

CA1: Data Mining (PEC-AIML601B)

Exploratory Data Analysis of Mall Customers Using Descriptive Statistics

1. Deliverable

Presentation file (PDF)

2. Assignment Description

Retail organizations rely heavily on data-driven insights to understand customer behavior and design effective marketing strategies. Before applying advanced data mining or machine learning techniques, it is essential to explore and summarize the data using **descriptive statistics**.

In this assignment, you will analyze the **Mall Customers Dataset** and apply appropriate descriptive statistical techniques to summarize customer demographics and spending behavior. You will then present your findings in the form of a structured presentation.

3. Objectives

After completing this assignment, students will be able to:

- Understand and describe the structure of a real-world dataset
 - Apply descriptive statistical measures to numerical data
 - Perform frequency analysis on categorical variables
 - Visualize data distributions effectively
 - Interpret statistical results in a business context
-

4. Dataset

Mall Customers Dataset (Customer Segmentation Dataset)

The dataset contains the following attributes:

- CustomerID
 - Gender
 - Age
 - Annual Income
 - Spending Score
-

5. Tasks to Perform

i) Data Understanding

- Briefly describe each attribute
- Identify numerical and categorical variables
- Report total number of records

ii) Measures of Central Tendency

For numerical attributes (Age, Annual Income, Spending Score):

- Mean
- Median
- Mode

Interpret what these values indicate.

iii) Measures of Dispersion

Compute and interpret:

- Range
- Variance
- Standard Deviation
- Interquartile Range (IQR)

iv) Distribution Analysis

- Comment on distribution shape (symmetric / skewed)
- Identify potential outliers

v) Categorical Data Analysis

- Frequency tables for Gender
- Percentage distribution

vi) Visualization

Include at least:

- Histogram (one numerical variable)
- Box plot (one numerical variable)
- Bar chart (Gender distribution)

- Scatter plot (Annual Income vs Spending Score)

vii) Insights & Interpretation

- Report at least **five meaningful insights**
 - Link statistics to possible business implications
-

6. Presentation Guidelines

- Duration: 5 minutes
 - Slides: less than 10
 - Use graphs and tables with brief explanations
 - Each figure must be interpreted
-