

# 1 Overview of Experiment

## 1.1 Data Sources

- **Customers.csv**: Profile information (e.g., Region, SignupDate).
- **Transactions.csv**: Transaction-level details (TotalValue, Quantity, Date).

## 1.2 Feature Engineering

- Aggregated transactions per customer: *Total\_Spend*, *Num\_Transactions*, *Avg\_Spend*, *Recency*, *Tenure*.
- Recency: days since last purchase.
- Tenure: days since signup.

## 1.3 Clustering Method

- Explored K-Means for k from 2 to 10.
- Evaluated with:
  - **Davies-Bouldin Index (DBI)** – lower is better.
  - **Silhouette Score** – higher is better.

# 2 Key Results

## 2.1 Clustering Metrics by k

k	DBI	Silhouette
2	1.493746	0.231352
3	<b>1.367294</b>	<b>0.233253</b>
4	1.535431	0.198701
5	1.483464	0.188024
6	1.458899	0.177327
7	1.584917	0.157673
8	1.561126	0.144893
9	1.485524	0.166461
10	1.393100	0.170809

- The **lowest DBI** is at **k = 3** (1.367294).
- Silhouette at k=3 is 0.233253, slightly higher than the other k values tested.

## 2.2 Final Model ( $k = 3$ )

- **DBI:** 1.3673
- **Silhouette:** 0.2333
- **Cluster Distribution:**
  - Cluster 0: 65 customers
  - Cluster 1: 121 customers
  - Cluster 2: 14 customers

## 3 Cluster Profiles

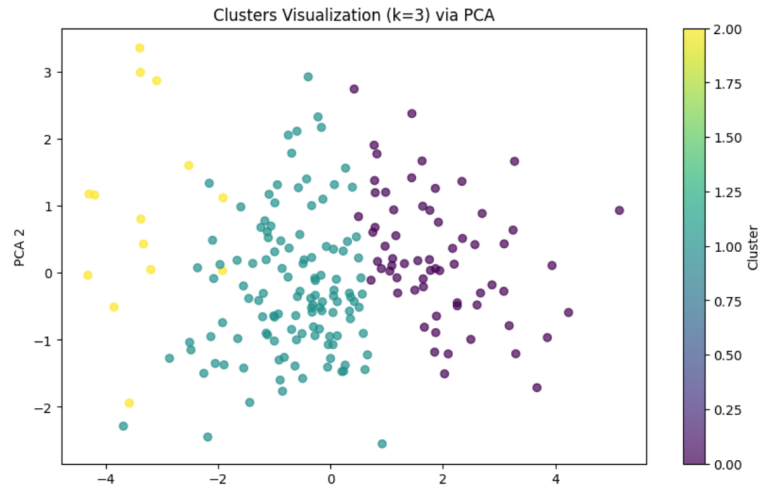
Cluster	Total_Spend	Num_Trans	Total_Qty	Avg_Spend	Recency_days	Tenure_days
0	5587.95	7.08	19.55	816.12	57.49	536.05
1	2610.58	4.30	10.04	631.30	66.49	537.60
2	778.53	1.43	3.64	528.13	259.71	459.71

Table 1: Cluster Profiles

### 3.1 Interpretation

- **Cluster 0:**
  - Highest total spend (avg 5588), most transactions, relatively recent activity.
  - Likely VIP/loyal customers; consider loyalty or premium programs.
- **Cluster 1:**
  - Mid-level spend and transaction count, moderate recency.
  - Good candidates for targeted promotions to move them into VIP status.
- **Cluster 2:**
  - Lowest spend, fewest transactions, longest recency (over 250 days).
  - Potentially lapsed customers; consider re-engagement or “win-back” campaigns.

## 4 PCA Visualization



*A 2D PCA projection of the clusters ( $k=3$ ). Each color is a cluster. Yellow (Cluster 2) lies mostly to the left, teal (Cluster 0) is central-left, and purple (Cluster 1) is more on the right.*

## 5 Conclusions and Recommendations

- **Number of Clusters:** 3, based on best Davies-Bouldin Index.
- **Davies-Bouldin Index:** 1.3673 (lower than other  $k$ 's).
- **Silhouette Score:** 0.2333, indicating modest separation.
- **Actionable Segmentation:**
  - **Cluster 0:** High-spend, frequent buyers → loyalty reinforcement.
  - **Cluster 1:** Mid-level spenders → cross-sell or up-sell campaigns.
  - **Cluster 2:** Low spend, infrequent → re-engagement or promotions to revive interest.