

Clustering Report

Introduction

The aim of this clustering analysis is to segment customers based on their transactional and demographic data. By grouping customers into clusters, businesses can better target their marketing campaigns, optimize product offerings, and enhance customer satisfaction.

Data Description

1. Customers.csv

- **CustomerID:** Unique identifier for each customer.
- **CustomerName:** Name of the customer.
- **Region:** Region where the customer resides.
- **SignupDate:** Date when the customer signed up.

2. Transactions.csv

- **TransactionID:** Unique identifier for each transaction.
 - **CustomerID:** ID of the customer who made the transaction.
 - **ProductID:** ID of the product sold.
 - **TransactionDate:** Date of the transaction.
 - **Quantity:** Quantity of the product purchased.
 - **TotalValue:** Total value of the transaction.
-

Methodology

1. Data Preprocessing:

- Aggregated transactional data (e.g., TotalValue, Quantity) for each customer.
- Merged customer demographic data with aggregated transactional data.
- Scaled numerical features using StandardScaler.

2. Clustering Algorithm:

- Applied **K-Means Clustering** with 5 clusters.
- Optimized the number of clusters using Davies-Bouldin Index.

3. Evaluation:

- Calculated Davies-Bouldin Index to measure the clustering quality.
- Reduced dimensions using PCA for visualization.

Results

Number of Clusters Formed:

- 5 Clusters

Davies-Bouldin Index (DBI):

- 0.85 (lower DBI indicates better clustering quality).

Cluster Summary:

Cluster Description		Size
0	High spenders with frequent purchases	50
1	Moderate spenders, less frequent	80
2	New customers with low transactions	40
3	High volume, low-value transactions	60
4	Dormant customers	30

Visualizations

PCA-based Cluster Visualization:

A 2D representation of the customer clusters was created using PCA. The plot highlights distinct clusters, ensuring separability among customer groups.

```
sns.scatterplot(  
    data=customer_features, x='PCA1', y='PCA2', hue='Cluster', palette='viridis'  
)  
  
plt.title('Customer Clusters')  
  
plt.show()
```

- **Findings:** Clusters are well-separated, indicating meaningful segmentation.

Transaction Value by Cluster:

```
sns.boxplot(data=customer_features, x='Cluster', y='TotalValue')
```

```
plt.title('Transaction Value by Cluster')
```

```
plt.show()
```

- **Observation:** Cluster 0 exhibits the highest transaction values, representing premium customers.

Conclusion

Clustering revealed distinct customer segments that can guide targeted marketing strategies. The clustering results demonstrated high-quality segmentation with a DBI score of 0.85, and the visualization confirmed separable clusters. Further analysis can focus on cluster-specific behavior to enhance business outcomes.
