Internship Final Report - Data Science & Visualization

Student Details

Student Name: Harsh Baviskar

University: Savitribai Phule Pune University

Major: Information Technology

Internship Duration: July 1st, 2025- July 31st,

2025

Company:Shadowfox

Domain: Data Science

Mentor:Mr. Hariharan

Coordinator: Mr. Aakash

Objectives

The primary goals of my internship were:

1. To acquire hands-on experience with real-world air quality data analysis.
2. To enhance proficiency in Python-based data science tools and visualization libraries.
3. To understand key pollutant patterns in urban air and develop meaningful insights.
4. To apply both static (Matplotlib) and interactive (Plotly) visual techniques for effective data storytelling.

Tasks and Responsibilities

Air Quality Data Processing & Analysis

- Imported and cleaned hourly air pollution datasets for Delhi using Pandas and NumPy.
- Resampled and segmented data to compute daily averages, time-wise means, and monthly summaries.
- Conducted descriptive statistics for pollutants like PM2.5, PM10, NO, SO, CO, O, and NH.

Exploratory Data Analysis (EDA)

- Created statistical tables for pollutant metrics, correlation analysis, and threshold exceedances.
- Identified high-pollution periods and plotted AQI trends over time.

Data Visualization Using Python

- Used Matplotlib for static plots: line, bar, histogram, box, scatter, and specialized plots.
- Used Plotly for interactive plots with zoom, hover, tooltips, and dynamic views.

Internship Final Report - Data Science & Visualization

Learning Outcomes

- Technical Mastery: Expertise in time-series analysis, grouped analytics, and Python scripting.
- Visualization Proficiency: Appropriate selection and implementation of visualization techniques.
- Problem-Solving Skills: Efficient data aggregation, transformation, and analysis.
- Reporting Acumen: Integrated statistical narratives into well-structured visual presentations.

Challenges and Solutions

- Handling Irregular Time Formats: Fixed using datetime parsing and reindexing.
- Memory Usage on Large Datasets: Optimized performance using column filtering and efficient aggregation.
- Pollutant Overlap in Plots: Applied color coding, legends, and plot segmentation.

Conclusion

This internship bridged academic knowledge with industry-relevant practices in environmental data science and Python visualization. I developed skills in EDA, statistical analysis, and visual storytelling, preparing me for real-world analytics roles.

Acknowledgments

I express my sincere gratitude to ShadowFox, especially my mentor, Mr. Hariharan, and coordinator, Mr. Aakash, for their guidance and support throughout my internship. I also thank ShadowFox for providing this internship opportunity, which has been instrumental in my personal and professional growth. This report reflects the integration of academic knowledge with practical skills gained during the internship, highlighting my journey of learning, growth, and development in the field of data science