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**

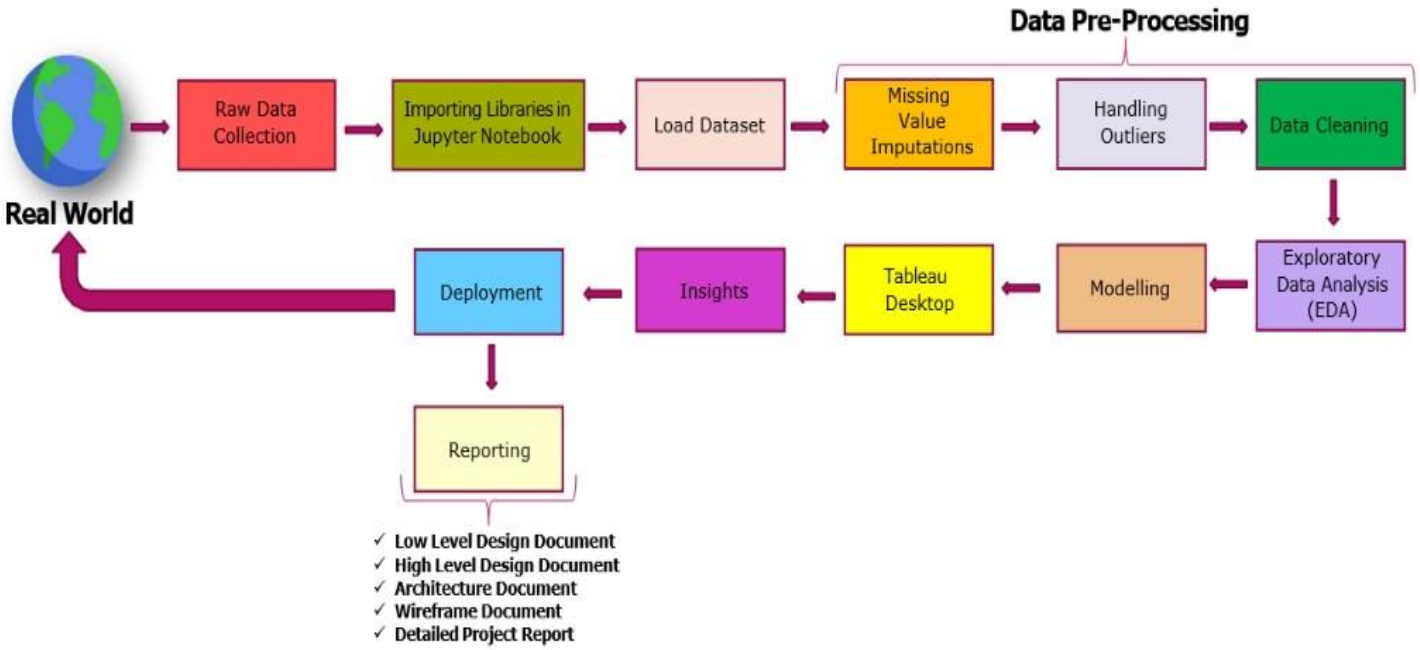
OBJECTIVE

- 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.

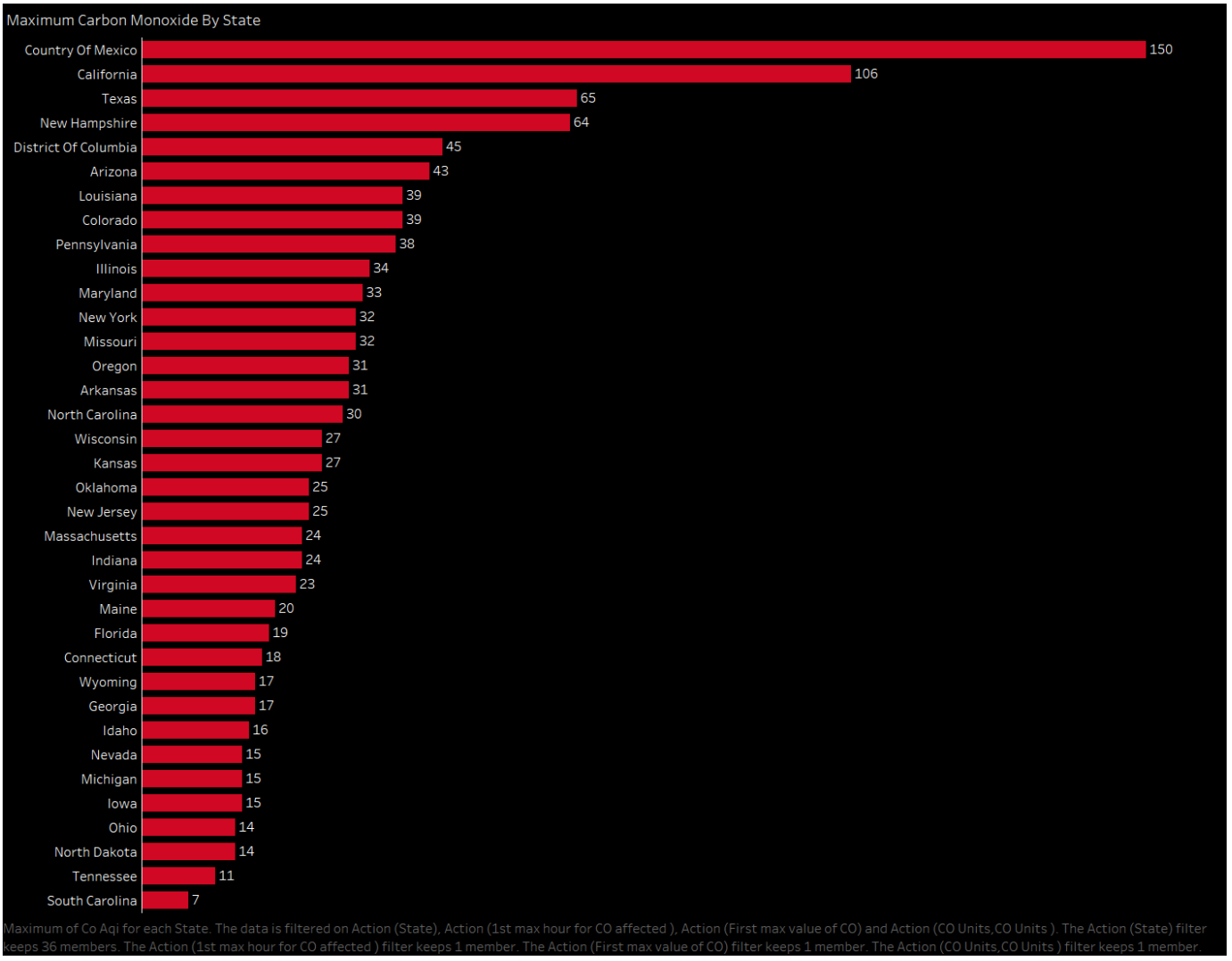
O_3 (Ozone): Ground-level ozone is a key component of smog; harmful to lungs.

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

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

INSIGHTS

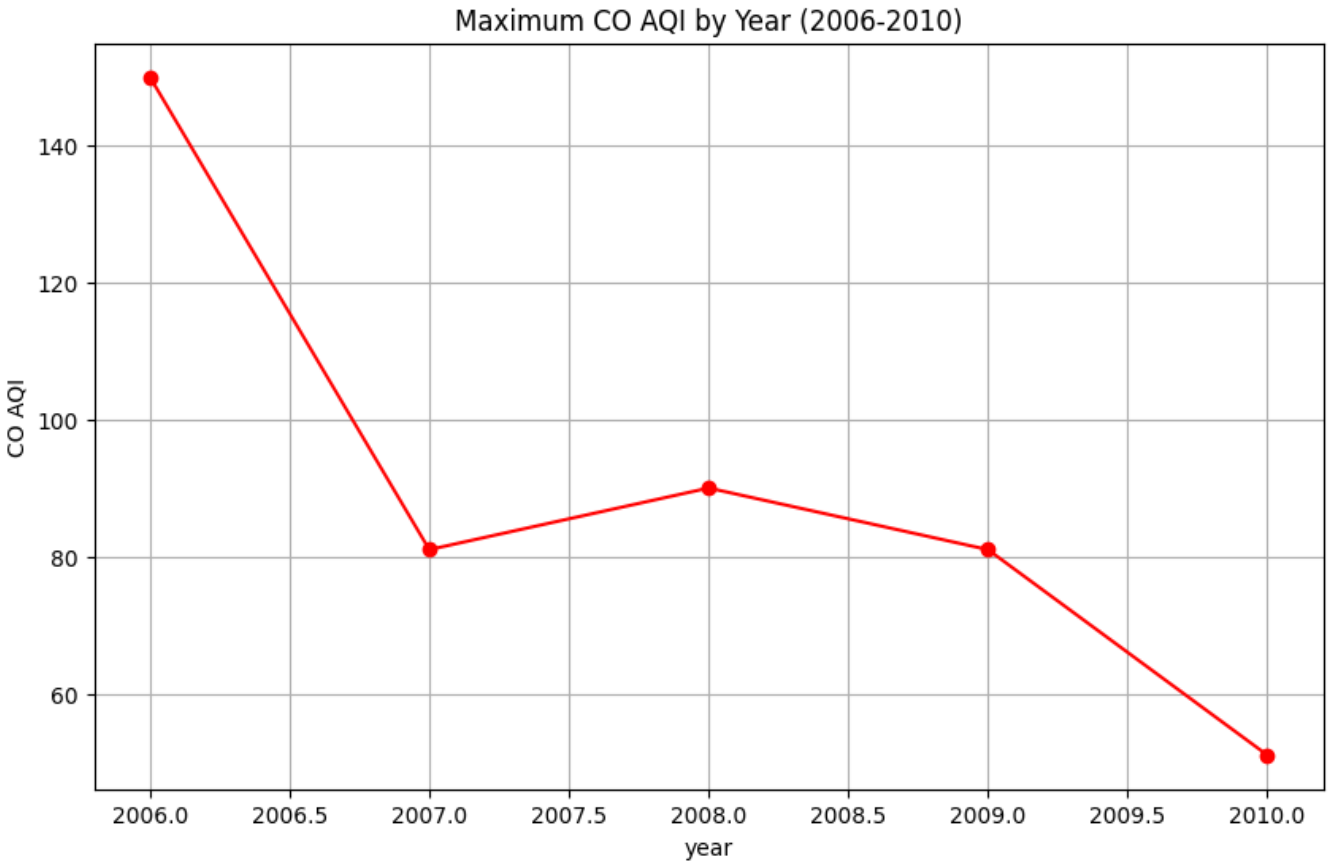
1. Which state have the maximum carbon monoxide CO AQI Present?



- 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.

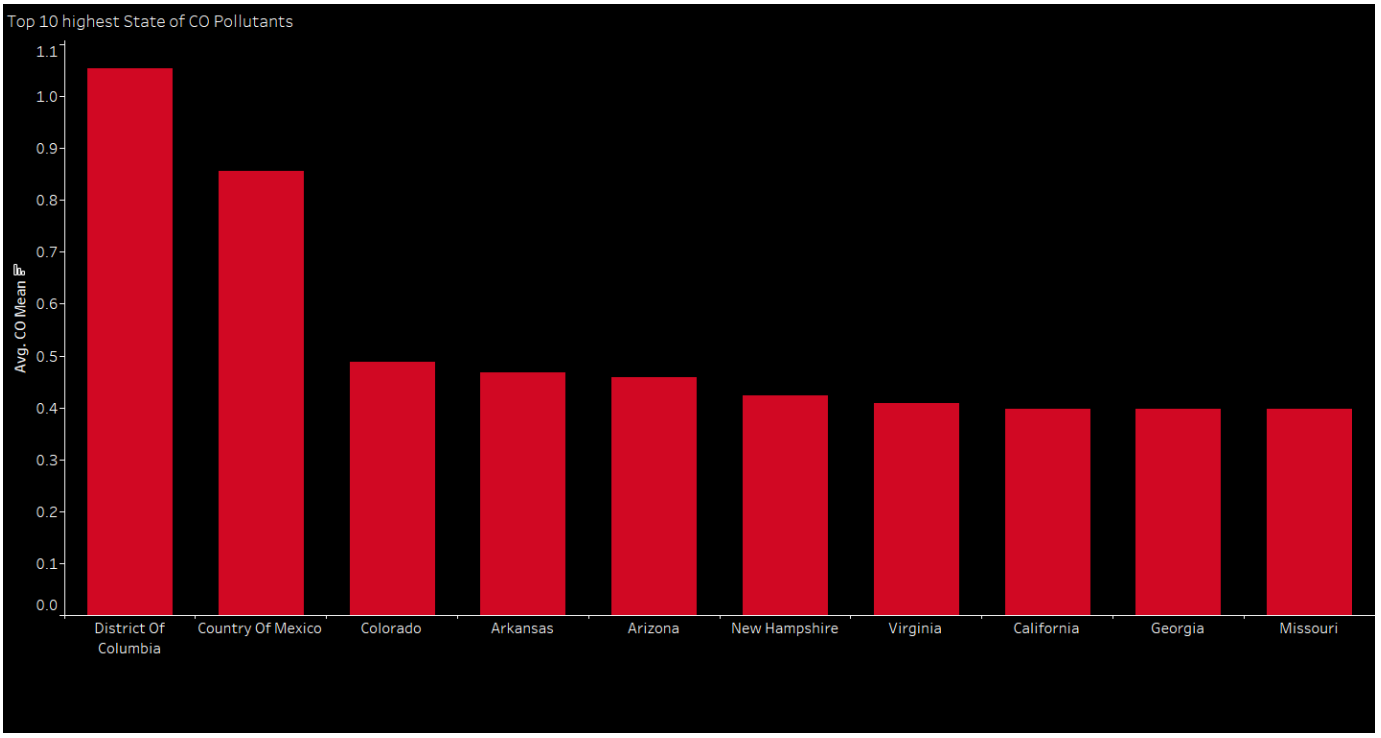
3. What are the yearly statistics for carbon monoxide CO?

Yearly CO Statistics

	2006	2007	2008	2009	2010
Avg. CO Mean	0.41	0.37	0.33	0.32	0.34
Avg. Co Aqi	3.48	3.00	2.67	2.58	2.69
Max. CO Mean	6.53	6.50	3.82	4.05	3.56
Max. Co Aqi	150.00	81.00	90.00	81.00	51.00
Min. CO Mean	0.00	0.00	0.00	0.00	0.00
Min. Co Aqi	0.00	0.00	0.00	0.00	0.00

- 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 ppm 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?



- Country of Mexico, District of Columbia, Colorado, Arizona, Georgia, Arkansas, Missouri, California, Virginia

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

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

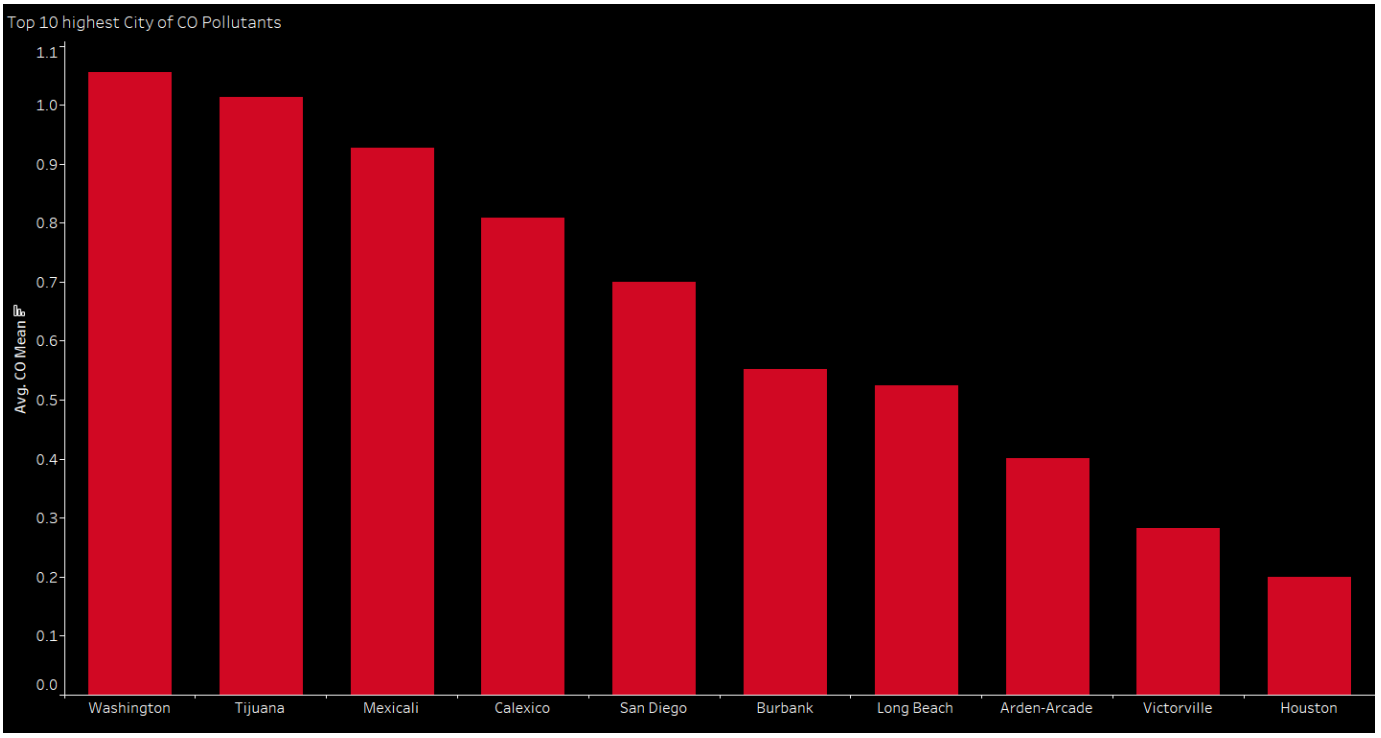
- The first maximum hour for CO is 23.

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

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

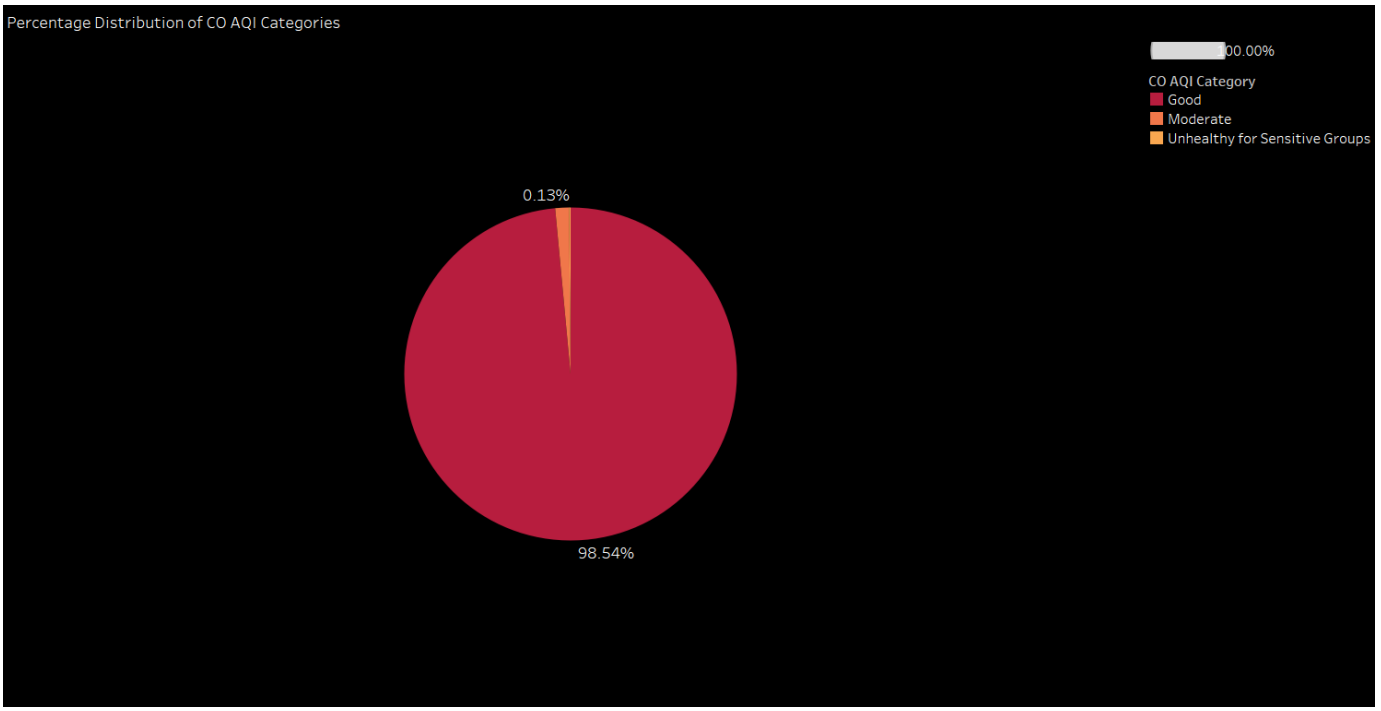
- The first maximum value for the CO is 16.5

7. Which are the top 10 city of carbon monoxide CO pollutants?



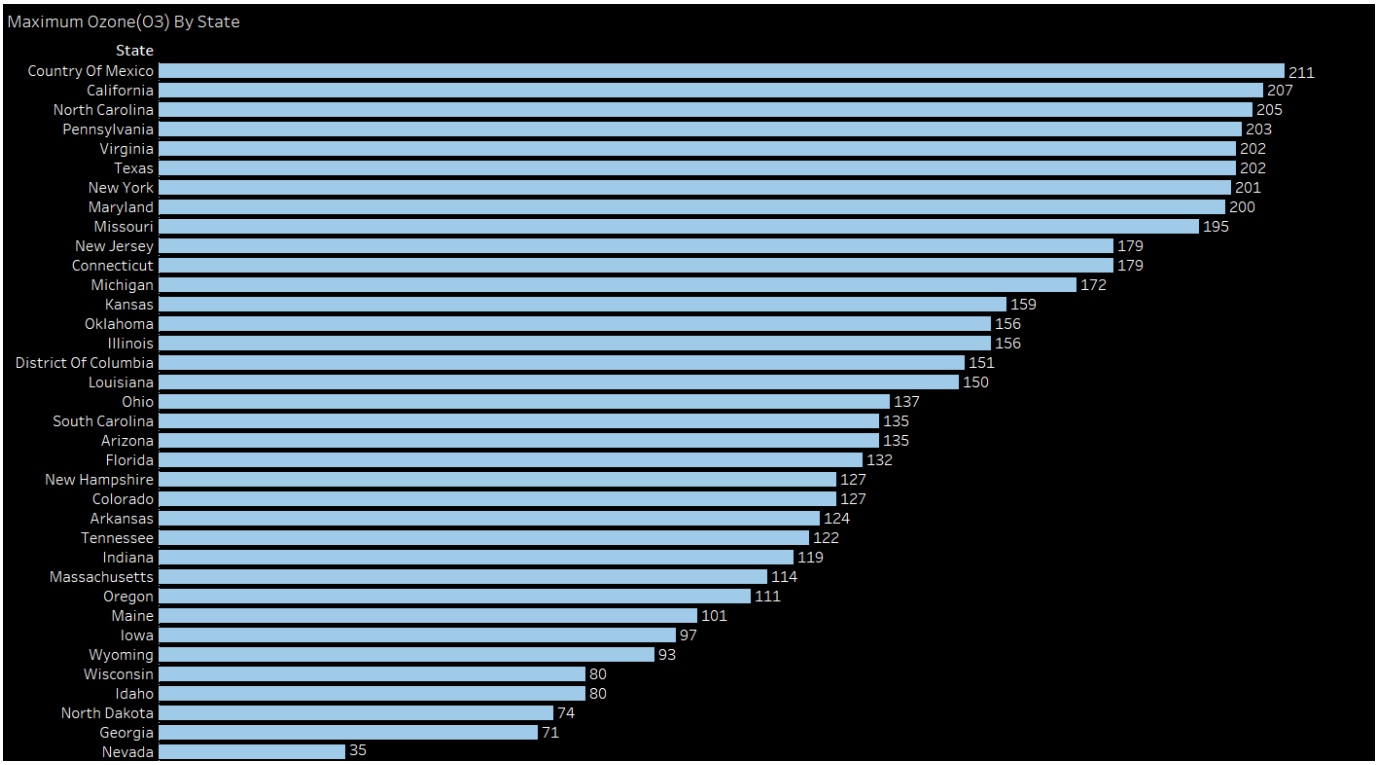
- Washington, Tijuana, Mexicali, Calexico, San Diego, Burbank, Long Beach, Arden-Arcade, Victorville, Houston are the top 10 cities

8. What is the Percentage Distribution of CO Aqi Categories?



- 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.

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.

10.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.

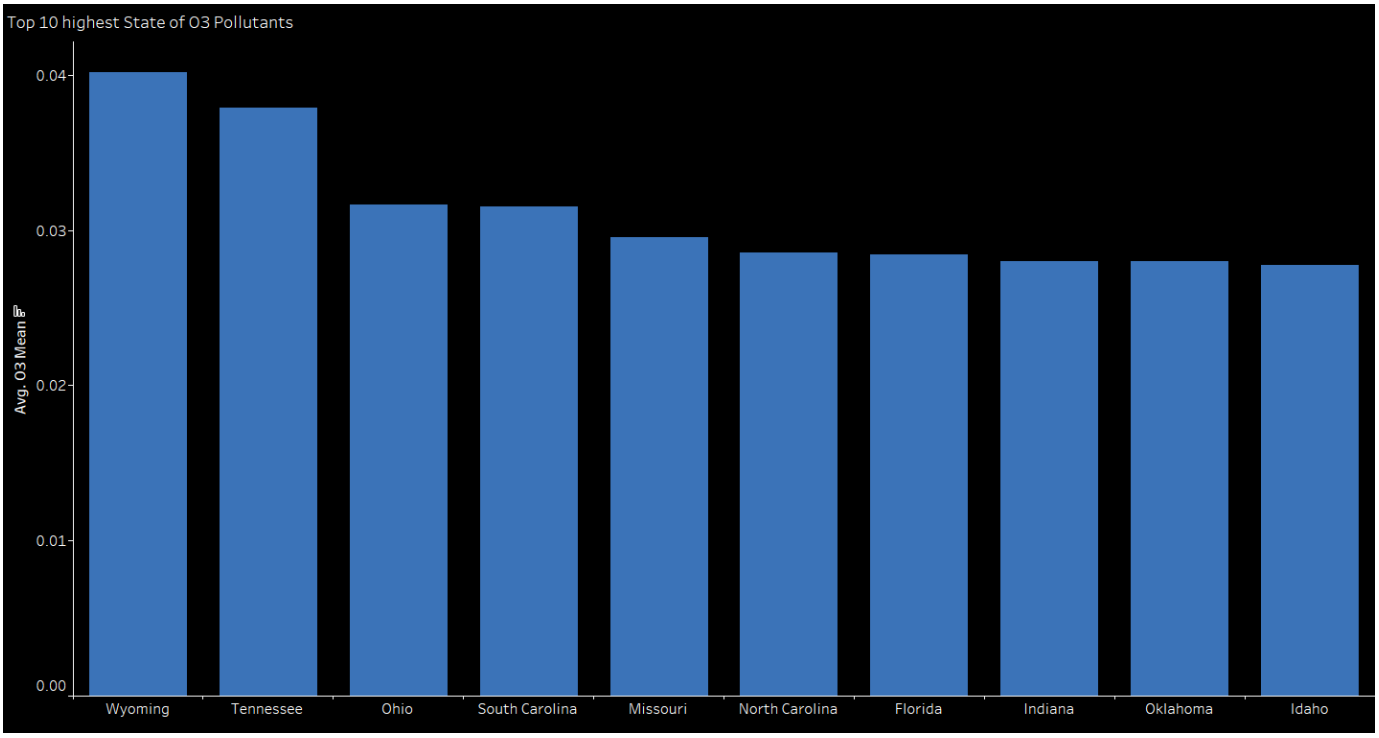
11.What are the yearly statistics for the ozone O3?

Yearly O3 Statistics

	2006	2007	2008	2009	2010
Avg. O3 Aqi	36.20	36.39	35.26	33.02	34.33
Avg. O3 Mean	0.03	0.03	0.03	0.03	0.03
Max. O3 Aqi	206	211	207	206	200
Max. O3 Mean	0.07	0.07	0.08	0.07	0.07
Min. O3 Aqi	0.00	0.00	0.00	0.00	0.00
Min. O3 Mean	0.00	0.00	0.00	0.00	0.00

- 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 ppm was recorded in 2007
- Recent years show lower maximum values, suggesting improved air quality

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



- Wyoming, Tennessee, Ohio, South Carolina, Missouri, North Carolina, Florida, Indiana, Oklahoma, Idaho

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

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

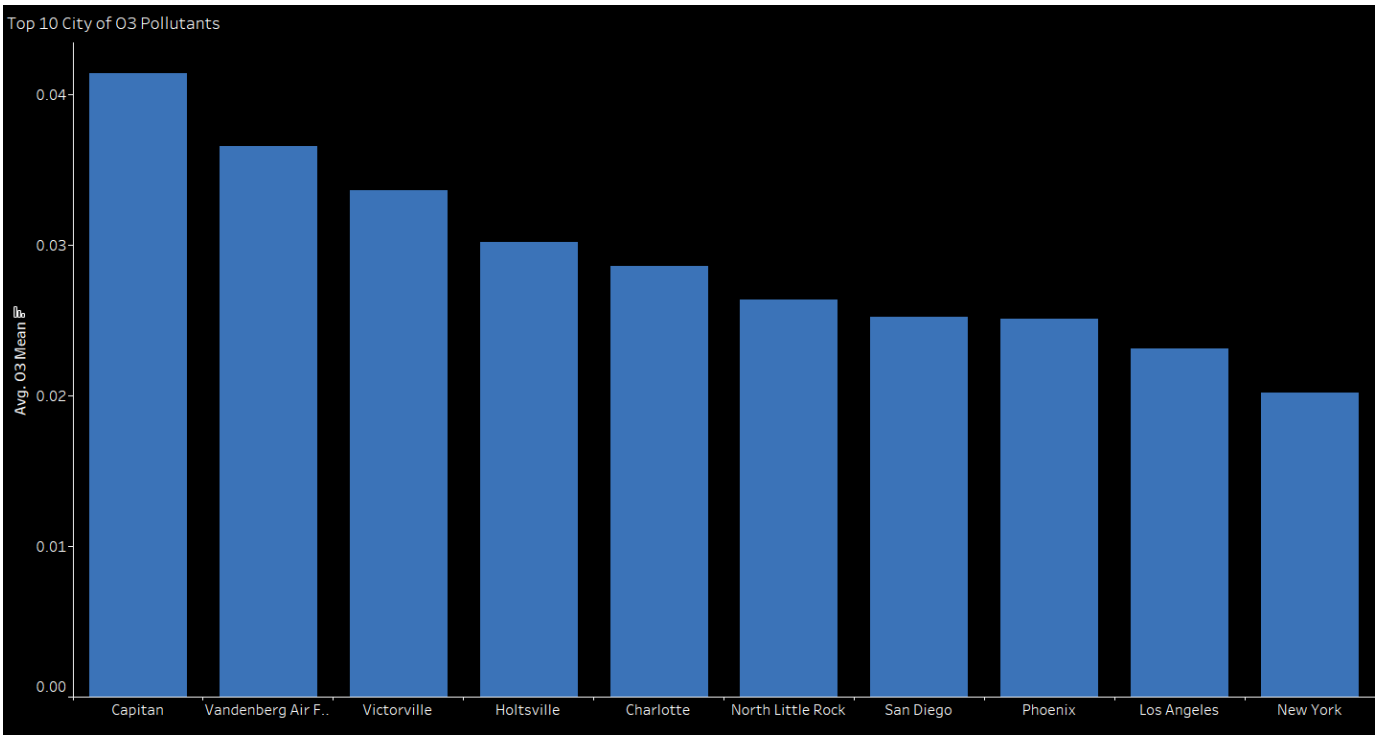
- **The first maximum hour for O3 is 10**

14.What is the first maximum value for O3?

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

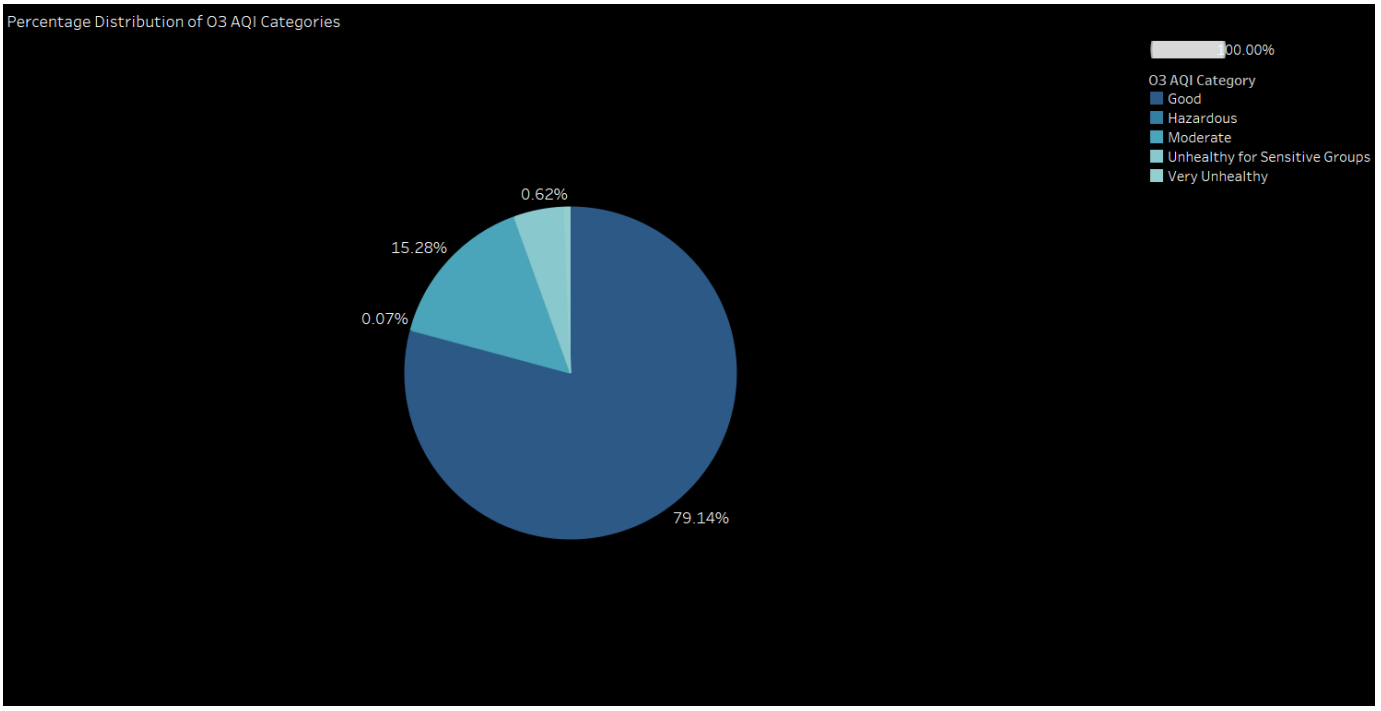
- **The first maximum value for O3 is 0.141**

15.Which are the top 10 Highest city of Ozone O3 pollutants?



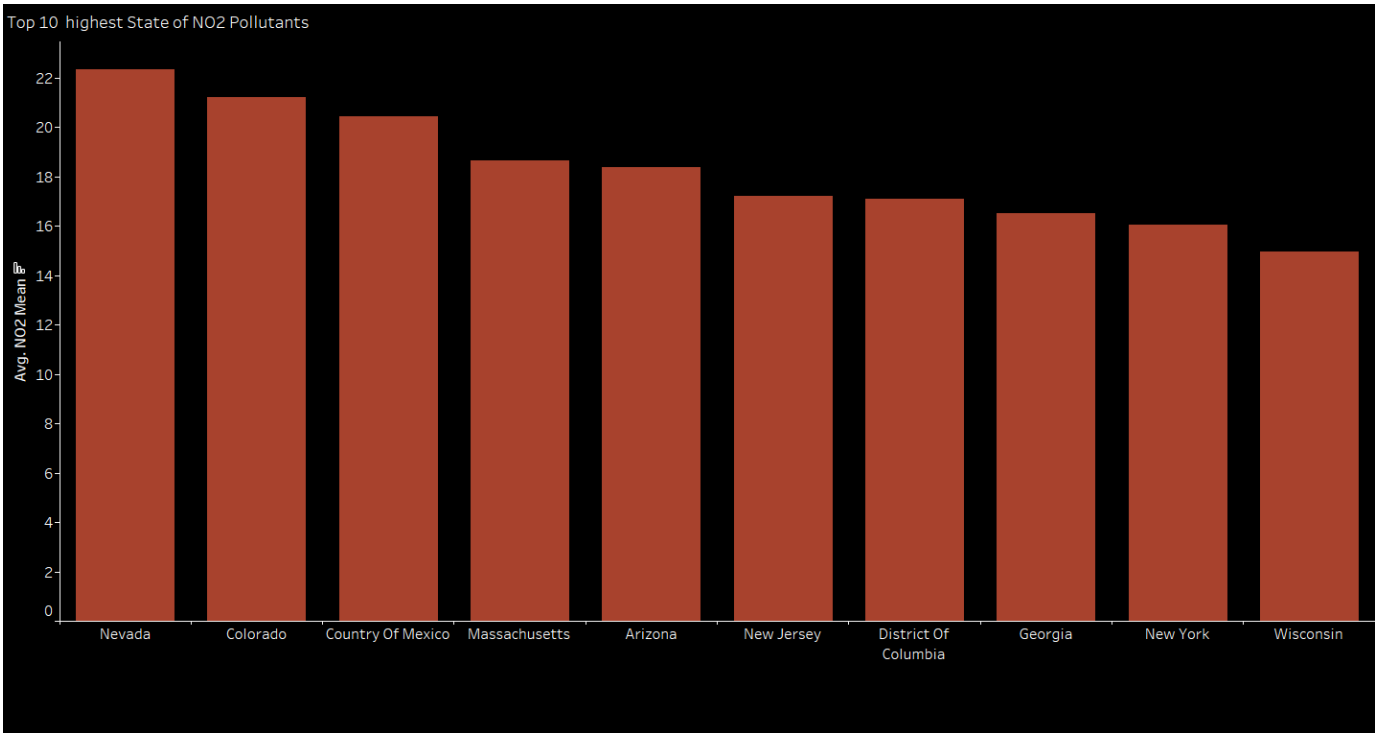
- Captain, Vandenberg, Holtsville, Charlotte, North little Rock, San Diego, Phoenix, Los Angeles, New York

16.What is the percentage distribution of O3 Categories?



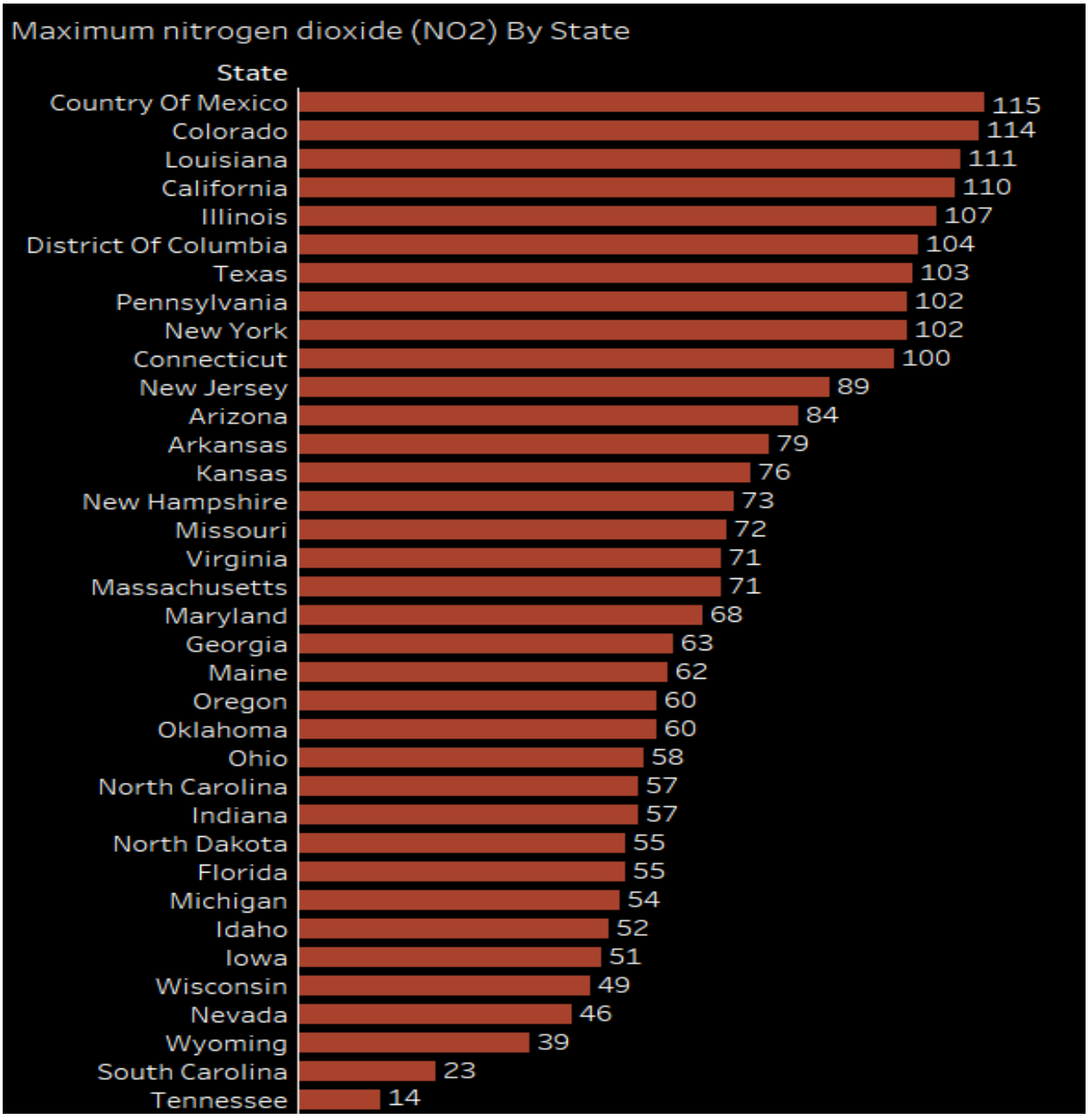
- Most diverse distribution among all pollutants 90.61% “Good" reading
- 7.76% 'Moderate' Levels
- 1.48% "Unhealthy for sensitive Groups"

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



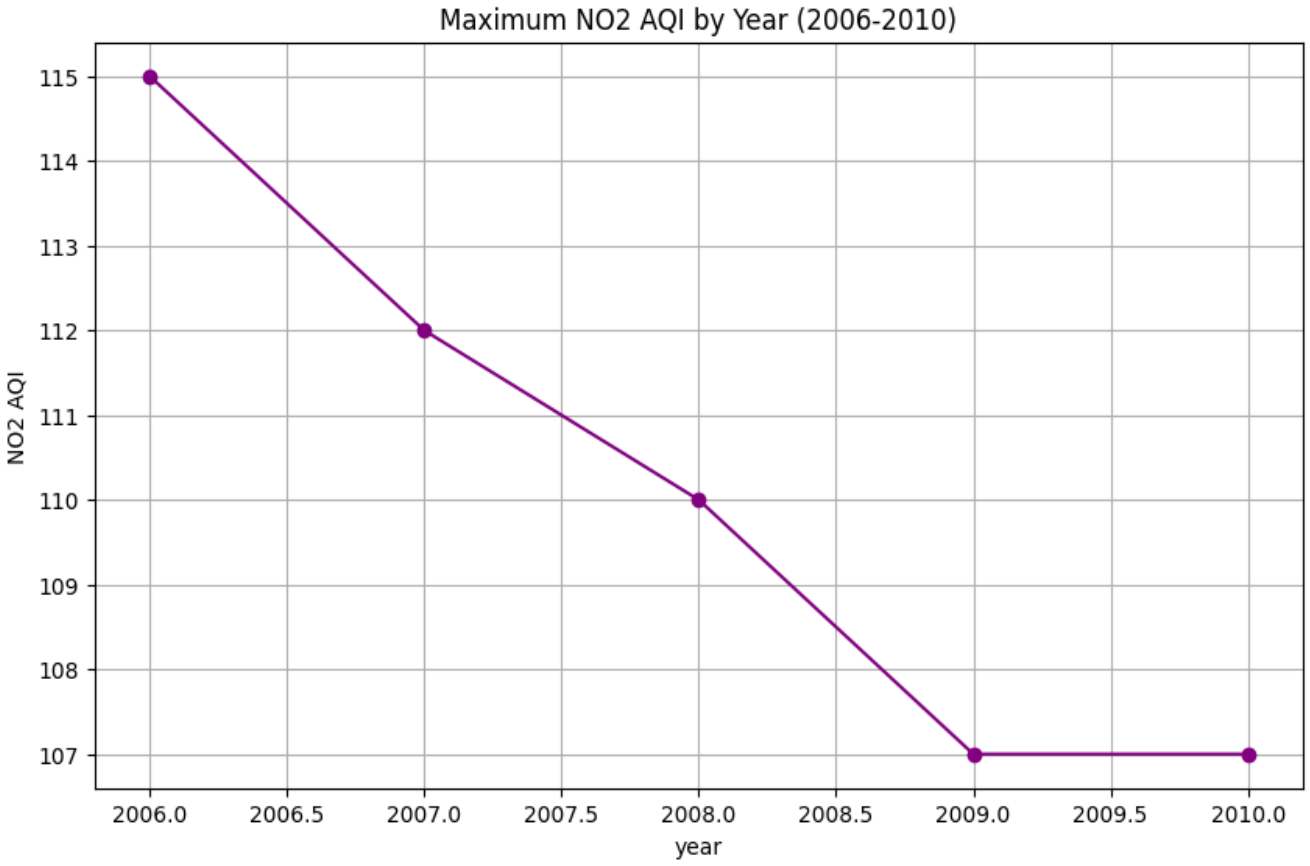
- Nevada, Colorado, Country of Mexico, Massachusetts, Arizona, New Jersey, District of Columbia, Georgia, New York, Wisconsin

18. 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.

19. 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.

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

Yearly NO2 Statistics

	2006	2007	2008	2009	2010
Avg. NO2 Mean	14.5	13.3	12.3	11.5	11.7
Avg. No2 Aqi	27.0	24.9	23.5	21.7	22.3
Max. NO2 Mean	70.6	92.0	76.5	60.6	56.9
Max. No2 Aqi	115	112	110	107	107
Min. NO2 Mean	0.0	0.0	0.0	0.0	0.0
Min. No2 Aqi	0.0	0.0	0.0	0.0	0.0

- 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 ppb was recorded in 2006
- Recent years show lower maximum values, suggesting improved air quality

21.What is the first maximum hour for NO2?

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

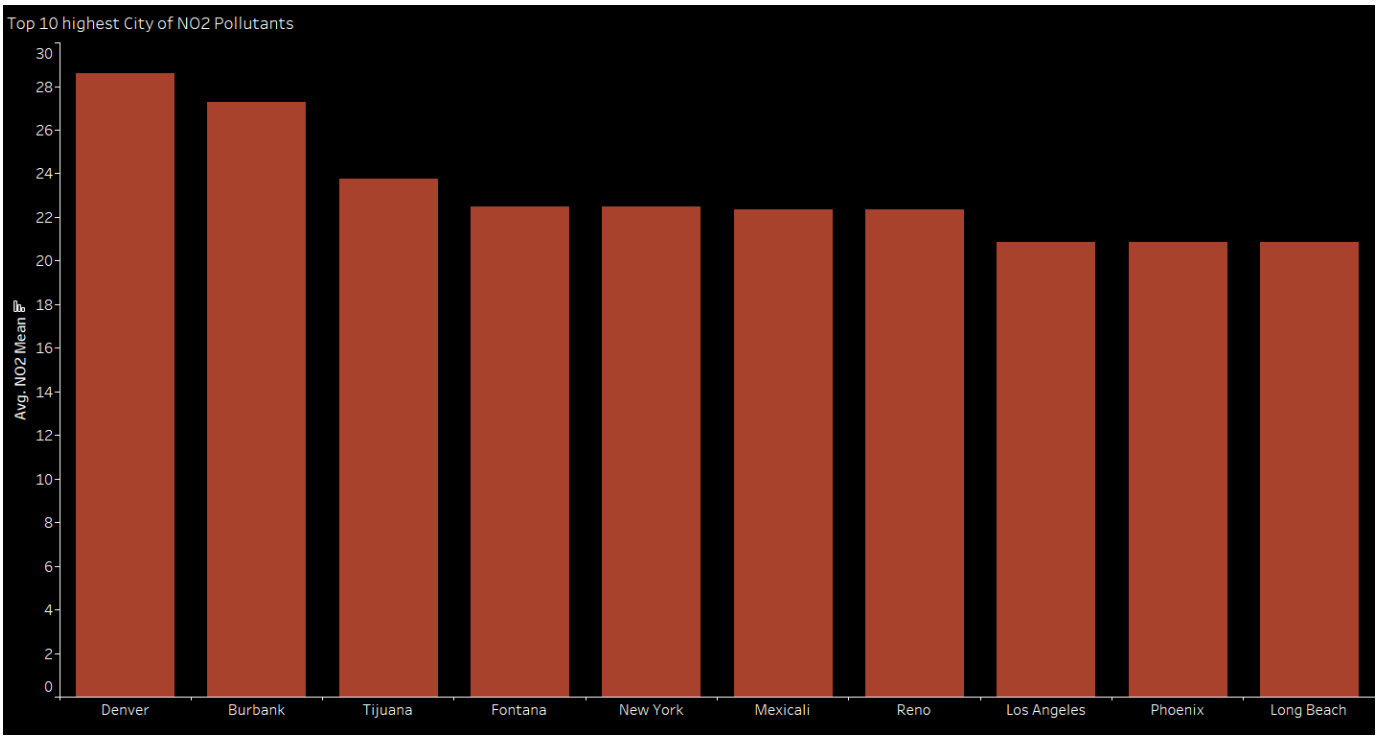
- The first Maximum hour for NO2 is 2.

22.What is the first maximum value for NO2?

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

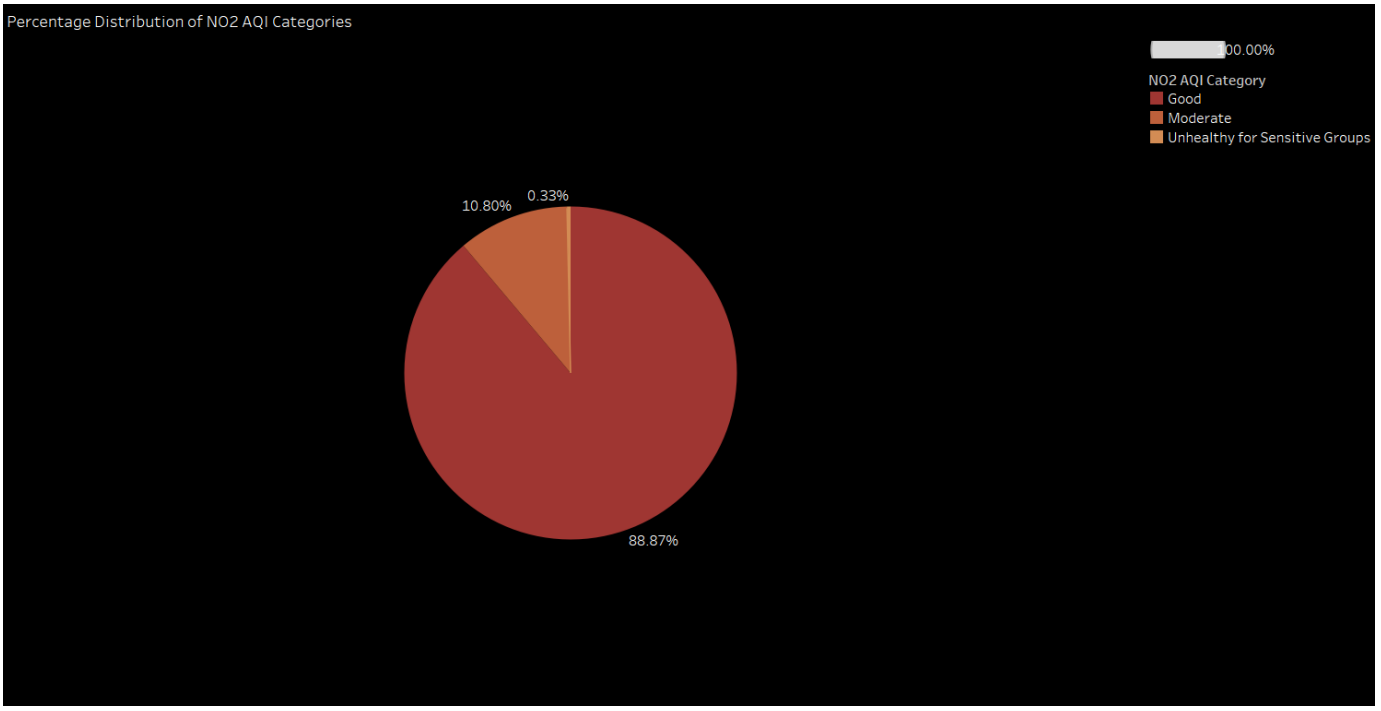
- The first maximum value for NO2 is 176

23. Which are the Top 10 highest City of Nitrogen Dioxide NO2 Pollutants?



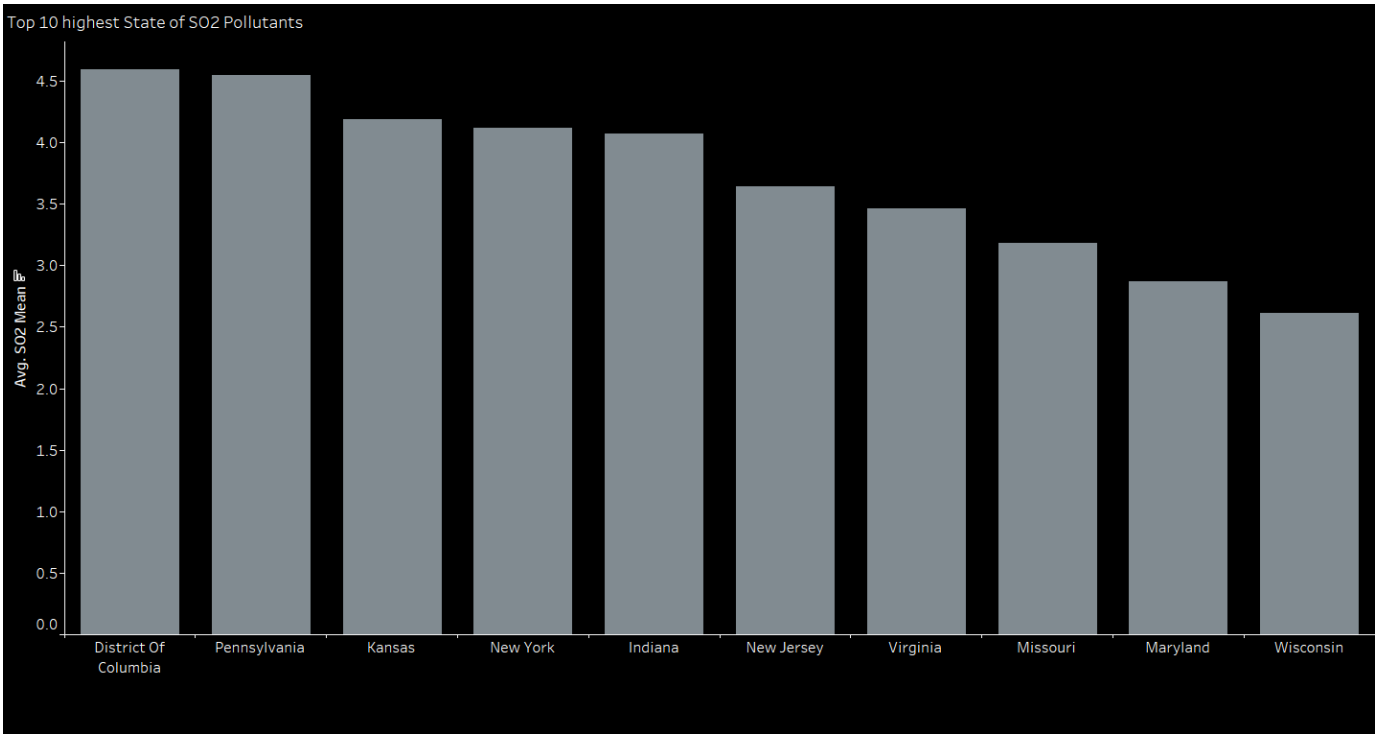
- Denver, Burbank, Tijuana, Fontana, New York, Mexicali, Reno, Los Angeles, Phoenix, Long Beach

24. What is the percentage distribution of Nitrogen Dioxide NO2 pollutants?



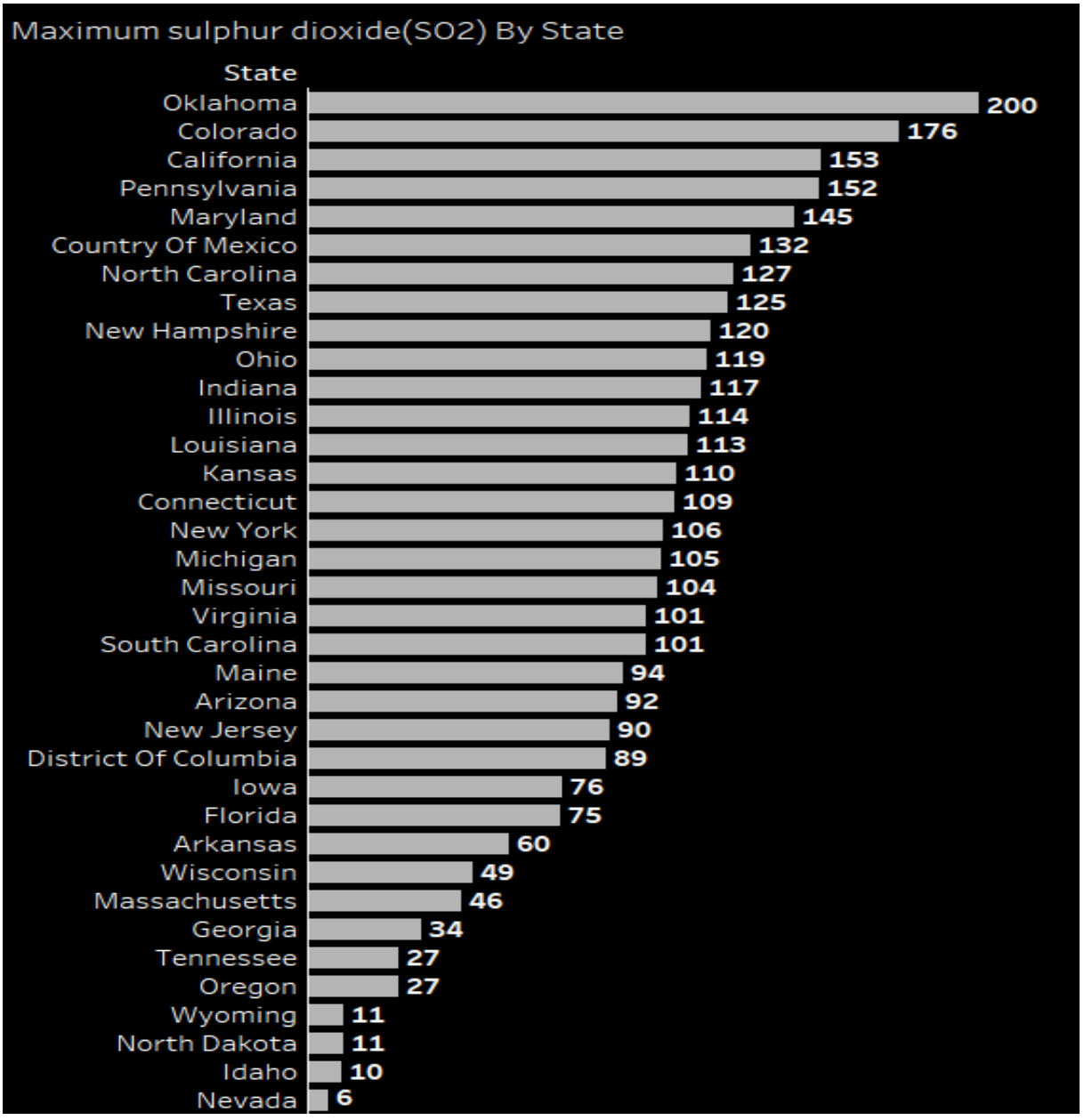
- 95.59% of readings fall in the "Good" category
- Only about 4.34% reach "Moderate" levels
- Very few instances (0.08%) of "Unhealthy for Sensitive Groups"

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



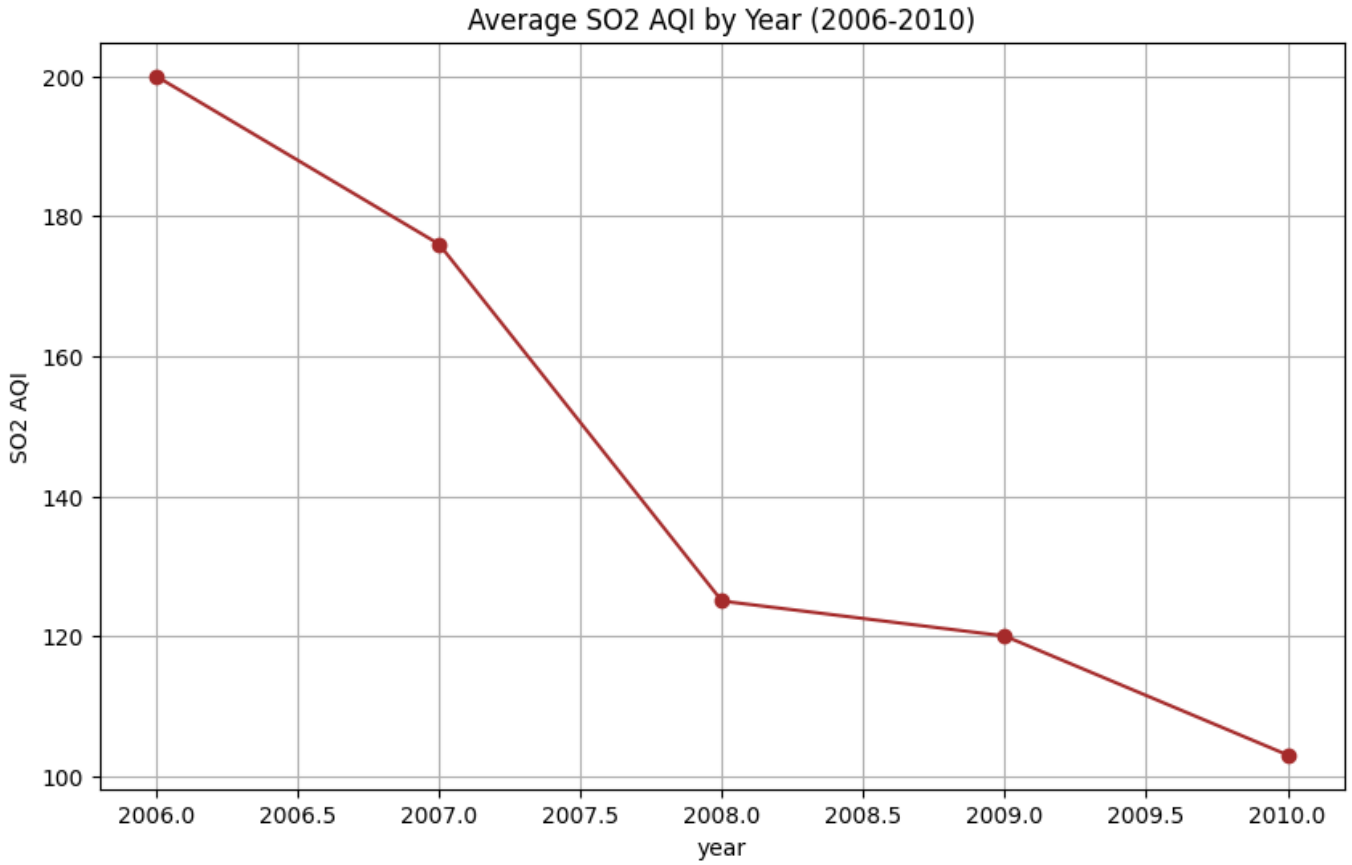
- District of Columbia, Pennsylvania, Kansas, New York, Indiana, New Jersey, Virginia, Missouri, Maryland, Wisconsin are the highest state of SO2 pollutants.

26. 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 the minimum SO2 AQI present.

27. 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

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

Yearly SO2 Statistics

	2006	2007	2008	2009	2010
Avg. SO2 Mean	2.6	2.3	1.9	1.6	1.2
Avg. So2 Aqi	5.0	4.5	3.7	2.8	1.9
Max. SO2 Mean	321.6	32.0	31.0	33.0	25.9
Max. So2 Aqi	200	176	125	120	103
Min. SO2 Mean	0.0	0.0	0.0	0.0	-0.5
Min. So2 Aqi	0.0	0.0	0.0	0.0	0.0

- 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 ppb was recorded in 2006
- Recent years show lower maximum values, suggesting improved air quality

29.What is the first maximum hour for SO2?

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

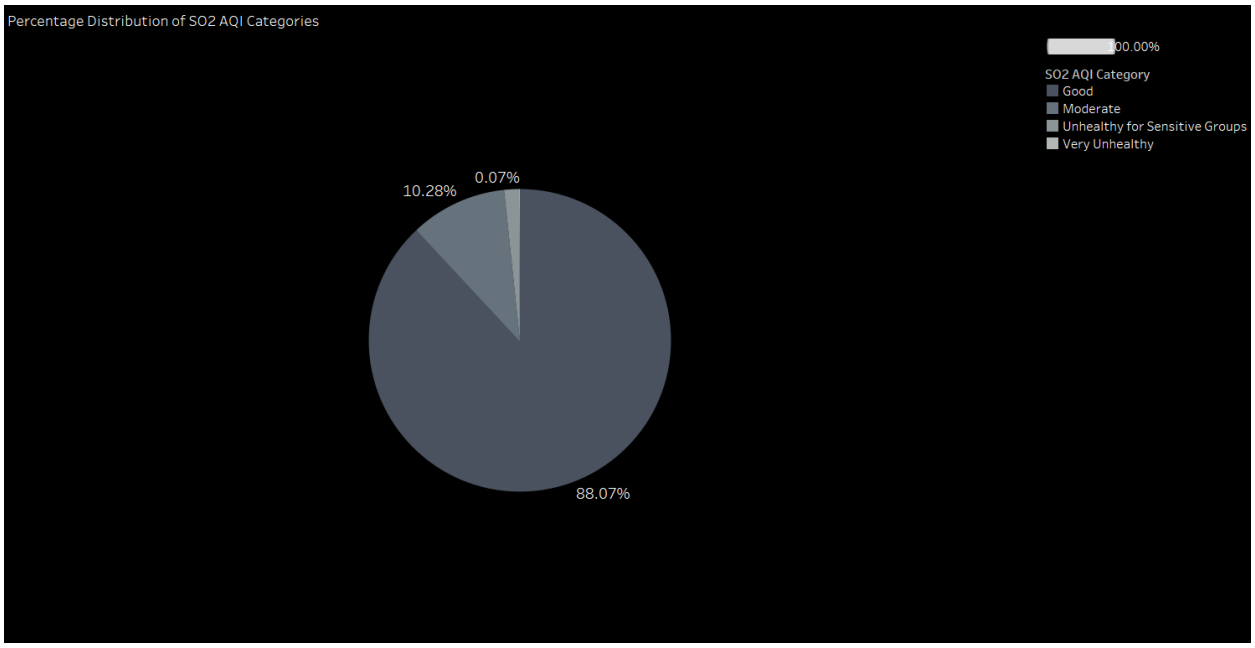
- **The First Maximum hour for SO2 is 0**

30.What is the first maximum value for SO2?

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

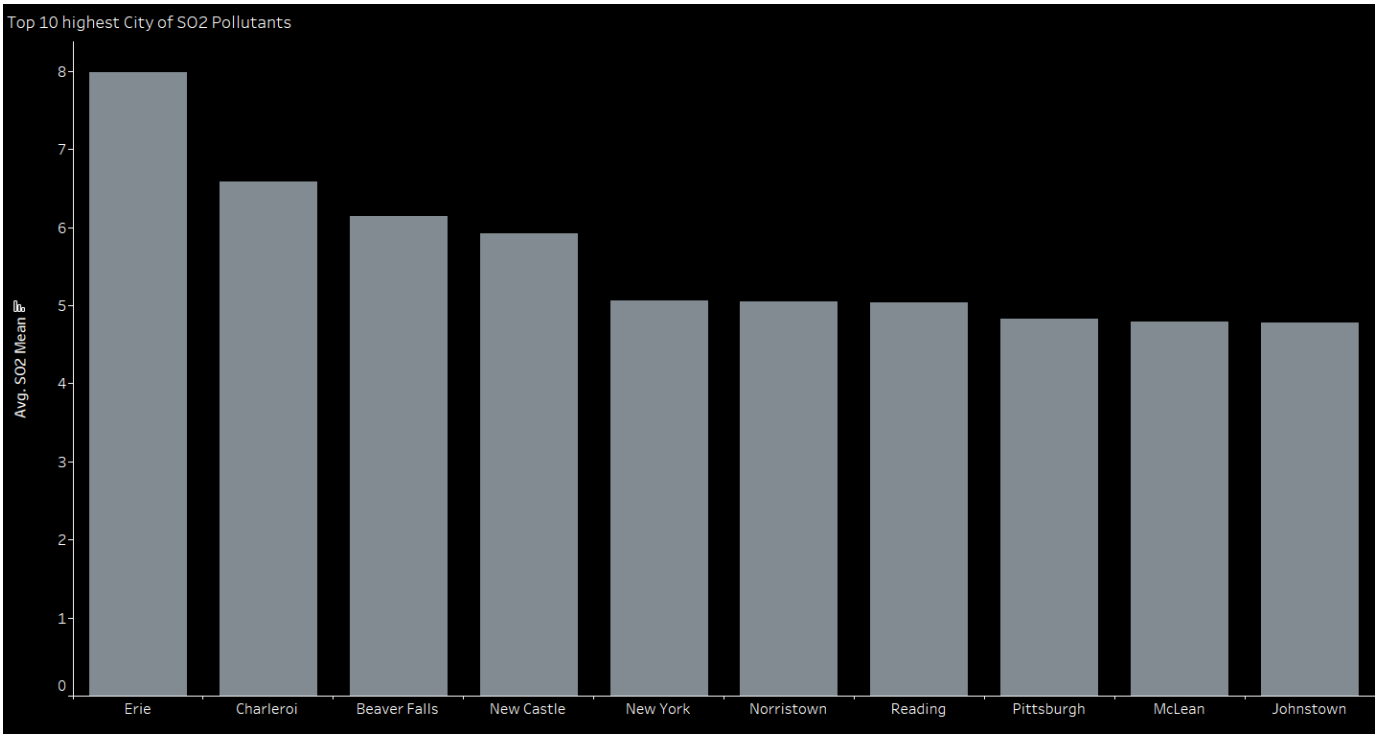
- **The First Maximum Value for SO2 is 351.**

31.What is the percentage distribution of SO2 AQI categories?



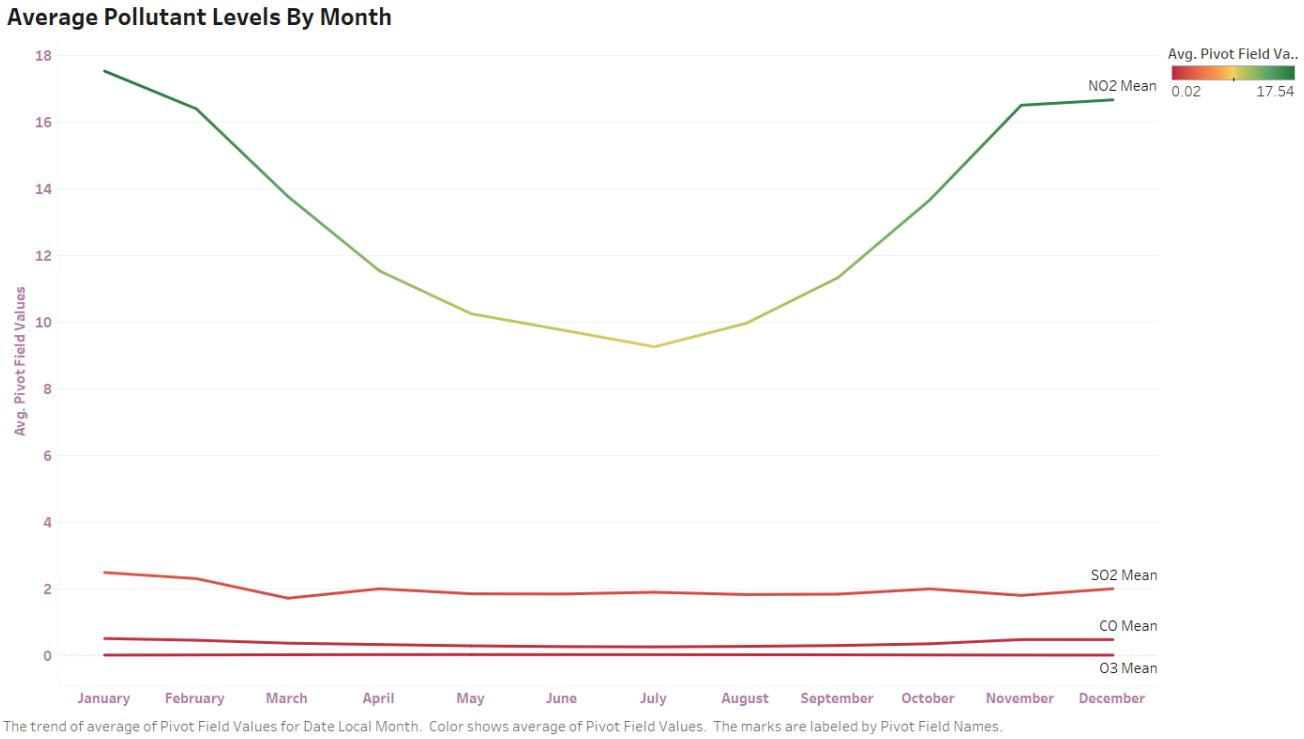
- SO2 (Sulfur Dioxide): About 99.37% "Good" readings values
  Very few instances (0.58%) of "Moderate" levels
  Minimal occurrences of higher categories

32.which are the top 10 highest City of Sulphur Dioxide SO2 pollutants?



- Erie, Charleroi, Beaver Falls, New Castle, New York, Norristown, Reading, Pittsburgh, McLean, Johnstown

33.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.

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.
- 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