Customer Segmentation Using K-Means Clustering

1. Introduction

This report presents the results of a customer segmentation analysis using **K-Means Clustering** on a dataset containing **Age, Gender, Quantity, Price, Payment Method, and Shopping Mall**. The goal is to identify customer groups based on purchasing behavior.

2. Data Preprocessing

2.1 Dataset Overview

The dataset consists of:

- **Customer Id** (Identification number of customer)
- **Age** (Customer age)
- **Quantity** (Number of items purchased)
- **Price** (Total price of purchases)
- **Category** (Category of items being purchased)
- **Payment Method** (Categorical: Encoded for analysis)
- **Shopping Mall** (Categorical: Encoded for analysis)

2.2 Data Cleaning & Preparation

- **Handled missing values** by checking and dropping rows if necessary.
- **Encoded categorical variables** (Gender, Payment Method, Shopping Mall) into numerical values.
- **Standardized numerical features** to ensure equal weighting in clustering.

3. Finding the Optimal Number of Clusters

We used the **Elbow Method**, which plotted the **Within-Cluster Sum of Squares (WCSS)** against the number of clusters. The optimal cluster count was found to be **4**, which balances cluster cohesion and separation.

4. K-Means Clustering Results

After running K-Means with **K=4**, the customers were segmented into four distinct clusters:

4.1 Cluster Insights

Cluster	Characteristics
Cluster 0	Young high-spenders, frequent shoppers
Cluster 1	Older customers, moderate spenders
Cluster 2	Budget-conscious shoppers, buy in bulk
Cluster 3	Shopping mall loyal customers, mid-range spenders

5. Data Visualization

Two key visualizations were generated:

- **Age vs. Price:** Shows spending behavior by age group.
- **Quantity vs. Price:** Identifies high-volume vs. low-volume buyers.

6. Conclusion

This segmentation provides actionable insights for marketing strategies and targeted promotions.