**Internship Final Report**

**Student Name:** Abirami V

**University:** SASTRA University

**Major:** Data Science

**Internship Duration:** January 2nd, 2025 - January 31st, 2025

**Company:** ShadowFox

**Domain:** Data Science

**Mentor:** Mr. Hariharan

**Coordinator:** Mr. Aakash

**Objectives**

**1. Document Python Visualization Libraries:** Develop an in-depth guide for two of the most widely used Python visualization libraries, **Matplotlib** and **Seaborn** with emphasis on the types of graphs they can produce. This documentation was intended to include practical examples with code snippets so that it could be used by users from beginners to intermediate levels.

**2. Visualize and Forecast AQI Categories :** Interpret air quality data by visualizing pollutant concentrations (PM2.5, PM10, O3, NO2, SO2, CO, and NH3) over a period and forecast Air Quality Index (AQI) categories as a function of these pollutant concentrations. This activity included employing data visualization methods to derive insights regarding air quality trends and classifying AQI on the basis of predetermined pollutant concentrations.

**Tasks and Responsibilities**

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

**1. Beginner Task - Visualization Library Documentation:**

**- Matplotlib and Seaborn Documentation:**

Investigated both Matplotlib and Seaborn recording their features, different plot types, and customization possibilities within the two libraries.

Made hands-on examples with code snippets illustrating how to create typical visualizations such as line plots, bar charts, scatter plots, histograms, heatmaps, and so on.

The documentation contained explicit explanations, step-by-step procedures, and best practices for employing these libraries efficiently.

**2. Intermediate Task - Visualization and AQI Prediction :**

**- Data Preprocessing:**

Operated on a dataset with historical air pollutant concentrations and preprocessed the data to get it into proper form for analysis.

**- AQI Category Prediction:**

Created a function to forecast the AQI category from pollutant concentrations, specifically PM2.5, and used it to classify the air quality.

The AQI levels were: "Good", "Moderate", "Unhealthy for Sensitive Groups", "Unhealthy", "Very Unhealthy", and "Hazardous".

**- Visualization :**

Utilized Matplotlib and Seaborn to produce line plots following the trend of pollutant concentrations (PM2.5, PM10, O3, NO2, SO2, CO, and NH3) over time.

These plots facilitated the visualization of air quality trends and the relationship between various pollutants.

**- Reporting :**

Reported my results, noting any trends seen in the data and explaining the relationship between pollutant concentrations and AQI ranges.

**Learning Outcomes**

**1. Visualization Library Knowledge Depth:**

Enhanced my knowledge of Matplotlib and Seaborn, understanding how to use them efficiently for a broad range of data visualizations, ranging from simple charts to complex statistical plots.

**2. Hands-on Experience in Data Analysis:**

Acquired practical experience in data preprocessing, dealing with time series data, missing values, and data consistency.

**3. AQI Prediction Skills :**

Acquired the ability to classify air quality with respect to pollutant concentrations and forecast AQI categories, which is essential for environmental data reporting and analysis.

**4. Data Visualization for Environmental Data :**

Developed my skills in visualizing environmental data, creating line plots and other visualizations that enable users to easily comprehend air quality trends.

**5. Technical Documentation :**

- Clear and informative technical writing for Matplotlib and Seaborn enhanced my communication skills and my technical skill to be able to present such complex technical ideas in simple terms.

**Challenges and Solutions**

**1. Complex Data Management :**

The air quality dataset included multiple pollutants, making it difficult to visualize all at once. In order to get around this problem, I made use of multiple color schemes, markers, and plot types in order to not clutter the graphs and to get a clear representation.

**2. Plot Customization for Matplotlib and Seaborn :**

Customizing the visualizations using **Matplotlib** involved understanding all the parameters and options for every plot type. I overcame this by checking the official documentation and experimenting with different options to achieve the desired output.

**Conclusion**

My internship at ShadowFox provided valuable hands-on experience in the field of data science.

It allowed me to acquire hands-on experience in technical documentation and data analysis. These tasks improved my skills in data preprocessing, visualization, and reporting, which are the essential skills for any data scientist.

**Acknowledgments**

I express my sincere thanks to my mentor, Mr. Hariharan, and coordinator, Mr. Aakash, for their constant guidance and support throughout the internship. Also, I appreciate the team for providing resources and valuable insights that have helped me improve both technical and communication skill. Lastly, I am grateful to the organization for offering me this wonderful opportunity to learn and grow.