Internship Final Report

Student Name: Tanmay Guha
University: Brainware University
Major: Computer Science and Engineering
Internship Duration: 1st July, 2025 - 31st July, 2025
Company: ShadowFox
Domain: Data Science
Mentor: Mr. Hariharan
Coordinator: Mr. Aakash

Objectives

During my internship at ShadowFox, my primary objectives were:

1. To gain hands-on experience in data cleaning, analysis, and visualization using Python libraries like Pandas, Matplotlib, and Plotly.

2. To understand the practical application of statistical modeling and exploratory data analysis (EDA) on real-world datasets.

3. To develop interactive dashboards and reports to communicate data-driven insights effectively.

Tasks and Responsibilities

Throughout the internship, I worked on the following tasks:

1. **Data Cleaning and Preprocessing:**
    - Cleaned and preprocessed datasets, handling missing values, outliers, and standardizing formats. For example, in the Delhi AQI dataset, I converted date columns and created new features like seasons and AQI categories.

2. **Exploratory Data Analysis (EDA):**

- Conducted EDA to uncover trends and patterns. For instance, I analyzed seasonal variations in PM2.5 levels and visualized correlations between pollutants using heatmaps.

3. **Data Visualization:**
    - Created static and interactive visualizations using Matplotlib, Seaborn, and Plotly. Examples include time-series plots of sales trends and box plots for seasonal PM2.5 distributions.
4. **Statistical Modeling:**
    - Explored relationships between variables, such as the impact of manufacturing and freight costs on profitability in the synthetic sales dataset.
5. **Report Generation:**
    - Compiled findings into comprehensive reports with visualizations and actionable insights.

Learning Outcomes

1. **Technical Proficiency:**
    - Enhanced my skills in Python libraries like Pandas for data manipulation and Matplotlib/Plotly for visualization.
    - Learned to create interactive dashboards and dynamic visualizations for better data storytelling.
2. **Understanding of Data Science Lifecycle:**
    - Gained a holistic understanding of the data science workflow, from data collection to model interpretation.
3. **Analytical Skills:**
    - Improved my ability to derive meaningful insights from complex datasets and present them clearly.
4. **Professional Development:**
    - Developed teamwork and communication skills by collaborating with mentors and peers.

Challenges and Solutions

1. **Handling Large Datasets:**
    - **Challenge:** Processing large datasets (e.g., Delhi AQI) was computationally intensive.
    - **Solution:** Optimized workflows by using efficient algorithms and chunking data where possible.
2. **Model Accuracy:**
    - **Challenge:** Ensuring statistical models (e.g., correlation analysis) were accurate and interpretable.
    - **Solution:** Used cross-validation and iterative refinement based on performance metrics.
3. **Tool Selection:**
    - **Challenge:** Deciding between Matplotlib (static) and Plotly (interactive) for specific use cases.
    - **Solution:** Referred to documentation and compared trade-offs (e.g., Matplotlib for publication-quality plots, Plotly for web-based interactivity).

Conclusion

My internship at ShadowFox was a transformative experience that bridged theoretical knowledge with practical application. The exposure to real-world datasets, visualization tools, and analytical techniques has solidified my passion for data science. I am now better equipped to tackle challenges in this field and look forward to applying these skills in future projects.

Acknowledgments

I extend my gratitude to my mentor, Mr. Hariharan, and coordinator, Mr. Aakash, for their guidance and support. I also thank ShadowFox and Brainware University for providing this invaluable opportunity to grow as a data science professional.

This report reflects my original work and learnings during the internship. All tasks and insights are based on my hands-on experience with the provided datasets and tools.