

# Report on AI Lab Manuals

By Khuram Iqbal (2022-CS-48)

October 2024

# 1 Introduction

This report summarizes the key learnings and observations from two manuals provided as part of my lab work. The manuals cover essential concepts in data analysis using Python, with a focus on libraries like *NumPy*, *Pandas*, and *Matplotlib*. Below is a detailed reflection on what I learned from each manual and the insights gained.

## 2 Manual 1: AI Lab with Numpy, Pandas, and Matplotlib

### 2.1 Outcome

After going through this manual, I gained hands-on experience with data analysis using Python. The manual introduced various case studies to apply concepts in real-world situations. Some of the key tasks I worked on included:

- **Data Generation:** Using *NumPy* and *Pandas* to create fake datasets such as random weather data or employee salaries. This was useful for practicing on data that resembles real-world scenarios.
- **Data Processing:** I learned to convert data into arrays using *NumPy* and performed basic statistics like calculating the **mean**, **median**, and **standard deviation**.
- **Data Filtering and Grouping:** With *Pandas*, I filtered data based on conditions (like finding days with temperatures over 30°C) and grouped data to calculate summaries, such as average sales or total quantities sold.
- **Visualization with Matplotlib:** I created various plots, such as line graphs showing temperature changes and bar charts representing different sales categories.

### 2.2 Observations

The main observation from this manual is how powerful and efficient these tools are when used together. *NumPy* handles complex calculations, *Pandas* is excellent for organizing data, and *Matplotlib* makes it easy to visualize the results. The combination of these tools allows for a smooth workflow when analyzing data.

## 3 Manual 2: Lab 6 - Data Preprocessing and Visualization

### 3.1 Outcome

This second manual focused more on preparing data for analysis, emphasizing the importance of cleaning and processing data before diving into the actual analysis. Here's what I learned:

- **Handling Missing Data:** I learned how to identify missing values in datasets and fill them with appropriate values, like the median or mean.
- **Encoding and Scaling Data:** Categorical data (like *Gender*) needs to be converted into numeric values for machine learning. I also practiced scaling numeric columns, such as income and age, to keep them consistent.
- **Visualization for Insight:** I created scatter plots and box plots to understand the relationships between different features. For instance, plotting performance ratings against years of experience helped me see if experience impacts performance.
- **Feature Engineering:** I practiced creating new features, such as *Income per Age*, to enhance the data for analysis.

### 3.2 Observations

This manual showed me that data cleaning is a critical step. Without cleaning and preparing data properly, analysis can give misleading results. I also realized that understanding the relationships between features, like through correlation analysis, can help improve the quality of insights.

## 4 Conclusion

Overall, these two manuals taught me a great deal about the data analysis process—from data generation and preprocessing to visualization and feature engineering. The first manual gave me hands-on experience with core Python libraries, while the second manual highlighted the importance of preparing data correctly for analysis. These skills are essential for real-world data science tasks, and I now feel more confident in using these tools in my own projects.