Delhi Air Pollution Trend Analysis (2020-2024)

1. Data Loading & Initial Exploration

- Load the dataset using `pandas.read_csv()` from the KaggleHub path.
- Use `.head()`, `.info()`, and `.describe()` to understand the structure.
- Identify missing values or incorrect data (e.g., negative or zero PM2.5).

2. Data Cleaning

- Convert the datetime column to proper datetime format.
- Filter rows between 2020-2024.
- Create `year` and `month` columns.
- Remove or fill missing values appropriately.

3. Analysis Tasks

a. Monthly Average Trends:

- Group by year and month; calculate mean for PM2.5, PM10, NO2, and AQI.

b. Pollution Severity Categorization:

- Use NumPy to classify AQI levels (Good, Satisfactory, Moderate, etc.).

c. Peak Pollution Periods:

- Top 5 most polluted days (highest AQI).
- Month with highest average PM2.5.

4. Visualizations

- Line Plot: Monthly average PM2.5 and PM10 (2020-2024).
- Bar Chart: Average AQI per year.

Delhi Air Pollution Trend Analysis (2020-2024)

- Boxplot: PM2.5 distribution per year.
- Heatmap: Correlation among pollutants.
- Histogram/KDE: AQI distribution.

5. Advanced Challenge

Create a function:

```
def get_pollution_summary(pollutant, year=None):
    Returns stats and visual for given pollutant/year.
```

6. Deliverable

- Submit as a Jupyter Notebook.
- Include markdown explanations and well-labeled plots.
- Conclude with insights (2-3 bullet points).