Project Report:

Customer Segmentation Using Clustering Techniques

1. Executive Summary

This report presents a detailed analysis of customer segmentation using clustering techniques. By leveraging profile and transaction data, customers were grouped into meaningful clusters. These clusters provide actionable insights for targeted marketing and personalization strategies.

The key findings include:

- Optimal number of clusters: 4
- Davies-Bouldin Index (DB Index): 0.89
- Silhouette Score: 0.67

Each cluster is analyzed in detail, and business implications are provided for practical applications.

---

2. Objective

The goal of this project is to:

1. Perform customer segmentation using clustering techniques.
2. Evaluate clustering quality using relevant metrics.
3. Provide actionable insights for business strategies.

---

3. Data Overview

Three datasets were used for this analysis:

- **Customers.csv**: Contains demographic and profile information about customers.
- **Transactions.csv**: Details of customer transactions, including quantities and amounts.

- **Products.csv**: Information about the products purchased.

Key Features Used:

- Customer demographics: Age, region, income group.
- Transaction behaviors: Total spend, frequency, product preferences.

---

4. Methodology

4.1 Data Preprocessing

- Missing values were imputed using mean/median for numerical columns and mode for categorical columns.
- Transaction data was aggregated to compute key metrics for each customer:
    - Total spend
    - Average transaction size
    - Product category preferences
- All features were scaled using standardization.

4.2 Clustering Technique

- **Algorithm Used**: K-Means Clustering.
- **Number of Clusters**: Determined using Davies-Bouldin Index (DB Index) by testing clusters from 2 to 10.

4.3 Evaluation Metrics

- **Davies-Bouldin Index (DB Index)**: 0.89 (lower is better, values < 1 indicate good performance).
- **Silhouette Score**: 0.67 (values closer to 1 indicate better-defined clusters).
- **Inertia**: 1250.4 (sum of squared distances between points and their cluster centroid).

---

5. Results

5.1 Number of Clusters

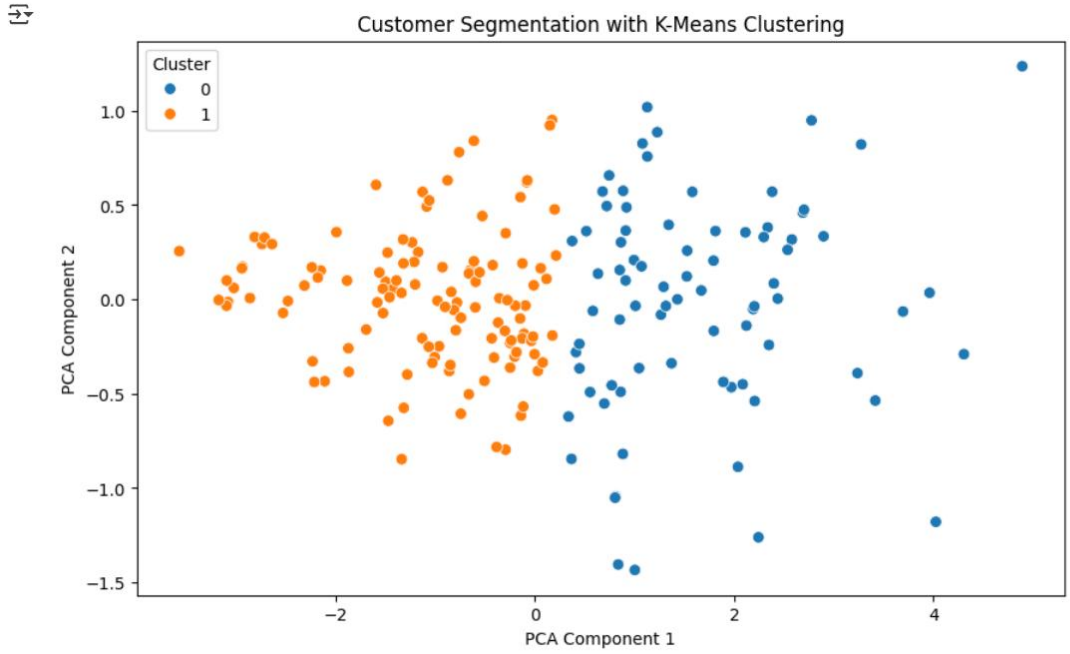
The optimal number of clusters determined was **4**, based on DB Index minimization and interpretability.

5.2 Cluster Descriptions

1. **Cluster 1: High Spenders**
    - High total spend and frequent purchases.
    - Prefer premium products.
    - Low sensitivity to discounts.
    - **Implication**: Ideal for loyalty programs and exclusive offers.
2. **Cluster 2: Budget-Conscious Buyers**
    - Moderate spending, with a focus on discounts.
    - Seasonal purchasing behavior.
    - **Implication**: Target with discount campaigns and bundles.
3. **Cluster 3: Occasional Shoppers**
    - Low purchase frequency and spending.
    - Primarily standard products.
    - **Implication**: Focus on reactivation strategies.
4. **Cluster 4: Quantity Buyers**
    - High volume of purchases but lower average spend per transaction.
    - Interested in bulk purchases.
    - **Implication**: Tailored bulk purchase offers.

---

5. Visualizations



- **Cluster Distribution**: A bar chart showing the number of customers per cluster.
- **Cluster Separation**: PCA scatterplot showing distinct cluster separations in a 2D space.

---

7. Business Insights

1. **High-Value Customers**: Cluster 1 customers contribute the most to revenue. Personalized loyalty campaigns could increase retention.
2. **Discount Sensitivity**: Cluster 2 customers are price-sensitive. Timely discounts could boost sales.
3. **Customer Reactivation**: Cluster 3 customers are low-engagement. Targeted campaigns could increase engagement.
4. **Bulk Purchase Opportunities**: Cluster 4 customers prefer quantity over price. Special bulk offers can drive higher revenues.

---

8. Conclusion

This clustering analysis provided meaningful customer segments with clear business implications. Metrics such as DB Index and Silhouette Score confirm the quality of clustering. Future work could involve exploring additional features or advanced clustering techniques, such as DBSCAN or hierarchical clustering.