**COMPARING THE AIR QUALITY INDEX OF DIFFERENT COUNTRIES TO ANALYZE THE IMPACT OF POLLUTION ON THE ENVIRONMENT**

# PYTHON PROJECT REPORT

**SUBMITTED** **BY:** **KANIKA** **PARASHAR AND TAMANNA PINKY**

**UNDER** **THE** **SUPERVISION** **OF**

**DR.** **CHONGLE PAN**

A red and white sign

Description automatically generated with low confidence

# DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY UNIVERSITY OF OKLAHOMA, NORMAN, OK 73019

**Objectives:**

The goal of this project is to answer the question: "What factors influence air quality variability across cities and countries?” Pollution and its impact on the environment are global concerns. This project is centered on the analysis of data from the Air Quality Index (AQI), considering pollutants such PM2.5 (particulate matter with a diameter of 2.5 micrometers), ozone, carbon monoxide (CO), and nitrogen dioxide (NO2). Ozone, NO2, and CO are important markers of several types of air pollution. A thorough evaluation of the primary pollutants from combustion processes (CO, NO2) as well as the secondary pollutants from intricate chemical reactions (Ozone) can be accomplished by keeping an eye on these metrics. When taken as a whole, they offer insightful information about the causes and effects of air pollution. By examining these contaminants in conjunction with other air quality metrics, one can obtain a thorough understanding of the prevailing environmental conditions and identify the factors and trends that contribute to variances in air quality between different regions.

By employing techniques such as exploratory data analysis, and correlation analysis, the project seeks to uncover patterns and relationships that contribute to air quality fluctuations. The outcomes of this research hold significance for informing environmental policies and promoting healthier living environments on a global scale. For this project pie charts and bar charts are strategically combined to create a comprehensive visualization strategy. The bar graph provides insights into variances in air quality by city and goes into detail within each country. In addition, the pie chart provides information with varying degrees of AQI categories as good or moderate with a reference by summarizing the overall distribution of AQI categories. By combining these two methods, the efficiency of visual interpretation is maximized, and air quality patterns may be thoroughly explored and understood at a high level.

**Significance:**

This project is significant due to the growing risks that air pollution poses to the environment and public health. Comprehending the worldwide variations in Air Quality Index (AQI) is important for formulating well-informed policy judgments and executing efficient pollution mitigation strategies. It expands the investigation to a worldwide scale, building on previous research that has mostly concentrated on regional air quality. This project's importance comes from the way it fills in the current research gaps. Most of the prior research has focused on local air quality; this study expands that analysis to a worldwide scale. Given the detrimental consequences air pollution has on ecosystems, human health, and the dynamics of climate change, a more comprehensive knowledge of the worldwide effects of air pollution is needed.

We provide a detailed interpretation of these results in the results and discussion section, looking at possible relationships, outliers, and significant patterns. The study addresses the implications of these findings in addition to summarizing the most important conclusions from the analysis and recommending environmental actions that could enhance the quality of the air in certain areas. Nevertheless, this study report contributes to a deeper comprehension of the worldwide effects of air pollution. It seeks to provide information for stimulate additional study and encourage cooperative efforts to lessen the global impact of air pollution. The research serves as evidence of how critical it is to address inequities in air quality as a first step toward a more sustainable and healthy future.

**Data Sources:** Air quality data collected from the epa.govt source.

****

****

**Methods:**

In the python code, we analyzed air quality data using pandas and matplotlib. The code appears to read a CSV file containing air quality information and then creates visualizations to compare AQI (Air Quality Index) values across different cities and countries. The code explores and presents important findings on air quality by fusing the visualization ability of matplotlib with the data manipulation capabilities of pandas. While pie charts provide a visual depiction of the distribution of AQI categories and values throughout the dataset, bar graphs enable comparison of AQI levels between cities and countries. The data on air quality can be better understood and patterns and trends can be found with the use of these graphic aids. The breakdown of the methods used in the code will discuss in the results and discussion section.

**Results:**

By employing the specified data analysis procedure and examining the resulting outputs, we gain the capability to evaluate countries systematically, utilizing their air quality index values as a critical metric. This approach empowers us to pinpoint the most polluted countries on a global scale, shedding light on the severity of air pollution around the world. Furthermore, the analysis enables us to delve deeper into the dynamics at play within these countries by identifying the specific cities that contribute significantly to this environmental challenge. In essence, this method not only facilitates a comprehensive understanding of the overall air quality landscape but also allows for targeted interventions and strategies to address the most pressing pollution hotspots, contributing to more effective environmental management and policy formulation.

**Discussion:**

**Code 1:**

**Step 1:**

A close-up of a computer

Description automatically generated

This Python code utilizes the Pandas and Matplotlib libraries for data manipulation and visualization, respectively. Let's break down the code step by step:

**a.** **Importing Libraries:** Here, the code imports the Pandas library and assigns it the alias 'pd'. It also imports the 'pyplot' module from Matplotlib and assigns it the alias 'plt'.

A close up of a website

Description automatically generated

**b.** **Reading CSV file:** The code reads a CSV file located at the specified 'file\_path' and creates a Pandas DataFrame ('df') to store the data. The 'read\_csv' function is used to load the data into the DataFrame.

A black text on a white background

Description automatically generated

**c.** **Displaying DataFrame Head:** This line prints the first few rows of the DataFrame to the console, allowing the user to verify that the data has been loaded correctly.

A close up of a text

Description automatically generated

**Step 2:**

**d.** **Bar graph-comparison of AQI values:** This section creates a bar graph that compares the Air Quality Index (AQI) values for different cities within each country. It iterates over unique countries in the DataFrame, extracts data for each country, and plots a bar for each city within that country. The resulting graph provides a visual comparison of AQI values across different places.

A white box with black text

Description automatically generated

**Output**

**A graph of different colored bars

Description automatically generated**

**Step 3:**

**e. Pie Chart -Distribution of AQI categories:** This section generates a pie chart that illustrates the distribution of AQI categories across all places in the dataset. It calculates the counts of each AQI category and creates a pie chart with corresponding labels, percentages, and colors.

A white rectangular sign with black text

Description automatically generatedIn summary, this code is designed to load and analyze air quality data from a CSV file, presenting insights through a bar graph comparing AQI values within countries and a pie chart visualizing the distribution of AQI categories across all places.

**Output**

**A pie chart with numbers and a red circle

Description automatically generated**

**Code 2:**

**A white rectangular box with black text

Description automatically generated**

This code is designed to analyze and visualize the distribution of Air Quality Index (AQI) categories across all places in a dataset. Let's break down the steps:

**a.** **Importing Libraries:** Here, the code imports the Pandas library and assigns it the alias 'pd'. It also imports the 'pyplot' module from Matplotlib and assigns it the alias 'plt'.

A close up of a website

Description automatically generated

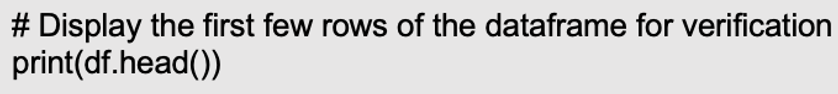
**Step 1:**

**b.** **Reading the CSV file:** Here, the code reads a CSV file located at the specified 'file\_path' using Pandas' read\_csv function. The data is loaded into a Pandas DataFrame named 'df,' which is a tabular data structure that allows for easy manipulation and analysis of the dataset.

A black text on a white background

Description automatically generated

**c.** **Displaying DataFrame Head:** This line prints the first few rows of the DataFrame to the console using head(). This is a common practice to quickly verify that the data has been loaded correctly and to get a glimpse of its structure.



**Step 2:** The step creates a pie chart using the matplotlib library in Python.

**d. Pie Chart -Distribution of AQI categories: Here's a breakdown of each part:**

1. **plt.figure(figsize=(8, 8)):**

This line creates a new figure (or window) for the plot and sets its size to 8 inches by 8 inches. The figsize parameter allows you to control the dimensions of the figure.

1. **plt.pie(df['AQI Value'], labels=df['City'], autopct='%1.1f%%', startangle=90):**

* This line creates a pie chart using the plt.pie function.
* The data for the pie chart comes from the 'AQI Value' column of the DataFrame (df) for each city, and the labels for each slice come from the 'City' column.
* **autopct='%1.1f%%'** adds percentage labels to each slice, displaying one decimal place.
* **startangle=90** sets the starting angle of the first slice at 90 degrees (which is typically the top of the pie chart).

**3.plt.title('Distribution of AQI Values Across Different Cities'):**

This line adds a title to the pie chart, describing its purpose or the data it represents.

**4.plt.show():**

This line displays the pie chart. If you are running this code in a Jupyter Notebook or an interactive environment, the chart will be displayed directly below the code cell. If you are running it in a script or a non-interactive environment, this line is necessary to actually show the chart.

The code, therefore, generates a pie chart illustrating the distribution of Air Quality Index (AQI) values across different cities. Each slice of the pie represents a city, and the size of each slice corresponds to the AQI value for that city. The percentage labels indicate the proportion of the total AQI values that each city contributes.

A black text on a white background

Description automatically generated

**A colorful circle with text

Description automatically generated**

**Future work:**

The basis for the analysis and visualization of air quality is the Python code that has been provided. Future research opportunities include investigating variations of not only air pollution as well as other key environmental pollution, comprehending correlations with external factors, creating predictive models, analyzing spatially, assessing health implications, integrating satellite data, putting real-time monitoring data source, encouraging citizen science and public awareness, comparing studies across regions, and extending analyses to considering multiple pollutants at once.

**Workload Distribution:**

Graduate Student (5783): Both Graduate students are responsible for statistical analysis, GIS visualization, regression analysis, and report writing. Expected workload: 60%.