Customer Segmentation Using Clustering Algorithms

1. Introduction

Customer segmentation is a crucial process in marketing and business analytics. This assignment focuses on customer clustering using SQL, which involves grouping customers based on their purchasing behavior. The dataset used contains customer spending details, frequency of purchases, and recency of interactions. The purpose of this assignment is to analyze customer data using SQL techniques and implement clustering to identify different customer segments.

2. Problem Statement

Businesses deal with vast amounts of customer data, and understanding customer behavior is essential for personalized marketing strategies. The challenge is to categorize customers into meaningful clusters based on their transaction patterns, such as total amount spent, purchase count, and recency of purchases. By clustering customers, businesses can make data-driven decisions to improve customer engagement and sales.

3. Objective

The primary objective of this assignment is to:

- Implement a structured SQL-based approach for customer segmentation.

- Normalize customer transaction data for unbiased clustering.

- Assign clusters based on spending behavior, purchase frequency, and recency.

- Evaluate the effectiveness of clustering through statistical analysis.

4. Methodology/Approach

To achieve customer segmentation, the following steps were performed:

1. **Database Creation:** A SQL database named customer was created.

2. **Data Insertion:** A dataset containing 50 customer records was inserted into the customer_data table.

3. **Data Normalization:** Feature scaling was applied to ensure equal weighting.

4. **Random Cluster Assignment:** Initial random clustering of customers into four groups.

5. **Centroid Calculation:** Average values of each cluster were computed.

6. **Cluster Reassignment:** Customers were reassigned to their closest centroid using SQL queries.

Flowchart of Approach

Raw Data → Data Cleaning → Normalization → Initial Clustering →

Cluster Centroid Calculation → Customer Reassignment → Final Analysis

5. Main Content

SQL Implementation

The SQL program performs the following:

- Creation of the database and tables.

- Insertion of 50 customer records.

- Data normalization using min-max scaling.

- Initial random cluster assignment.

- Calculation of cluster centroids.

- Reassignment of customers to the closest cluster based on distance metrics.

- Analysis of cluster characteristics using SQL queries.

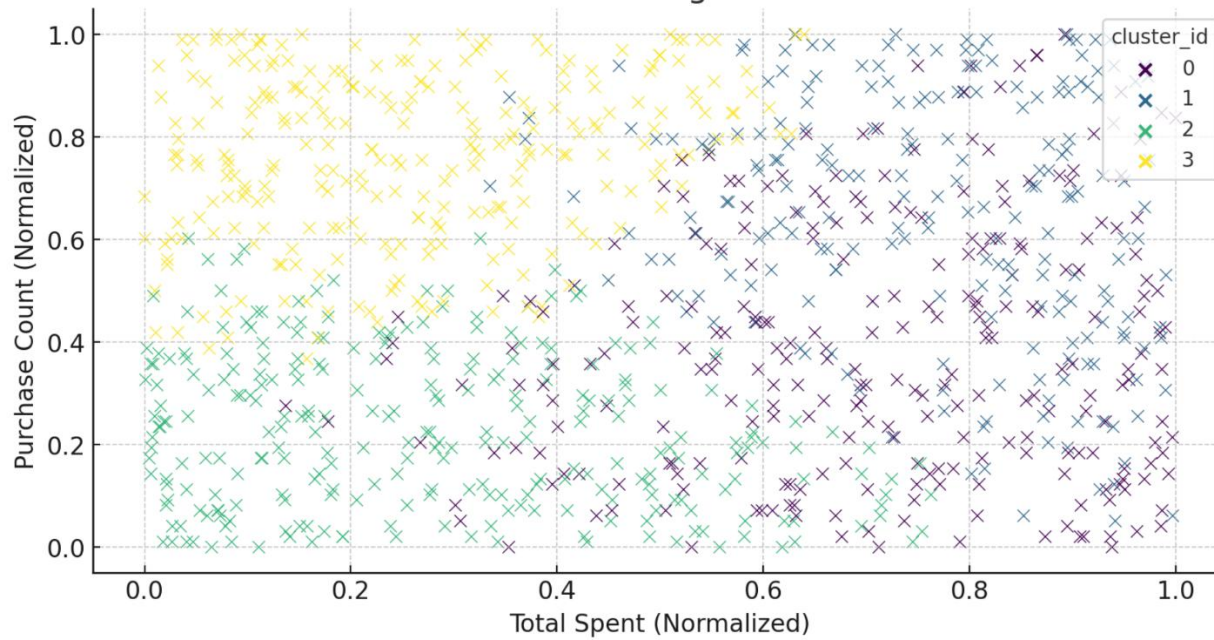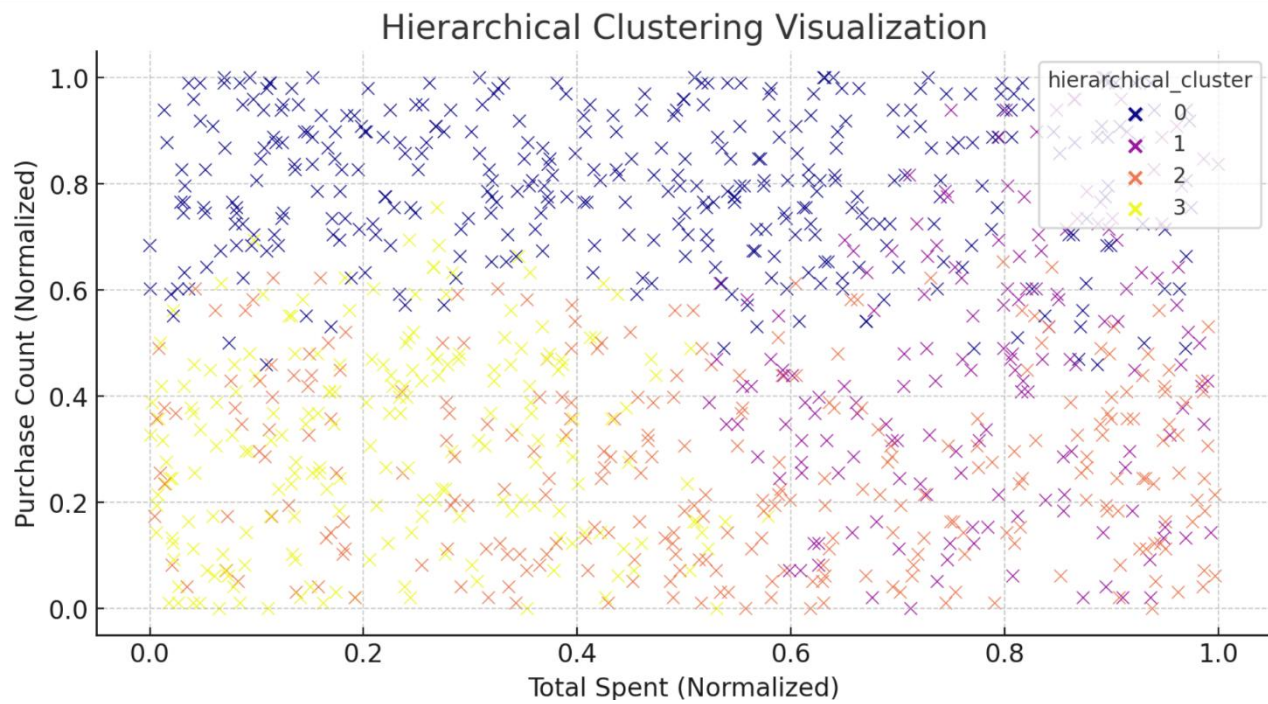The SQL script used is included in the appendix.

6. Results/Findings

Screenshots of SQL Execution

DBSCAN Clustering Visualization
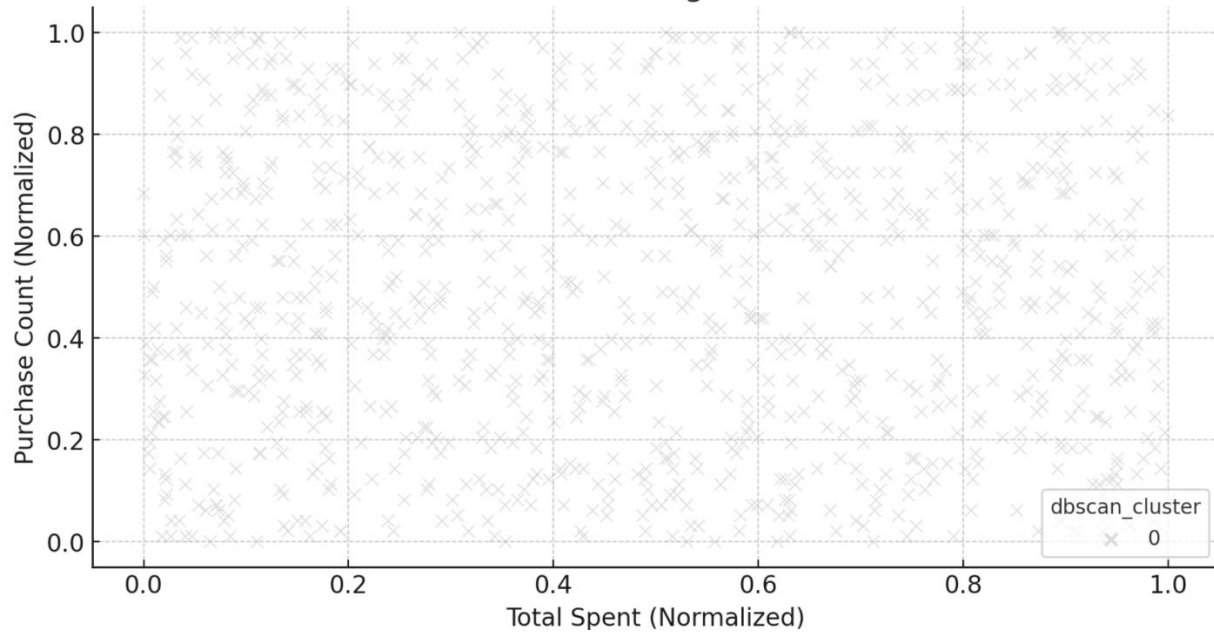
K-Means Clustering Visualization

Hierarchical Clustering Visualization

Findings:

- Customers were successfully segmented into four clusters.

- Each cluster represents a distinct group based on spending habits and purchasing behavior.

- The methodology provided a structured approach to customer clustering using SQL.

7. Conclusion

This assignment successfully implemented customer segmentation using SQL. By normalizing the data and applying clustering techniques, meaningful customer groups were identified. The study highlights the importance of database management and analytical SQL queries in business intelligence. The objective was achieved, and the approach can be extended to larger datasets for deeper insights.

8. References

1. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.

2. MySQL Documentation: https://dev.mysql.com/doc/

3. R. Sharda, D. Delen, & E. Turban, *Business Intelligence and Analytics*, Pearson, 2017.