Customer Segmentation Analysis Report

Dataset Source:
The dataset used in this project is sourced from Kaggle (Customer Segmentation - Credit Cards), containing various features of credit card customers, including their spending, balance, cash advances, and other demographic variables.

Data Exploration and Preprocessing

Initial Exploration

The dataset contains various columns, such as:

- **Balance**: The balance of the credit card.
- **Purchases**: The total purchases made using the credit card.
- **Cash_Advance**: The total amount of cash advances taken by the customer.
- **Credit_Limit**: The total credit limit assigned to the customer.
- **Education**: The education level of the customer.

Upon initial exploration, the dataset contained missing values, which were handled using **mean imputation** for continuous variables and **mode imputation** for categorical variables. Several outliers were detected, particularly in high-purchase clusters, and were addressed by applying **Winsorization** to limit extreme values.

Feature Engineering

Several derived features were created to better represent customer behavior:

- **Purchase-to-Balance Ratio**: The ratio of purchases to the credit card balance, which helps identify heavy spenders.
- **Cash Advance-to-Balance Ratio**: The ratio of cash advances to the balance, providing insights into how frequently customers take cash advances.

Data Normalization

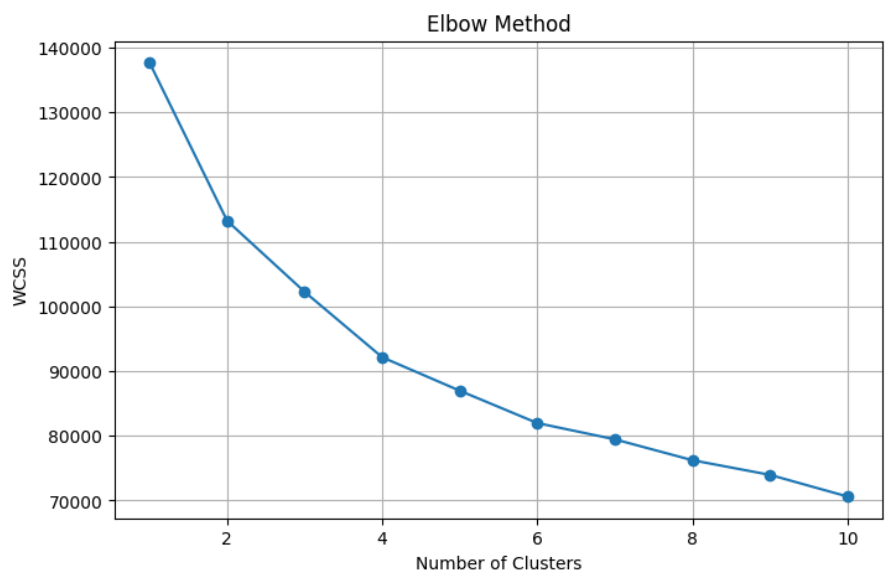
Since the dataset includes features on different scales (e.g., balance, purchases), **StandardScaler** was used to normalize the data before clustering.

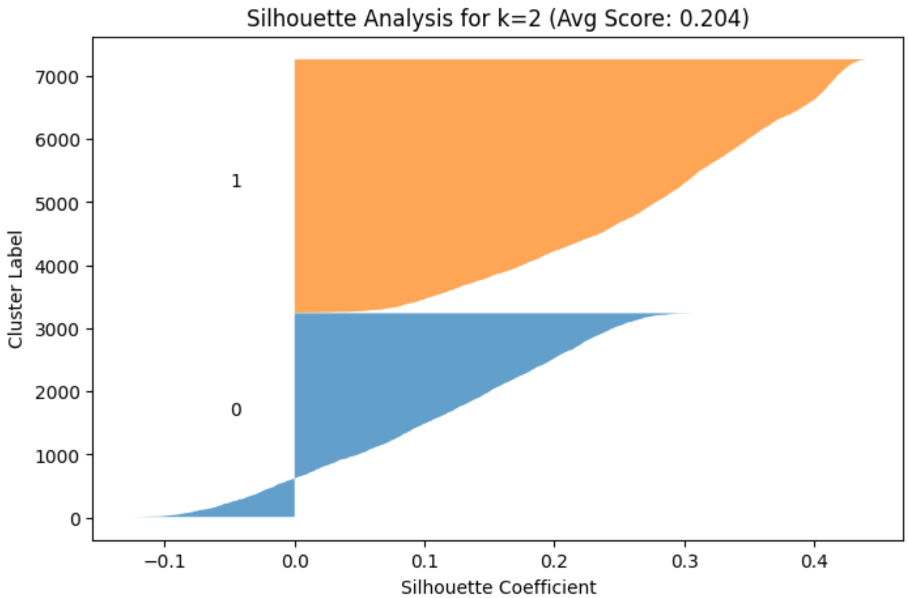
1. Summary of Findings

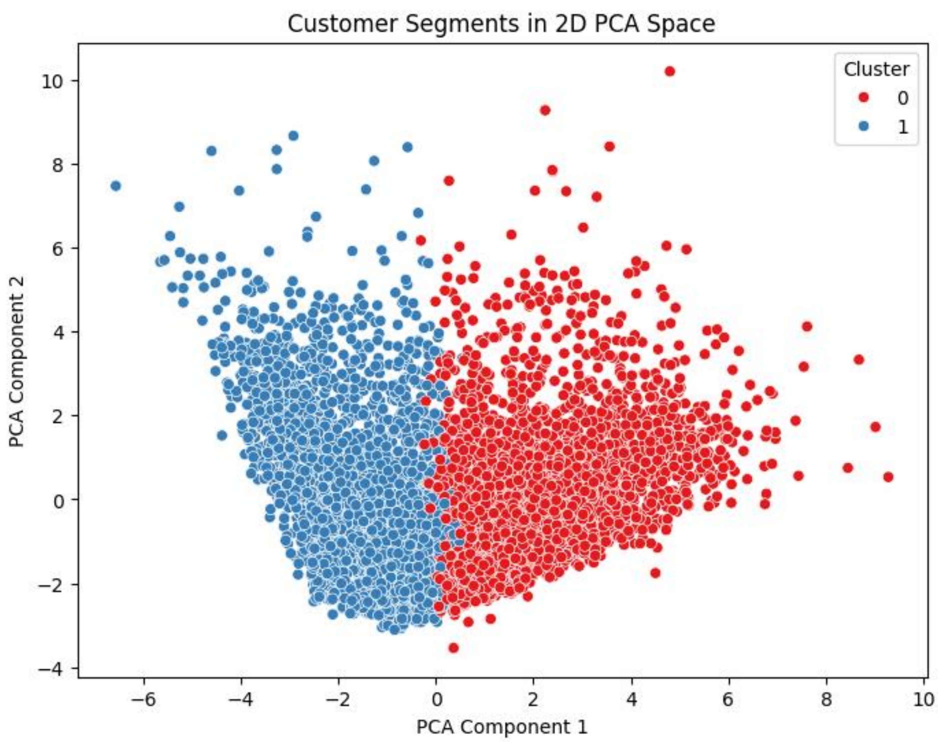
Using KMeans clustering on preprocessed credit card customer data, we successfully segmented customers into **2 distinct clusters**. The dataset included features such as **credit limit**, **balance**, **payment behavior**, **purchase patterns**, and **cash advances**. Key steps involved:

- **Data normalization (StandardScaler)**
- **Clustering (KMeans with K=2)**
- **Dimensionality reduction (PCA) for 2D visualization**
- **Evaluation using Silhouette Score**

Silhouette Score: 0.279 — suggests fair clustering structure, with moderate separation between clusters.




Silhouette Analysis for k=2 (Avg Score: 0.204)
Cluster Label
Silhouette Coefficient
1
0


Customer Segments in 2D PCA Space
Cluster
0
1
PCA Component 2
PCA Component 1

2. Cluster Profiles

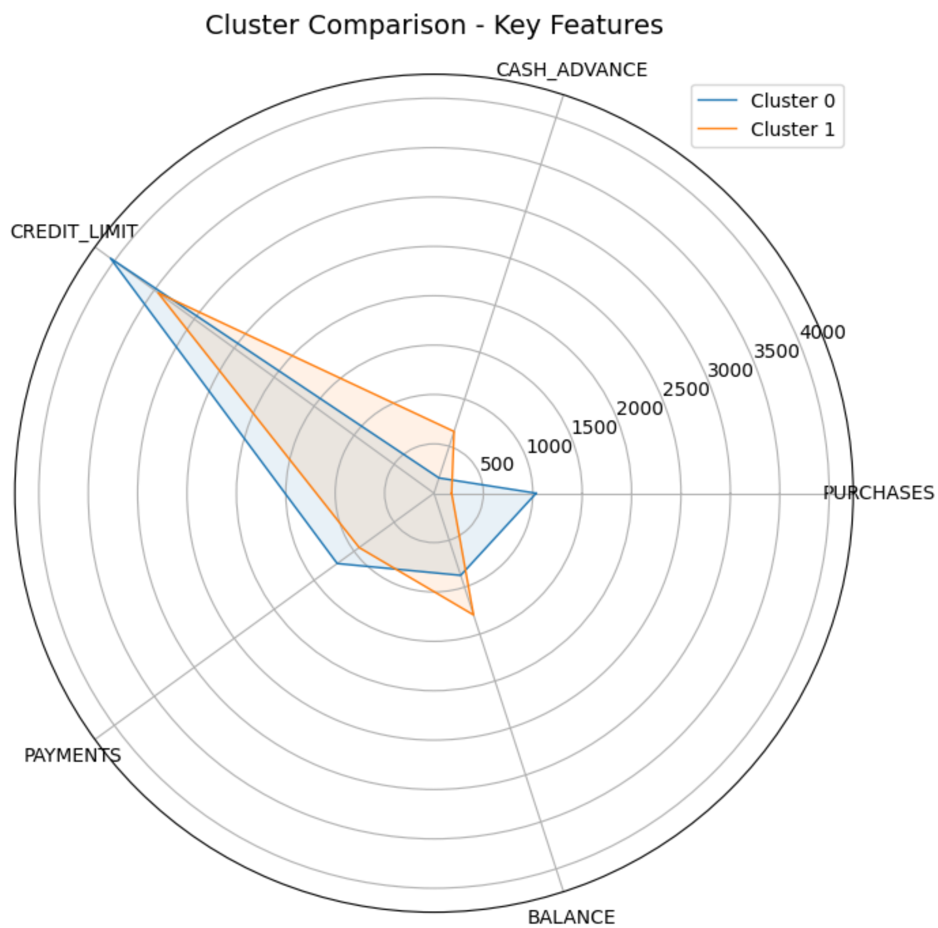
After examining the cluster centroids and means, we characterized the clusters as follows:

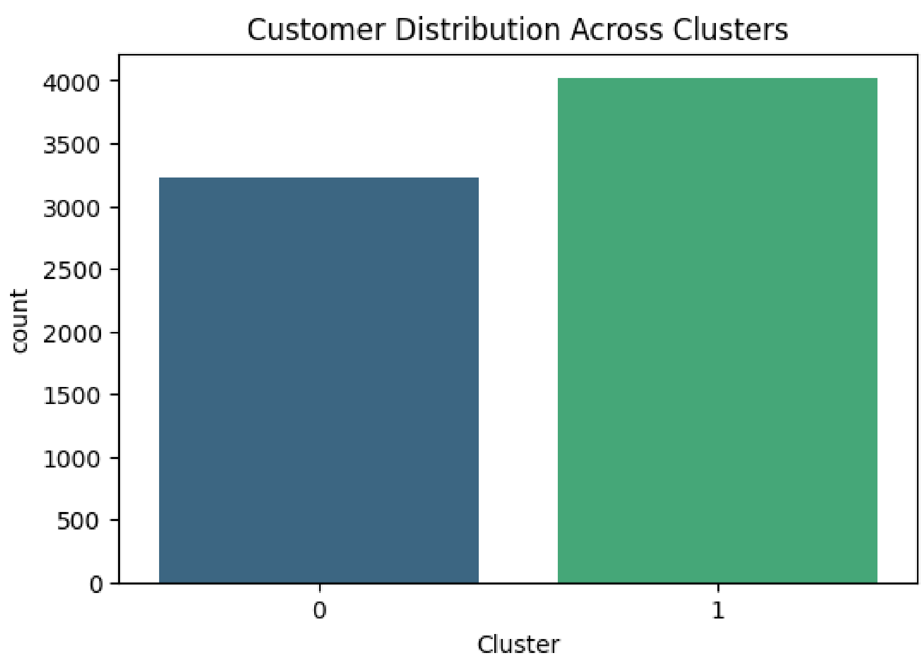
Cluster 0: Financially Active Customers

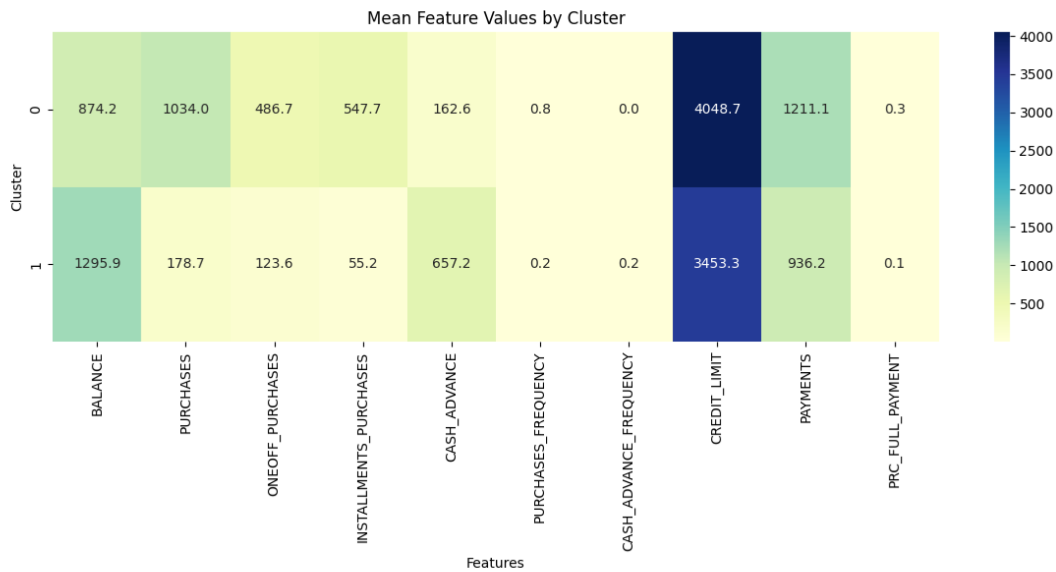
- **High average balance** and **credit limit**
- **High number of purchases** (both one-off and installment)
- **Higher payments and full payment ratios**
- **Low cash advance usage**
- Tend to **pay minimum or full amounts regularly**
- Likely represent **profitable and low-risk customers**

Cluster 1: Low Activity or Risk-Prone Customers

- **Lower balance** and **credit usage**
- Fewer **transactions** and **monthly purchases**
- High **cash advance frequency** (possible financial stress)
- Lower **payment-to-balance ratios**
- May represent **credit risk**, or low engagement customers


Cluster Comparison - Key Features
CASH_ADVANCE
CREDIT_LIMIT
PURCHASES
PAYMENTS
BALANCE
Cluster 0
Cluster 1
500
1000
1500
2000
2500
3000
3500
4000


Customer Distribution Across Clusters
count
Cluster
0
1
0
500
1000
1500
2000
2500
3000
3500
4000



3. Key Visualizations

The following visualizations were critical in interpreting the cluster structures:

PCA Scatter Plot (2D)

- Shows two well-formed groups.
- Red (Cluster 0) is more dispersed, indicating behavioral diversity.
- Blue (Cluster 1) is compact, suggesting similar, restricted behavior.

Bar Plot of Feature Means by Cluster

- Highlights which features most strongly differentiate the clusters (e.g., payments, purchases, balance).

Heatmap of Feature Differences

- Visually emphasizes the magnitude of difference across dimensions.

.

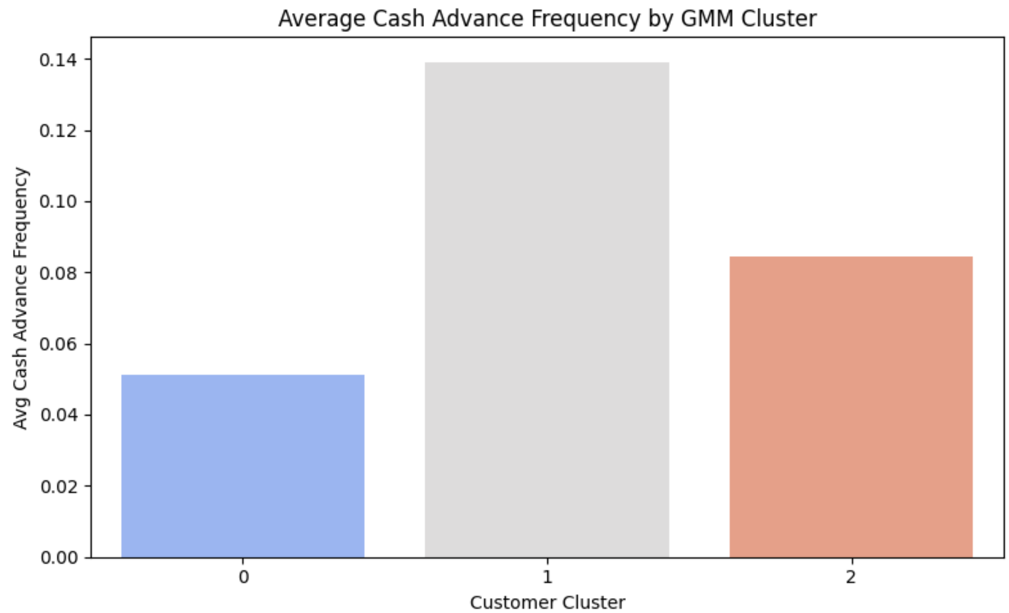
4. Business Insights & Recommendations

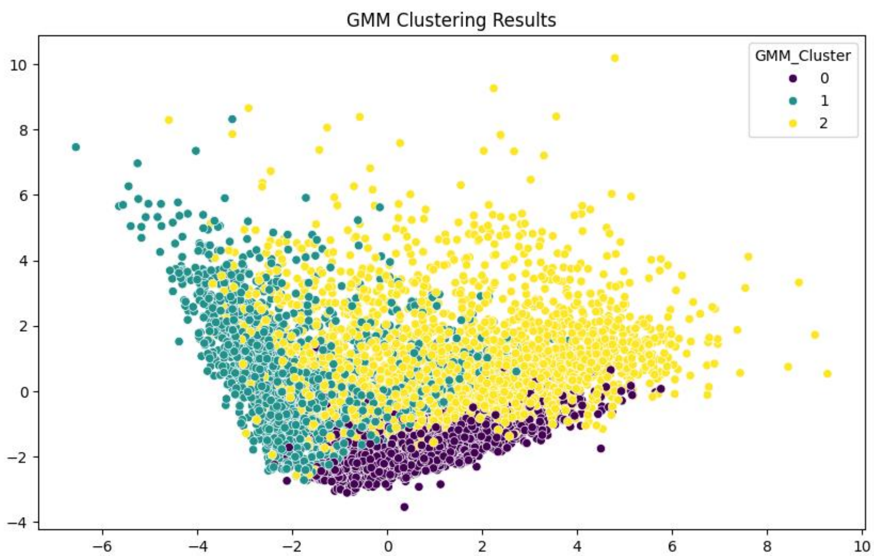
Cluster 0 – High-Value Customers

- Likely loyal and profitable.
- **Recommendations**:
  - Offer **cashback**, **reward programs**, or **premium cards**
  - **Cross-sell**: investments, insurance, or personal loans
  - Use **loyalty campaigns** to increase retention
  - Track behavioral changes to avoid churn

Cluster 1 – At-Risk or Inactive Customers

- May be financially constrained or disengaged.
- **Recommendations**:
  - Use **educational content** or **financial wellness tools**
  - Offer **small credit line increases** based on performance
  - Provide **incentives for usage** (installment offers, discounts)
  - Monitor for **delinquency risks**





5. Explanation of Segmentation Results

Why PCA?

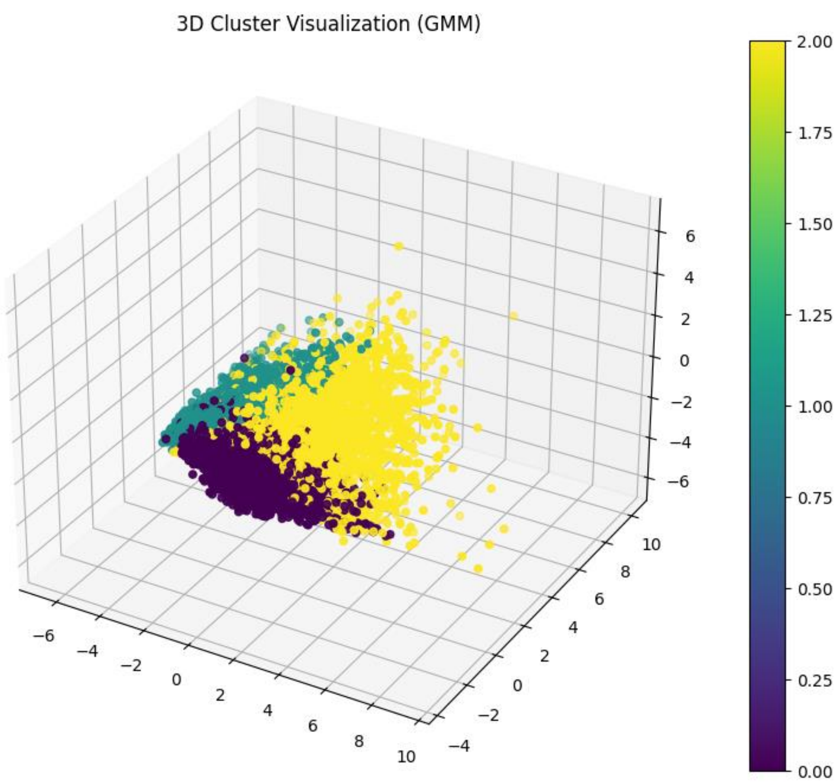
- PCA reduces dimensionality while preserving variance.
- Helped **visualize complex customer behavior** on 2D plane.

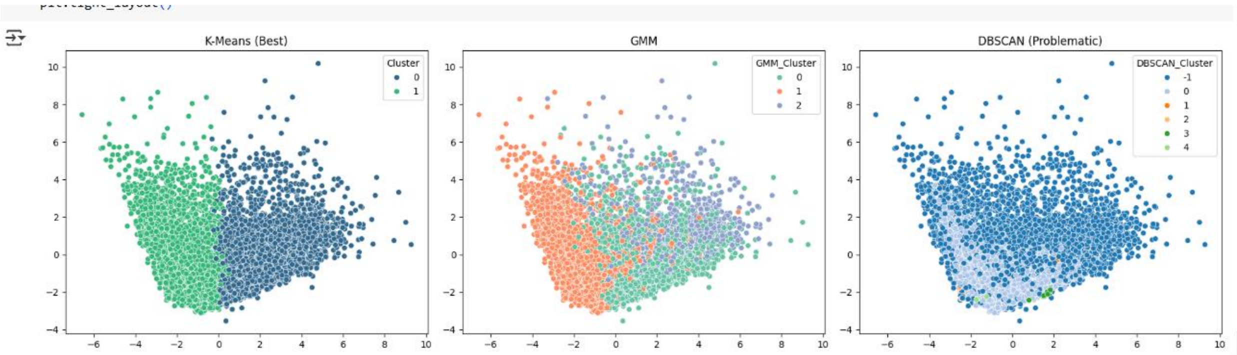
Why KMeans?

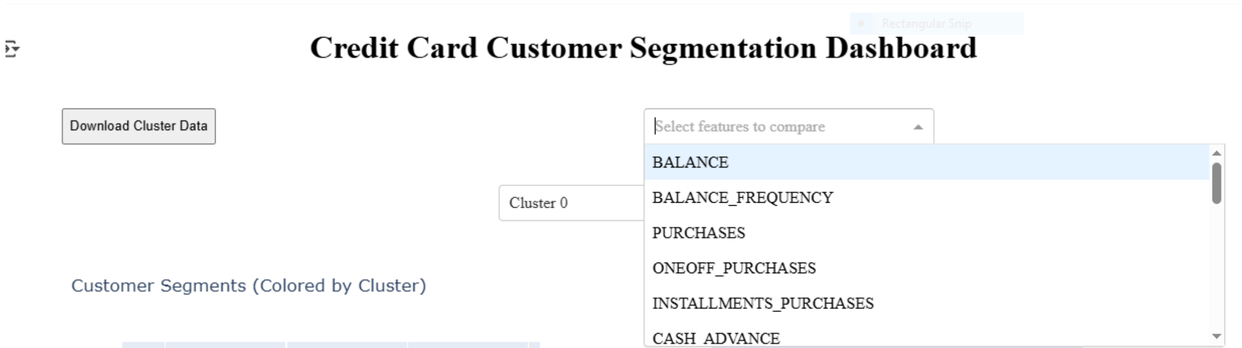
- Simple, fast, and effective for numerical data.
- Suitable when the number of clusters is known or tested via elbow/silhouette methods.

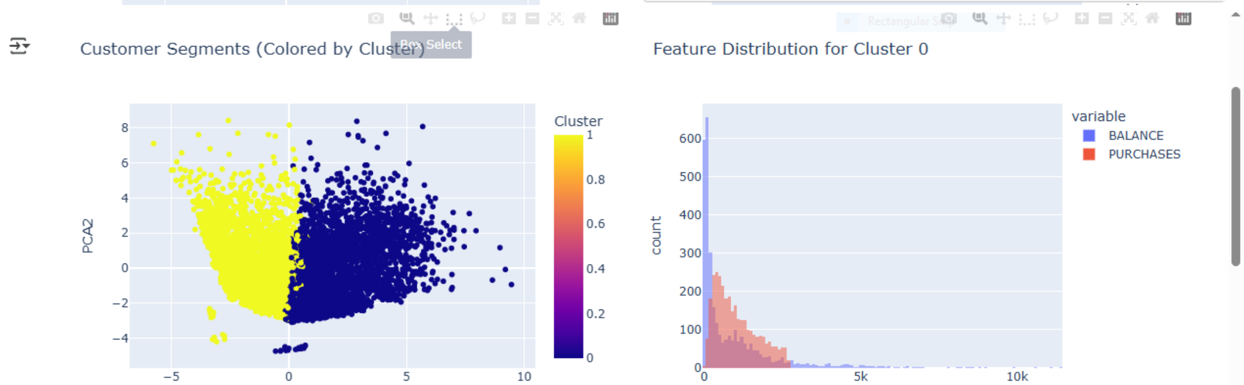
Interpretation

- The clustering revealed **two distinct behavioral groups**.
- Despite a **moderate silhouette score**, the clusters provide useful differentiation for business strategy.




K-Means (Best)
Cluster
0
1
GMM
GMM_Cluster
0
1
2
DBSCAN (Problematic)
DBSCAN_Cluster
-1
0
1
2
3
4


Credit Card Customer Segmentation Dashboard
Download Cluster Data
Select features to compare
BALANCE
BALANCE_FREQUENCY
PURCHASES
ONEOFF_PURCHASES
INSTALLMENTS_PURCHASES
CASH_ADVANCE
Cluster 0
Customer Segments (Colored by Cluster)


Customer Segments (Colored by Cluster)
Box Select
PCA2
Cluster
Feature Distribution for Cluster 0
count
variable
BALANCE
PURCHASES


Feature Correlation Heatmap
Feature Correlation Heatmap
BALANCE
PURCHASES
CREDIT_LIMIT
1
0.02317974
0.3541741
0.02317974
1
0.2463108
0.3541741
0.2463108
1

Cluster Statistics

BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FR
3231	3231	3231	3231	3231	3231	3231	3231
874.2477902011761	0.9184432482203652	1034.0389569792635	486.73522438873414	547.6509285051068	162.64387869761686	0.8402299538842465	0.27689797833488083
1303.0706868943578	0.1837247691181051	679.9880567611672	617.6156194376307	506.6803995100667	449.2476929086365	0.205654530125504	0.3330503248532037
0	0	8.4	0	0	0	0.083333	0
74.4552305	0.909091	468.6	0	185.12	0	0.727273	0
280.142117	1	874.94	192.95	417.96	0	0.916667	0.142857
1191.2315145	1	1494.955	838.01	777.165	0	1	0.5
11416.64736	1	2711.9	2711.9	2693.95	2854.551358	1	1

Cluster Comparison