Task Level (Intermediate):

1. Conduct an in-depth analysis of the Air Quality Index (AQI) in Delhi, addressing the specific environmental challenges faced by the city. Define research questions centered around key pollutants, seasonal variations, and the impact of geographical factors on air quality. Utilize statistical analyses and visualizations to gain insights into the dynamics of AQI in Delhi, offering a comprehensive understanding that can inform targeted strategies for air quality improvement and public health initiatives in the region.

Dataset-Link:

https://drive.google.com/drive/folders/1Bjn2YEmafyYckkJAwsOSqngpEMGWyIXc? usp=drive_link

Aim of the Task

The main aim of this task is to conduct an in-depth analysis of the Air Quality Index (AQI) in Delhi, identify the major pollutants contributing to poor air quality, examine seasonal and diurnal variations, and assess the influence of environmental and geographical factors. This study also aims to provide insights that can inform targeted strategies for air quality improvement and public health protection.

Objectives

1. To understand the overall trend of AQI in Delhi.
2. To identify the key pollutants contributing to poor air quality.
3. To analyze seasonal and diurnal variations in pollutant concentrations.
4. To examine geographical and meteorological factors influencing Delhi's air quality.
5. To provide data-driven recommendations for air quality management and health initiatives.

Methodology

- **Dataset Used:** AQI dataset from Delhi monitoring stations (Jan 1–24, 2023).
- **Analysis Performed:**
  - Data cleaning and preprocessing (standardizing timestamps, column names).

- Statistical summaries (mean, median, minimum, maximum values for pollutants).
- Visualizations of pollutant trends (daily and hourly averages).
- Seasonal analysis of AQI and pollutant behavior (Winter focus, as only Jan data available).
- Correlation analysis among pollutants and meteorological variables.
- Predictive modeling using a Random Forest to estimate pollutant concentrations from co-pollutants and environmental factors.
- Event analysis (comparison of high-pollution days vs lower-pollution days).

All analysis outputs are supported with **visualizations** generated from the dataset (time-series plots, boxplots, diurnal cycle plots, correlation matrix, and feature importance bar charts).

Analysis & Findings

1. Dataset Scope

- Records: **561 rows**.
- Coverage: **2023-01-01 to 2023-01-24** (limited to January only).
- Pollutants available: **PM10, NO_2, SO_2, CO, O_3**.
- Missing: **PM2.5 and AQI**, which limits direct AQI calculation.

---

2. City-Level Summary (Descriptive Statistics)

- **PM10:** Mean ~421 µg/m³ (Median ~341) → *Extremely high, well above Indian standards*.
- **NO_2:** Mean ~75 µg/m³ (Median ~64).
- **SO_2:** Mean ~65 µg/m³ (Median ~67).
- **CO:** Mean ~3815 (units unclear).
- **O_3:** Mean ~30 µg/m³ (Median ~0, skewed distribution).

Interpretation:

- PM10 is the dominant pollutant in this dataset, consistently at hazardous levels.
- NO_2 and CO also show high concentrations, indicating strong combustion sources.
- O_3 values appear inconsistent, possibly due to data gaps or unit mismatches.

---

3. Diurnal (Hour-of-Day) Variations

- **PM10:** Peaks at **17:00 (5 PM)**.
- **NO_2:** Peaks at **13:00 (1 PM)**.
- **SO_2:** Peaks at **14:00 (2 PM)**.
- **CO:** Peaks at **15:00 (3 PM)**.
- **O_3:** Peaks at **09:00 (9 AM)**.

Interpretation:

- Evening peaks of PM10 and CO indicate traffic congestion and resuspension of road dust.

- Midday NO_2 and SO_2 suggest strong industrial/vehicular influence.
- O_3's morning peak is atypical and may reflect measurement issues or local photochemistry.

(Visualizations included: Diurnal cycle plots for each pollutant)

---

4. Correlation Analysis

- PM10 strongly correlated with CO (**$\rho \approx 0.92$**) and NO_2 (**$\rho \approx 0.71$**).
- O_3 negatively correlated with PM10, NO_2, and CO (e.g., PM10 vs $O_3 \approx -0.69$).

Interpretation:

- High correlation between PM10, CO, and NO_2 indicates shared sources (traffic, combustion).
- Negative relationship with O_3 suggests chemical titration or contrasting diurnal cycles.

(Visualization included: Heatmap of correlation matrix)

---

5. Predictive Modeling (Random Forest)

- **Target variable:** PM10 (since PM2.5 was absent).
- **Top predictive features:** CO, NO_2, and SO_2.
- Model performed with MAE ≈ *moderate error*; results confirm that co-pollutants strongly explain PM10 variations.

(Visualization included: Random Forest feature importance plot)

---

6. Seasonal & Event Analysis

- **Limitation:** Only January data is available → full seasonal comparison not possible.
- Within January, elevated PM10 concentrations indicate winter inversion conditions.
- Event comparisons (Diwali, stubble burning) cannot be assessed from this dataset.

Conclusion

The analysis highlights that **Delhi's air quality in January 2023 was dominated by extremely high PM10 levels**, with NO_2 and CO as strong co-pollutants. Diurnal cycles reveal traffic-related peaks, while correlation patterns confirm shared combustion sources. Seasonal and event analysis was constrained by the short dataset, but the findings align with known winter pollution phenomena in Delhi.

Health Implications

- High PM10 levels increase risks of **asthma, bronchitis, lung damage, cardiovascular disease, and premature mortality**.
- Vulnerable groups (children, elderly, people with respiratory conditions) are most at risk.

Recommendations

1. **Transport Emissions:** Expand public transport and shift toward electric vehicles.
2. **Dust Control:** Implement road dust suppression and regulate construction site emissions.
3. **Industrial Regulation:** Enforce stricter standards for industrial emissions.
4. **Stubble Management:** Introduce alternatives to crop burning in neighboring states.
5. **Awareness:** Conduct campaigns on air pollution health risks and personal protection.
6. **Monitoring Expansion:** Increase the number of monitoring stations for real-time alerts.