Project Overview:

Air Pollution is contamination of the indoor or outdoor environment by any chemical, physical or biological agent that modifies the natural characteristics of the atmosphere. Household combustion devices, motor vehicles, industrial facilities and forest fires are common sources of air pollution. Pollutants of major public health concern include particulate matter, carbon monoxide, ozone, nitrogen dioxide and sulfur dioxide. Outdoor and indoor air pollution cause respiratory and other diseases and are important sources of morbidity and mortality.

Data overview:

This dataset provides geolocated information about the following pollutants:

Nitrogen Dioxide [NO2] : Nitrogen Dioxide is one of the several nitrogen oxides. It is introduced into the air by natural phenomena like entry from stratosphere or lighting. At the surface level, however, NO2 forms from cars, trucks and buses emissions, power plants and off-road equipment. Exposure over short periods can aggravate respiratory diseases, like asthma. Longer exposures may contribute to develoment of asthma and respiratory infections. People with asthma, children and the elderly are at greater risk for the health effects of NO2.

Ozone [O3] : The Ozone molecule is harmful for outdoor air quality (if outside of the ozone layer). At surface level, ozone is created by chemical reactions between oxides of nitrogen and volatile organic compounds (VOC). Differently from the good ozone located in the upper atmosphere, ground level ozone can provoke several health problems like chest pain, coughing, throat irritation and airway inflammation. Furthermore it can reduce lung function and worsen bronchitis, emphysema, and asthma. Ozone affects also vegetation and ecosystems. In particular, it damages sensitive vegetation during the growing season.

Carbon Monoxide [CO] : Carbon Monoxide is a colorless and odorless gas. Outdoor, it is emitted in the air above all by cars, trucks and other vehicles or machineries that burn fossil fuels. Such items like kerosene and gas space heaters, gas stoves also release CO affecting indoor air quality.

Breathing air with a high concentration of CO reduces the amount of oxygen that can be transported in the blood stream to critical organs like the heart and brain. At very high levels, which are not likely to occur outdoor but which are possible in enclosed environments. CO can cause dizziness, confusion, unconsciousness and death.

Particulate Matter [PM2.5] : Atmospheric Particulate Matter, also known as atmospheric aerosol particles, are complex mixtures of small solid and liquid matter that get into the air. If inhaled they can cause serious heart and lungs problem. They have been classified as group 1 carcinogen by the International Agengy for Research on Cancer (IARC). PM10 refers to those particules with a diameter of 10 micrometers or less. PM2.5 refers to those particles with a diameter of 2.5 micrometers or less.

Data source: Global Air Pollution Dataset

Data Preparation:

The data was cleaned and transformed in Python notebook.

before being imported into Power BI. The steps taken to

prepare the data included removing null values, filtering out.

unnecessary columns, treating outliers and some other preprocessing concepts.

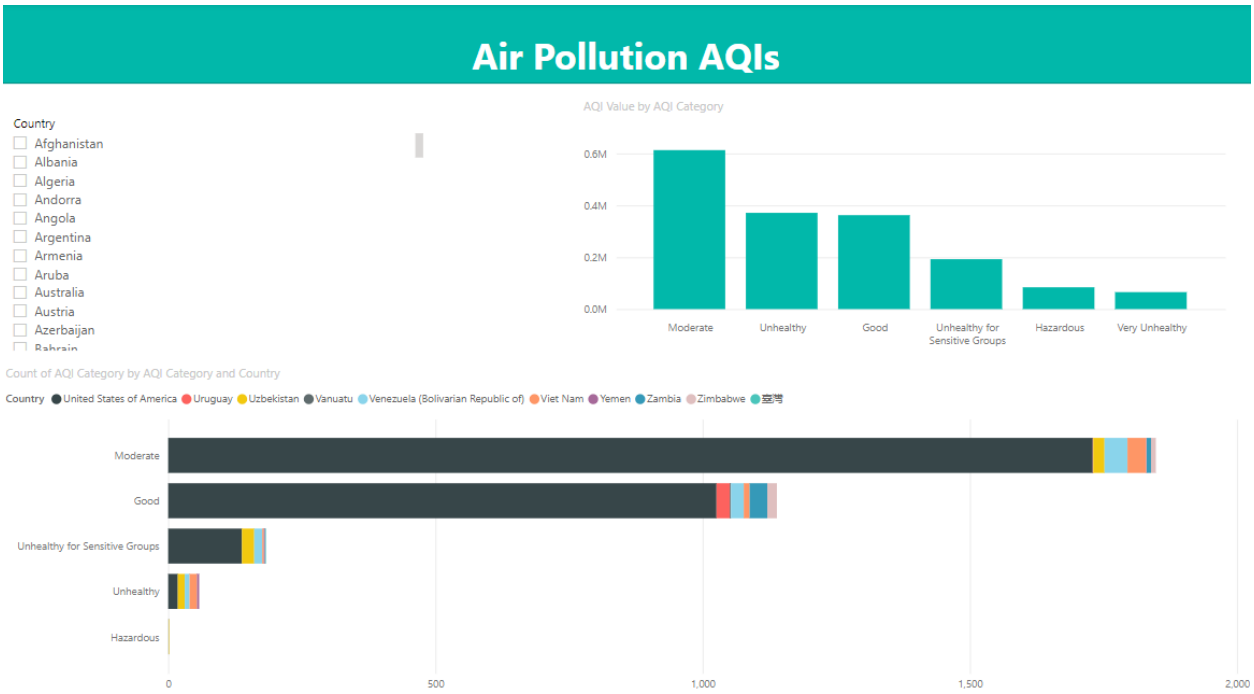
Questions we are trying to find its answers:

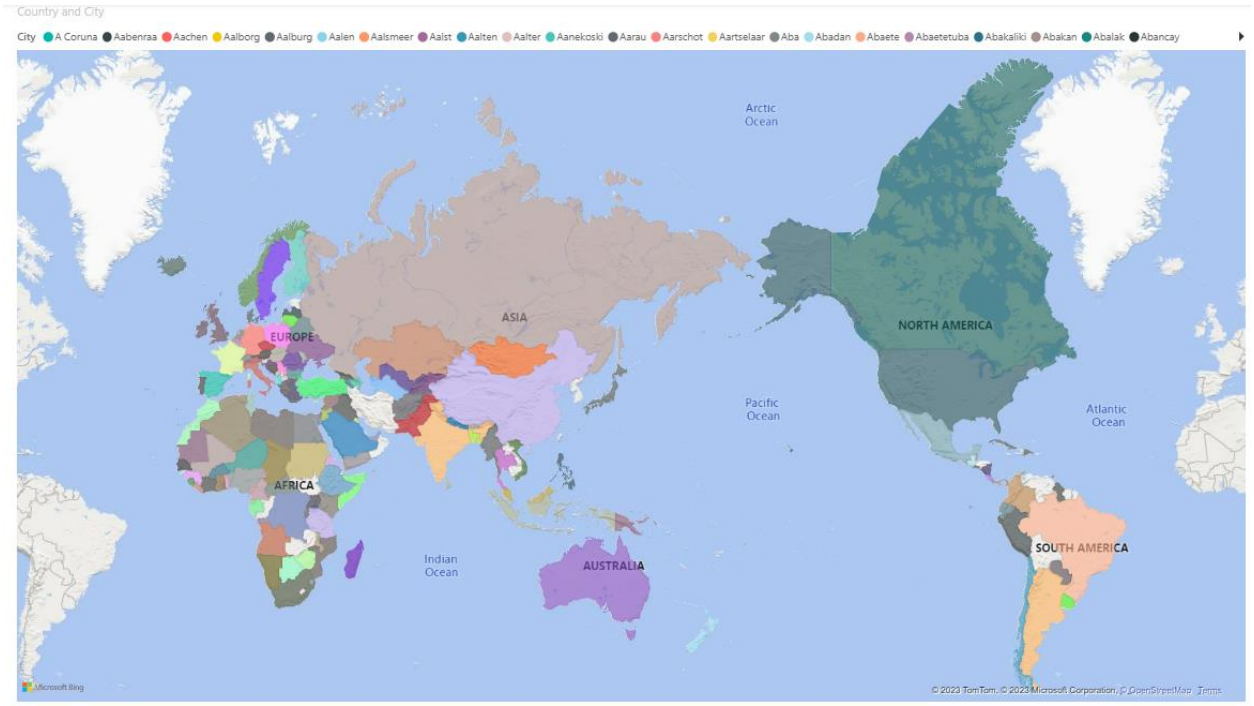
- What does the difference in standard deviation and mean in the AQI values indicates for?
- At what ranges does the AQI concentrated at?
- What does the fact that "good" is the most common category indicates for?
- What are the most countries that have "good" air global and vice versa?
- How features vary along with different type of pollutants?

Conclusion and insights we've gained from our analysis:

- The AQI ranges from 0 to 500, with higher values indicating greater levels of pollution and corresponding health concerns. An AQI value of 100 is considered moderate, while values above 150 are considered unhealthy for sensitive groups and values above 200 are considered unhealthy for the general public.
- The AQI Value column has the highest range, meaning that it can have higher spikes and more variation in values. In contrast, the CO AQI Value column has the lowest range, suggesting that it may be less variable over time. This information can be useful when deciding which AQI value type to prioritize in environmental management and policy-making.
- The fact that "good" is the most common category in all the categorical data suggests that the air quality in the area represented in the data is generally good
- The pivot plots show that as the AQI category gets worse (i.e., from "good" to "unhealthy"), the AQI values tend to increase. This is expected since the AQI categories are based on increasing levels of air pollution, and thus, higher AQI values correspond to worse air quality.
- The AQI categories are based on increasing levels of air pollution, and thus, higher AQI values correspond to worse air quality.
- These conclusions suggest that the dataset represents a region with generally good air quality but some instances of worse air quality, and that the AQI values are related to the AQI category but with some variability in the relationships between the different AQI value types. These insights could be useful for environmental management and policy-making decisions in the represented region.
- The difference in standard deviation and mean in the AQI values indicating that they may have different levels of variability and central tendency.

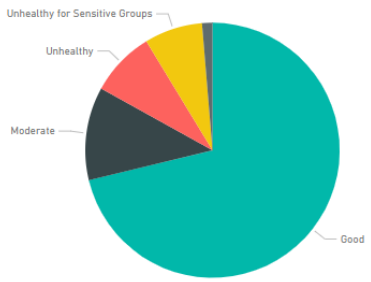
Visualizations:



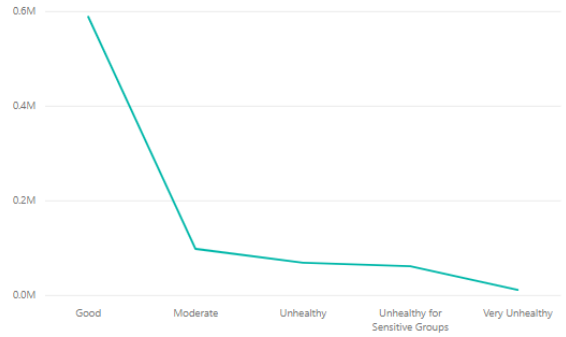


Ozone AQI

Ozone AQI Value by Ozone AQI Category



Ozone AQI Value by Ozone AQI Category



Ozone AQI Category	Ozone AQI Value
Good	587970
Moderate	97452
Unhealthy	68184
Unhealthy for Sensitive Groups	60966
Very Unhealthy	10866
Total	**825438**

0
Min of Ozone AQI Value

35.19
Average of Ozone AQI Value

235
Max of Ozone AQI Value

PM2.5 AQI

PM2.5 AQI Value by PM2.5 AQI Category

PM2.5 AQI Value by PM2.5 AQI Category

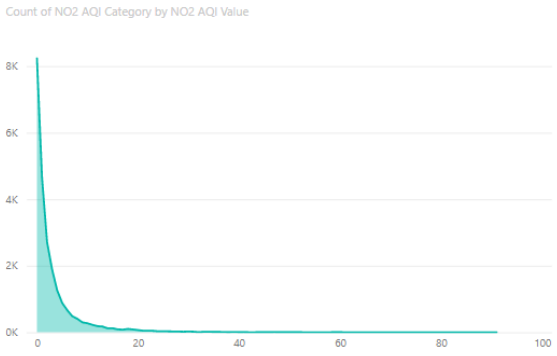
PM2.5 AQI Category	PM2.5 AQI Value	PM2.5 AQI Categ
Good	323613	Cluster5
Hazardous	71222	Cluster1
Moderate	603737	Cluster3
Unhealthy	352481	Cluster2
Unhealthy for Sensitive Groups	196016	Cluster4
Very Unhealthy	60120	Cluster1
Total	**1607189**	

0
Min of PM2.5 AQI Value

68.52
Average of PM2.5 AQI Value

500
Max of PM2.5 AQI Value

NO2 AQI



Good
70978
NO2 AQI Value

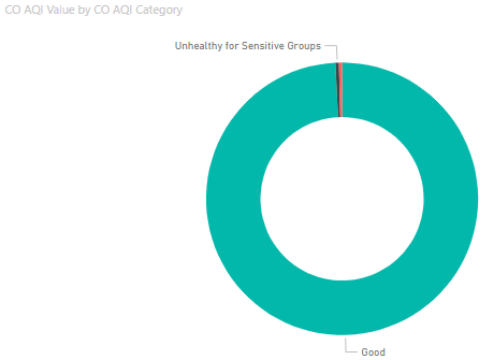
Moderate
893
NO2 AQI Value

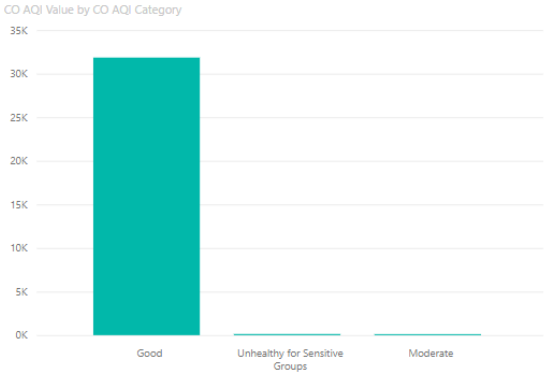
0
Min of NO2 AQI Value

3.06
Average of NO2 AQI Value

91
Max of NO2 AQI Value

CO AQI





CO AQI Category	CO AQI Value
Good	31848
Unhealthy for Sensitive Groups	133
Moderate	118
Total	**32099**

0
Min of CO AQI Value

1.37
Average of CO AQI Value

133
Max of CO AQI Value

For more visualizations have a look at the notebook:[Global Air Pollution Notebook]

All Rights Reserved to ApplAi

Under the supervision of:

- Director of Machine Learning and Training Department:Eslam Shouman

Team Leader:

- Nour Ahmed

Team Members:

- Jonathan Mounir
- Sarah Ahmed
- Youmna Muhammed
- Habiba Alaa