

**Friedrich-Alexander-Universität Erlangen-Nürnberg**

**Methods of Advanced Data Engineering**

**Summer Semester 2024**

**03.07.2024**

**Correlation Between World Population and Air Pollution Analysis Report**

**Berkant Cinar**

Student ID: 23224469

---

# 1 Introduction

Population and air pollution are critical issues for global sustainability and public health. The goal of this study is to investigate the relationship between population size and air pollution levels across different countries. Our aim is to identify trends and correlations that provide insights into how population size impacts air quality, helping to inform policies for sustainable development and pollution control.

## 2 Main Question

1. Is there a direct connection between world population size and air pollution levels?

## 3 Data Format

### 3.1 World Population Dataset

- Metadata URL: [World Population Dataset](#)
- Source: Kaggle
- Content: This dataset offers detailed information on population trends, including rates of urbanization, for all nations in the world. This dataset is really important for comprehending the size and rate of population growth in various countries.
- Data Format: CSV format with columns including 'Country', 'Continent', '2022 Population', 'Area', 'Density', 'Growth Rate', 'World Population Percentage'.

### 3.2 Global Air Pollution Dataset

- Metadata URL: [Global Air Pollution Dataset](#)
- Source: Kaggle
- Content: This dataset provides comprehensive data on air pollution levels, including AQI values and concentrations of various pollutants for many locations around the world. It is essential for assessing air quality and understanding the impact of different pollution sources.
- Data Format: CSV format with columns including 'Country', 'City', and various pollution metrics like Nitrogen Dioxide [NO<sub>2</sub>], Ozone [O<sub>3</sub>], Carbon Monoxide [CO], Particulate Matter [PM<sub>2.5</sub>] .

## 4 Data Pipeline Output

### 4.1 Output Data

The pipeline output data has:

- Cleaned and transformed world population data for 2022.
- Aggregated air pollution data with mean values for each country because there were multiple values for some countries.

### 4.2 Description

After the data processing with pipeline, this is the final data that we will analyse and answer our research question.

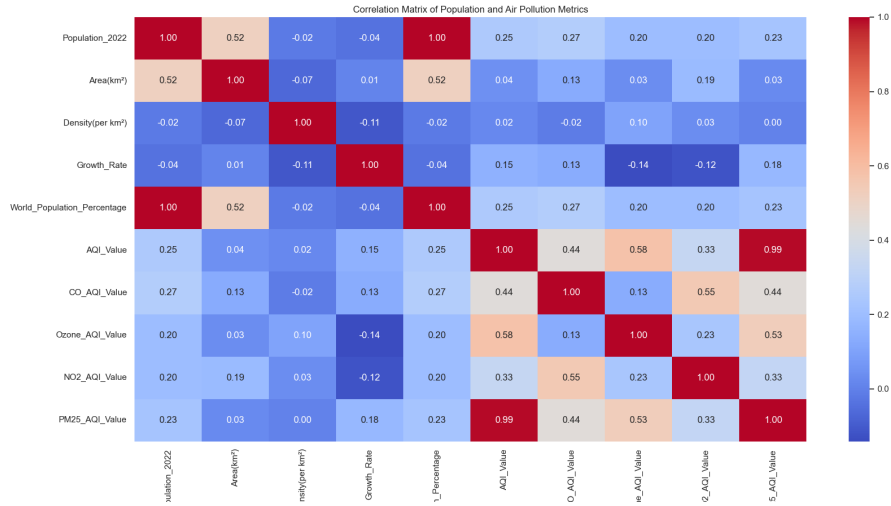
Merged Data:												
	Country	Continent	Population 2022	Area(km²)	Density(per km²)	Growth_Rate	World_Population_Percentage	AQI_Value	CO_AQI_Value	Ozone_AQI_Value	NO2_AQI_Value	PM25_AQI_Value
0	Afghanistan	Asia	41128771	652230	63.0587	1.0257	0.52	95.959184	0.591837	40.183673	0.020408	95.918367
1	Albania	Europe	2842321	28748	98.8702	0.9957	0.04	68.250000	1.000000	42.062500	0.718750	68.093750
2	Algeria	Africa	44903225	2381741	18.8531	1.0164	0.56	88.166667	1.916667	47.166667	9.333333	80.250000
3	Andorra	Europe	79824	468	170.5641	1.0100	0.00	29.333333	0.666667	29.333333	0.000000	22.000000
4	Angola	Africa	35588987	1246700	28.5466	1.0315	0.45	83.925926	3.148148	22.703704	1.888889	81.925926
5	Argentina	South America	45518318	2780400	16.3683	1.0052	0.57	28.235294	0.352941	15.485294	1.544118	24.867047
6	Armenia	Asia	2780469	29743	93.4031	0.9962	0.03	53.636364	0.863636	34.363636	0.000000	52.500000
7	Aruba	North America	106445	180	591.3611	0.9991	0.00	163.000000	0.000000	23.000000	0.000000	163.000000
8	Australia	Oceania	26177413	7692024	3.4032	1.0099	0.33	33.600000	0.211765	22.058824	3.117647	27.364705
9	Austria	Europe	8939617	83871	106.5877	1.0020	0.11	53.720588	0.941176	35.970588	1.102941	52.352941

Figure 1: Final Data

## 5 Analysis

### 5.1 Correlation Matrix

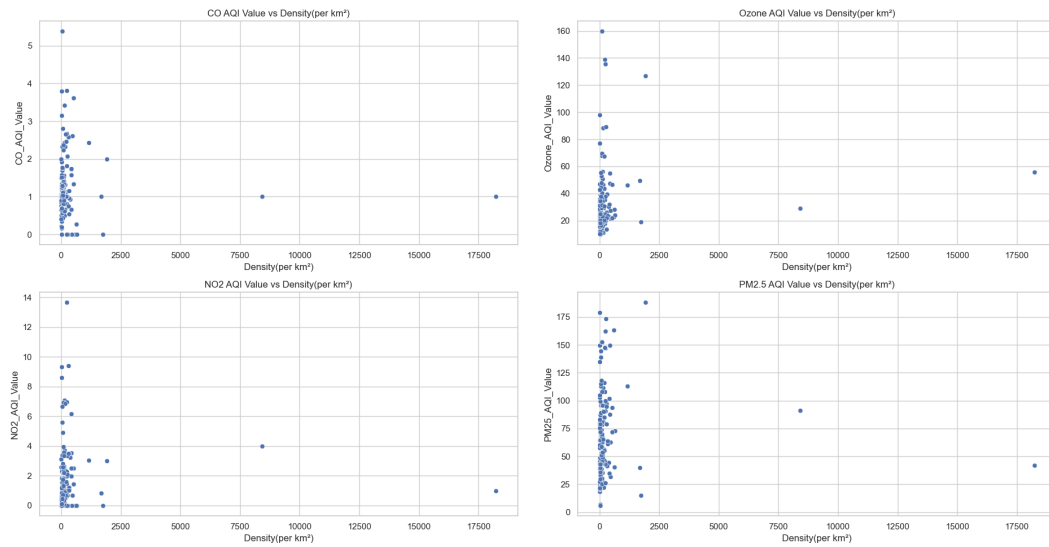
The correlation matrix provides an overview of the relationships between various population metrics and air pollution metrics.



In the Correlation Matrix, these observations are notable:

- **Population and CO Levels:** There is a moderate positive correlation between population size and CO levels (0.27). This indicates that countries with larger populations tend to have higher carbon monoxide levels. The primary sources of CO include vehicle emissions and industrial activities, which are more in densely populated areas.
- **Population and O3 Levels:** Ozone levels also show a positive correlation with population size. Larger populations contribute to these emissions, resulting in higher ozone levels.
- **Population and Particulate Matter (PM2.5):** The correlation between population size and PM2.5 levels is moderate (0.23). PM2.5 particles, which include dust, dirt, soot, and smoke, are often generated by industrial activities, construction, and transportation, all of which are more intense in crowded regions.

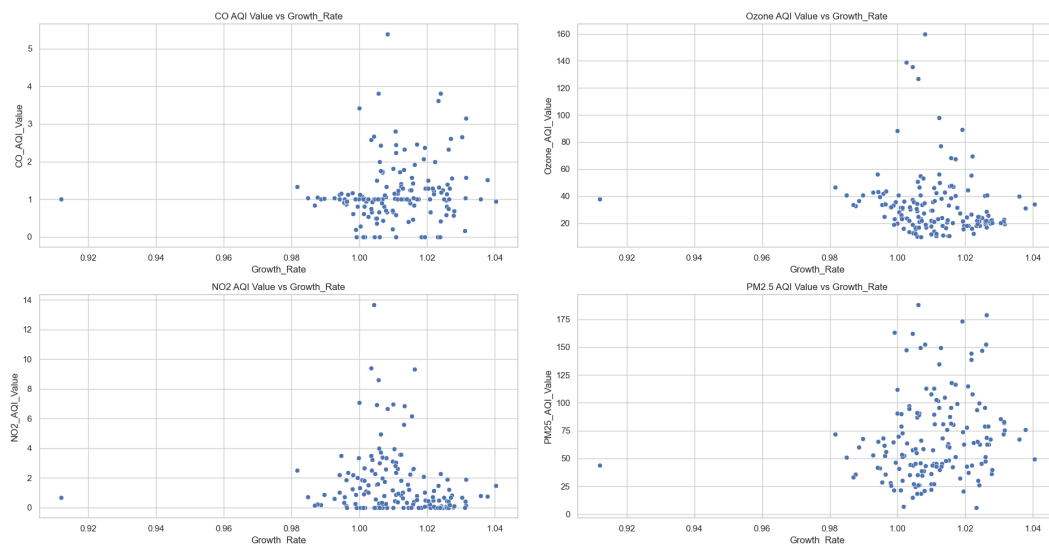
## 5.2 Density Analysis



Key observations:

- **CO Levels:** The plot shows a significant number of data points clustered at lower density values with high CO levels, suggesting that even countries with lower population densities can experience high CO pollution, possibly due to high vehicle usage or industrial activities.

## 5.3 Growth Rate Analysis



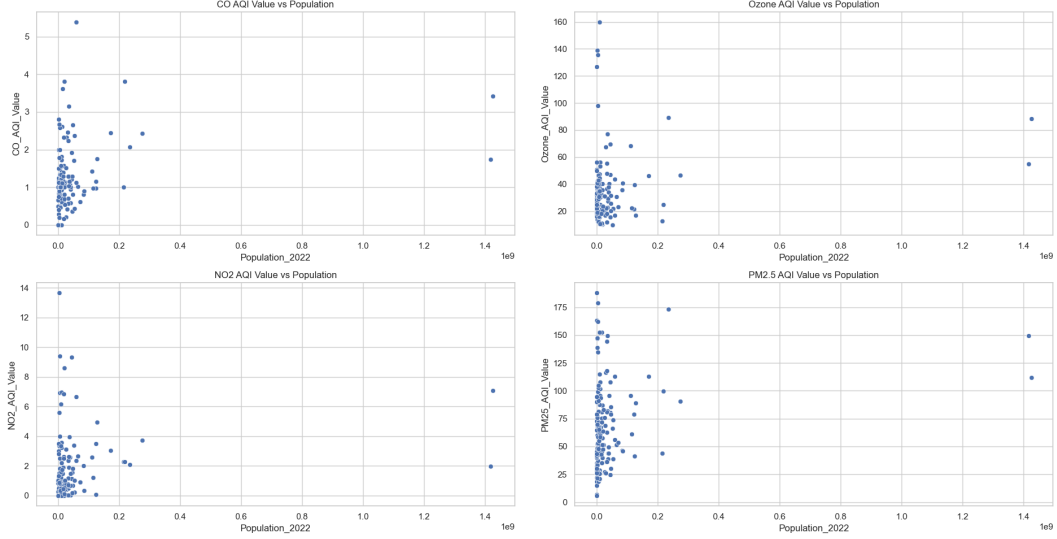
Key Observations:

- **PM2.5 Levels:** Higher growth rates correlate with higher PM2.5 levels, which can be

---

attributed to increased construction, industrial activities, and use of fossil fuels as the population and economy expand.

## 5.4 Population Analysis



### Key Observations:

- **CO and Ozone Levels:** Larger populations tend to have higher levels of CO and ozone, indicating that these pollutants are significantly influenced by the total number of people and the associated activities such as transportation and energy consumption.
- **NO2 and PM2.5 Levels:** The correlation is less but still present.

## 6 Conclusions

In conclusion, the analysis highlights the complex relationships between population metrics and air pollution levels. While larger populations and higher growth rates are generally associated with increased pollution levels, population density alone does not exhibit a strong correlation with air pollution. Instead, a combination of factors, including industrial activities, vehicular emissions, and urban planning, plays a important role in determining air quality.