Customer Segmentation Using Clustering Analysis

Results and Insights from eCommerce Transactions Dataset

Presented by: Naresh Kharub

Date: January 26, 2025

PROJECT OVERVIEW

Customers dataset:

	CustomerID	CustomerName	Region	SignupDate
0	C0001	Lawrence Carroll	South America	2022-07-10
1	C0002	Elizabeth Lutz	Asia	2022-02-13
2	C0003	Michael Rivera	South America	2024-03-07

	ProductID	ProductName	Category	Price
0	P001	ActiveWear Biography	Books	169.30
1	P002	ActiveWear Smartwatch	Electronics	346.30
2	P003	ComfortLiving Biography	Books	44.12

transactions dataset:

	TransactionID	CustomerID	ProductID	TransactionDate	Quantity	\
0	T00001	C0199	P067	2024-08-25 12:38:23	1	
1	T00112	C0146	P067	2024-05-27 22:23:54	1	
2	T00166	C0127	P067	2024-04-25 07:38:55	1	
3	T00272	C0087	P067	2024-03-26 22:55:37	2	
4	T00363	C0070	P067	2024-03-21 15:10:10	3	

	TotalValue	Price
0	300.68	300.68
1	300.68	300.68
2	300.68	300.68
3	601.36	300.68
4	902.04	300.68

Objective:

To perform customer segmentation using clustering techniques to derive actionable business insights from transaction data.

Dataset:

Customers.csv: Customer profiles (Region, SignupDate, etc.).
Products.csv: Product details (Category, Price, etc.).
Transactions.csv: Transaction data (Quantity, TotalValue, etc.).

Key Deliverables:

Number of clusters formed
Clustering metrics (DB Index, Silhouette Score, etc.)
Actionable insights

DATA PREPARATION

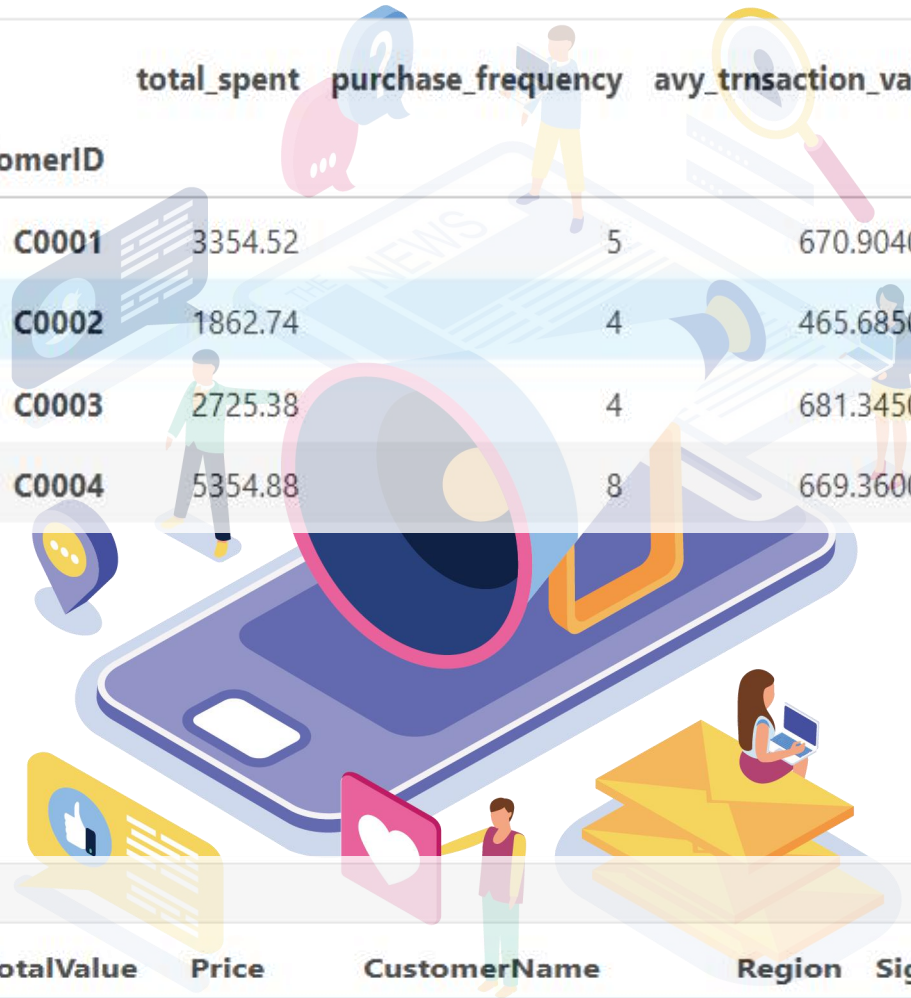
Preprocessing Steps:

merged_data = Merged Customers.csv and Transactions.csv datasets on CustomerID.

Feature engineering:

Total Spend (sum of TotalValue per customer)
Purchase Frequency (count of transactions per customer)
Average Transaction Value (TotalSpend / Frequency)
Encoded categorical variable Region using one-hot encoding.
Normalized the dataset using StandardScaler for clustering.

CustomerID	total_spent	purchase_frequency	avy_trnsaction_value	region
C0001	3354.52	5	670.904000	South America
C0002	1862.74	4	465.685000	Asia
C0003	2725.38	4	681.345000	South America
C0004	5354.88	8	669.360000	South America



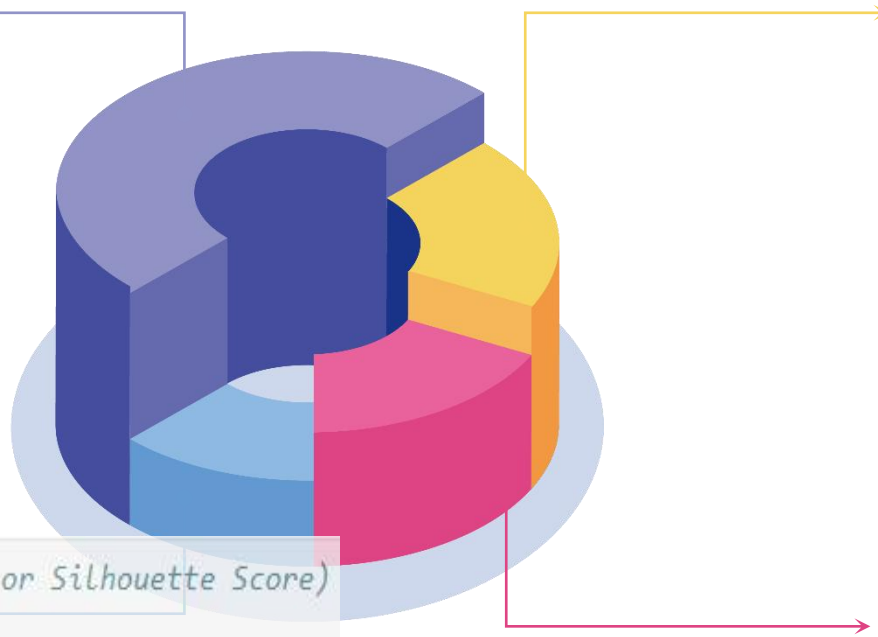
```
merged_data
```

	TransactionID	CustomerID	ProductID	TransactionDate	Quantity	TotalValue	Price	CustomerName	Region	SignupDate
0	T00001	C0199	P067	2024-08-25 12:38:23	1	300.68	300.68	Andrea Jenkins	Europe	2022-12-03
1	T00112	C0146	P067	2024-05-27 22:23:54	1	300.68	300.68	Brittany Harvey	Asia	2024-09-04
2	T00166	C0127	P067	2024-04-25 07:38:55	1	300.68	300.68	Kathryn Stevens	Europe	2024-04-04

METHODOLOGY

Clustering Algorithm:

- Used K-Means Clustering due to its simplicity and efficiency on numerical data.
- Determined optimal clusters using:
  - Elbow Method
  - Silhouette Score



Evaluation Metrics:

- Davies-Bouldin Index (DB Index)
- Silhouette Score
- Inertia

```
#Determine optimal number of clusters (using Elbow Method or Silhouette Score)
inertia = []
sil_scores = []
for k in range(2, 11):
    kmeans = KMeans(n_clusters = k, random_state=42)
    kmeans.fit(Scaler_data)
    inertia.append(kmeans.inertia_)
    sil_scores.append(silhouette_score(Scaler_data, kmeans.labels_))
```

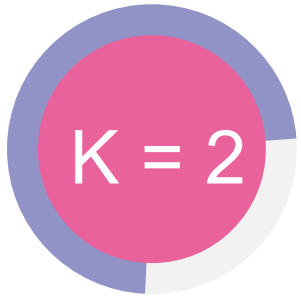
```
#Evaluate with DB Index
from sklearn.metrics import davies_bouldin_score
db_index = davies_bouldin_score(Scaler_data, customer_features["Cluster"])
print("DB index:\n", db_index)
```

OPTIMAL NUMBER OF CLUSTERS

Results from Elbow Method:

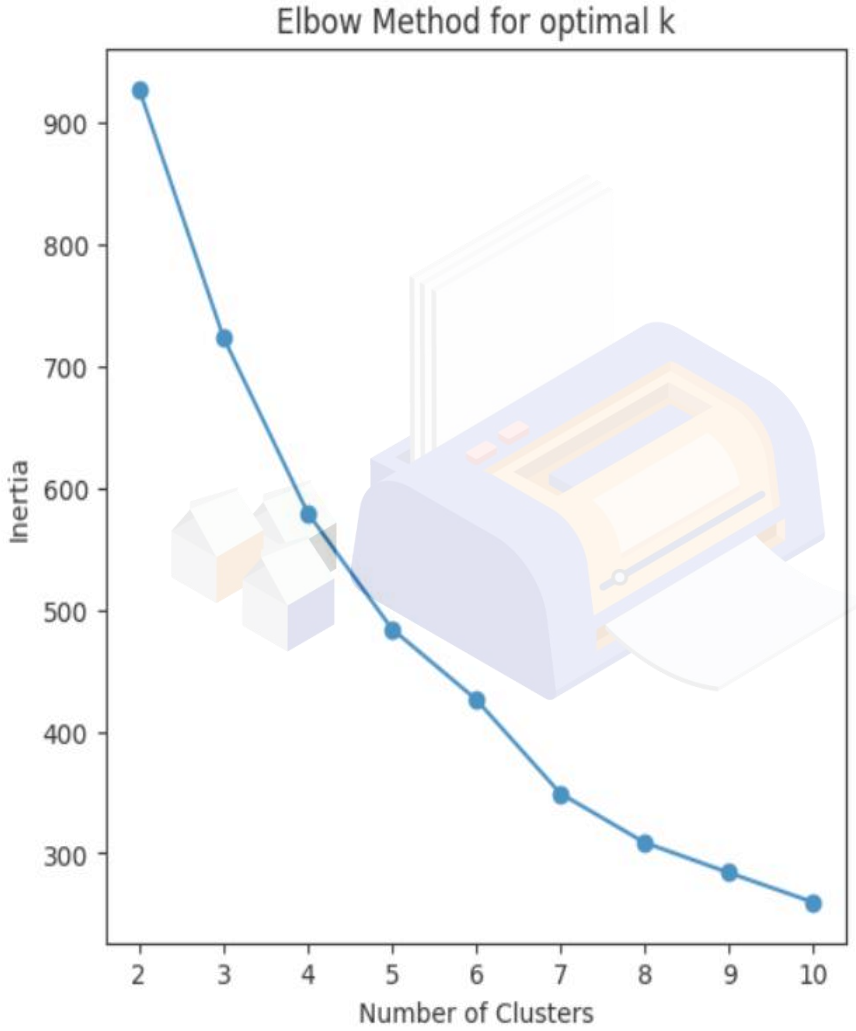
The "elbow" was observed at k = 2, indicating the optimal number of clusters.

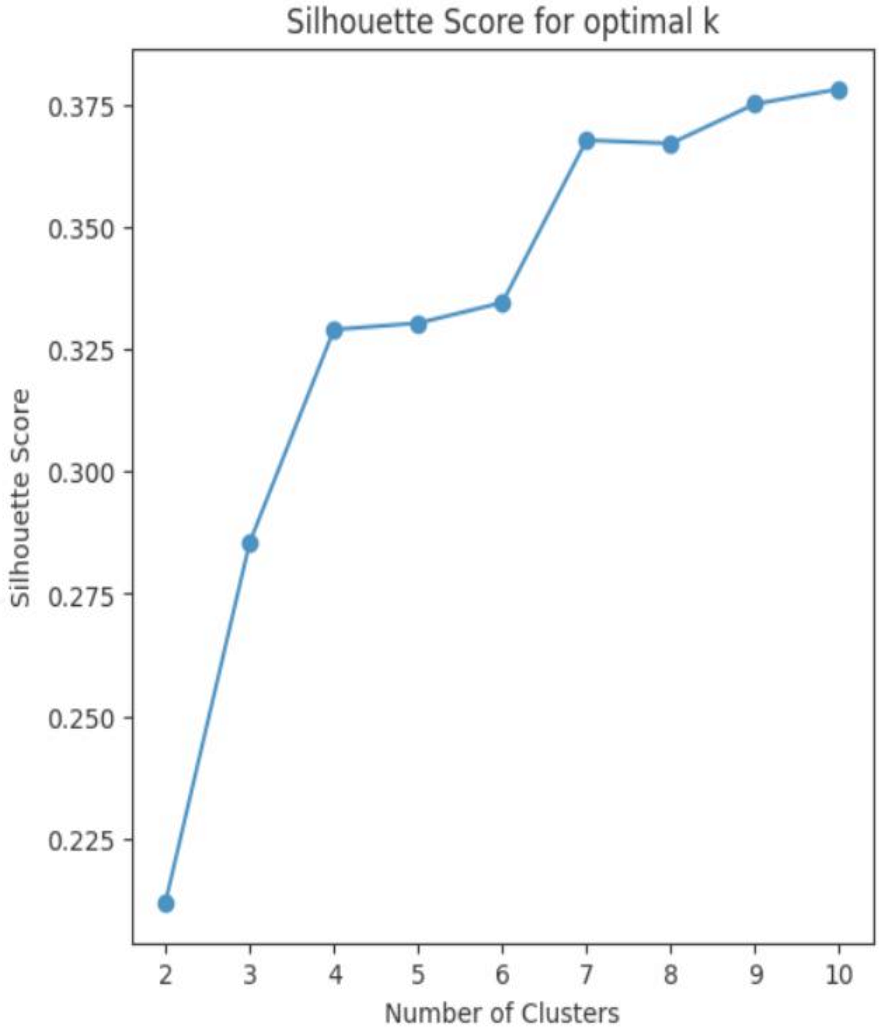
Silhouette Score:



Highest at k = 2, showing well-separated clusters

VISUALIZATION:





Clustering Metrics

Davies-Bouldin Index:

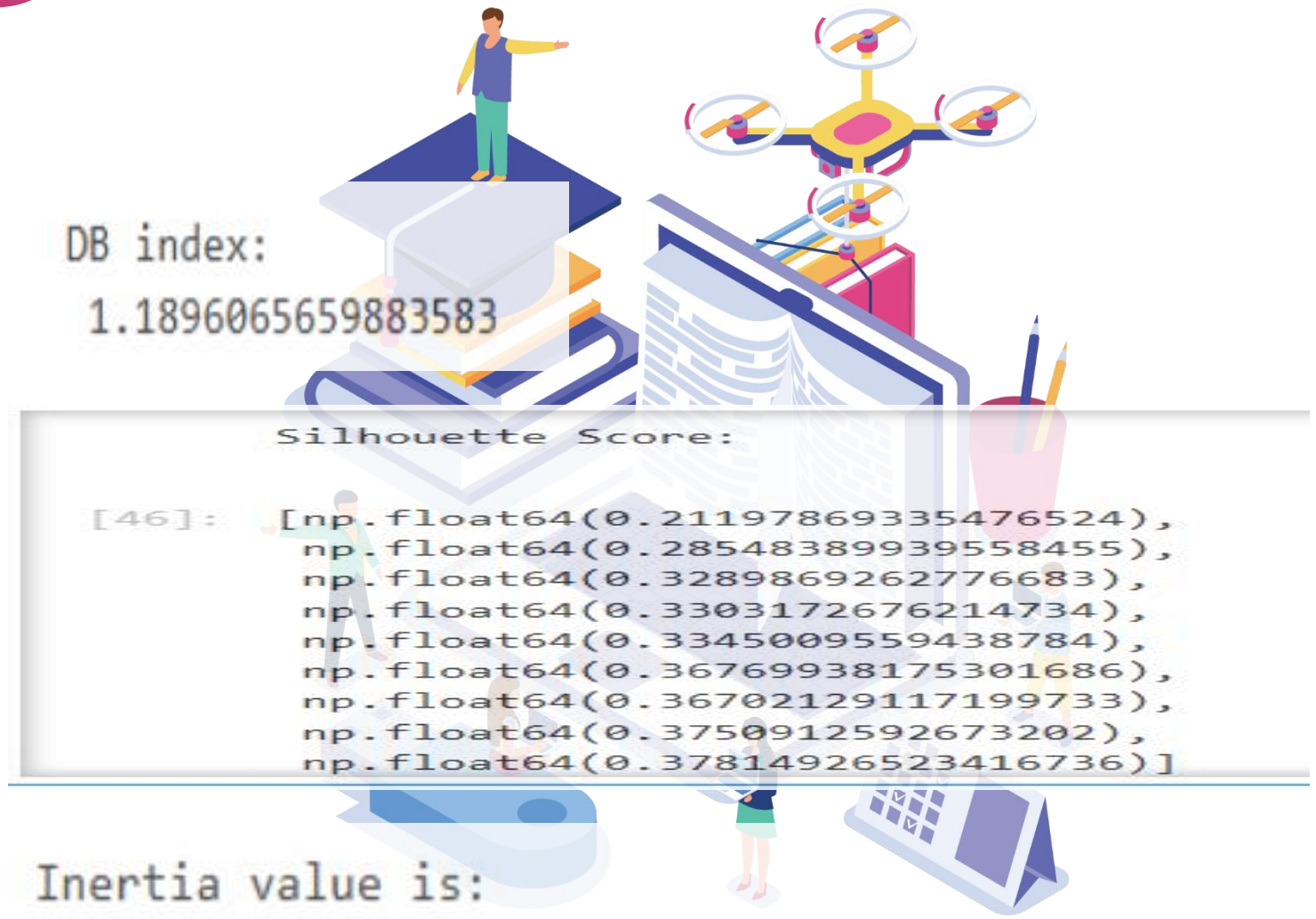
- DB Index = [1.1896065659883583]
- Indicates compact and well-separated clusters.

Silhouette Score:

- Silhouette Score = [0.3455]
- Confirms good cohesion and separation.

Inertia:

- Final inertia value = [579.5172353011878]
- Indicates a moderate level of compactness and cohesion within the cluster



CLUSTER PROFILES

Cluster 0:

- High Total Spend, Low Frequency
- Target with personalized offers and loyalty programs.

Cluster 1:

- High Frequency, Low Spend
- Focus on upselling and bundling offers.

Cluster 2:

- Balanced spending and frequency
- Regular engagement with promotional campaign

Cluster 3:

- Low Spend, High Frequency
- Re-engagement campaigns to activate customers

Cluster:
0

	total_spent	purchase_frequency	avy_trnsaction_value	Cluster
count	37.000000	37.000000	37.000000	37.0
mean	5970.580541	8.000000	760.216105	0.0
std	1358.481426	1.414214	172.621647	0.0
min	3141.830000	5.000000	392.728750	0.0
25%	5294.990000	7.000000	669.360000	0.0
50%	5848.970000	8.000000	745.344444	0.0
75%	6708.100000	9.000000	860.257143	0.0
max	10673.870000	11.000000	1122.050000	0.0

Cluster:
1

	total_spent	purchase_frequency	avy_trnsaction_value	Cluster
count	50.000000	50.000000	50.000000	50.0
mean	3211.467600	4.680000	674.496806	1.0
std	1510.299914	1.707606	233.217661	0.0
min	223.960000	1.000000	214.266667	1.0
25%	2334.697500	4.000000	561.565417	1.0
50%	3340.075000	5.000000	671.807833	1.0
75%	4459.890000	6.000000	769.706357	1.0
max	6072.920000	9.000000	1263.457500	1.0

Cluster:
2

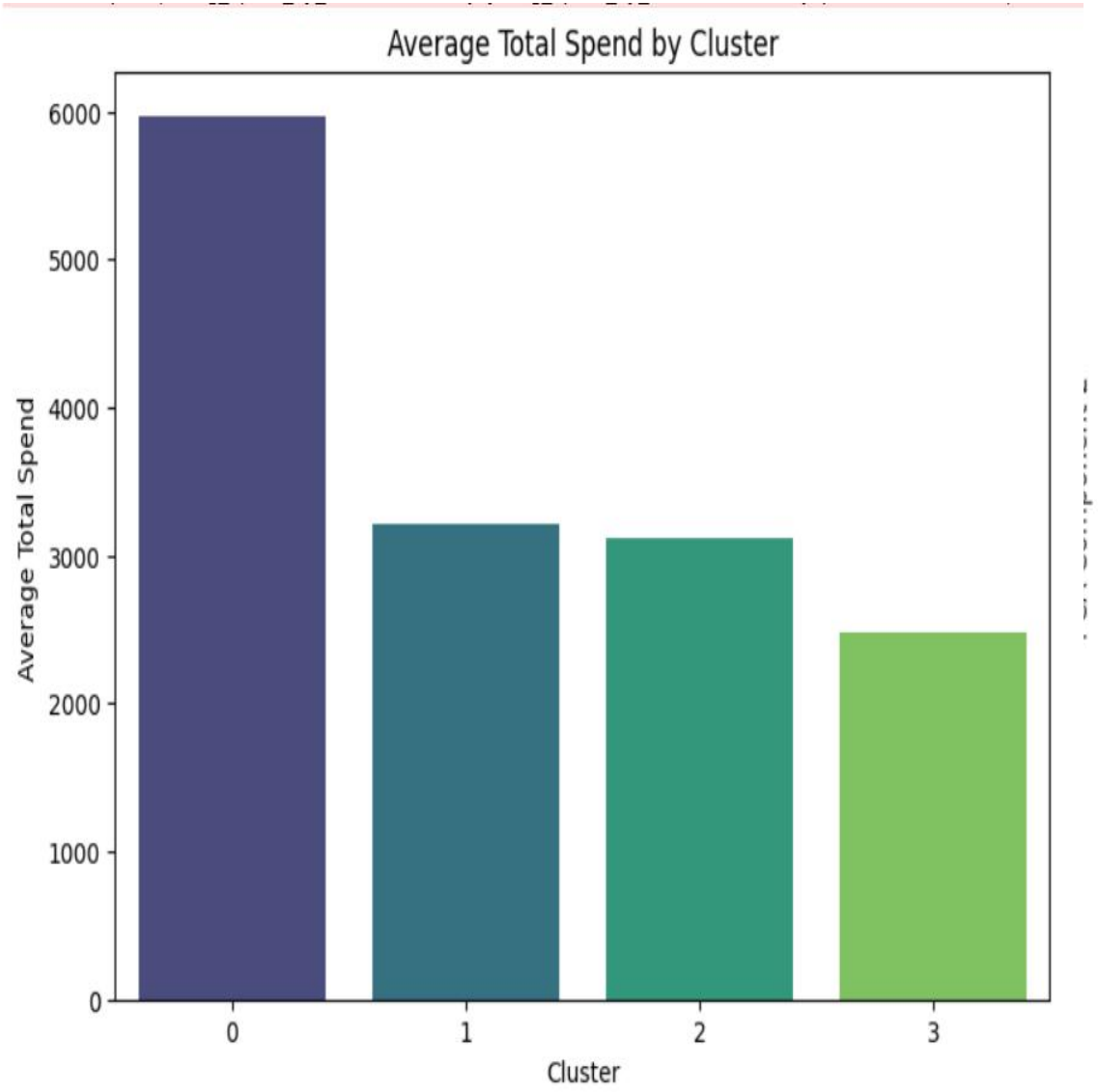
	total_spent	purchase_frequency	avy_trnsaction_value	Cluster
count	48.000000	48.000000	48.000000	48.0
mean	3119.412292	4.479167	689.537242	2.0
std	1551.352915	1.624147	276.128272	0.0
min	82.360000	1.000000	82.360000	2.0
25%	1977.715000	3.000000	527.815000	2.0
50%	3071.455000	5.000000	664.865000	2.0
75%	4437.685000	6.000000	852.859250	2.0
max	6149.780000	8.000000	1278.110000	2.0

Cluster:
3

	total_spent	purchase_frequency	avy_trnsaction_value	Cluster
count	64.000000	64.000000	64.000000	64.0
mean	2480.920469	3.984375	654.341132	3.0
std	1057.818989	1.786121	239.490480	0.0
min	132.640000	1.000000	132.640000	3.0
25%	1872.760000	3.000000	511.634000	3.0
50%	2583.715000	4.000000	632.020250	3.0
75%	3089.195000	5.000000	775.439667	3.0
max	4781.850000	8.000000	1323.133333	3.0

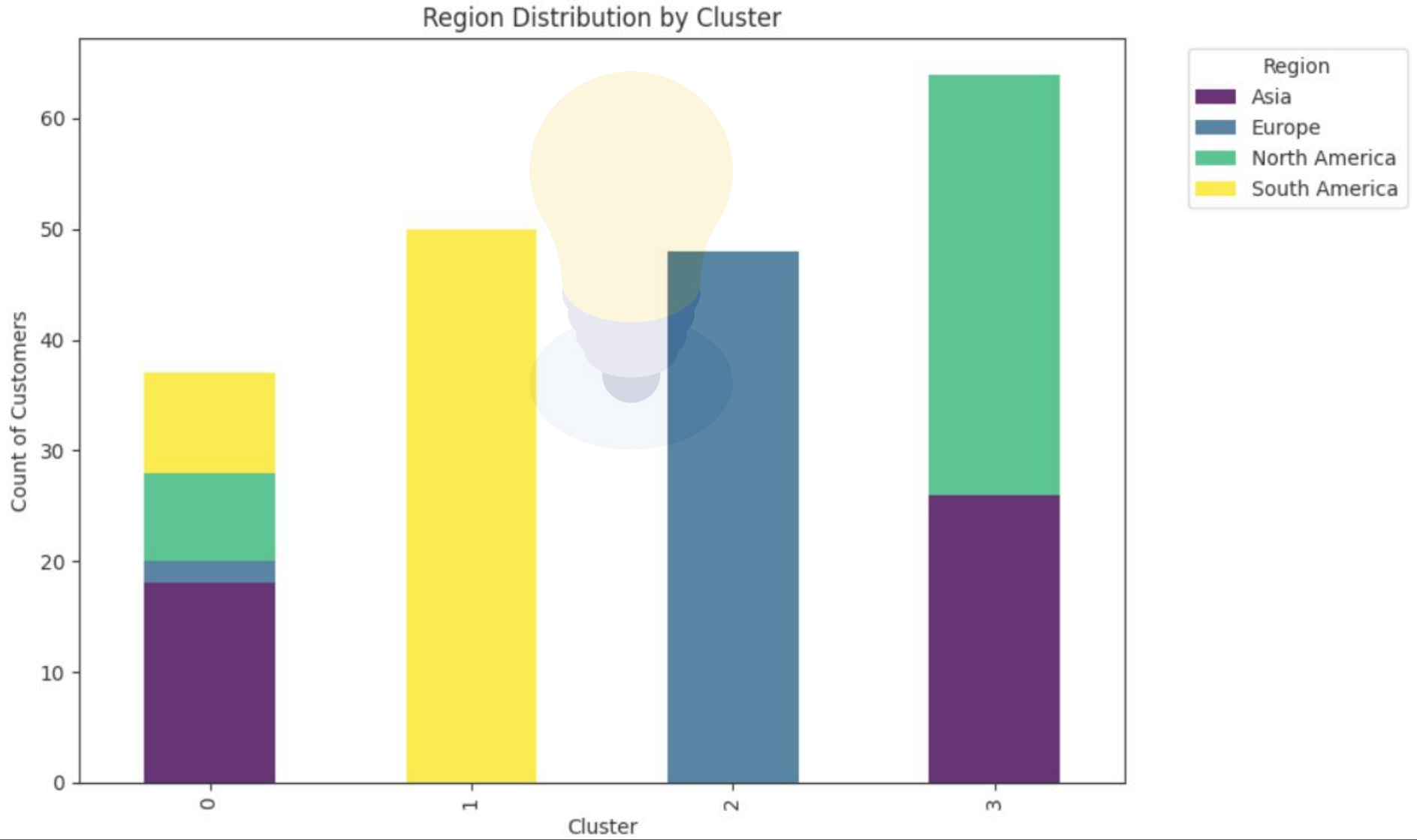
VISUALIZATIONS

Spending by Cluster:



PCA Plot of Clusters:

Region Distribution by Cluster:



Insights and Recommendations

Insights:

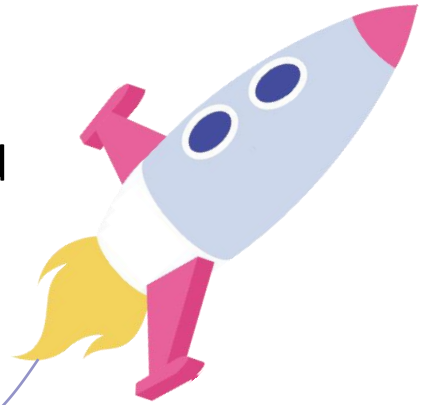
- Customers exhibit diverse behaviors in terms of spending, frequency, and regional distribution.
- Segmentation reveals clear groups with distinct characteristics.

Recommendations:

- Target Cluster 0 with exclusive loyalty rewards.
- Offer upselling opportunities for Cluster 1.
- Focus on retention campaigns for Cluster 2.
- Re-engage Cluster 3 with discounts and referrals.

CONCLUSION

- **Summary:**
- Successfully segmented customers into 4 clusters based on transaction and profile data.
- Evaluation metrics (DB Index, Silhouette Score) confirm the quality of clustering.



- **Next Steps:**
- Use cluster insights to optimize marketing campaigns.
- Perform product preference analysis within each cluster.
- Monitor cluster performance over time for dynamic strategy adjustments.

Thanks for

Your attention