Pune Smart City Environmental Sensor Data Analysis (2019)

Exploratory Data Analysis using Smart City Environmental Sensor Data

Agenda

- Problem Statement
- Dataset Overview and Features Classification
- Tools and Technologies Used
- Data Cleaning and Preprocessing
- Data Quality Assessment
- Monitoring Locations Analysis
- Distribution of Key Air Pollutants
- Environmental Factors Distribution
- Variability Analysis
- Correlation Analysis
- AQI Priority Parameters
- Spatial Analysis – Pollution Hotspots
- Temporal Analysis
- Most Polluted Locations
- Stability Analysis
- Clustering Locations
- Challenges and Data Limitations
- Conclusion

Problem Statement

- Pune faces increasing air pollution due to:
  - Rapid urbanization
  - Traffic congestion
  - Industrial & commercial activity
- City authorities need data-driven insights to:
  - Identify pollution hotspots
  - Understand pollution behavior over time
  - Take targeted corrective action

Dataset Overview and Feature Classification

Dataset: Pune Smart City Environmental Sensor Dataset (2019)

Records: 103,205

Features: 28

Data Type: Time-series + Location-based sensor data

Source: PSCDCL & IISc Bangalore

Location Data: Name, Latitude, Longitude

Air Pollutants: PM2.5, PM10, NO_2, CO, SO_2, Ozone

Environmental Parameters: Temperature, Humidity, Sound, Light, Air Pressure

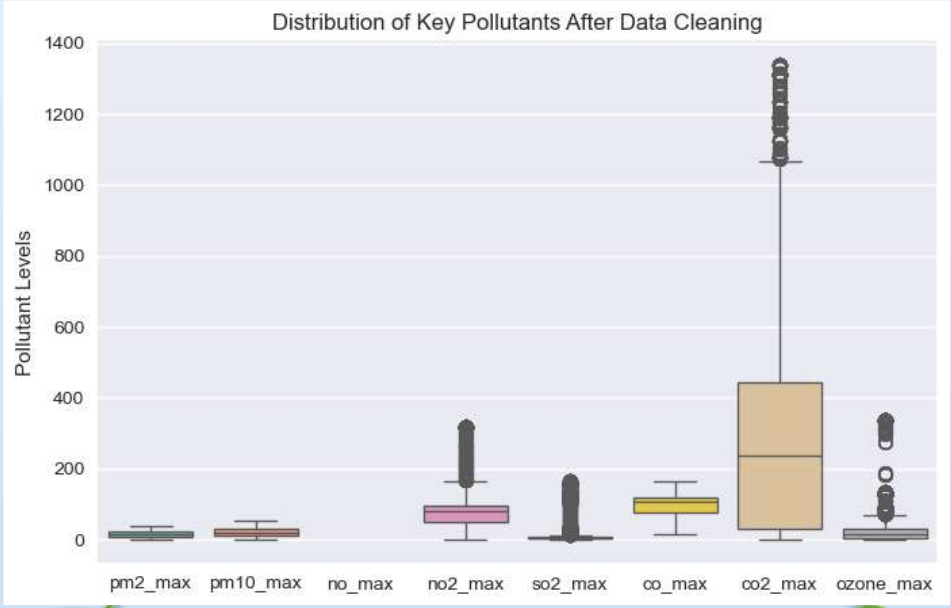
Time Feature: Timestamp, Hour, Day, Weekday

TOOLS & TECHNOLOGIES USED

- Python (Pandas, NumPy) – Data Cleaning and Preprocessing.
- Matplotlib and Seaborn – Exploratory Data Analysis.
- Scikit-learn – Clustering and Pattern Detection.
- Jupyter Notebook – Analytical workflow.
- Smart City Environmental Dataset (PSCDEL, 2019).

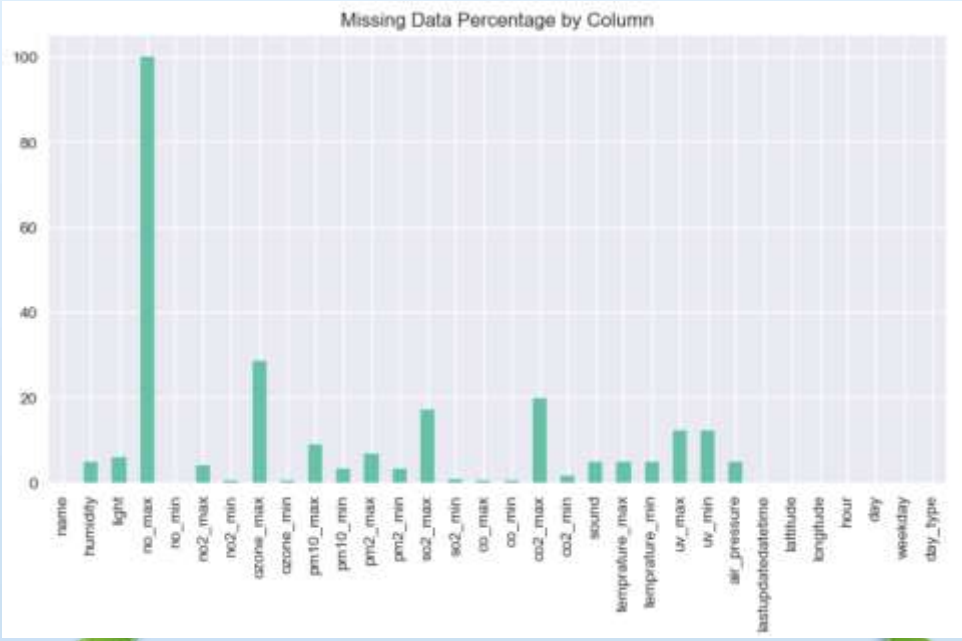
Data Cleaning and Preprocessing

- Handled Missing Values.
- Removed zero & unrealistic sensor readings.
- Treated outliers using IQR method.
- Converted timestamp into meaningful time features.



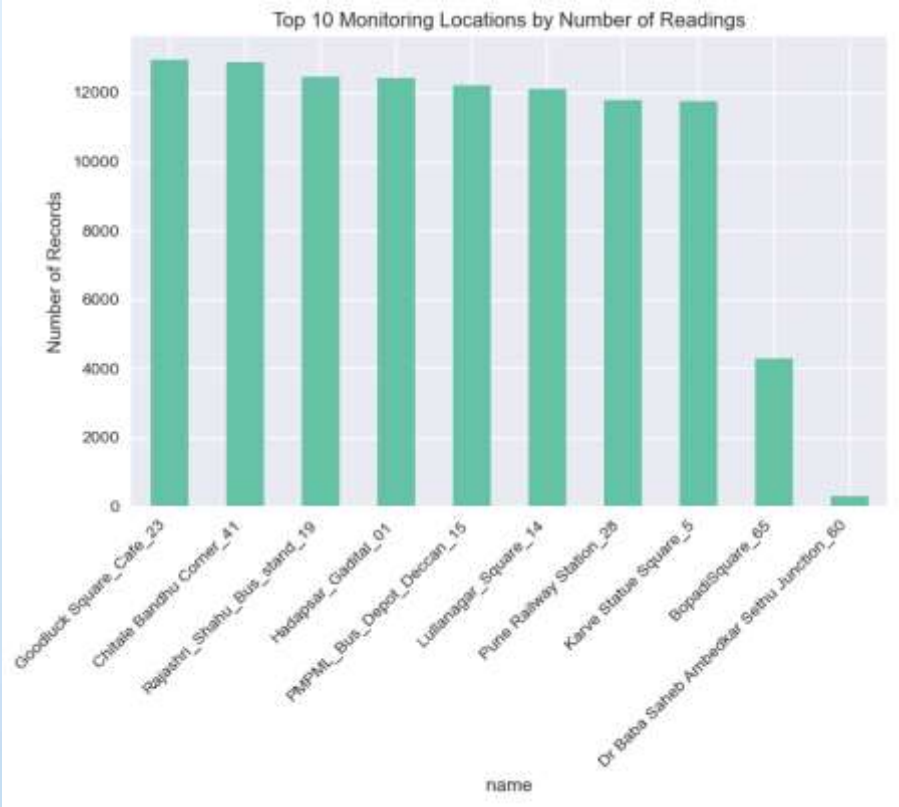
Data Quality Assessment

- Some pollutant sensors showed higher missing values.
- PM2.5 & PM10 had the most quality issues.
- Environmental parameters were relatively stable.



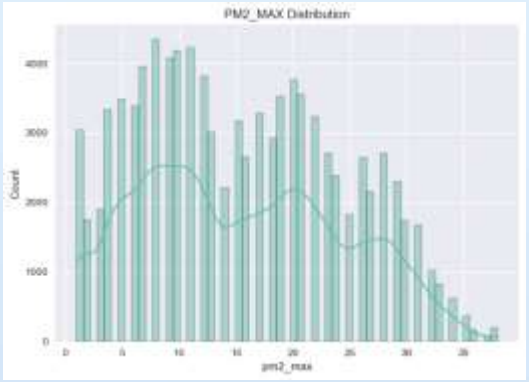
MONITORING LOCATIONS ANALYSIS

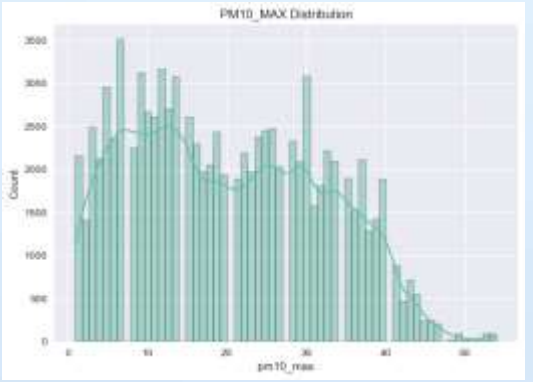
- Multiple monitoring locations across Pune.
- Few locations contributed most readings.
- Uneven sensor coverage observed.

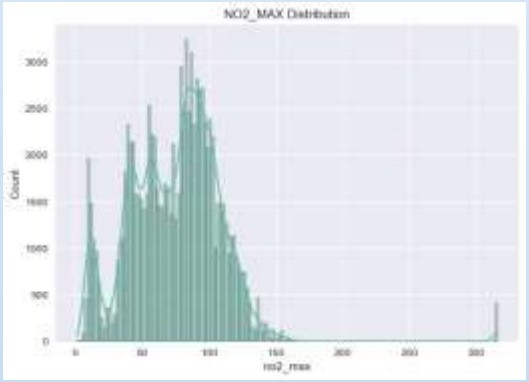


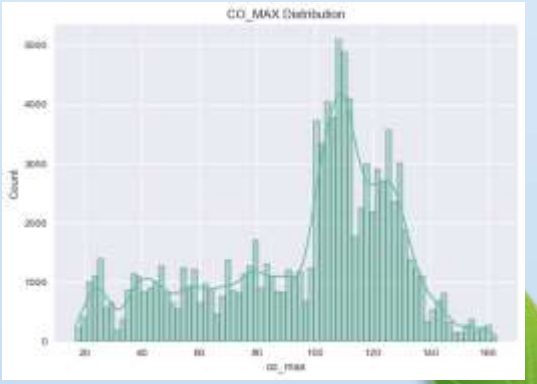
DISTRIBUTION OF KEY AIR POLLUTANTS

- PM2.5 and PM10 show right-skewed distribution.
- Frequent high pollution episodes.
- Presence of extreme values.



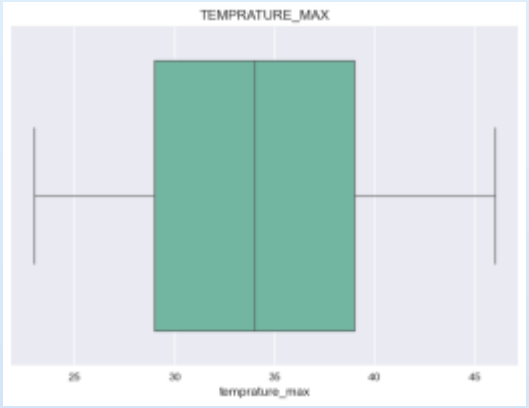


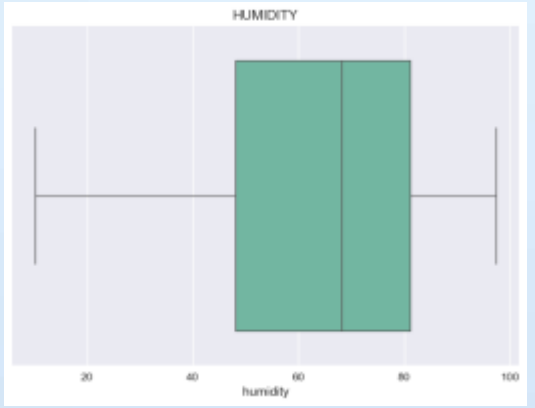


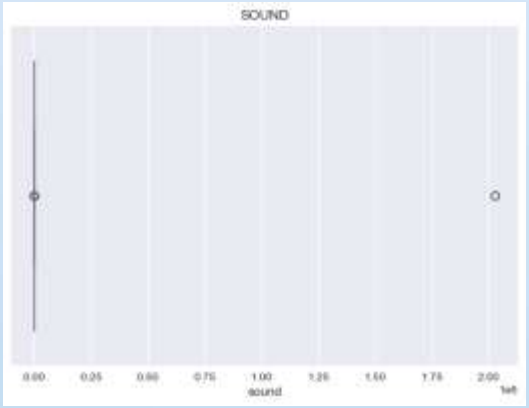


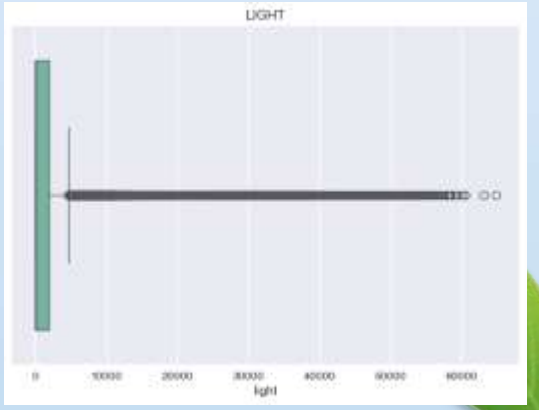
ENVIRONMENTAL FACTORS DISTRIBUTION

- Temperature and Humidity shows normal trends.
- Sound and light indicate urban activity patterns.



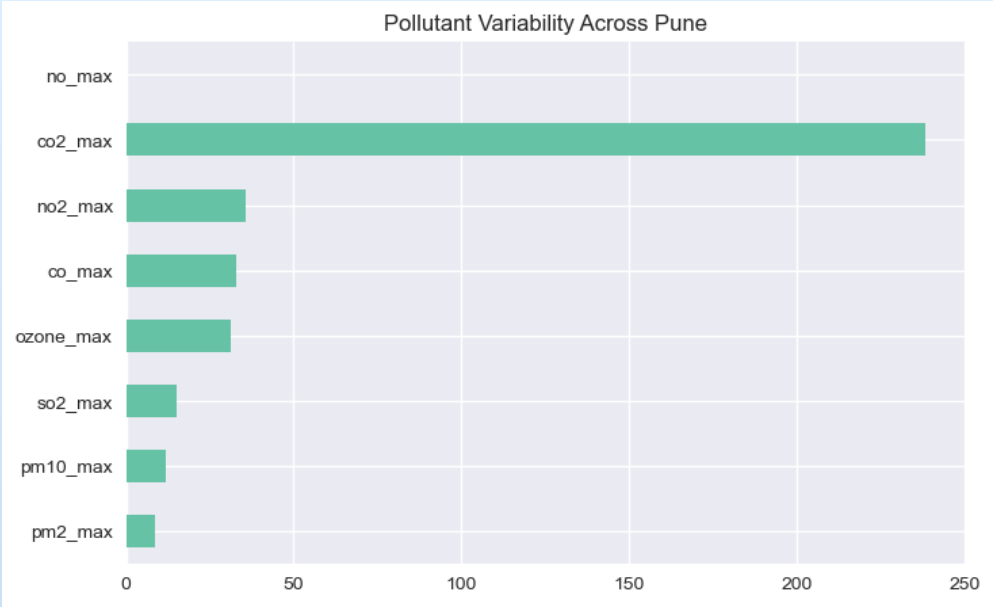






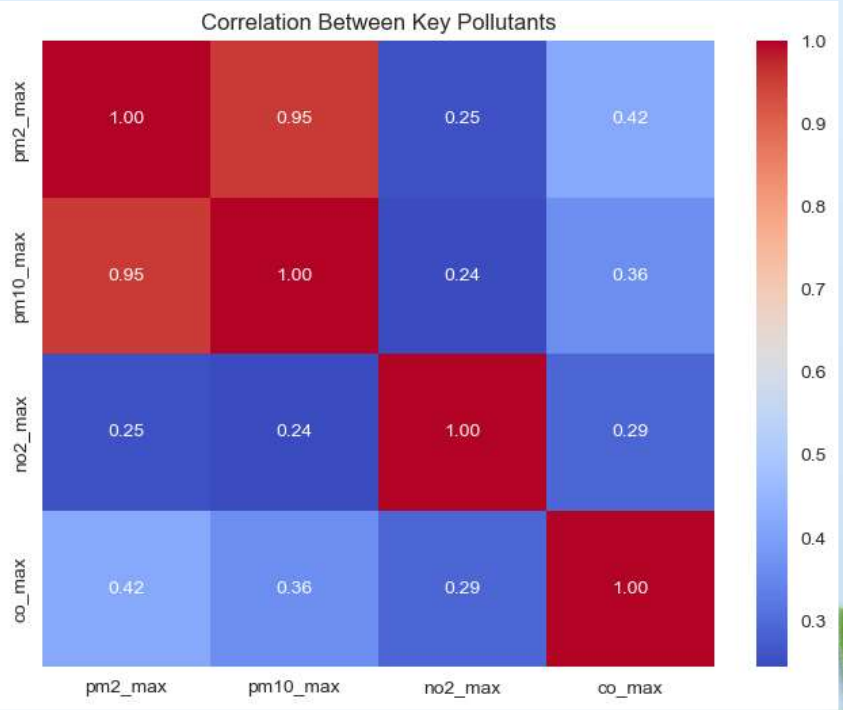
VARIABILITY ANALYSIS

- PM2.5 and PM10 show highest variability.
- Indicates uneven pollution exposure across Pune.



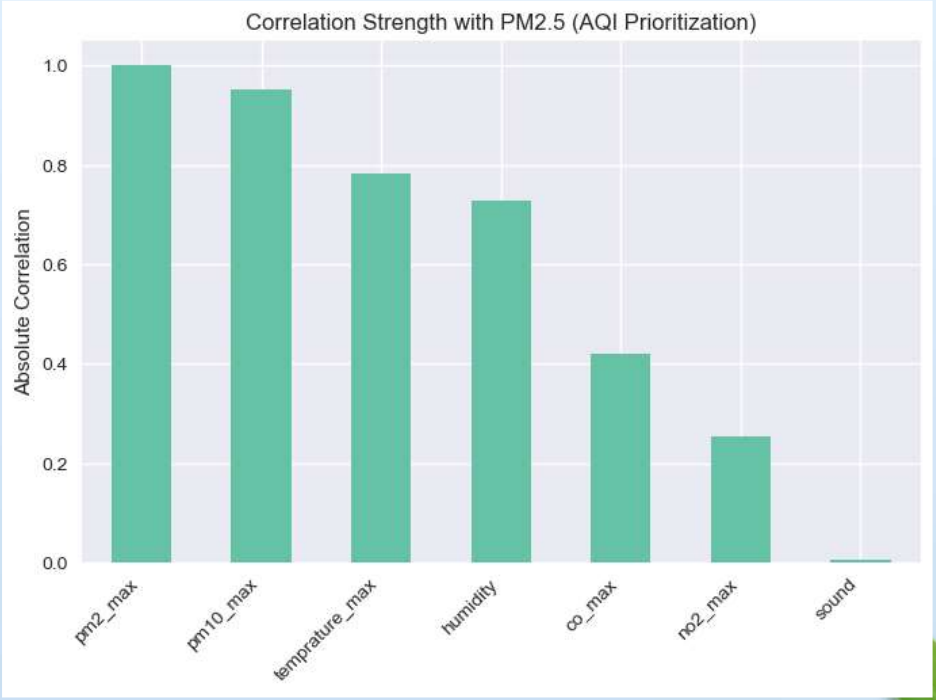
CORRELATION ANALYSIS

- Strong correlation between PM2.5 and PM10.
- CO and NO_2 correlated – Traffic impact.
- Environmental factors moderately influence pollutants.



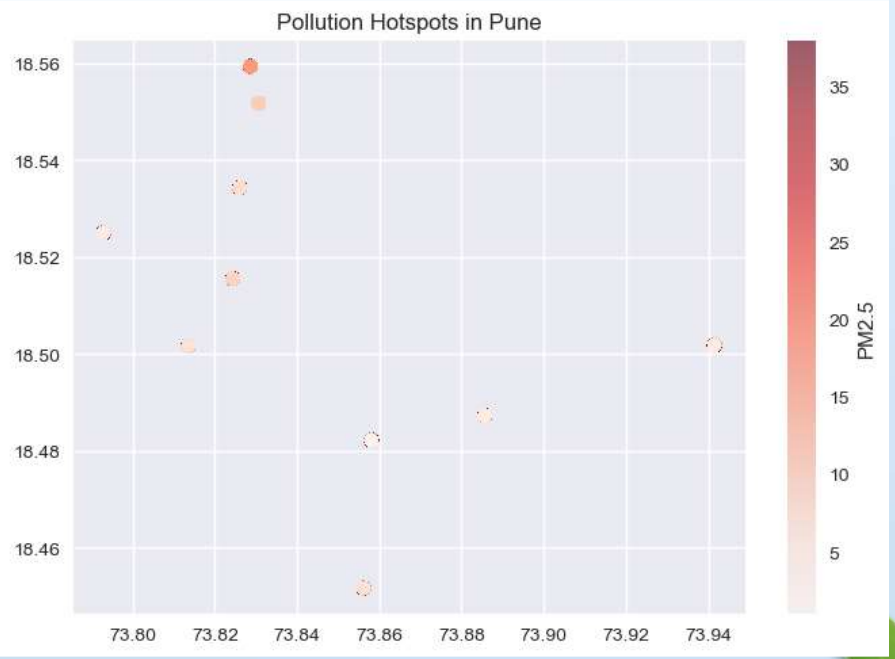
AQI PRIORITY PARAMETERS

- **Top 5 parameters to monitor:**
  - PM2.5
  - PM10
  - NO_2
  - CO
  - Ozone

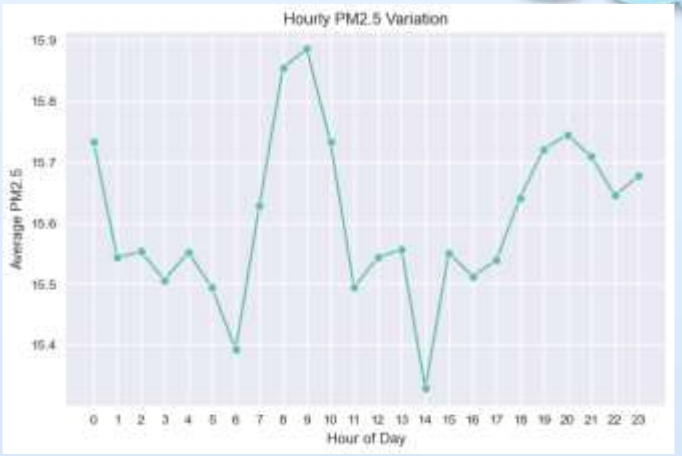


SPATIAL ANALYSIS – POLLUTION HOTSPOTS

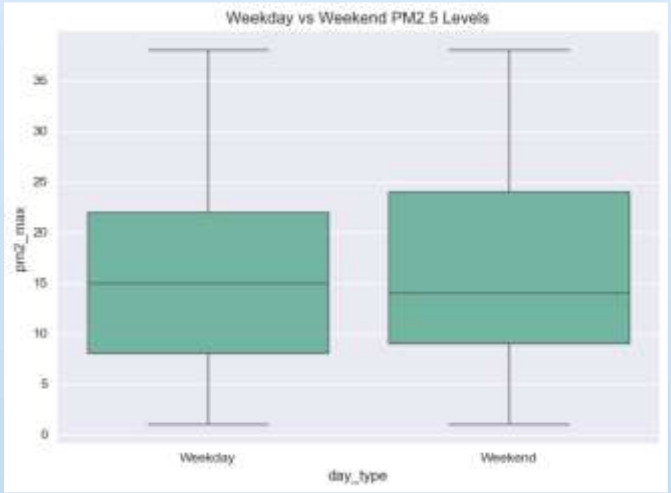
- Clear pollution hotspots identified.
- Transport hubs and dense areas most affected.



TEMPORAL ANALYSIS

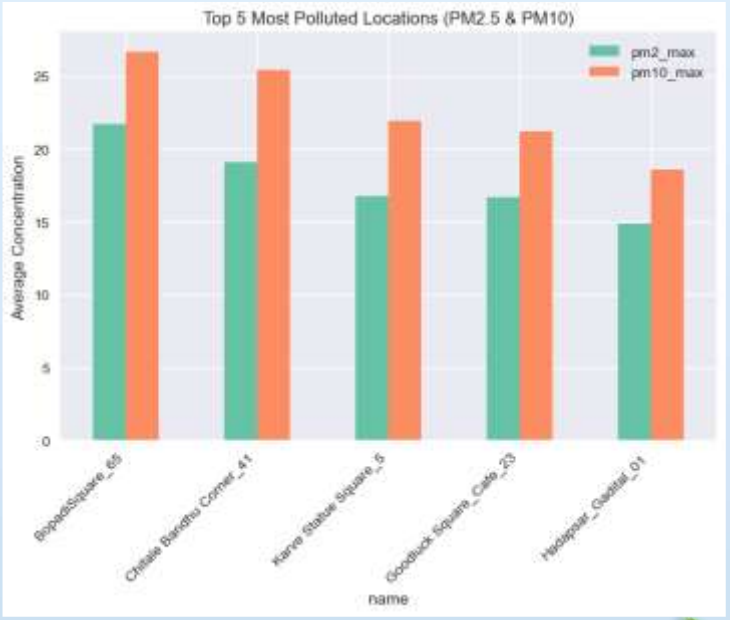


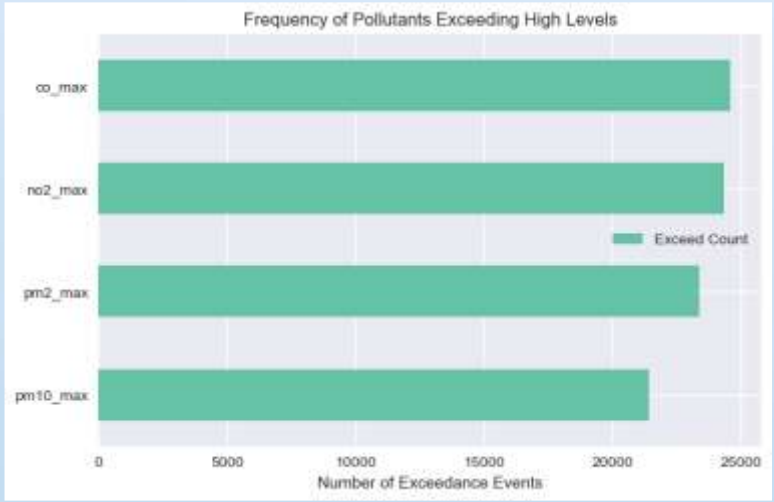
- Pollution peaks during morning and evening hours.
- Weekdays shows higher pollution than weekends.
- Strong link with human activity.



MOST POLLUTED LOCATIONS

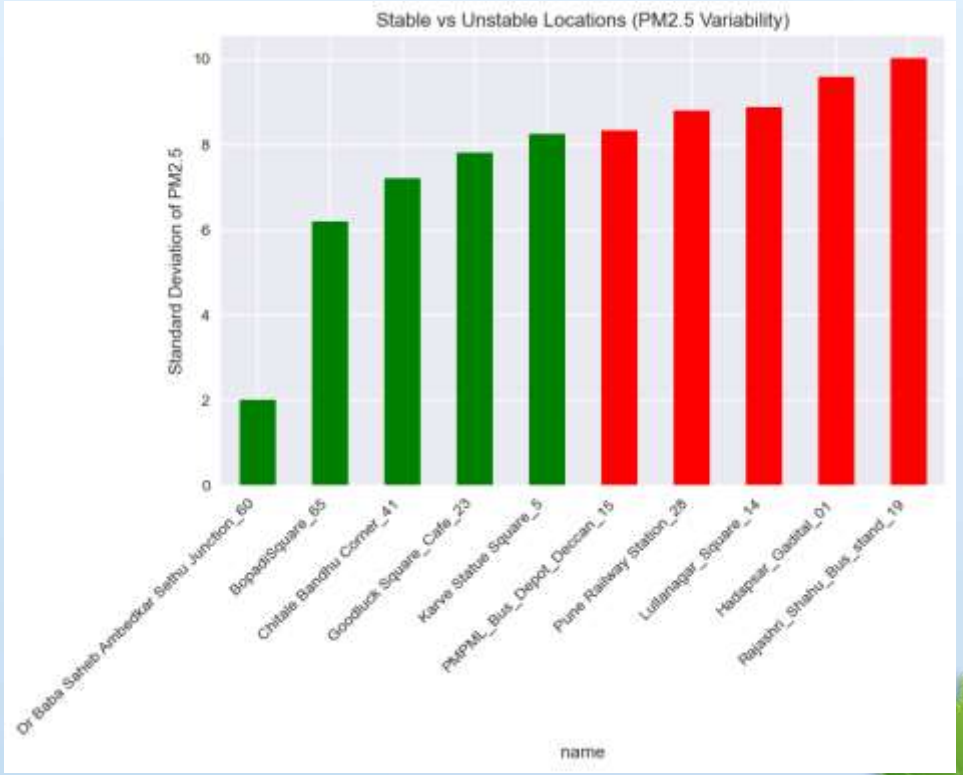
- Ranked top 5 polluted locations.
- PM2.5 and PM10 exceed safe limits most frequently.





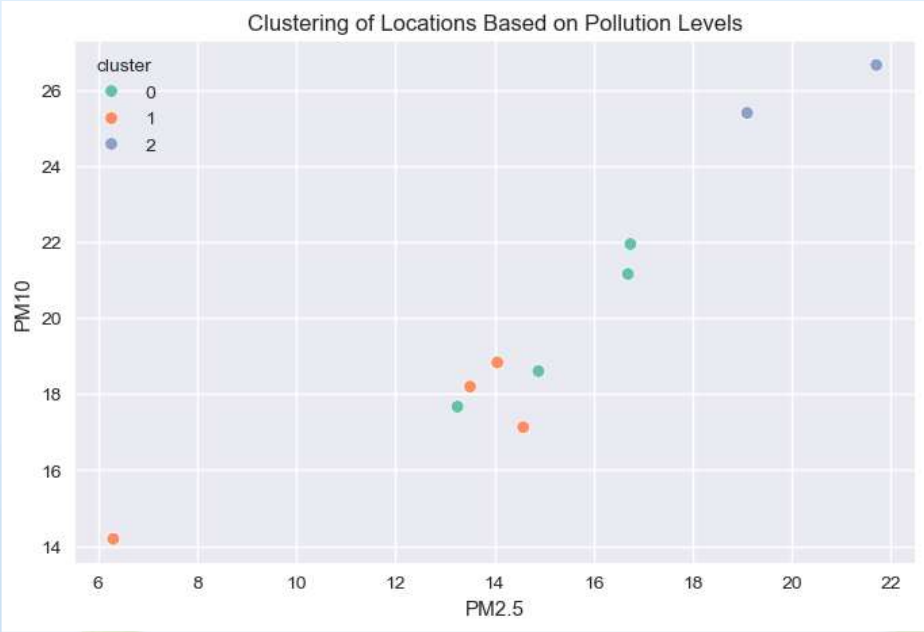
STABILITY ANALYSIS

- Some location shows stable pollution.
- Others show high fluctuation – risk zones.



CLUSTERING LOCATIONS

- Locations clustered into :
  - High pollution
  - Moderate pollution
  - Low pollution
- Helps in targeted interventions.



CHALLENGES AND DATA LIMITATIONS

- Missing values in pollutants sensor (especially UV readings – 11%).
- Presence of Zero & unrealistic environmental readings.
- High variability and extreme outliers in PM2.5 and PM10.
- Uneven distribution of sensors readings across locations.
- Time-series feature extraction and trend alignment.

CONCLUSION

- Pune's environmental sensor data reveals significant variability in air pollution across locations and time.
- PM2.5 and PM10 emerged as the most critical pollutants with strong correlation patterns.
- Pollution peaks were observed during high human activity hours, especially on weekdays.
- Spatial analysis identified key pollution hotspots requiring targeted intervention.
- Data-driven monitoring can significantly improve urban environmental planning and decision-making under the Smart City initiative.

Thank You !