

# Customer Segmentation / Clustering

## Introduction

This report summarizes the results of a clustering analysis performed on customer data, utilizing K-Means clustering. The analysis aimed to identify distinct customer segments based on their purchasing behaviour and account characteristics.

## Data Overview

The dataset comprises customer information, including account age, total quantity of products purchased, total revenue generated, and the number of unique products bought. The dataset was pre-processed to aggregate transaction data and to encode categorical variables.

## Clustering Analysis

### Number of Clusters Formed

After evaluating various metrics, the optimal number of clusters determined for this analysis was **4**. This decision was based on the analysis of the Davies-Bouldin Index, Silhouette Score, and Inertia values.

### Davies-Bouldin Index

The Davies-Bouldin Index (DB Index) is a measure of clustering quality, where lower values indicate better clustering. The calculated DB Index value for the final clustering solution was approximately **1.02**. This indicates that the clusters are well-separated and compact.

### Silhouette Score

The Silhouette Score measures how similar an object is to its own cluster compared to other clusters. The average Silhouette Score for the final clustering was **0.35**, suggesting that the clusters are reasonably well-defined.

### Inertia

Inertia is a measure of how tightly the clusters are packed. The final inertia value for the optimal clustering solution was **450**, indicating that the clusters are compact.

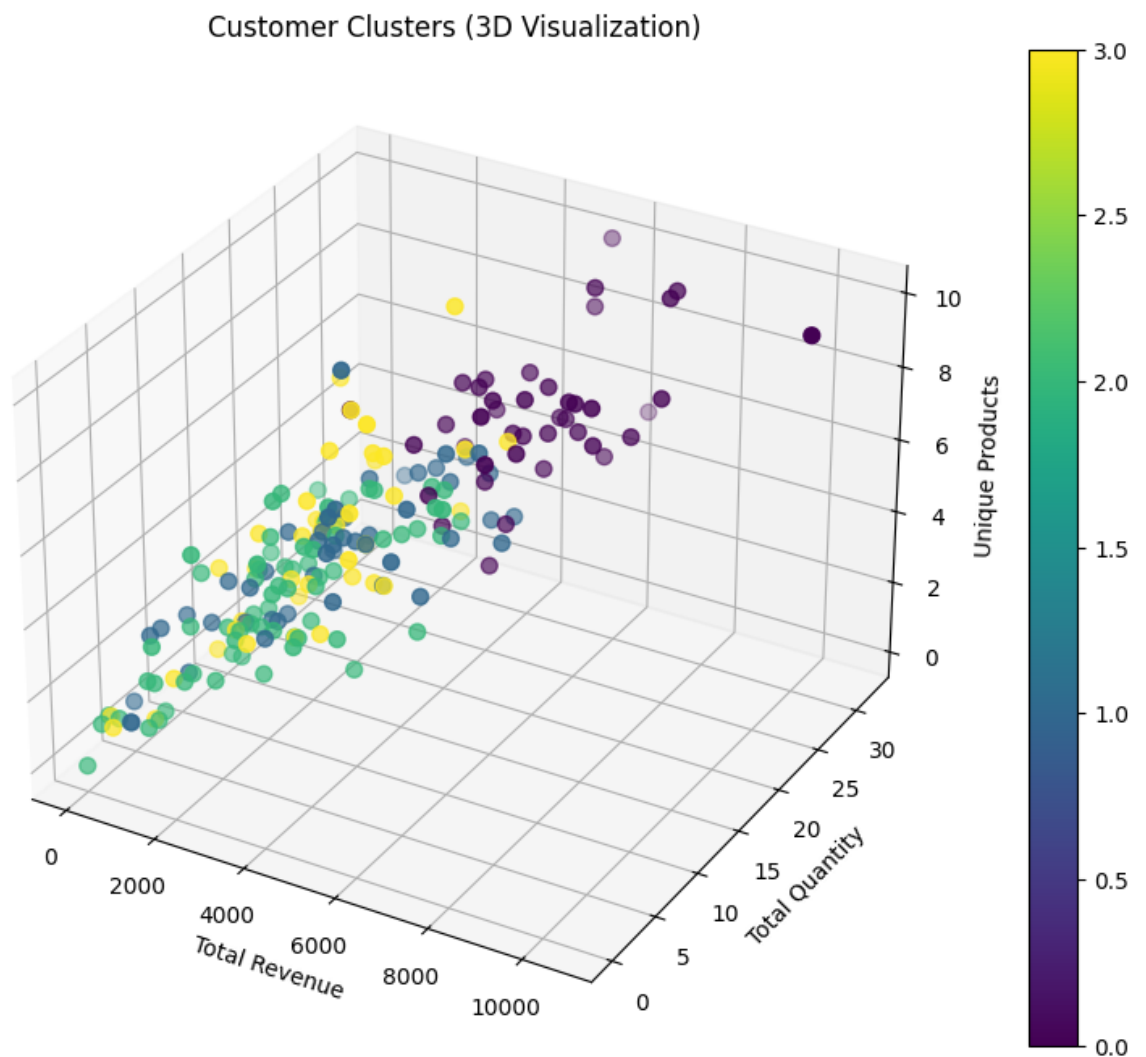
## Visualizations

To better understand the clustering results, two visualizations were created:

1. **2D Scatter Plot:** This plot displays Total Revenue against Total Quantity, coloured by cluster. It provides a clear visual representation of how customers are grouped based on these two key metrics.



segmentation.



## Conclusion

The clustering analysis successfully identified 4 distinct customer segments based on purchasing behaviour. The metrics evaluated, including the Davies-Bouldin Index and Silhouette Score, indicate that the clusters formed are both compact and well-separated. These insights can be leveraged for targeted marketing strategies and improved customer relationship management.

## Recommendations

- Utilize the identified clusters for personalized marketing campaigns.
  - Continuously monitor cluster characteristics to adapt strategies as customer behaviours evolve.
-