**Shadowfox Data Science Task 2**

**Report on Air Quality Analysis Code**

The provided code conducts an analysis of air pollutant levels in Delhi using a dataset containing various air quality metrics over time. It leverages libraries such as Pandas for data manipulation, Matplotlib for plotting, and Seaborn for correlation visualization.

**1. Data Loading and Initial Processing:**

\* The dataset is loaded from a CSV file, and the first few rows are printed for an initial overview.

\* Dates are converted to a `datetime` format for easier manipulation.

\* Missing values are removed to ensure a clean dataset for analysis.

**2. Descriptive Statistics:**

\* The `describe()` function generates summary statistics (count, mean, std, min, max, etc.) for the pollutant levels, offering insight into the distribution of the data.

**3. Time Series Visualization:**

\* A line plot is created to visualize the concentrations of various pollutants over time. This helps identify trends, spikes, and seasonal patterns in air quality.

**4. Hourly Average Analysis:**

\* The dataset is enriched by extracting the hour and month from the date.

\* Average pollutant levels are calculated by hour, allowing for a deeper understanding of daily pollution patterns.

\* A bar plot visualizes these hourly averages.

**5. Correlation Analysis:**

\* A correlation matrix is computed to analyze the relationships between different pollutants.

\* A heatmap is generated to visualize these correlations, helping to identify which pollutants may be associated with one another.

**Graphs Used and Their Definitions**

**1. Time Series Plot:**

A line graph displaying the levels of air pollutants (e.g., CO, NO2, PM2.5) over time.

Where it is used: To visualize trends and variations in air quality over the specified period.

**2. Hourly Averages Plot:**

A bar graph showing average pollutant concentrations for each hour of the day.

Where it is used: To identify peak pollution hours and understand daily cycles in air quality.

**3. Correlation Heatmap:**

A graphical representation of the correlation coefficients between pairs of pollutants.

Where it is used: To assess the relationships among pollutants, indicating how changes in one may be associated with changes in another.

**Overall Results of the Code**

1. **Trends**: The time series plot highlights fluctuations in pollutant levels, indicating periods of higher pollution. Seasonal variations may also be visible, suggesting environmental or anthropogenic factors influencing air quality.
2. **Daily Patterns**: The hourly average plot reveals typical pollution trends throughout the day, which can be critical for public health and policy decisions, indicating times when pollution is at its peak.
3. **Correlations**: The heatmap shows the relationships between various pollutants, which can inform further studies on sources of pollution and their combined effects on health and the environment.

**Conclusion**

This analysis provides valuable insights into air quality trends and patterns in Delhi. The visualization techniques employed not only enhance understanding of the data but also support further research and decision\*making related to air quality management. By continuing to analyze these trends and correlations, stakeholders can better address air pollution issues in the region.



Embedded the assignment in the above pdf.