Air Quality Project – Summary Report

This short report summarizes key findings and contains quick links to generated outputs and visuals.

Executive summary
- Dataset: Air quality data for Indian cities (2015–2024).
- Purpose: Clean, analyze, and visualize pollutant trends, hotspots, and station coverage.

Key findings
- Top polluted cities (by average AQI):
 - Mumbai
 - Delhi
 - Kolkata
 - Chennai
 - Bangalore

- Seasonal patterns: Winter and spring show elevated PM2.5 and PM10 across many cities (see `output/seasonal_means_by_pollutant.csv`).
- Correlations: PM2.5 and PM10 show strong positive correlation across the dataset.
- Stations: Station counts are available in `output/stations_summary.csv`.

Important outputs
- CSV reports: `output/top_polluted_cities.csv`, `output/city_yearly_avg.csv`, `output/stations_summary.csv`, `output/seasonal_means_by_pollutant.csv`, `output/missing_values_summary.csv`.
- Visuals: See the `visuals/` folder for charts and the interactive `visuals/city_comparison_dashboard.html`.

How to regenerate
1. Create a virtual environment and install dependencies:

```powershell
python -m venv .venv
.\.venv\Scripts\Activate.ps1
pip install -r requirements.txt
```

2. Run the pipeline:

```powershell
python scripts/clean_data.py
python scripts/analyze_data_enhanced.py
python scripts/analyze_stations.py
python scripts/seasonal_trends.py
python scripts/top_polluted_cities.py
python scripts/pollution_hotspots.py
python scripts/city_pollution_over_years.py
python scripts/generate_summary.py
```

Notes
- This summary replaces the previous `output/air_quality_report.pdf`. The markdown is lightweight and easy to view on GitHub.
- To produce a printable PDF, you can export this markdown to PDF externally.