Air Quality Index of Indian Cities - Comprehensive EDA

Presented By- Team MATRIX

TEAM MEMBERS

Names	E-Mail IDs
Manmath Mohanty	manmath.mohanty@adypu.edu.in
Mehebub alom	Mehebub.alom@adypu.edu.in
Yash Sharma	Yash.sharma@adypu.edu.in
Lomesh Sonkeshriya	lomesh.sonkeshriya@adypu.edu.in

INDIVIDUAL CONTRIBUTION

Data Visualisation and Graph Plotting

- Distribution Analysis: Used histograms and boxplots to visualize pollutant_avg distributions across pollutant types.
- Faceting: Plotted data by pollutant type to improve clarity and comparability.
- Heatmap Insights: Revealed strong correlations (e.g., PM2.5 & PM10) and weaker links among other pollutants.Computed pairwise correlations to identify relationships between pollutants.
- Scatter Matrix: Showed pairwise pollutant relationships, with visible clustering patterns influenced by geography

By- Manmath Mohanty

INDIVIDUAL CONTRIBUTION

Bandfort’s Law Implimentation

- Benford’s Law Applied: Tested the distribution of leading digits in pollutant_avg, pollutant_min, and pollutant_max.
- Expected Pattern: Compared actual first-digit frequencies to Benford’s expected distribution (e.g., ~30% should start with ’1’).
- Observed Deviations: Noted significant differences—certain digits like ’1’ appeared less often, while others like ’7’ or ’8’ were more frequent.
- Interpretation: Deviations likely result from bounded pollutant values and measurement/reporting practices, not naturally unbounded data.

By- Yash Kumar Sharma

INDIVIDUAL CONTRIBUTION

Data Cleaning Overview

- Handled missing values and removed duplicates to ensure clean, reliable data.
- Formatted dates correctly and standardized city/station names for consistency
- Created a pivot table to aggregate pollutant levels by city.
- Computed pairwise correlations to identify relationships between pollutants.
- Performed descriptive analysis of pollutant metrics and examined categorical variable distributions.

By- Lomesh Sonkeshriya

INDIVIDUAL CONTRIBUTION

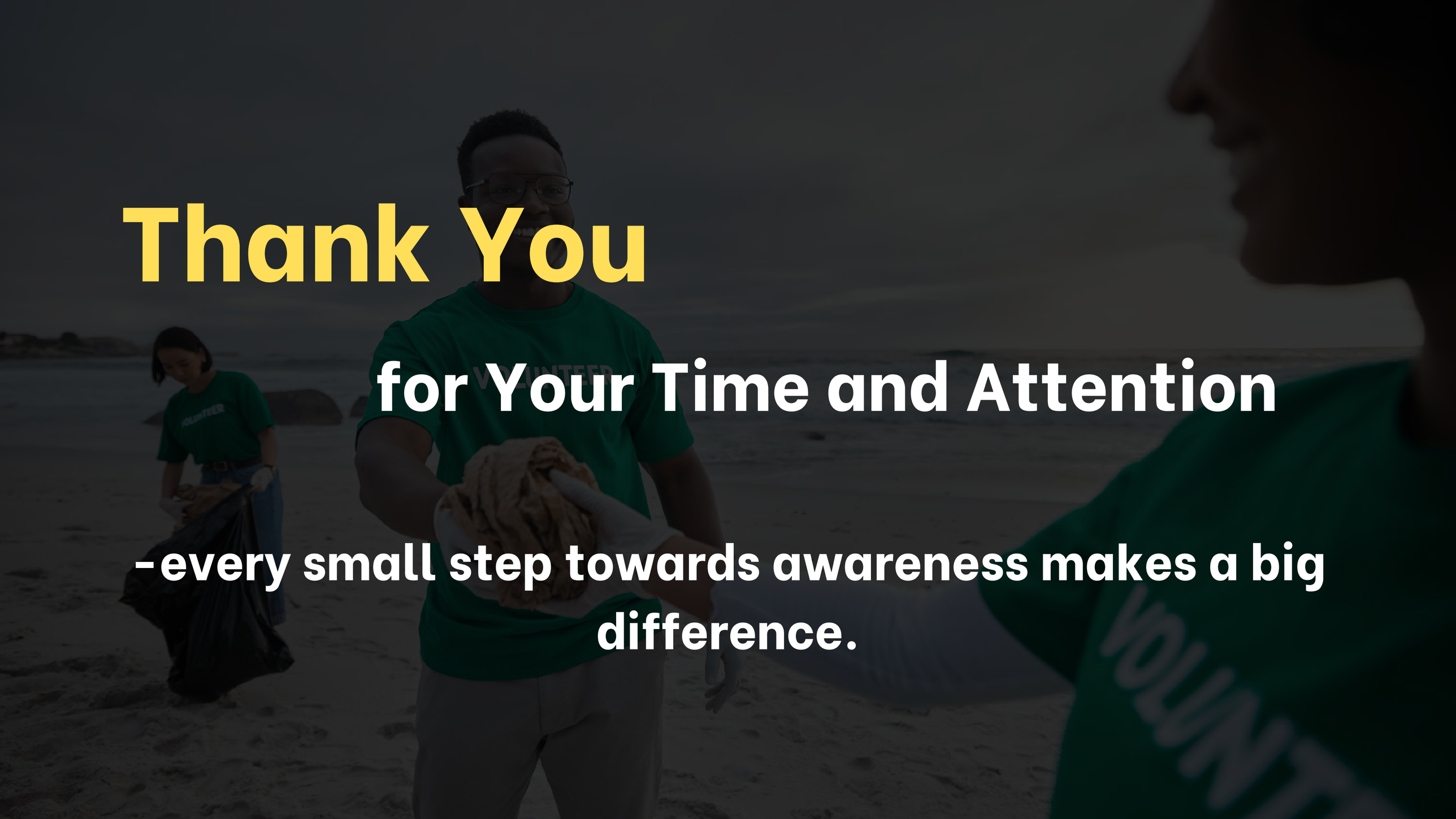
Presentation Making and Statics Analysing

- Overall Structure Designing.
- Helped in Bandfort’s Law implimentation.
- Presentation making and analysing statistics
- |

By- Mehebub Alom

CONCLUSION

We performed extensive exploratory analysis of air quality data across Indian cities. We cleaned the data, computed summary statistics, and visualized pollutant distributions. Using PCA and clustering, we identified major patterns in pollution profiles. An Isolation Forest helped flag outlier cities with unusual pollutant levels. Finally, we engineered features such as pollutant ratios and a composite pollution level to aid interpretation.



Thank You

for Your Time and Attention

-every small step towards awareness makes a big difference.