

## Task 3: Customer Segmentation / Clustering

### Customer Clustering Report

1. Overview Customer segmentation is a critical process in understanding customer behavior and optimizing marketing strategies. This report details the clustering process, key metrics, and insights from customer data analysis.

---

## 2. Methodology

### 2.1 Data Preparation

- Aggregated transaction data at the customer level using total spending (TotalValue), total items purchased (Quantity), and transaction count.
- Merged transactional features with customer profiles.
- One-hot encoded categorical features such as Region.
- Normalized numerical features (TotalValue, Quantity, TransactionCount) to ensure proper scaling for clustering.

### 2.2 Clustering Technique

- Applied the KMeans algorithm with n\_clusters = 4 (adjustable).
  - Evaluated clustering quality using:
    - Davies-Bouldin Index (DBI): Measures intra-cluster similarity and inter-cluster differences (lower is better).
    - Silhouette Score: Measures separation between clusters (higher is better).
  - Used PCA for dimensionality reduction to visualize the clusters in two dimensions.
- 

## 3. Results

### 3.1 Number of Clusters

- Formed 4 clusters based on customer behavior and transactional data.

### 3.2 Clustering Metrics

- Davies-Bouldin Index (DBI): **1.3253565994399066**
  - Indicates that the clusters are reasonably well-separated.
- Silhouette Score: **0.24055228830667674**
  - Suggests that the clusters are distinct and well-formed.

### 3.3 Cluster Visualization

- PCA visualization shows distinct clusters with minimal overlap.
- Each cluster represents a unique customer **group based on behavior and spending patterns**.

---

## 4. Insights and Recommendations

### 4.1 Insights

- **High Spenders (Cluster X):** Customers in this cluster exhibit high spending and frequent transactions. Focus marketing efforts on loyalty programs or premium services.
- **Moderate Spenders (Cluster Y):** These customers show moderate spending behavior. Offering targeted discounts can encourage higher engagement.
- **Budget Buyers (Cluster Z):** Customers in this cluster have low spending but high transaction counts, indicating price sensitivity. Promote affordable products and bundles.
- **Dormant Customers (Cluster W):** Customers with low transaction counts and spending may require re-engagement campaigns.

### 4.2 Recommendations

1. **Personalized Marketing:** Leverage cluster insights to design targeted campaigns for each customer group.
2. **Loyalty Programs:** Create tiered rewards to retain high-value customers.
3. **Promotional Offers:** Offer discounts or bundles tailored to moderate and budget-conscious customers.
4. **Re-engagement Strategies:** Identify dormant customers and initiate personalized outreach campaigns.

---

## 5. Deliverables

1. **Code:** A Python script implementing the clustering process.
2. **Visualizations:** PCA-based scatterplot showing customer clusters.
3. **Output:** A CSV file containing customer data with assigned cluster labels for further analysis.

---

## 6. Conclusion

This analysis provides a comprehensive view of customer behavior and segmentation. By leveraging these insights, the organization can optimize marketing strategies, improve customer retention, and increase revenue.