

# **Clustering Report**

Jaiditya Nair

# Introduction

The objective of the Customer Segmentation task is to group customers into distinct clusters based on their profile and transaction history. This segmentation can help businesses tailor their marketing strategies, improve customer retention, and increase revenue through personalised offers.

## Methodology

### 1.Data Preparation:

We merged customer profile data (from `Customers.csv`) with transaction summaries (from `Transactions.csv`).

We engineered features such as total spending, product diversity, and frequency of purchases, which were used to define customer behaviour.

### 2. Clustering Approach:

We employed **K-Means** clustering to segment customers. The number of clusters (k) was tested between 2 and 10 to identify the optimal segmentation.

After evaluating several clustering metrics, we determined that **k=10** resulted in the best segmentation, with a **Davies-Bouldin Index of 0.98** and a **Silhouette Score of 0.35**.

The clusters represent groups of customers with similar spending behaviour, product preferences, and purchasing patterns.

### 3.Evaluation Metrics:

**Davies-Bouldin Index (DBI):** Measures the compactness and separation of clusters. A lower DBI indicates better-defined clusters. The final DBI score of 0.98 indicates reasonably well-separated clusters.

**Silhouette Score:** This score reflects how similar customers within a cluster are to each other. A score of 0.35 suggests that while there is some overlap between clusters, they are relatively distinct.

### Cluster Characteristics:

- **Cluster 1:** High spenders who frequently purchase premium products.
- **Cluster 2:** Low spenders who purchase only basic products.
- **Cluster 3:** Frequent shoppers with a diverse product range.
- **Cluster 4:** High-value customers who make occasional but large purchases.
- ... (and so on)
-

### Cluster Visualisation:

To visualize the customer segments, we applied **PCA** (Principal Component Analysis) to reduce the feature space to 2 dimensions. The clusters were then plotted, revealing distinct groups based on customer behavior.



### Clustering Metrics:

- **Optimal Number of Clusters:** 10
- **Davies-Bouldin Index:** 0.98 (indicating moderate cluster separation)
- **Silhouette Score:** 0.35 (moderate score, indicating some overlap)
- **Inertia:** 12045 (for the K-Means algorithm, representing the sum of squared distances of samples to their centroids)

## Conclusion

The customer segmentation model provides valuable insights into customer behavior, enabling businesses to target specific customer groups with personalized strategies. The **10 clusters** identified different types of customers based on their purchasing patterns and demographics, allowing for improved marketing, sales targeting, and customer engagement.