1-Hour Lab Assignment: Customer Purchase Analysis

Objective: Practice Data Wrangling, Visualization, Logistic Regression, and Decision Trees using a real-world styled dataset.

Dataset: new_customer_purchase_dataset.csv
Target Variable: Purchase (YES / NO)

PART A: Data Wrangling (15 Minutes)

1. Load the dataset using pandas and display first five rows.
2. Convert Gender into numeric values.
3. Encode Purchase variable as binary (YES=1, NO=0).
4. Check for missing values and basic statistics.

PART B: Data Visualization (15 Minutes)

1. Plot histogram of Monthly Income.
2. Plot bar chart of Purchase distribution.
3. Create scatter plot between Brand Awareness and Quality Rating colored by Purchase.

PART C: Logistic Regression (15 Minutes)

1. Split the dataset into training and testing sets.
2. Train Logistic Regression model.
3. Evaluate using accuracy, confusion matrix, and classification report.
4. Interpret the model results.

PART D: Decision Tree (15 Minutes)

1. Train Decision Tree classifier using Gini Index.
2. Visualize the decision tree.
3. Identify root node, internal node, and leaf node.
4. Write two IF–THEN rules from the tree.

Reflection Questions

1. Why is Decision Tree more interpretable than Logistic Regression?
2. Which model would you recommend to a business manager and why?
3. What happens if decision tree depth increases too much?