Internship Final Report

Student Name: Srusti M
University: BNM Institute of Technology
Major: Computer Science
Internship Duration: August 1st, 2025 - August 26th, 2025
Company: ShadowFox
Domain: Data Science
Mentor: Mr. Hariharan
Coordinator: Mr. Aakash

Objectives

My primary objectives for this internship are to:

1. Strengthen my knowledge of Python-based data visualization libraries and their practical applications.
2. Gain hands-on experience in analyzing datasets using beginner, intermediate, and advanced-level projects.
3. Develop skills in documenting technical processes, conducting exploratory analysis, and applying statistical/machine learning methods.
4. Enhance my ability to communicate technical findings clearly through visualizations, reports, and structured documentation.

Tasks and Responsibilities

During my internship, I was involved in the following key tasks:

Beginner Level

- **Visualization Library Documentation**:
  I prepared a detailed documentation guide covering two Python visualization libraries. The guide focused on the types of graphs supported (line, bar, scatter, histogram, etc.), along with use cases and code snippets.
- **Library Overview**: Each library's background, features, and typical use cases were highlighted.
- **Graph Types & Examples**: Multiple visualization types were documented with code demonstrations.
- **Comparison**: A comparative section was included to show strengths and weaknesses of each library in terms of ease of use, customization, interactivity, and handling large datasets.

Intermediate Level

- **Air Quality Index (AQI) Analysis (Delhi)**:
  I conducted a comprehensive study on Delhi's air quality by analyzing AQI data.
  - Examined pollutants such as PM2.5, PM10, and NO_2.
  - Identified seasonal variations in AQI values.
  - Explored the influence of environmental and geographical factors on pollution trends.
  - Created visualizations (line plots, bar charts, and heatmaps) to present trends clearly.

    - Applied statistical methods to uncover relationships between pollutants and health risks.

Advanced Level

- **Cricket Fielding Analysis**:
  As part of advanced learning, I worked on a cricket fielding performance analysis project.
    - Collected and organized data on player fielding efforts from a T20 match.
    - Recorded details such as clean picks, catches, drop catches, run outs, direct hits, and runs saved.
    - Computed performance scores using a weighted formula to quantify fielding impact.
    - Developed a performance matrix to compare players and highlight their contributions.
    - Identified areas of improvement in defensive play and recommended strategic insights.

Learning Outcomes

- Technical Skills: Acquired strong knowledge of Python libraries like Matplotlib, Seaborn, Pandas, and Plotly, as well as advanced visualization and reporting techniques.
- Analytical Thinking: Gained practical experience in handling real-world datasets, performing statistical analysis, and interpreting patterns effectively.
- Domain Knowledge: Understood environmental challenges linked to air pollution and the dynamics of cricket fielding in performance analytics.
- Documentation & Communication: Improved ability to write structured documentation and present findings using clear visuals and formal reports.
- Problem Solving: Learned to approach tasks at varying levels of complexity, from beginner documentation to advanced sports analytics.

Challenges and Solutions

- Data Complexity: Handling incomplete or noisy datasets in AQI and cricket analysis was challenging. I resolved this by applying preprocessing techniques such as handling missing values and standardizing formats.

- Visualization Customization: Initially, creating visually appealing plots was time-consuming. This was solved by leveraging Seaborn's themes and customizing Matplotlib outputs.

- Model Accuracy: In AQI prediction, ensuring statistical accuracy required iterative validation. I applied cross-validation methods and refined models based on evaluation metrics.

- Time Management: Balancing tasks across three levels (beginner, intermediate, advanced) required proper planning and prioritization.

Conclusion

This internship experience provided me with exposure to different dimensions of data analysis — from documentation and visualization to statistical and sports-based analytics. By working on projects at multiple difficulty levels, I not only improved my technical knowledge but also learned how to connect theoretical concepts with real-world applications. This experience has motivated me to continue building expertise in data science and has prepared me to tackle more advanced projects in the future.

Acknowledgments

I would like to sincerely thank **ShadowFox**, my mentor **Mr. Hariharan**, and coordinator **Mr. Aakash** for their constant guidance, encouragement, and constructive feedback throughout this journey. Their support has been invaluable in helping me improve my technical as well as professional skills. I am also deeply grateful to My Teachers for providing me with the platform and resources to pursue this internship.

This experience has not only enhanced my knowledge and confidence but also contributed immensely to my overall personal and career growth. I truly cherish the learnings and exposure I have gained through this opportunity.

This internship has broadened my perspective on real-world applications of my academic knowledge. It has also motivated me to continuously upskill and stay curious in the field. I look forward to applying these learnings in future projects and professional endeavors.