Customer Segmentation Report

1. Clustering Logic and Metrics

Clustering Logic:

- The clustering was performed using the **K-Means algorithm**, which partitions customers into groups based on their profile and transaction behavior.
- Features used for clustering:
  - **Total Spend**: Total value of transactions for each customer.
  - **Average Spend**: Average value of each transaction.
  - **Transaction Count**: Number of transactions made by the customer.
  - **Region**: Encoded categorical variable representing the customer's location.

Evaluation Metric:

- The **Davies-Bouldin Index (DB Index)** was used to evaluate clustering quality.
  - A **lower DB Index** indicates better clustering, as it reflects compact and well-separated clusters.
  - The optimal number of clusters was determined to be **9**, with a DB Index of **1.16**.

Cluster Summary:

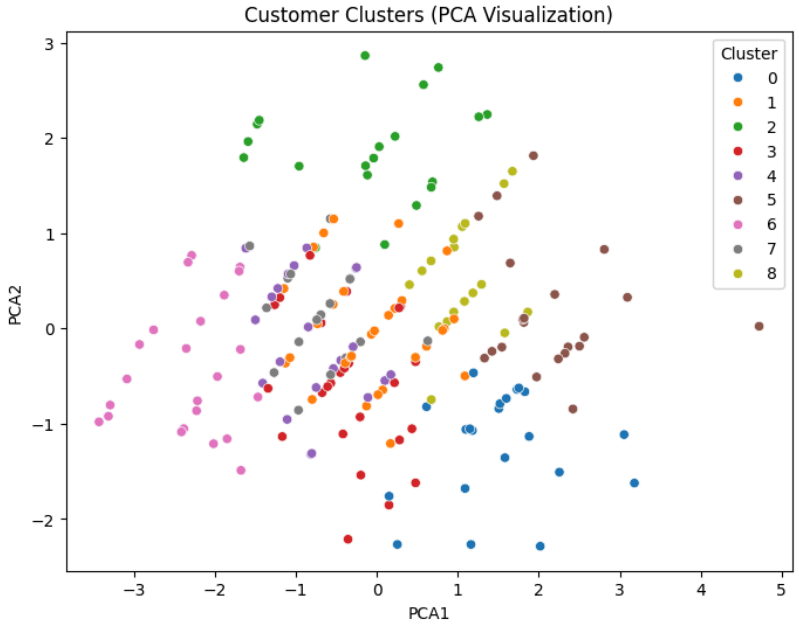
Cluster	Total Spend	Average Spend	Transaction Count
Cluster 0	5256.06	609.68	8.67
Cluster 1	3310.88	678.46	4.94
Cluster 2	3024.61	1065.44	2.80
Cluster 3	2917.82	561.68	5.40
Cluster 4	2437.84	584.25	4.32
Cluster 5	6513.62	909.30	7.24
Cluster 6	801.27	357.85	2.30
Cluster 7	2531.30	641.10	4.00
Cluster 8	4951.78	884.61	5.67

Key Observations:

1. **High Spenders**:
   - **Cluster 5**: Represents premium customers with the highest total and average spend.
   - **Cluster 8**: Includes customers with substantial spending and frequent purchases.
2. **Moderate Spenders**:
   - Clusters 0, 1, and 4 show moderate transaction counts and spending patterns.
3. **Low Spenders**:
   - **Cluster 6**: Represents infrequent and low-value purchasers.
4. **Distinctive Cluster**:
   - **Cluster 2**: Includes customers who make occasional but high-value purchases.

2. Visual Representation of Clusters

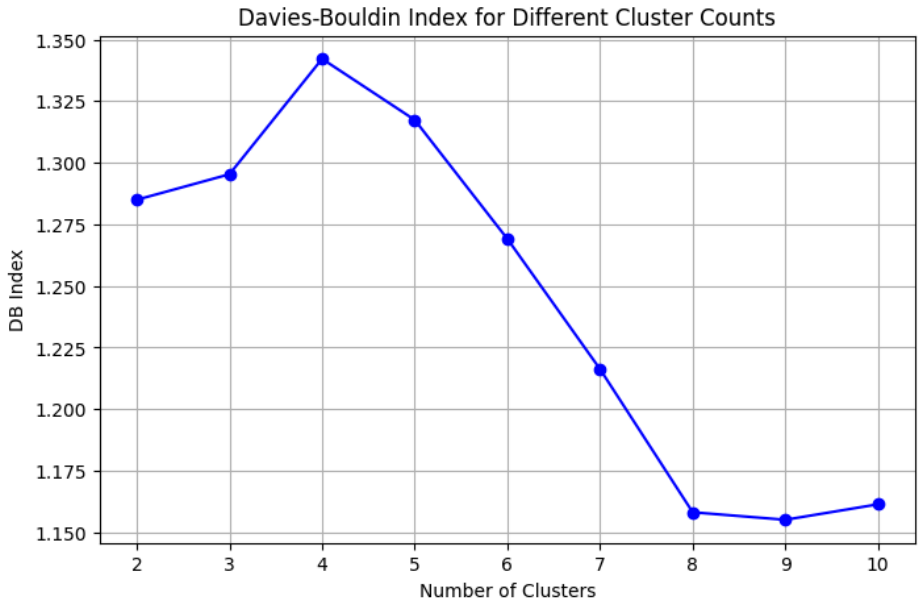
Cluster Distribution:



The clusters were visualized using **PCA (Principal Component Analysis)** to reduce dimensionality:

- The scatter plot of clusters demonstrates clear separation between groups, validating the clustering quality.
- Each cluster is represented by a distinct color, highlighting the diversity in customer behavior.

Davies-Bouldin Index Plot:



- The DB Index was computed for cluster counts ranging from 2 to 10.
- The optimal cluster count was identified as **9**, where the DB Index reached its lowest value (1.16).

P.T.O.

3. Recommendations and Next Steps

Recommendations:

1. **High Spenders (Clusters 5 and 8)**:
    - Focus on loyalty programs and exclusive offers to retain these valuable customers.
2. **Low Spenders (Cluster 6)**:
    - Target low spenders with promotional campaigns to increase their engagement.
3. **Occasional High-Value Purchasers (Cluster 2)**:
    - Provide personalized deals to encourage more frequent purchases.

Next Steps:

- Enhance clustering by incorporating additional demographic and behavioral features (e.g., product preferences, signup dates).
- Use advanced clustering techniques, such as DBSCAN or hierarchical clustering, to validate findings.
- Monitor cluster behavior over time to adjust marketing strategies dynamically.