US Pollution Analysis

DETAILED PROJECT REPORT

Shubham Tembhurne

PROJECT DETAIL

Project Title	US Pollution Analysis
Technology	**Business Intelligence**
Domain	**Environment**
Project Difficulty level	**Advance**
Programming Language Used	**Python**
Tools Used	**Jupyter Notebook, MS-Excel, Tableau**

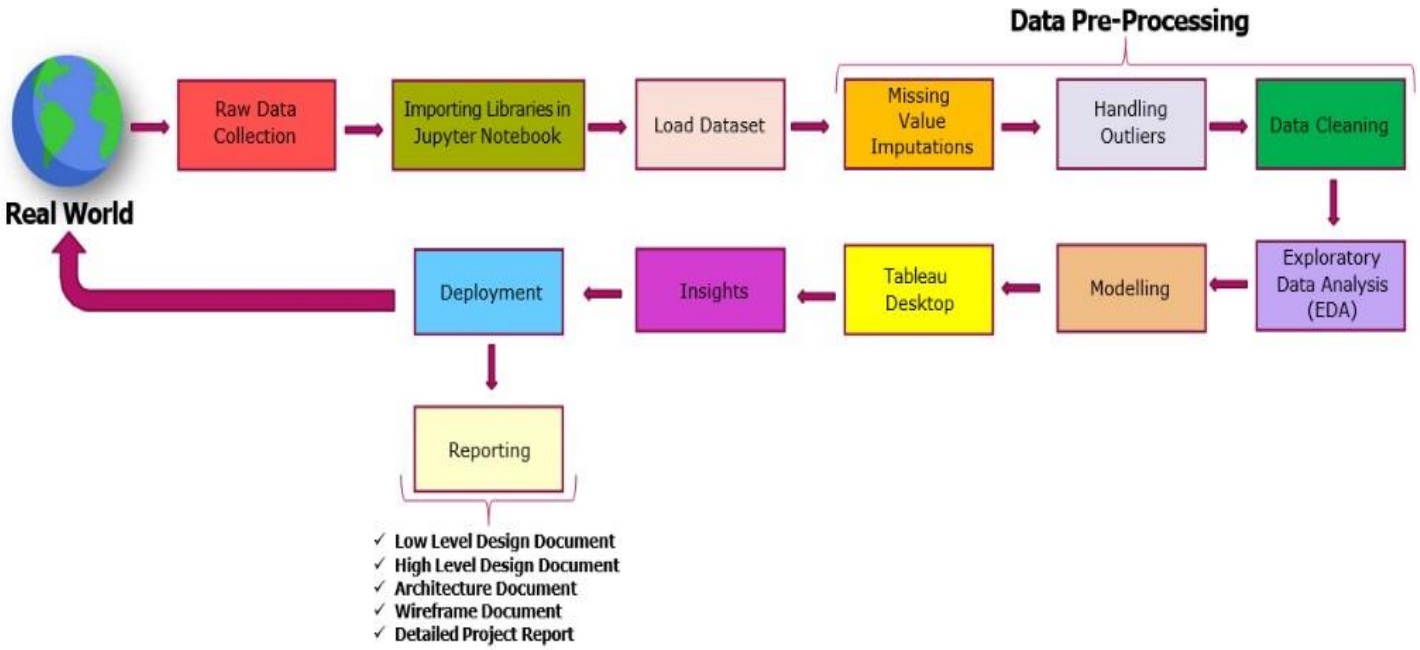
OBJECTIVE

- The goal of this project is to analyze the air pollution is the contamination of air due to the presence of substances in the atmosphere that are harmful to the health of humans and other living beings, or cause damage to the climate or to materials.

PROBLEM STATEMENT

- Air pollution is the contamination of air due to the presence of substances in the atmosphere that are harmful to the health of humans and other living beings, or cause damage to the climate or to materials.
- This dataset provides you the pollution in the U.S. It contains daily data for the four major pollutants NO2, O3, SO2 and CO each has 5 specific columns during 2006 and 2010.

Architecture



DATASET INFORMATION

Abbreviations: Carbon Monoxide (CO), Ozone (O3), Sulphur Dioxide (SO2), Nitrogen Dioxide (NO2) The four pollutants (NO2, O3, SO2 and O3) each has 5 specific columns

State names: Lists of state names in US
State code: Code numbers belong to each state
County code: Unique code numbers specified for county
Site number: numbers given to the sites
Address: The Address of each locality where pollutant has been detected
Country names: List of county names
City names: Names of city
Date local: Day Month and the year of every local area when pollutant took place and the hierarchy
NO2 units: Units represented by Parts Per Billion, NO2 mean, NO2 1st MAX value, NO2 1st MAX hour, NO2 AQI
O3 units: Units represented by parts per million, O3 mean, O3 1st MAX value, O3 1st MAX hour, O3 AQI
SO2 units: Units represented by Parts Per Billion, SO2 mean, SO2 1st MAX value, SO2 1st MAX hour, SO2 AQI
CO units: Units represented by parts per million, CO2 mean, CO2 1st MAX value, CO2 1st MAX hour, CO2 AQI

Why these Parameter are important?

1.Location Metadata

State Code, County Code, Site Num: Help uniquely identify the monitoring location.

Address, State, County, City: Provide human-readable location information for spatial analysis, policy targeting, and regional pollution trends.

Importance: These are crucial for geospatial analysis and identifying pollution hotspots.

2.Date

Date Local: Specifies the date of observation.

Importance: Enables temporal analysis to detect trends, seasonal patterns, or year-over-year changes in air quality.

3.Pollution Metrics by Gas For each of these gases, the dataset includes:

Units (e.g., parts per billion/million)

Mean concentration

1st Max Value: Peak value observed during the day.

1st Max Hour: Hour of peak pollution.

AQI (Air Quality Index): Standardized index to represent pollution severity

4.Gases Tracked:

NO_2 (Nitrogen Dioxide): Harmful to respiratory health; indicates traffic and industrial emissions.

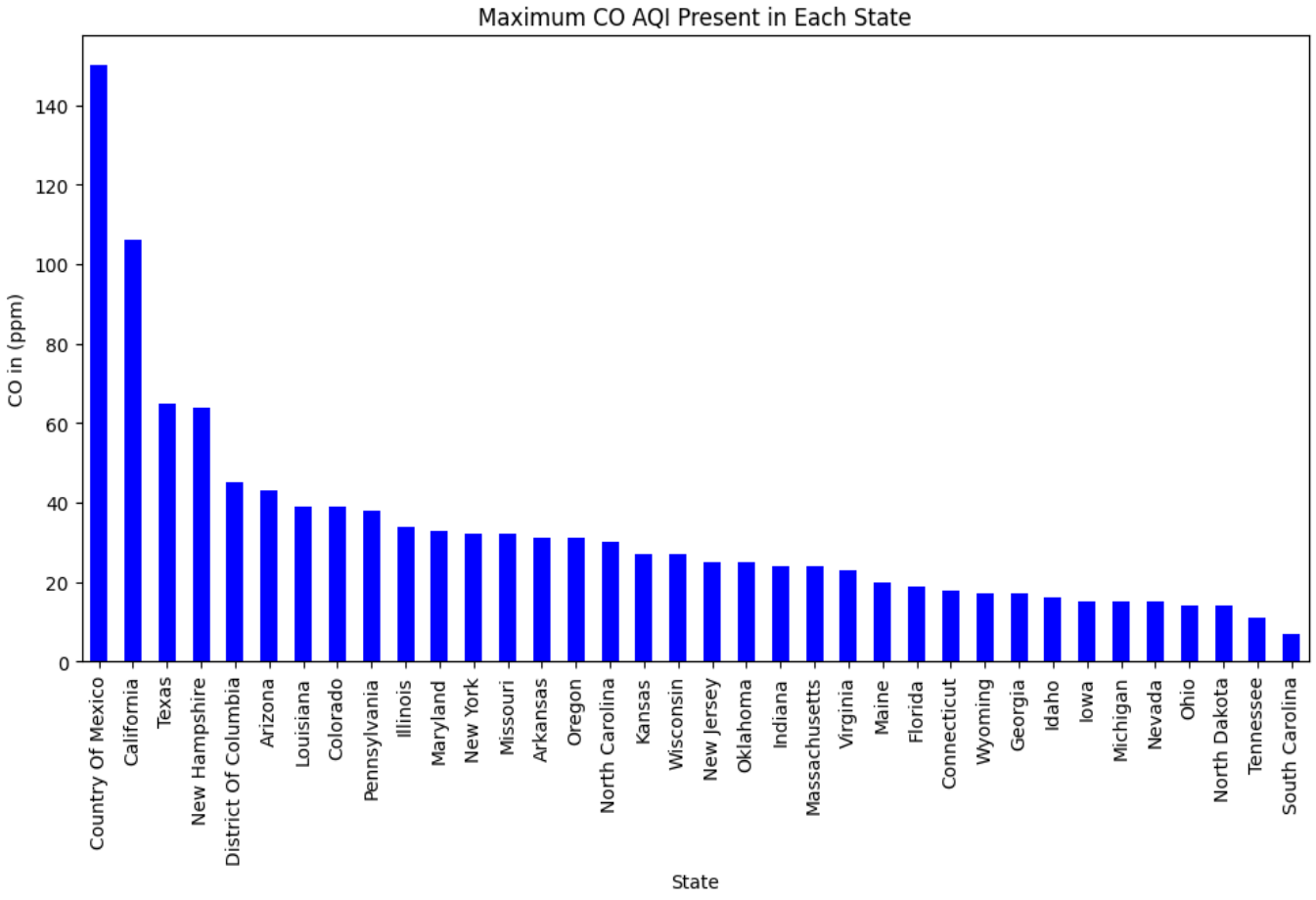
O_3 (Ozone): Ground-level ozone is a key component of smog; harmful to lungs.

SO_2 (Sulfur Dioxide): Results from burning fossil fuels; can cause acid rain.

CO (Carbon Monoxide): Emitted from combustion; dangerous at high concentrations.

INSIGHTS

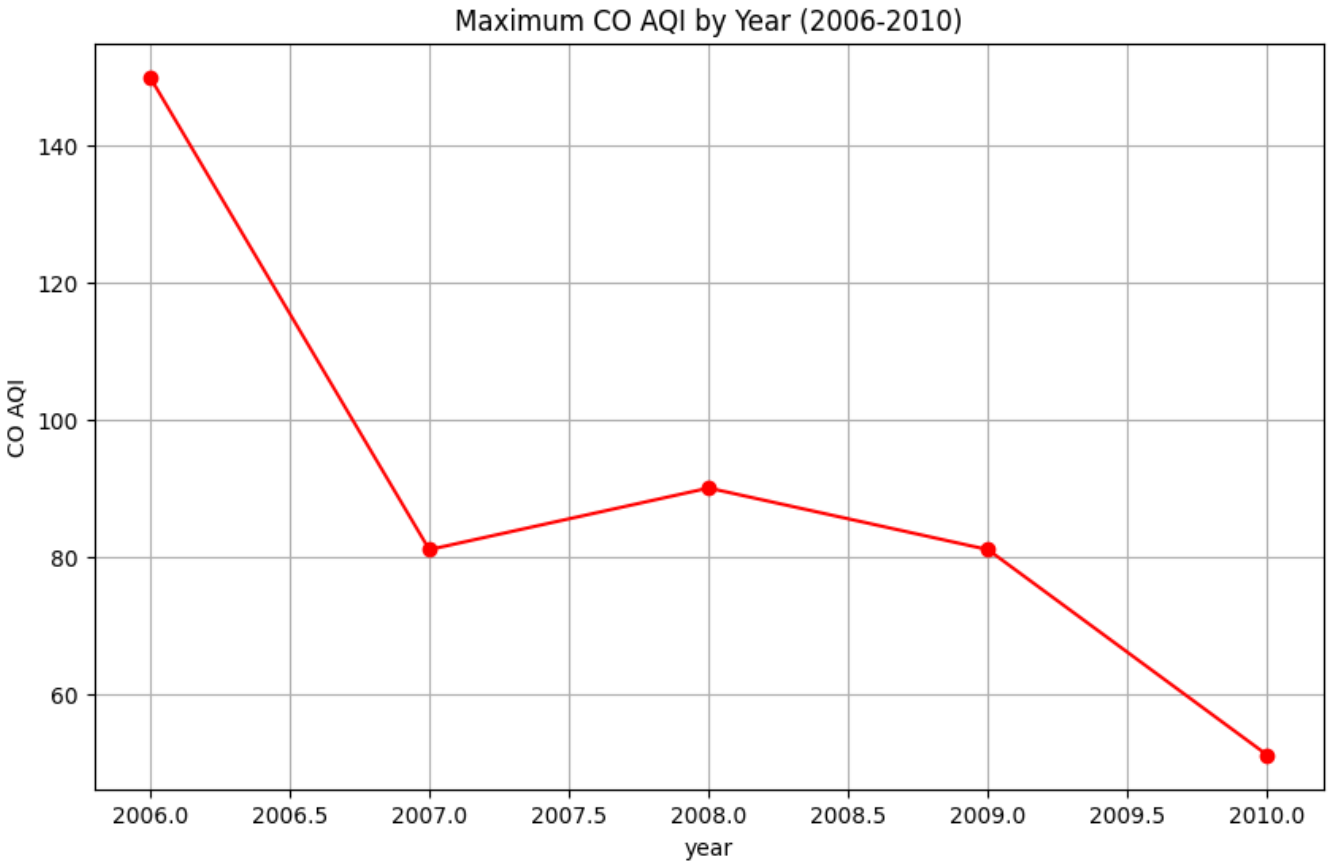
1. Which state have the maximum carbon monoxide CO AQI Present?



- Country of Mexico, California, Texas, New Hampshire are the states with maximum CO AQI present and highly polluted with CO.

- South Carolina, North Dakota, Tennessee are the states with minimum CO AQI present and least polluted with CO.

2. **How the maximum carbon monoxide CO changes with respective to year?**



- Over the year from 2006 to 2010 CO AQI gradually depleting and it came close to 80 ppm in 2009 and slightly Decreased in 2010.Overall change in air quality.

3. What are the yearly statistics for carbon monoxide CO?

Yearly CO Statistics:

	CO Mean			CO AQI		
	mean	min	max	mean	min	max
Date Local						
2006	0.41	0.0	6.53	3.48	0.0	150.0
2007	0.37	0.0	6.50	3.00	0.0	81.0
2008	0.33	0.0	3.82	2.67	0.0	90.0
2009	0.32	0.0	4.05	2.58	0.0	81.0
2010	0.34	0.0	3.56	2.69	0.0	51.0

- Highest CO levels were recorded in 2006 (max 6.53 ppm)
- There's a general declining trend in average CO levels from 2006 to 2010
- The highest AQI value of 150 was recorded in 2006
- Recent years show lower maximum values, suggesting improved air quality.

4. Which are the Top 10 Highest State of CO Pollutants?

```
Top 10 Highest state of CO Pollutants :
State
Country Of Mexico       9.188815
District Of Columbia    7.076041
Colorado                4.280659
Arizona                 4.250470
Georgia                 3.396739
Arkansas                3.355925
Missouri                3.303599
California              3.235519
Kansas                  3.166492
Nevada                  3.150000
Name: CO AQI, dtype: float64
```

- Country of Mexico, District of Columbia, Colorado, Arizona, Georgia, Arkansas, Missouri, California, Kansas, Nevada.

5. What is the first maximum hour for the Carbon monoxide CO?

```
First Maximum Hours for CO
23
```

- The first maximum hour for CO is 23.

6. What is the maximum value for the carbon monoxide CO?

```
First Maximum Value for CO
16.5
```

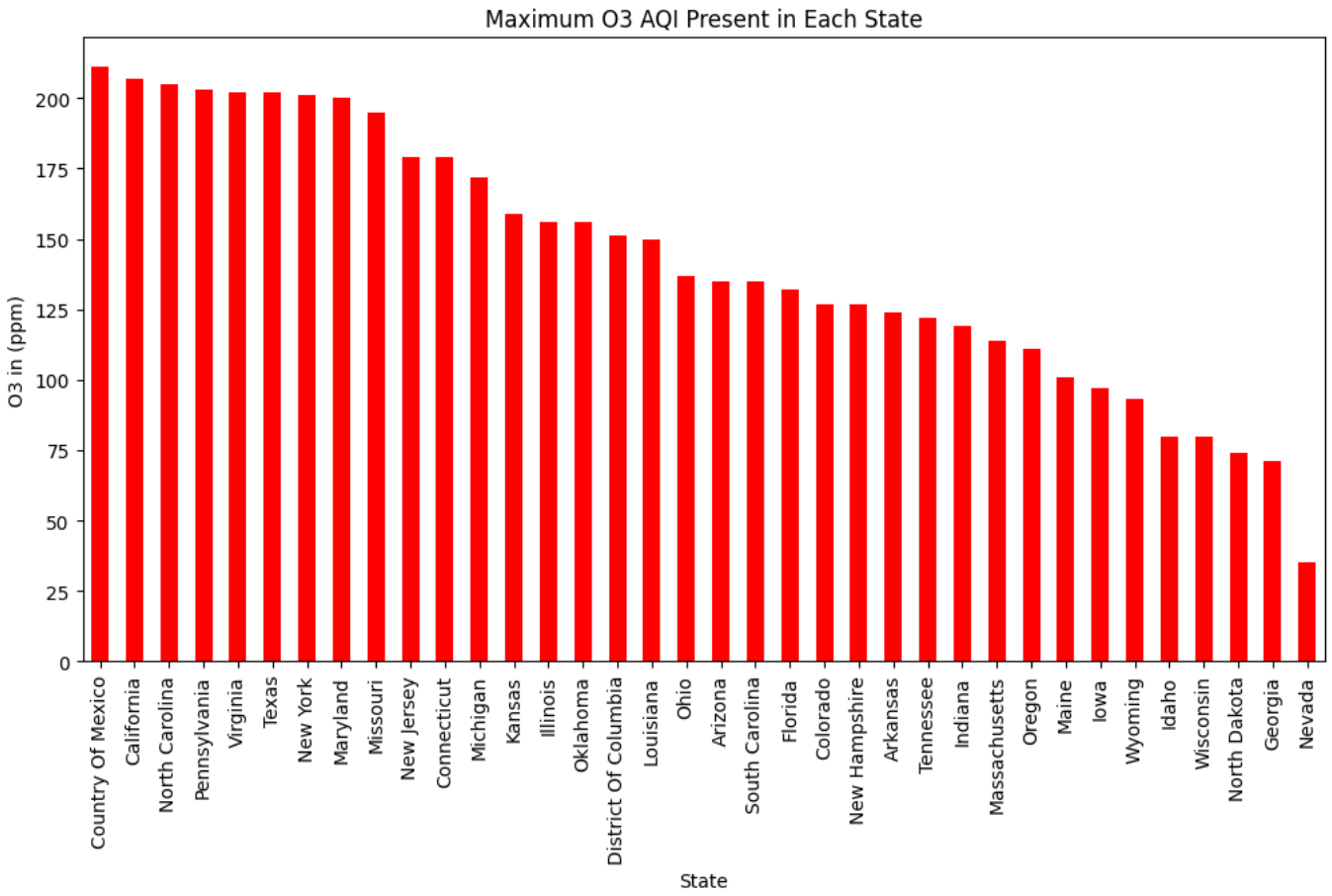
- The first maximum value for the CO is 16.5

7. Which is the highest city with CO pollutants?

```
Highest City of CO Pollutants :
City
Mexicali    11.257726
Name: CO AQI, dtype: float64
```

- The Highest City of CO pollutants is Mexicali.

9.Which State have the maximum ozone O3 AQI present?



- The Country of Mexico, California, North California, Pennsylvania and Virginia have the maximum O3 AQI present and highly polluted with O3.
- Nevada, Georgia, North Dakota, Wisconsin are the minimum O3 AQI present and least polluted with O3.

11.How the maximum ozone O3 changes with respective to year?

- Over the year from 2006 to 2010 O3 AQI gradually depleting and it came to 206 in 2009 and in 2010 decrease to 200.Overall change in air quality.

12.What are the yearly statistics for the ozone O3?

Yearly O3 Statistics:

	O3 Mean			O3 AQI		
	mean	min	max	mean	min	max
Date Local						
2006	0.03	0.0	0.07	36.20	0	206
2007	0.03	0.0	0.07	36.39	0	211
2008	0.03	0.0	0.08	35.26	0	207
2009	0.03	0.0	0.07	33.02	0	206
2010	0.03	0.0	0.07	34.33	0	200

- Highest O3 levels were recorded in 2008 (max 0.08 ppm)
- There's a general Constant trend in average O3 levels from 2006 to 2010
- The highest AQI value of 211 was recorded in 2007
- Recent years show lower maximum values, suggesting improved air quality

8. Which are the top 10 Highest State of O3 Pollutants?

```
Top 10 Highest state of O3 Pollutants :
State
North Carolina    42.733760
Missouri          42.663537
Wyoming           40.619048
Pennsylvania      39.745689
Indiana           38.481654
Arizona           38.314290
Ohio              37.996132
Virginia          37.780835
Tennessee         37.469716
South Carolina    37.157018
Name: O3 AQI, dtype: float64
```

- North Carolina, Missouri, Wyoming, Pennsylvania, Indiana, Arizona, Ohia, Virginia, Tennessee, South Carolina .

10. What is the first maximum hour for Ozone O3?

```
First Maximum Hours for O3
10
```

- **The first maximum hour for O3 is 10**

11.What is the first maximum value for O3?

```
First Maximum Value for O3
0.141
```

- **The first maximum value for O3 is 0.141**

12.Which is the highest city of O3 pollutants?

```
Highest City of O3 Pollutants :
City
Rubidoux    53.597239
Name: O3 AQI, dtype: float64
```

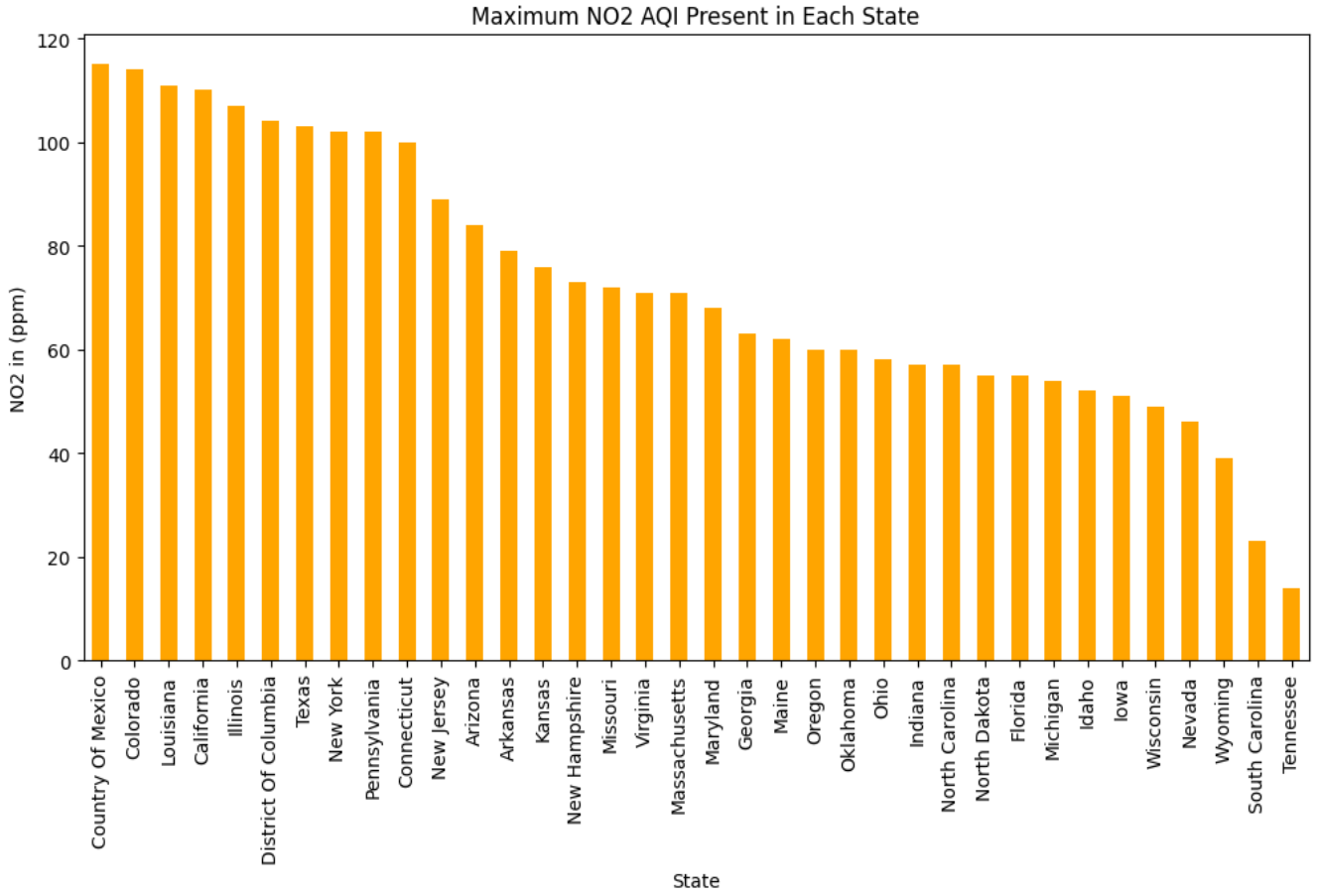
- **The highest city of O3 pollutants is Rubidoux**

13.Which are the top 10 highest state of nitrogen dioxide NO2 pollutants?

```
Top 10 Highest state of NO2 Pollutants
State
Country Of Mexico       38.888704
Colorado                38.734909
Arizona                 35.388709
Nevada                  34.400000
Georgia                 30.739130
District Of Columbia    30.678781
New Jersey              29.903587
Massachusetts           29.729207
Missouri                29.322379
New York                27.565746
Name: NO2 AQI, dtype: float64
```

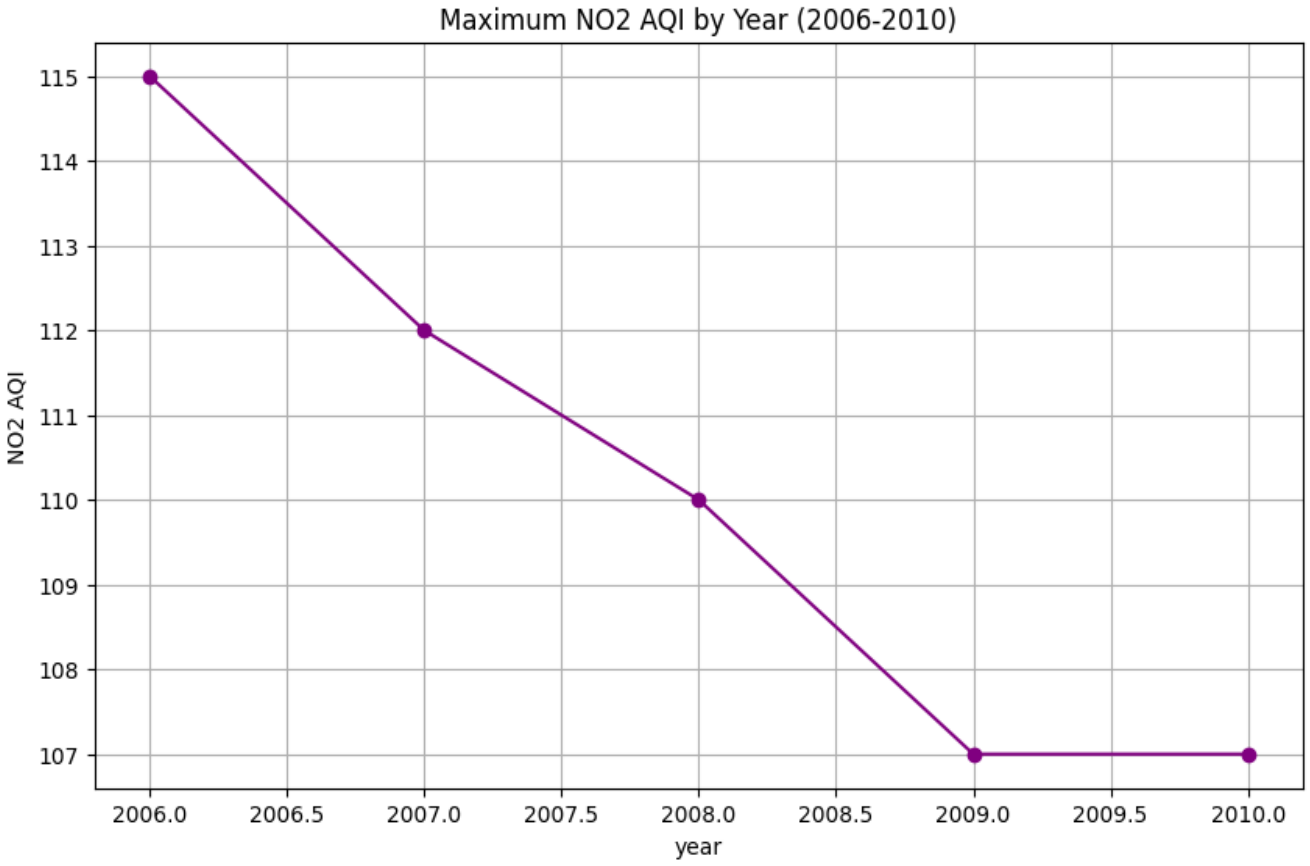
- Country of Mexico, Colorado, Arizona, Nevada, Georgia, District of Columbia, New Jersey, Massachusetts, Missouri, New York

14. Which state have the maximum nitrogen Dioxide NO2 AQI present?



- Country Of Mexico, Colorado, Louisiana, California have the maximum NO2 AQI present and Wyoming, South Carolina, Tennessee with least NO2 AQI present.

15. How the maximum Nitrogen Dioxide NO2 changes with respective to year?



- Over the year from 2006 to 2009 NO2 AQI gradually depleting and it's came to 107 in 2009 and from 2019 it remains constant. Overall change in air quality.

16.Which is the Highest city of NO2 pollutants?

```
Highest City of NO2 Pollutants :
City
Denver    44.030486
Name: NO2 AQI, dtype: float64
```

- The highest city of NO2 Pollutant is Denver

17. What are the yearly statistics for nitrogen dioxide NO2?

Yearly NO2 Statistics:

	NO2 Mean			NO2 AQI		
	mean	min	max	mean	min	max
Date Local						
2006	14.53	0.0	70.59	26.99	0	115
2007	13.25	0.0	92.00	24.91	0	112
2008	12.25	0.0	76.45	23.50	0	110
2009	11.46	0.0	60.61	21.73	0	107
2010	11.66	0.0	56.91	22.29	0	107

- Highest NO2 levels were recorded in 2007 (max 92 ppm)
- There's a general Decrease trend in average NO2 levels from 2006 to 2010
- The highest AQI value of 115 was recorded in 2006
- Recent years show lower maximum values, suggesting improved air quality

18.What is the first maximum hour for NO2?

```
First Maximum Hours for NO2
2
```

- The first Maximum hour for NO2 is 2.

19.What is the first maximum value for NO2?

```
First Maximum Value for NO2
176.0
```

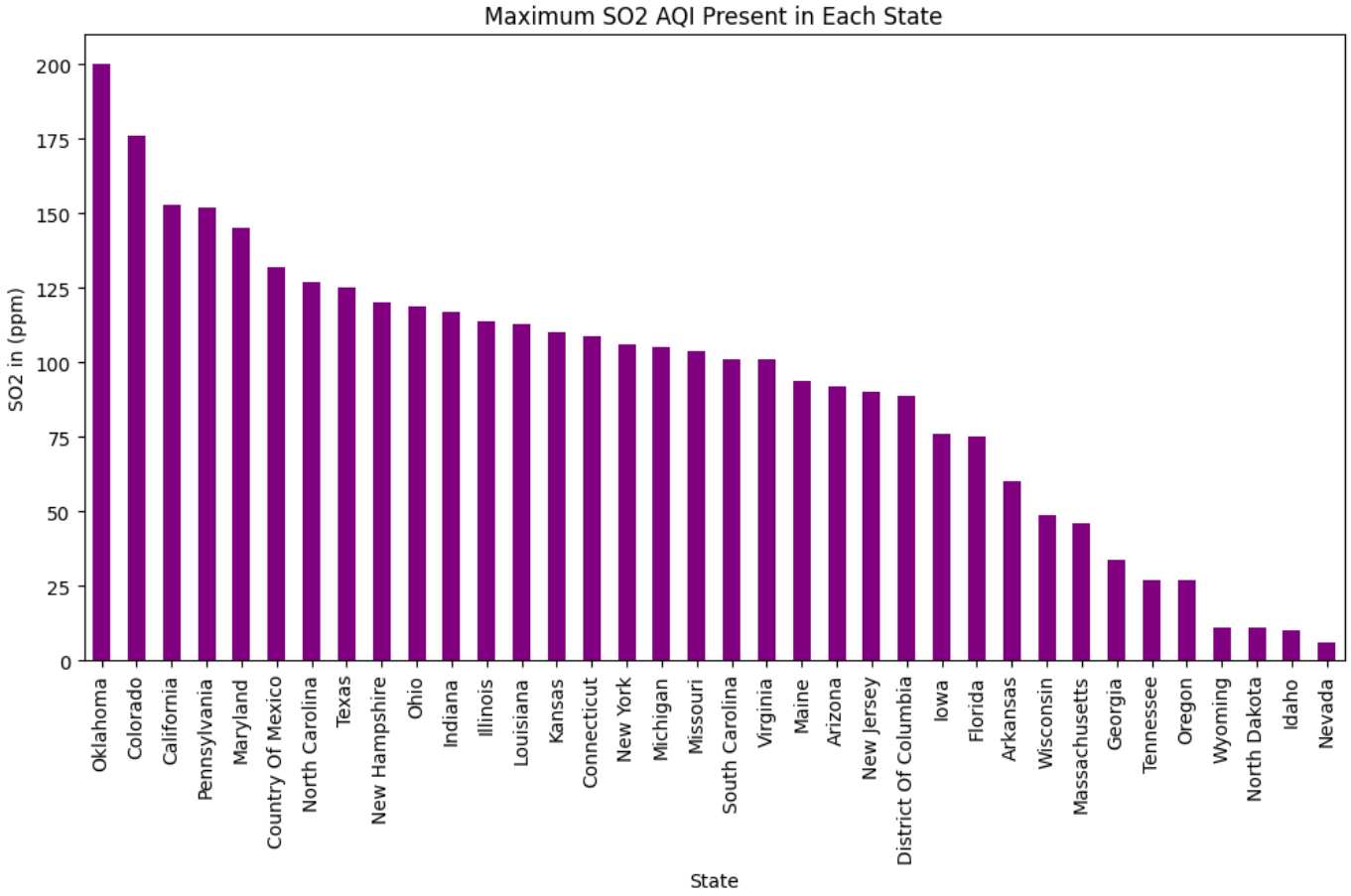
- The first maximum value for NO2 is 176

20. Which are the top 10 highest state of Sulphur dioxide SO2 pollutants?

```
Top 10 Highest state of SO2 Pollutants :
State
Missouri                9.840376
Pennsylvania            8.586706
Maryland                7.080318
New Jersey              7.018570
District Of Columbia    6.830165
Indiana                 6.786255
Kansas                  6.699790
Michigan                6.368272
Ohio                    6.253385
New York                6.251107
Name: SO2 AQI, dtype: float64
```

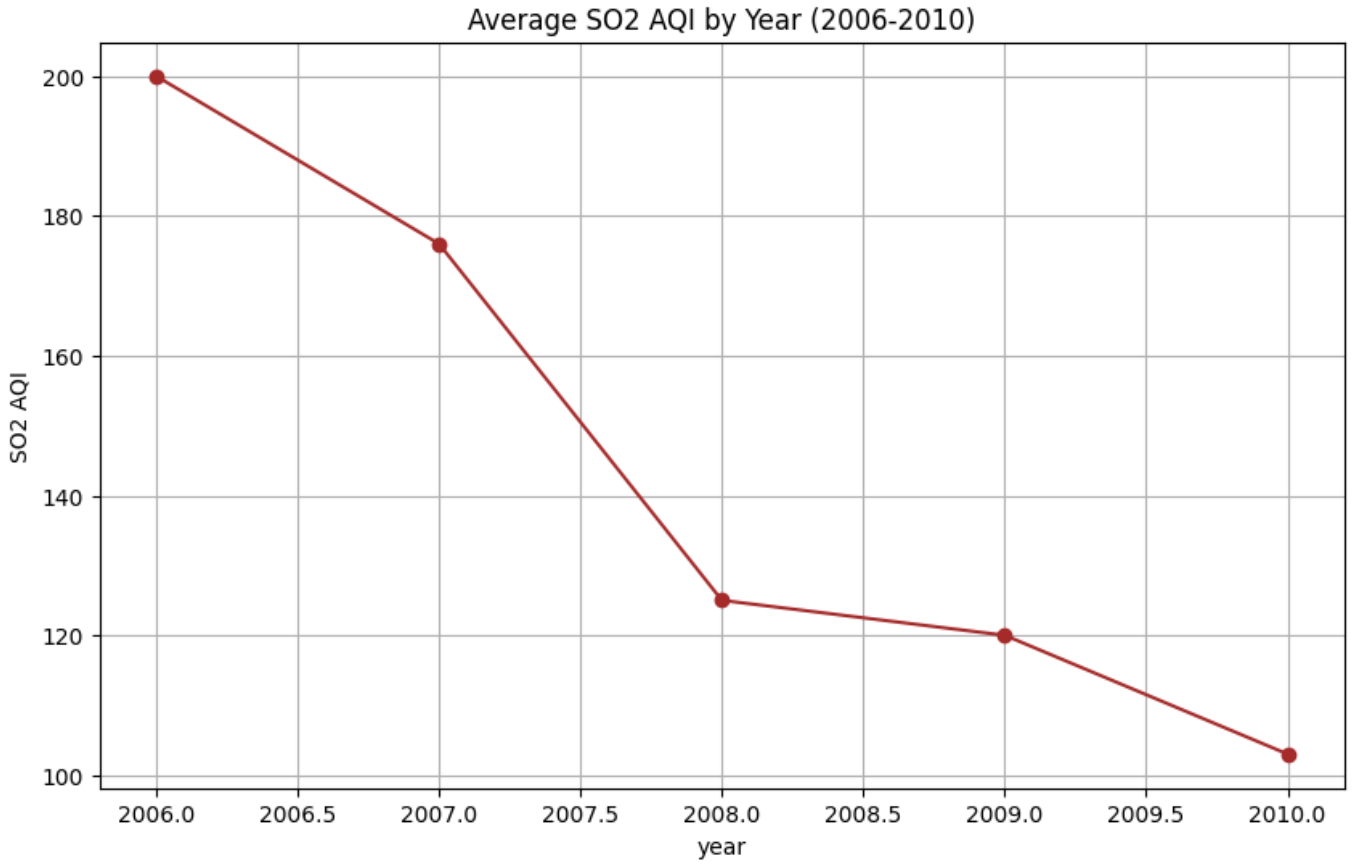
- Missouri, Pennsylvania, Maryland, New Jersey, District of Columbia, Indiana, Kansas, Michigan, Ohio, New York are the highest state of SO2 pollutants.

21. Which state have the maximum Sulphur dioxide SO2 AQI present?



- The Oklahoma, Colorado, California, Pennsylvania have the maximum SO2 AQI present and Idaho, Nevada, North Dakota are with minimum SO2 AQI present.

22. How the maximum Sulphur dioxide SO2 changes with respective to year?



- Over the year from 2006 to 2010 SO2 AQI gradually depleting and it came close to 120 in 2009 and above 100 in 2010. Overall change in air quality

23. Which is the Highest city of SO2 pollutants?

```
Highest City of SO2 Pollutants :
City
Beaver Falls    12.606383
Name: SO2 AQI, dtype: float64
```

- **The highest city of SO2 pollutants is Beaver Falls**

24. What are the yearly statistics for Sulphur dioxide SO2?

```
Yearly SO2 Statistics:
           SO2 Mean                SO2 AQI
               mean   min     max    mean  min    max
Date Local
2006           2.60  0.00  321.62    5.00  0.0  200.0
2007           2.34  0.00   32.04    4.53  0.0  176.0
2008           1.91  0.00   31.00    3.68  0.0  125.0
2009           1.56  0.00   33.00    2.75  0.0  120.0
2010           1.18 -0.45   25.91    1.91  0.0  103.0
```

- Highest SO2 levels were recorded in 2006 (max 321.62 ppm)
- There's a general decrease trend in average SO2 levels from 2006 to 2010
- The highest AQI value of 200 was recorded in 2006
- Recent years show lower maximum values, suggesting improved air quality

25.What is the first maximum hour for SO2?

```
First Maximum Hours for SO2
0
```

- **The First Maximum hour for SO2 is 0**

26.What is the first maximum value for SO2?

```
First Maximum Value for SO2
351.0
```

- **The First Maximum Value for SO2 is 351.**

26.What is the percentage distribution of AQI categories?

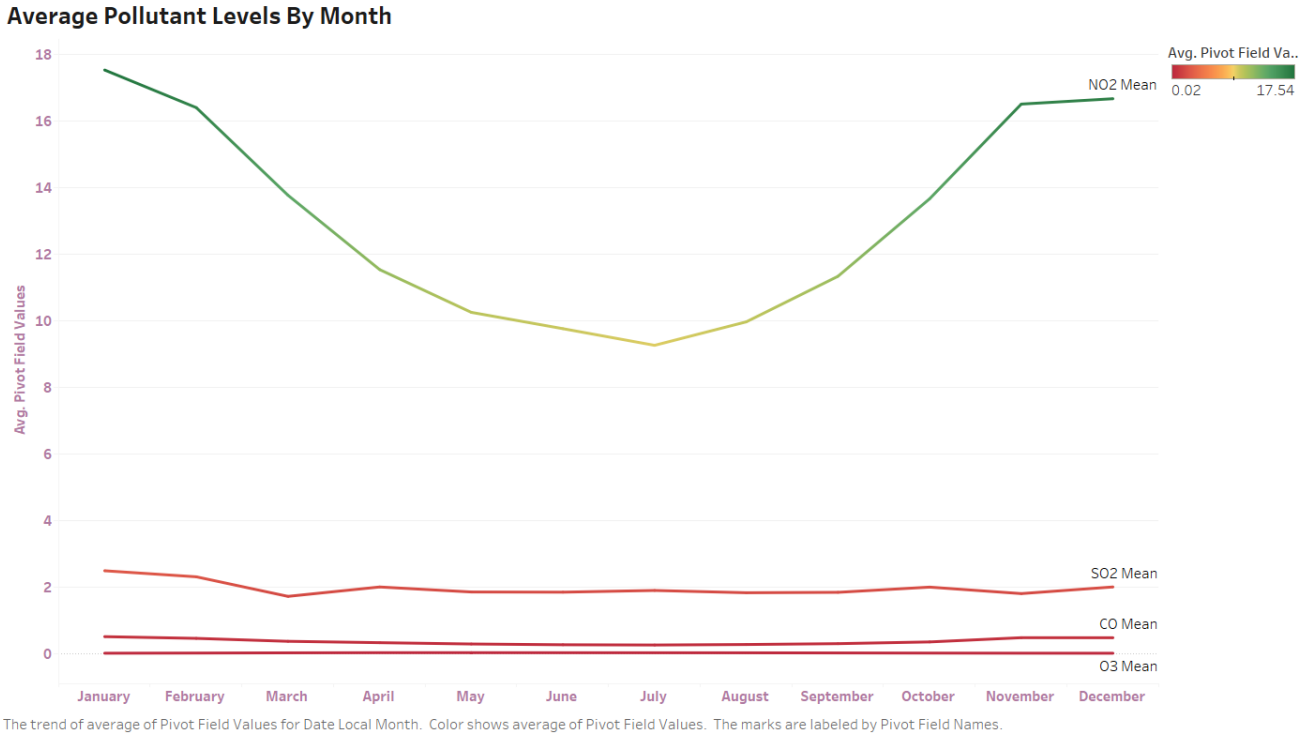
```
Percentage Distribution of AQI Categories:
                                  NO2     O3    SO2     CO
Good                            95.59  90.61  99.37  99.93
Hazardous                         NaN   0.01    NaN    NaN
Moderate                         4.34   7.76   0.58   0.06
Unhealthy for Sensitive Groups   0.08   1.48   0.05   0.00
Very Unhealthy                    NaN   0.13   0.00    NaN
```

- NO2 (Nitrogen Dioxide):95.59% of readings fall in the "Good" category
  Only about 4.34% reach "Moderate" levels
  Very few instances (0.08%) of "Unhealthy for Sensitive Groups"
- O3 (Ozone): Most diverse distribution among all pollutants
  90.61% “Good" reading
  7.76% 'Moderate' Levels
  1.48% "Unhealthy for sensitive Groups"

Small but notable occurrences of "Unhealthy" (0.13%) and "Very Unhealthy" conditions

- SO2 (Sulfur Dioxide): About 99.37% "Good" readings values
  Very few instances (0.58%) of "Moderate" levels
  Minimal occurrences of higher categories
- CO (Carbon Monoxide): Similar to SO2 with nearly equal between "Good"
  No recorded instances in higher AQI categories
- This analysis suggests that while most readings fall within safe levels, Ozone (O3) shows the most concerning pattern with more frequent occurrences of unhealthy conditions compared to other pollutants.

27.what is the Average pollutant level by month?



- NO2 and CO are highest in winter months and lowest in summer.
- Ozone (O3) peaks in late spring and early summer, then drops in winter.
- SO2 is slightly higher in winter and early spring.

28.What is top 10 highest city of Carbon monoxide CO pollutants?

```
Top 10 Highest City of CO Pollutants :
City
Mexicali      11.257726
Tijuana        9.460539
Calexico       7.133277
Washington     7.076041
San Diego      5.589106
Denver         5.344913
Phoenix        5.209843
Kenner         4.944444
Burbank        4.796969
Rosarito       4.504878
Name: CO AQI, dtype: float64
```

- Mexicali, Tijuana, Calexico, Washington, San Diego, Denver, Phoenix, Kenner, Burbank, Rosarito are the highest CO pollutant

29. What is top 10 highest city of ozone O3 pollutants?

```
Top 10 Highest City of O3 Pollutants :
City
Rubidoux       53.597239
Fresno         52.027516
Fontana        50.740816
Victorville    49.132291
Norristown     47.654971
Lancaster      47.453366
Bristol        44.568507
Westport       43.505415
York           42.927570
Charlotte      42.733760
Name: O3 AQI, dtype: float64
```

- Rubidoux, Fresno, Victorville, Norristown, Lancaster, Bristol, Westport, York, Charlotte highest city of O3 pollutant.

30. What is top 10 highest city of nitrogen dioxide NO2 pollutants?

```
Top 10 Highest City of NO2 Pollutants :
City
Denver         44.030486
Mexicali       43.570262
Tijuana        42.143918
Burbank        40.069328
Phoenix        39.170227
Fontana        37.197959
Scottsdale     37.056813
Welby          36.289129
New York       36.220152
Los Angeles    35.142508
Name: NO2 AQI, dtype: float64
```

- Denver, Mexicali, Tijuana, Burbank, Phoenix, Fontana, Scottsdale, Welby, New York, Los Angeles highest city of NO2 Pollutant.

31.What is the top 10 highest city of Sulphur dioxide SO2 pollutant?

```
Top 10 Highest City of SO2 Pollutants :
City
Beaver Falls    12.606383
Reading         12.424242
Charleroi       11.574881
Greensburg      11.456471
New Castle      11.194400
York            11.089369
Detroit         10.738095
Johnstown       10.352976
Pittsburgh      10.009772
St. Louis        9.840376
Name: SO2 AQI, dtype: float64
```

- Beaver Falls, Reading, Charleroi, Greensburg, New Castle, York, Detroit, Johnstown, Pittsburgh, St. Louis are the Highest City of SO2 Pollutant

KEY PERFORMANCE INDICATOR(KPI)

1. Average Pollutants by Month
2. Percentage distribution of AQI Categories.
3. Yearly Statistics for CO, NO2, SO2 and O3.
4. Highest Cities for CO, NO2, SO2, O3 Pollutants
5. Maximum CO, NO2, SO2 and O3 AQI by Year (2006-2010)
6. Maximum CO, NO2, SO2 and O3 AQI in each state
7. Highest States for CO, NO2, SO2 and O3 Pollutants

CONCLUSION

1. Geographic Hotspots

- The most polluted cities are mainly large urban centers, especially in Colorado (Denver), California (Burbank, Los Angeles), and along the US-Mexico border (Tijuana, Mexicali).
- These cities consistently show higher average NO2 and CO levels, which are linked to heavy traffic and urban activity.
- Western states generally have higher pollution levels than many eastern states, likely due to a combination of population density, geography, and industrial activity.

---

2. Seasonal Patterns

- NO2 and CO concentrations are highest in winter (December–February). This is likely due to increased heating, more vehicle idling, and atmospheric conditions (like temperature inversions) that trap pollutants near the ground.
- Ozone (O3) peaks in the summer months, driven by sunlight and heat, which promote ozone formation.
- SO2 patterns are more variable, often spiking in areas with specific industrial activities.

3. Yearly Pattern

From 2006 to 2010, the Air Quality Index (AQI) for major pollutants showed a general downward trend, indicating an overall improvement in air quality:

- **Carbon Monoxide (CO)** AQI steadily decreased, nearing 80 in 2009 and slightly dropping further in 2010.
- **Ozone (O_3)** AQI showed a consistent decline, reaching around 206 in 2009 and dropping below 200 in 2010.
- **Sulfur Dioxide (SO_2)** AQI gradually decreased to about 120 in 2009 and remained slightly above 100 in 2010.

- **Nitrogen Dioxide (NO_2)** AQI declined from 2006 to 2009, reaching 107, and then dropped significantly to 22 by 2010, maintaining that level through 2019.
- These trends suggest effective air pollution control measures during this period, with the most significant long-term improvement observed in NO_2 levels.

4.Yearly Statistic for Pollutants

Between 2006 and 2010, air quality showed a noticeable improvement across multiple pollutants:

- Carbon Monoxide (CO): CO levels peaked in 2006, but both average and maximum values declined steadily through 2010, indicating a consistent improvement.
- Ozone (O_3): Although the highest levels were recorded in 2008, average O_3 levels remained relatively constant. The highest AQI related to O_3 occurred in 2007, but subsequent years saw reduced extremes.
- Nitrogen Dioxide (NO_2): NO_2 levels peaked in 2007, yet average concentrations declined from 2006 to 2010, reflecting better air quality control over time.
- Sulfur Dioxide (SO_2): SO_2 levels were highest in 2006, with a clear decreasing trend in both average levels and peak values in later years.
- Air Quality Index (AQI): The highest AQI values for each pollutant were mostly concentrated in 2006–2007, with noticeably lower values in later years.

5. Public Health Implications

- Urban populations are exposed to higher and more consistent levels of multiple pollutants.
- Seasonal changes mean that different groups are at risk at different times of year (e.g., winter for NO2/CO, summer for O3).
- Border regions face unique challenges, as pollution can cross international lines.

- Industrial areas show distinct pollution spikes, especially for SO2.

---

6. Recommendations

- Focus emission controls and public health interventions on winter months in urban areas.
- Implement stricter vehicle emission standards in cities with high NO2 and CO.
- Develop cross-border pollution management strategies, especially for cities like Tijuana and Mexicali.
- Enhance monitoring and regulation in industrial zones with high SO2.

7.Percentage distribution of AQI categories

- The overall air quality is generally safe, with the vast majority of pollutant readings—particularly for **NO_2**, **SO_2**, and **CO**—falling within the "Good" category.
- However, **Ozone (O_3)** stands out as the pollutant of most concern, showing the widest distribution across air quality categories and the highest percentage of readings in unhealthy ranges
- . This indicates that while short-term exposure to most pollutants may not pose a significant health risk, **O_3 requires closer monitoring and potential mitigation efforts**, especially due to its more frequent impact on sensitive groups and occasional spikes into unhealthy zones.

Q & A

Q1) What's the source of data?

Ans) The Dataset was taken from Kaggle

Q2) What was the type of data?

Ans) The data was a combination of numerical and Categorical values.

Q3) What's the complete flow you followed in this Project?

Ans) Refer slide 5th for better Understanding

Q4) What techniques were you using for data?

Ans) -Removing unwanted attributes

-Visualizing relation of independent variables with each other and output variables

-Removing outliers

-Cleaning data and imputing if null values are present.

-Converting Numerical data into Categorical values.

Q6) What were the libraries that you used in Python?

Ans) I used Pandas, NumPy, Matplotlib, **and Seaborn libraries in Pandas.**

THANK YOU