Credit Card Customer Segmentation Using K Means Clustering Report

# Overview

This project aimed to segment credit card customers into distinct groups based on their spending and payment behavior. The ultimate goal is to help businesses understand customer profiles and tailor strategies accordingly for improved customer relationship management and targeted marketing.

# Methodology Overview

To segment the customers, I followed a structured and analytical pipeline involving data preparation, feature transformation, and clustering. Here's a high-level overview of the steps followed:

**Data Loading & Cleaning**

* Loaded the Credit Card Dataset from Kaggle.
* Removed duplicate entries.
* Handled missing values using KNN Imputer, which intelligently fills in missing numeric data.

**Outlier Removal**

* Detected and removed outliers using Local Outlier Factor (LOF) with a contamination rate of 5%. This step ensured clustering quality by removing noise and anomalies.

**Feature Engineering**

* Created new features to better represent user behavior:
* BALANCE\_TO\_LIMIT: Ratio of balance to credit limit.
* PAYMENT\_UTILIZATION: Ratio of payments to balance.
* PURCHASE\_INTENSITY: Product of total purchases and frequency.
* CASH\_ADVANCE\_FREQUENCY: Ratio of cash advance transactions to purchases.

**Redundancy Reduction**

* Removed highly correlated features (correlation > 0.95) to avoid multicollinearity and improve clustering performance.

**Transformation & Scaling**

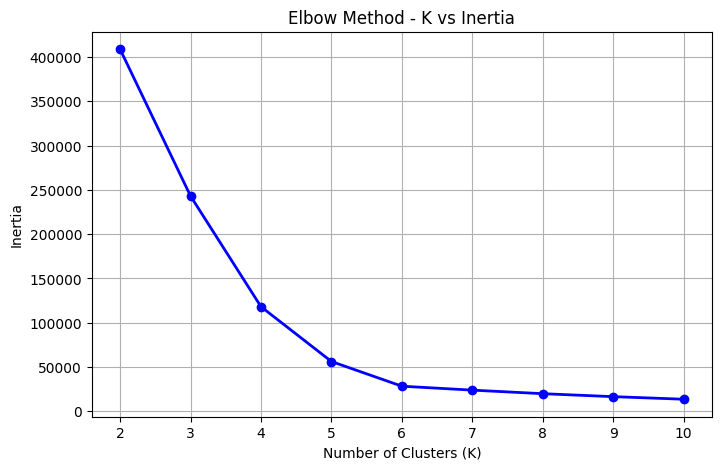
* Applied Power Transformation (Yeo-Johnson) for normalization.
* Scaled data using StandardScaler for uniformity.

**Dimensionality Reduction**

* Reduced dimensionality in two stages:
* PCA to retain 95% variance and remove noise.
* UMAP to embed high-dimensional data into a 10-dimensional space suitable for clustering.

**Clustering using K means**

* Performed KMeans Clustering using k-means++ initialization and the elkan algorithm.
* Evaluated performance using Silhouette Score and Elbow Method for K values from 2 to 10.
* Selected K = 6 based on:
* Highest Silhouette Score: 0.7765
* Inertia plot (Elbow Curve) suggesting the "knee" point at K=6.

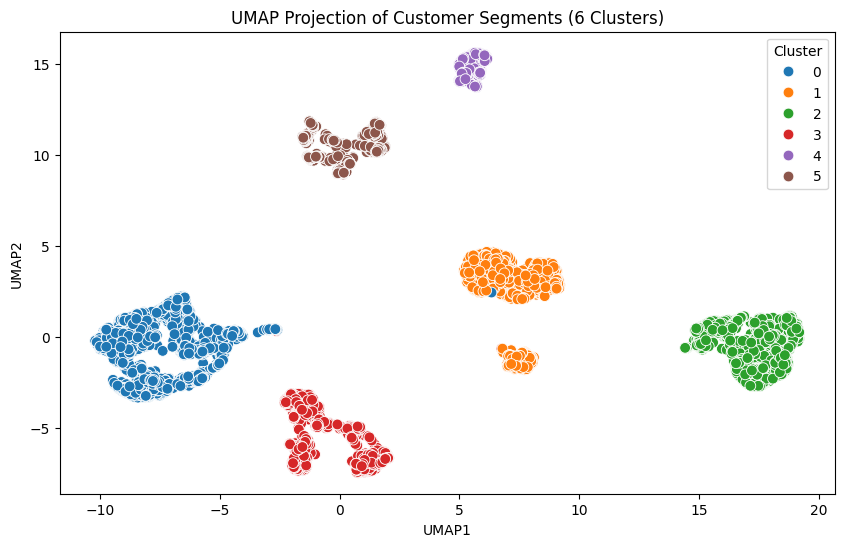


# 

# 

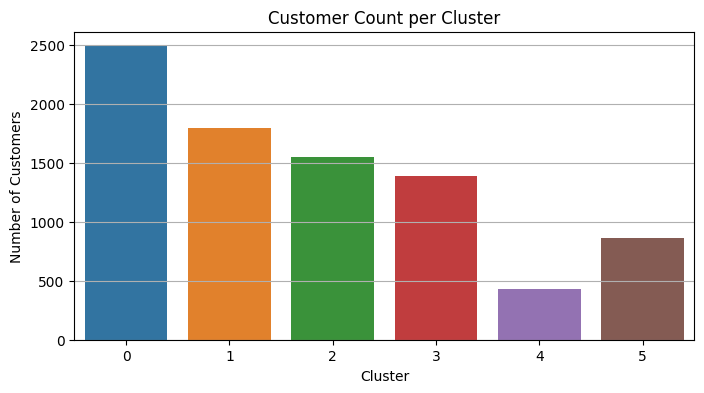
# Cluster profiles & Key visualizations

To visualize the clustering results, the high-dimensional feature space was projected into two dimensions using UMAP, enabling a clear representation of the six customer segments. The scatter plot below illustrates the distribution of customers across these clusters.



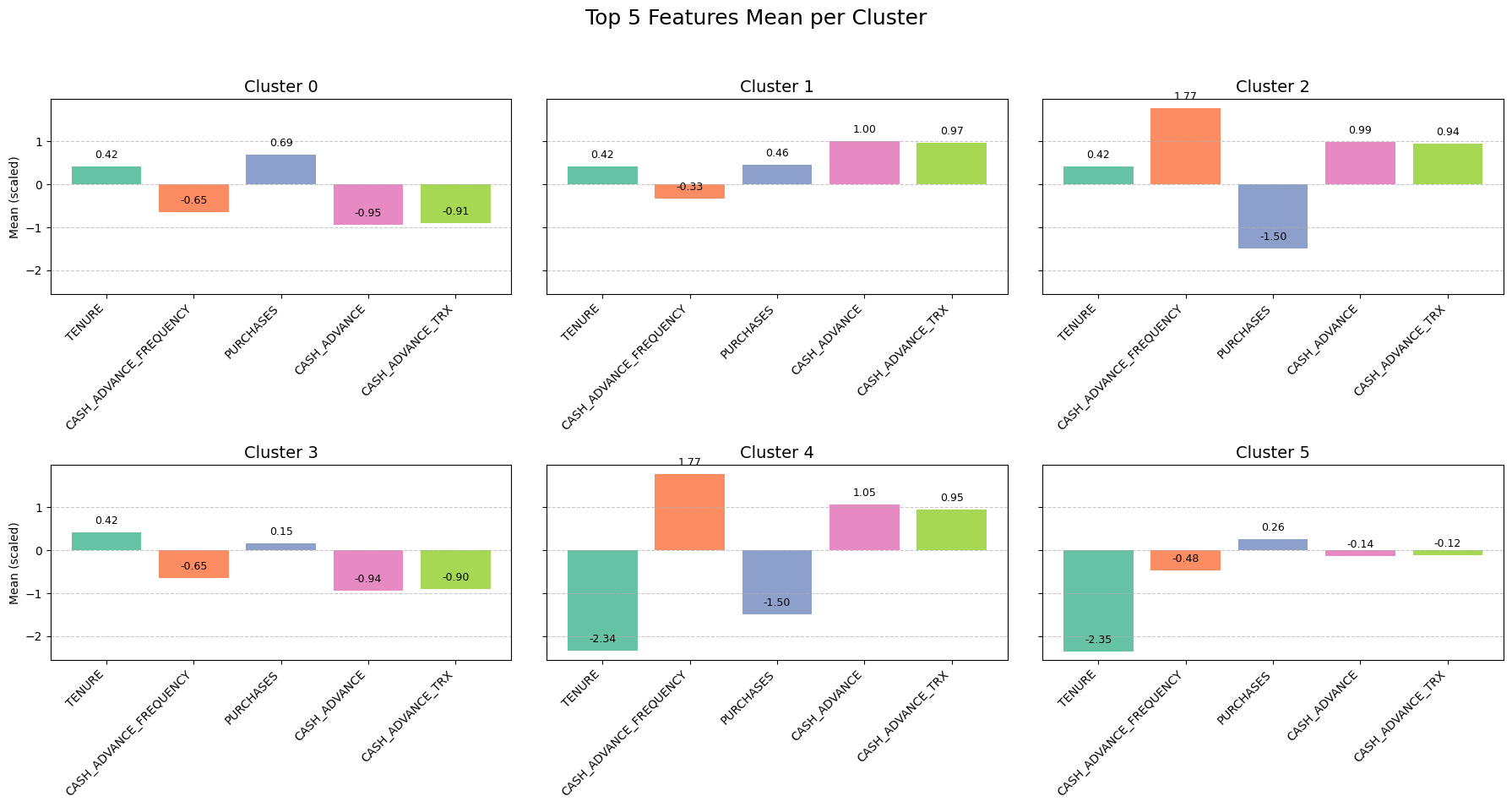
# 

The UMAP scatter plot revealed well-separated clusters, indicating that the customers have been segmented based on distinct behavioral patterns. This provides confidence that the segmentation process has captured meaningful differences among customer groups.



In addition, a bar plot was used to show the number of customers in each cluster, offering insight into the size and distribution of the segments as follow :

* Cluster 1 : **2486** customer
* Cluster 2 : **1793** customer
* Cluster 3 : **1547** customer
* Cluster 4 : **1387** customer
* Cluster 5 : **430** customer
* Cluster 6 : **859** customer



To better understand what drives differences between the segments, the mean values of all features were calculated per cluster. The **top 5 features** with the highest variation across clusters were:

* **TENURE** (Duration of being a customer)
* **CASH\_ADVANCE\_FREQUENCY** (Ratio of cash advance transactions to purchases)
* **PURCHASES** (Total purchase amount)
* **CASH\_ADVANCE** (Total amount of cash advances)
* **CASH\_ADVANCE\_TRX** (Number of cash advance transactions)

These features were used to profile each cluster and interpret their behavioral patterns. So based on the feature distributions, visual patterns from clustering projections, and the **cluster profiling function**, the following business insights and interpretations are proposed

# Business Insights and Recommendations

**Cluster 1: Purchase-Oriented Customers**

* + **High in**: One-off purchases, purchase frequency, and total purchases.
  + **Low in**: Cash advances and related transaction frequency.
  + **Insight**: This segment consists of highly engaged customers who prefer traditional purchase behavior over cash advances.
  + **Recommendation**: Offer loyalty rewards, cashback programs, or exclusive discounts to retain these valuable, low-risk customers.

**Cluster 2: Cash Advance-Heavy Users**

* + **High in**: Cash advance amount, transaction count, and overall balance.
  + **Low in**: Payment utilization and full payments.
  + **Insight**: These users rely heavily on cash advances, potentially indicating financial stress or dependency on short-term credit.
  + **Recommendation**: Introduce educational material on financial wellness, or targeted offers to encourage purchases instead of cash advances. Consider credit risk monitoring.

**Cluster 3: Frequent Cash Advance Users**

* + **High in**: Cash advance frequency and amount.
  + **Low in**: Purchase activity (both volume and frequency).
  + **Insight**: This group uses the card almost exclusively for cash advances and avoids retail purchases.
  + **Recommendation**: These may be high-risk customers. Re-evaluate credit policies or offer alternative credit products to reduce reliance on cash withdrawals.

**Cluster 4: Installment-Oriented Customers**

* + **High in**: Installment purchases and payment utilization.
  + **Low in**: One-off purchases and cash advances.
  + **Insight**: Customers in this cluster prefer structured payments and steady usage.
  + **Recommendation**: Promote installment plans, zero-interest EMI schemes, or long-term usage benefits tailored to this group’s behavior.

**Cluster 5: Short-Term, Cash-Oriented Users**

* + **High in**: Cash advance transactions and frequency.
  + **Low in**: Tenure and purchase-related activity.
  + **Insight**: These are likely new or transient customers, primarily using the card for cash access rather than purchases.
  + **Recommendation**: Consider onboarding campaigns to educate them on the card’s purchasing benefits, or evaluate for early risk signs.

**Cluster 6: Consistent Purchasers with Lower Tenure**

* + **High in**: Purchase frequency and total purchases.
  + **Low in**: Tenure and cash advance frequency.
  + **Insight**: Likely newer customers who actively use their card for purchases and avoid cash advances.
  + **Recommendation**: These are promising new customers—initiate retention strategies like welcome offers, upgrade suggestions, and early loyalty perks.

By uncovering key behavioral patterns across customer clusters, this analysis provides a foundation for smarter decision-making. These insights not only highlight diverse usage habits but also point to targeted opportunities for growth, risk management, and customer loyalty. Moving forward, aligning business strategies with these profiles can help deliver more personalized, effective, and sustainable customer experiences.