

# eCommerce Customer Segmentation using Clustering

## Task 3: Customer Segmentation / Clustering

### 1. Clustering Methodology:

- Algorithm: K-Means Clustering
- Number of Clusters Evaluated: 2 to 5
- Features Used: Customer Region (One-Hot Encoded), Transaction Count, Total Transaction Value, Avg. Quantity per Transaction, Recency, Category Purchase Value and Quantity features for Books, Clothing, Electronics, Home Decor.
- Scaling: StandardScaler

### 2. Optimal Number of Clusters and Metrics:

- Optimal Number of Clusters (based on DB Index): 5
- Davies-Bouldin Index: 1.7691
- Silhouette Score: 0.1811
- Calinski-Harabasz Index: 25.7705

### 3. Cluster Profiles and Interpretations:

- Cluster 0: North America - Focused Value Buyers: This segment consists exclusively of customers from North America who are valuable, making frequent purchases with good transaction value and a moderate recency. While they purchase across categories, they show a relatively higher preference for Books and Clothing compared to Electronics and Home Decor.
- Cluster 1: Asia - Focused Moderate Value Buyers: This cluster is solely comprised of customers from Asia, exhibiting moderate transaction values and frequencies. Their purchase behavior is consistent, but their overall transaction value is lower compared to some other segments. There is a relatively high preference for Books within this segment.  
Cluster 2: Europe - Focused Moderate Value Buyers: Similar to Cluster 1, this segment is exclusive to European customers and shows moderate transaction value and purchase frequency. Their category preference is relatively balanced, possibly with a slight inclination towards Electronics.
- Cluster 3: High-Value, High Transaction Rate, Diverse Regional Mix: This represents the most valuable customer segment. Characterized by the highest total transaction value and transaction count, these highly active buyers demonstrate strong purchasing power across all product categories, especially in Electronics and Home Decor. They exhibit good recency and come from a diverse mix of regions.
- Cluster 4: South America - Lower Moderate Value Buyers: This segment is exclusive to South American customers and represents a moderate, but lower value segment. They

are characterized by less frequent purchases and are the least recent buyers compared to other clusters. Their spending across all categories is also moderately lower.

Cluster Sizes:

Cluster	Customer Count
0	42
1	38
2	43
3	29
4	48

Cluster Feature Means:

	Region_Asia	Region_Europe	Region_North America
Cluster			
0	0.000000	0.000000	1.000000
1	1.000000	0.000000	0.000000
2	0.000000	1.000000	0.000000
3	0.241379	0.241379	0.137931
4	0.000000	0.000000	0.000000

	Region_South America	TransactionCount	TotalTransactionValue
Cluster			
0	0.00000	4.857143	2989.942381
1	0.00000	4.210526	2883.438684
2	0.00000	4.232558	2841.617209
3	0.37931	7.862069	6206.343103
4	1.00000	4.708333	3180.704583

AvgQuantity Per Transaction   Recency   CategoryValue\_Books

Cluster

0	2.442262	91.833333	791.031905
1	2.459127	79.657895	599.346579
2	2.463123	83.139535	752.822326
3	2.766043	42.620690	2379.092414
4	2.530771	66.833333	724.665000

CategoryValue\_Clothing CategoryValue\_Electronics

Cluster

0	771.075238	849.492619
1	981.689737	534.475526
2	684.074884	602.326279
3	1444.620690	1838.935862
4	524.418125	949.282708

CategoryValue\_Home Decor CategoryQuantity\_Books

Cluster

0	578.342619	2.880952
1	767.926842	2.210526
2	802.393721	2.720930
3	543.694138	7.551724
4	982.338750	2.916667

CategoryQuantity\_Clothing CategoryQuantity\_Electronics

Cluster

0	2.880952	2.976190
---	----------	----------

1	3.500000	1.815789
2	2.255814	2.232558
3	4.896552	6.448276
4	2.020833	3.125000

CategoryQuantity\_Home Decor

Cluster

0	2.785714
1	3.157895
2	3.209302
3	2.758621
4	3.833333

