**Lab 7: Data Cleaning – Handling Missing Values and Outliers**

**Prelab Questions**

1. Why is data cleaning important before analysis?
2. What are the different types of missing data?
3. What are some techniques for handling missing values in a dataset?
4. How do outliers affect statistical analysis and machine learning models?
5. What methods can be used to detect outliers in a dataset?

**In-Lab Details**

**Objective**:

* Identify and handle missing values in a dataset.
* Detect and treat outliers to improve data quality.

**Resources**:

* Python (Jupyter Notebook).
* Libraries: Pandas, NumPy, Seaborn, Matplotlib.
* Dataset: customer\_data.csv containing customer age, income, and purchase frequency.

**Expected Output**:

1. **Missing Values Summary**: Displaying columns with missing data.
2. **Boxplot Visualization**: Detecting income outliers.
3. **Cleaned Data Summary**: Dataset with missing values handled and outliers removed.

**Postlab Questions**

1. Why is it necessary to check for missing values before analysis?
2. Compare different strategies for handling missing data (mean, median, mode, or removal).
3. What is the interquartile range (IQR), and how is it used for outlier detection?
4. How can extreme outliers affect predictive models like regression?
5. What alternative techniques can be used to handle outliers instead of removal?