

Task 3: Customer Segmentation / Clustering

Perform customer segmentation using clustering techniques. Use both profile information (from Customers.csv) and transaction information (from Transactions.csv).

- You have the flexibility to choose any clustering algorithm and any number of clusters in between (2 and 10)
- Calculate clustering metrics, including the DB Index (Evaluation will be done on this).
- Visualise your clusters using relevant plots.

Deliverables:

- A report on your clustering results, including:
 - The number of clusters formed.
 - DB Index value.
 - Other relevant clustering metrics.
- A Jupyter Notebook/Python script containing your clustering code.

Evaluation Criteria:

- Clustering logic and metrics.
- Visual representation of clusters.

Data Preparation

Datasets Used:

1. **Customers.csv:** Contains customer profiles, including CustomerID, CustomerName, Region, and SignupDate.
2. **Transactions.csv:** Includes transaction details such as TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, and Price.
3. **Products.csv:** Provides product information such as ProductID, ProductName, Category, and Price.

Data Processing Steps:

1. Merged Transactions.csv with Products.csv to include product price (Price_y).
2. Aggregated transaction data by CustomerID to compute:
 - **Total_Spent:** Sum of TotalValue for each customer.
 - **Total_Quantity:** Total quantity of items purchased.

- **Average_Price:** Mean price of items purchased.
 - **Transaction_Count:** Number of transactions made.
3. Standardized the aggregated features using StandardScaler to normalize the data.

Clustering Implementation

Algorithm:

K-Means clustering was selected for this task due to its simplicity and effectiveness for segmenting numerical data. The number of clusters was determined using the Elbow Method and Davies-Bouldin Index.

Steps:

1. **Feature Selection:** Used aggregated features (Total_Spent, Total_Quantity, Average_Price, Transaction_Count) for clustering.
2. **Number of Clusters:**
 - Explored clusters ranging from 2 to 10.
 - Selected 4 clusters based on optimal Davies-Bouldin Index and visual inspection.
3. **Dimensionality Reduction:** Applied Principal Component Analysis (PCA) to reduce features to 2 dimensions for visualization.

Evaluation Metrics

Davies-Bouldin (DB) Index:

- **DB Index Value:** 0.78 (Lower values indicate better clustering).

Additional Metrics:

1. **Inertia:** 645.32 (Sum of squared distances within clusters).
2. **Silhouette Score:** 0.59 (Measures the separation between clusters; higher is better).

Cluster Insights

Cluster	Total_Spent	Total_Quantity	Average_Price	Transaction_Count	PCA1	PCA2
0	6136.10	22.03	282.06	8.16	2.54	-0.20
1	2504.78	7.15	346.55	2.98	-1.21	1.29
2	1670.60	8.31	196.47	3.35	-1.56	-0.89
3	3836.65	14.11	274.02	5.68	0.42	-0.04

Cluster Characteristics:

1. **Cluster 0:** High spenders with frequent transactions and moderate average prices.
2. **Cluster 1:** Low spenders purchasing high-priced items with minimal transactions.
3. **Cluster 2:** Budget-conscious customers with low spending and lower average prices.
4. **Cluster 3:** Moderate spenders with frequent transactions and mid-range prices.

Visualization

PCA-Based Cluster Plot:

A 2D scatter plot was created using PCA-reduced features (PCA1 and PCA2). Clusters are distinctly separated, demonstrating effective segmentation.

```
plt.figure(figsize=(10, 6))  
sns.scatterplot(  
    x='PCA1', y='PCA2', hue='Cluster', data=customer_features, palette='viridis', s=100  
)  
plt.title('Customer Segmentation Clusters')  
plt.xlabel('PCA1')  
plt.ylabel('PCA2')  
plt.legend(title='Cluster')  
plt.show()
```

Business Implications

1. **Targeting High Spenders:**
 - Customers in Cluster 0 are prime candidates for premium offers and loyalty programs.
2. **Promoting High-Value Products:**
 - Cluster 1 customers prefer high-priced items. Marketing exclusive, luxury products may resonate with them.
3. **Budget-Conscious Offers:**
 - Cluster 2 customers are price-sensitive. Discounts and bundle offers could drive sales.
4. **Upselling Opportunities:**

- Cluster 3 customers show balanced behavior and could respond well to targeted upselling campaigns.

Conclusion

The customer segmentation task identified four distinct clusters, each with unique purchasing behaviors. These insights enable targeted marketing strategies to improve customer satisfaction and maximize revenue

