couple of peaks at (August,Sunday) and (August,Wednesday)
- We could also confirm the conclusion drawn before that the SO2 emission in August and September was high and the same in December and January was low

Showing how the mean of 'Mean' value varies according to the day of the week using a bar plot for the concatenated dataframe.

```
data7.groupby('Day').Mean.mean().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e288af7fc8>
```

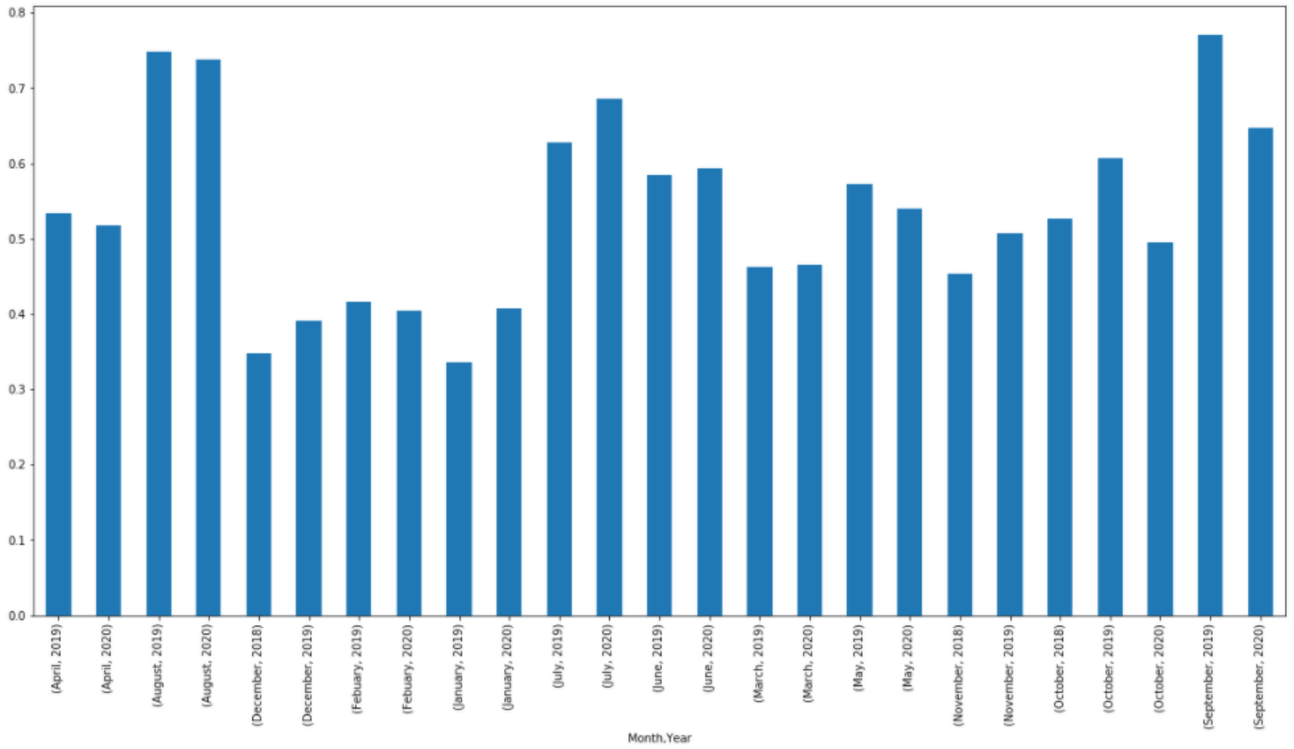


Inferences

- SO2 level on all days in a week was roughly the same
- Mean value on each day was approximately 0.5

```
plt.figure(figsize=(20,10))
data7.groupby(['Month','Year']).Mean.mean().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1e2897fb3c8>
```

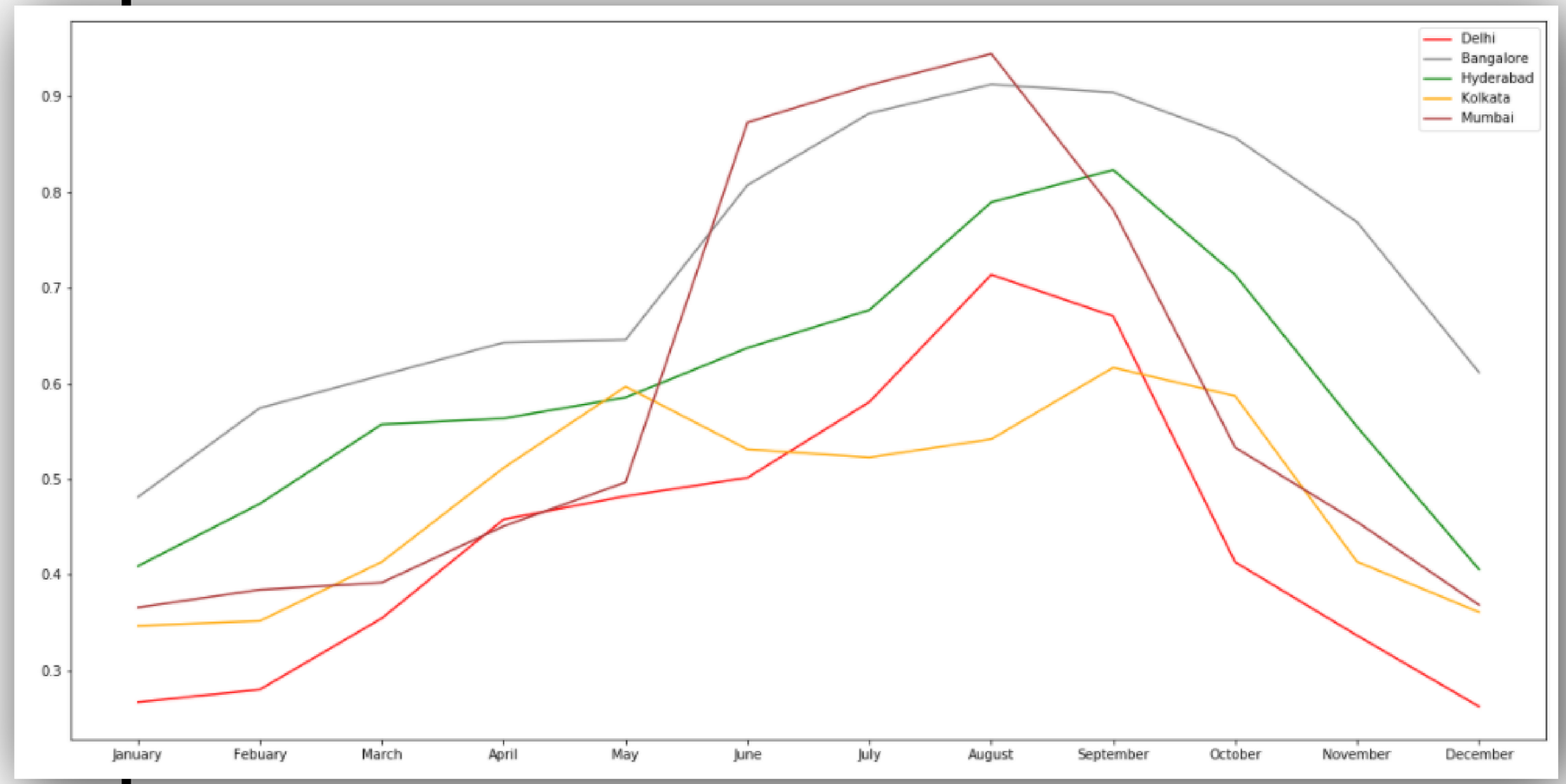


Inferences

- SO2 level was the usually highest during August-September
- SO2 level was the lowest during December-February
- SO2 levels in September and October 2020 are significantly lower than those in September and October 2020.
- Highest monthly mean - September 2019, Lowest monthly mean - January 2019

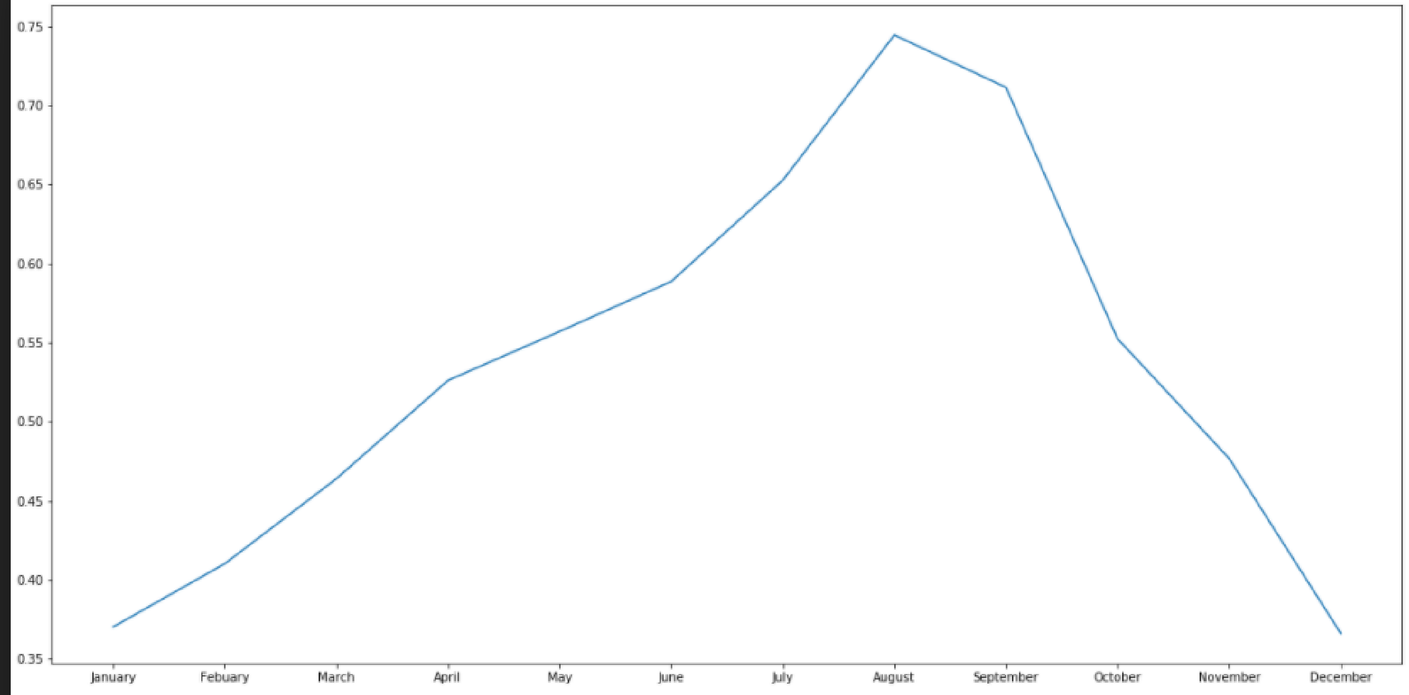
Showing how the monthly mean of 'Mean' value varies for the concatenated dataframe using a bar plot.

MJ



Visualizing how the 'Mean' values of all 5 cities varies during the months.

```
[<matplotlib.lines.Line2D at 0x1e28b3e2f08>]
```

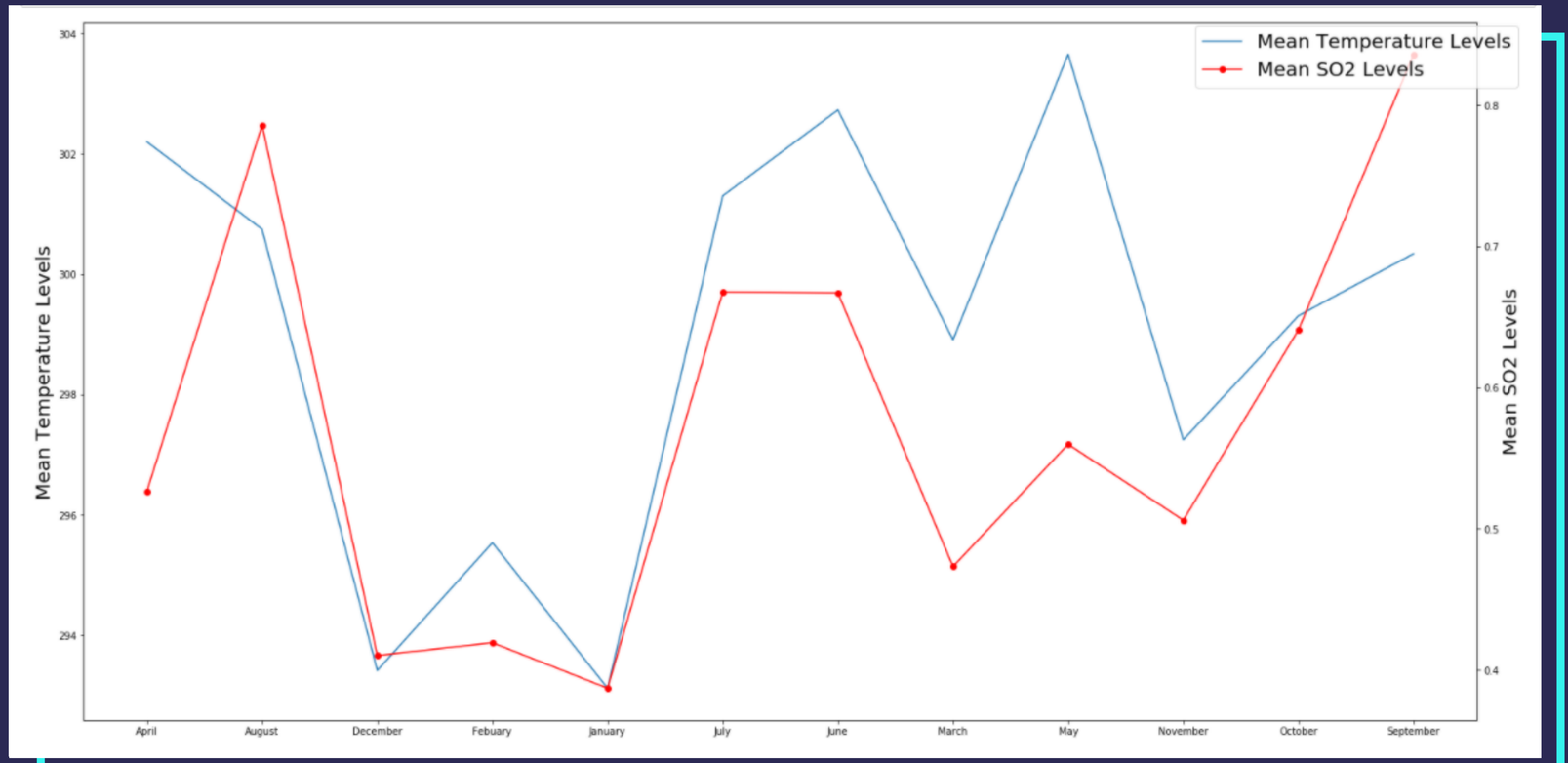


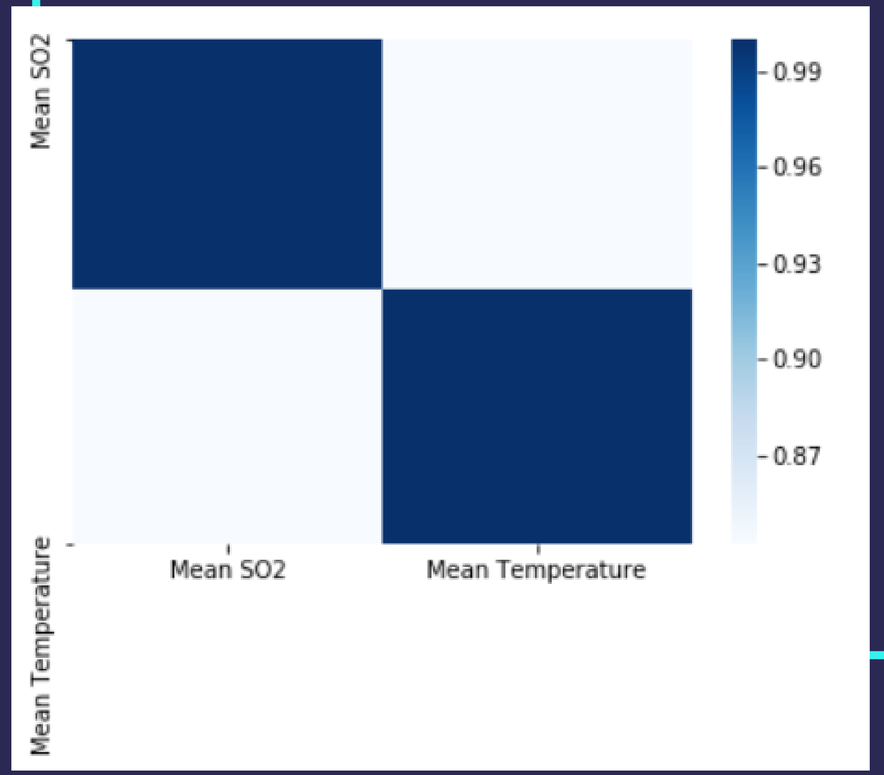
- SO2 level was the usually highest during August-September
- SO2 level was the lowest during December-February

Showing how the mean of 'Mean' value varies according to the month of the year using a line graph for the concatenated dataframe.

CORRELATIONAL ANALYSIS

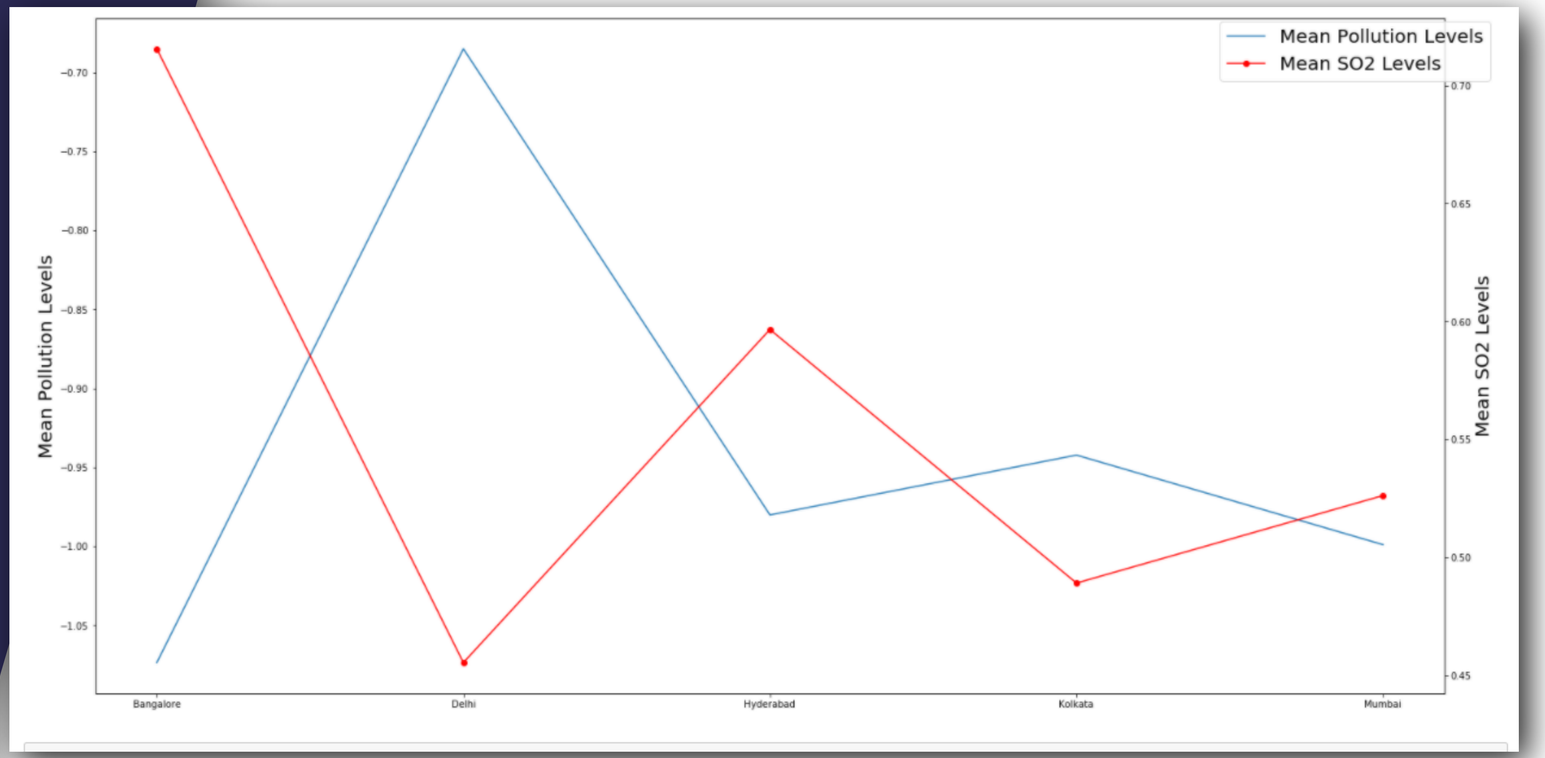
SO2 vs TEMPERATURE
SO2 vs POLLUTION
SO2 vs CARBON MONOXIDE

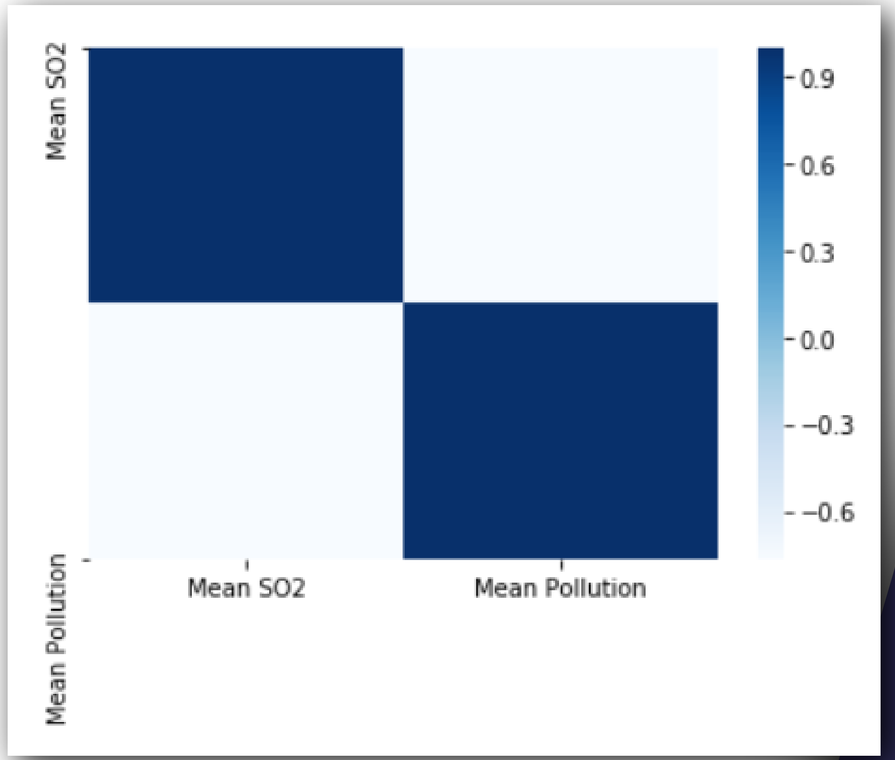




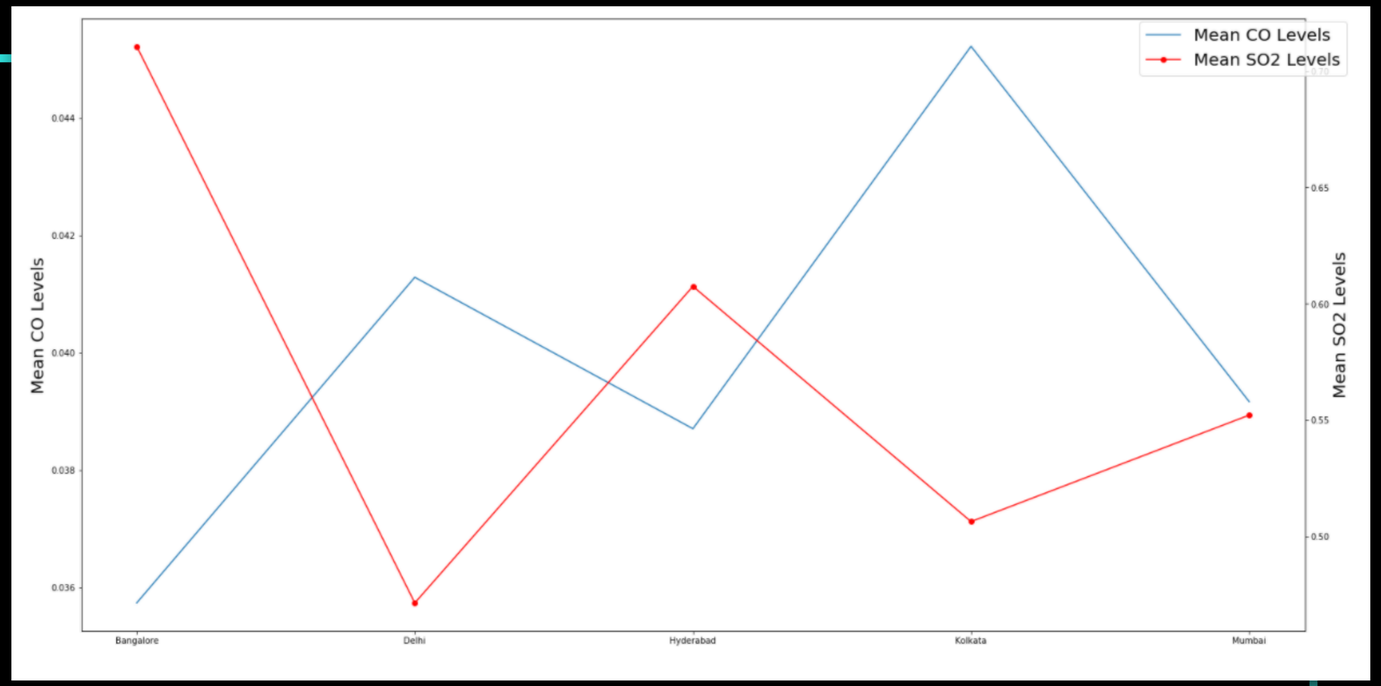
From the above graphs, it is clear that there is a high correlation between the Mean SO2 value and the Temperature.

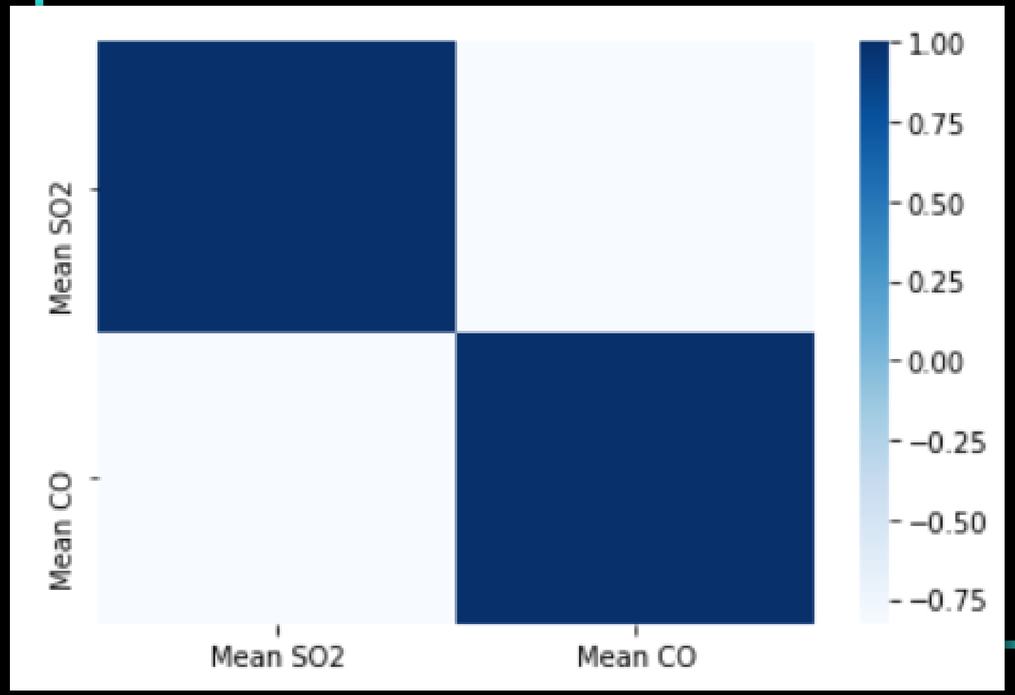
Correlation R value = 0.65927





From the above graphs, it is clear that there is a high correlation between the Mean SO2 value and the Pollution.
Correlation R value = -0.763477





From the above graphs, it is clear that there is a high correlation between the Mean SO2 value and the Carbon Monoxide.
Correlation R value = -0.826897

MACHINE LEARNING MODEL

VS

Random Forest Regression
K-Neighbour Regression
Multivalued Linear Regression
Gradient Boosting

Due to the very strong correlation between Sulphur Dioxide and Pollution we can make confident predictions of Sulphur Dioxide values for a particular day using the pollution levels of that day as an input.
The Top 4 Models along with their Accuracies are:

Gradient Boosting Regression Model:
Accuracy is 0.670

K-Neighbour Regression Model:
Accuracy is 0.669

Random Forest Regression Model:
Accuracy is 0.580

Multivalued Linear Regression Model:
Accuracy is 0.302

Due to the very strong correlation between Sulphur Dioxide and Temperature we can make confident predictions of Sulphur Dioxide values for a particular day using the temperature levels of that day as an input.
The Top 4 Models along with their Accuracies are:

Random Forest Regression Model:
Accuracy is 0.625

Gradient Boosting Regression Model:
Accuracy is 0.605

K-Neighbour Regression Model:
Accuracy is 0.594

Multivalued Linear Regression Model:
Accuracy is 0.296

Due to the very strong correlation between Sulphur Dioxide and Carbon Monoxide we can make confident predictions of Sulphur Dioxide values for a particular day using the Carbon Monoxide levels of that day as an input.
The Top 4 Models along with their Accuracies are:

Gradient Boosting Regression Model:
Accuracy is 0.756

Random Forest Regression Model:
Accuracy is 0.746

K-Neighbour Regression Model:
Accuracy is 0.639

Multivalued Linear Regression Model:
Accuracy is 0.299

Due to the very strong correlation of Sulphur Dioxide with Carbon Monoxide, Pollution and Tempertaure we can make confident predictions of Sulphur Dioxide values for any particular day using these three attributes of that day as an input. As all the variables are related to each other, we can see that our accuracy using all models also increases. The Top 4 Models along with their Accuracies are:

Gradient Boosting Regression Model:
Accuracy is 0.722

Random Forest Regression Model:
Accuracy is 0.998

K-Neighbour Regression Model:
Accuracy is 0.999

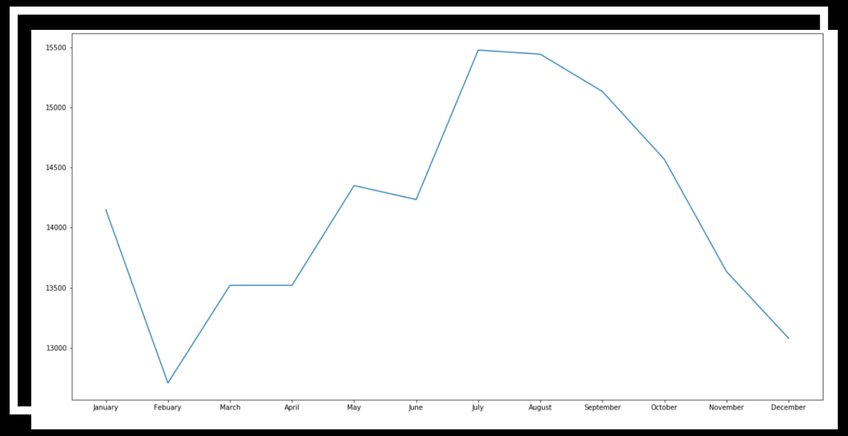
Multivalued Linear Regression Model:
Accuracy is 0.401

VS

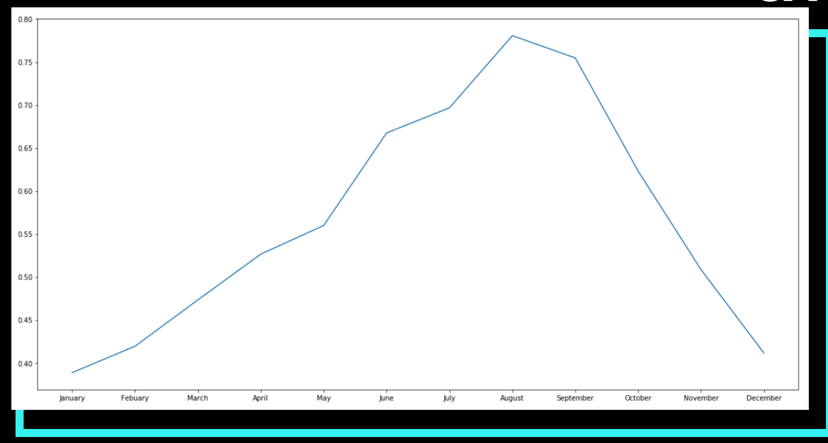
The Data shows that there is high positive corelation between the concentration levels of SO2 in the five cities that we studied, and the number of children born in the Netherlands during the time period of October 2018 to May 2020.

SPURIOUS RELATIONS

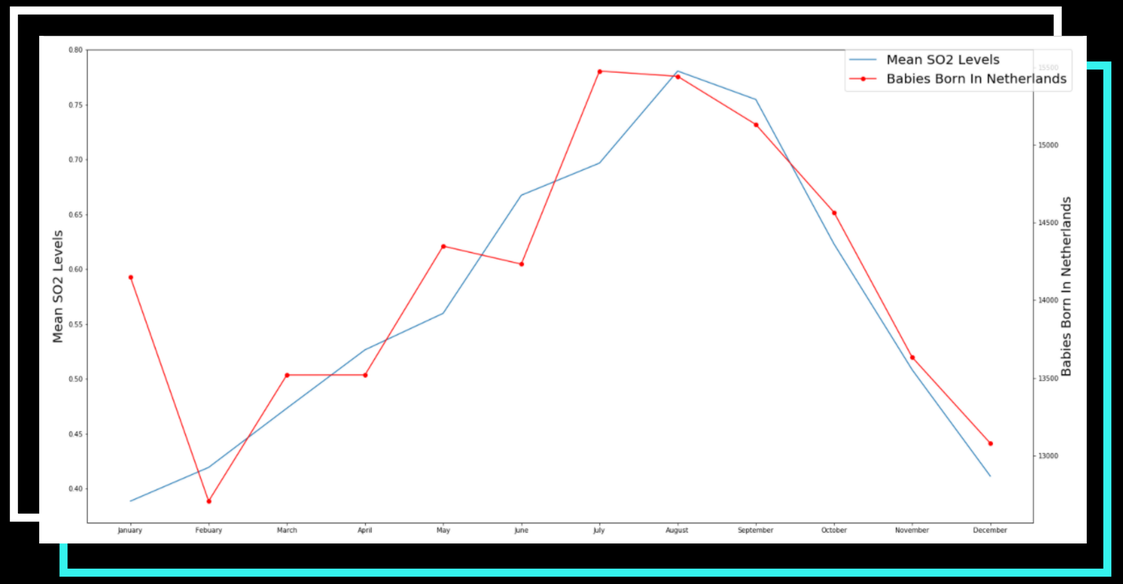
We know there are a lot of young couples in urban cities. When they saw our statistics of how levels of SO2 are very strongly related to Pollution, Temperature and Greenhouse gases like Carbon Monoxide they knew it will be impossible to live in these conditions. So they decided to move to Netherlands being a clean and safe country. But the weather of Netherlands was so cold that they spent most of their time indoors and it has been shown that the chance of having a baby are more if a couple spends more time indoors. Hence we can find a relation between the two statistics.



Number of children born in the Netherlands



Concentration level of SO2 in the cities



Correlation = 82.93%

The Data record of the number of Children born was collected from http://data.un.org/Data.aspx?d=POP&f=tableCode%3A55#f_1

EM

IS CLIMATE CHANGE A MYTH?

SHOULD THE GOVERNMENT BE CONCERNED?

CLIMATE CHANGE: HOW DO WE KNOW?

Scientific evidence for warming of the climate system is unequivocal.
Earth's climate has changed throughout history. Just in the last 650,000 years, there have been seven cycles of glacial advance and retreat, with the abrupt end of the last ice age about 11,700 years ago marking the beginning of the modern climate era — and of human civilization. The study of whether climate change is a myth by studying the pollution levels considering SO2 as a proxy has opened our eyes to the truth.

It was drawn to our attention that the SO2 levels are increasing at an alarming rate and this has its peak during the later part of August in the summer. SO2 is the component of greatest concern and is used as the indicator for the larger group of gaseous sulfur oxides (SOx). SO2 can affect both health and the environment. Short-term exposures to SO2 can harm the human respiratory system and make breathing difficult. People with asthma, particularly children, are sensitive to the effects of SO2. At high concentrations, gaseous SOx can harm trees and plants by damaging foliage and decreasing growth. SO2 and other sulfur oxides can contribute to acid rain which can harm sensitive ecosystems. With our study we have been able to draw strong correlations between Temperature and SO2, CO and SO2 and also between Pollution and SO2. We were able to visualise the effect and the impact each of these attributes has on SO2 levels.

We believe that Climate Change is not a myth and is and is a reality that will hit us hard if we don't react immediately.

VS

Should the government start considering the effects of climate change now? We think, ‘NO’. The government should have been concerned decades ago about Climate Change.

But it's better late than never.

Let’s act now!

THANK YOU