## **Project Report: Customer Segmentation for Credit Card Customers**

### **Executive Summary**

This project aimed to segment credit card customers based on their usage patterns using unsupervised learning. After testing both K-Means and Gaussian Mixture Models (GMM), K-Means provided superior clustering quality with a silhouette score of 0.498 (at k=5) compared to GMM’s weaker performance (score = 0.216). Five distinct customer segments were identified, each reflecting unique behaviors in purchases, repayments, and credit utilization. These insights pave the way for targeted marketing, product development, and retention strategies.

### **Introduction**

#### **Dataset**

Source: Customer Segmentation - Credit Cards  
 (<https://www.kaggle.com/code/des137/customer-segmentation-credit-cards>)

#### **Project Objectives**

1. Explore and understand the dataset.
2. Identify the best number of clusters to group the customers.
3. Segment the customers into distinct clusters based on relevant features.
4. Analyze and interpret the characteristics of each customer segment.
5. Provide business insights and strategic recommendations based on the analysis.

### **1. Data Exploration and Preprocessing**

* Conducted an initial review of the dataset to understand structure and key features.
* Removed irrelevant columns (e.g., CUST\_ID).
* Handled missing values using median imputation.
* Check for duplicated rows (none found).
* Standardized all features using z-score normalization to eliminate scale bias.
* Visualized distributions and detected outliers using boxplots.

#### 

#### **Feature Engineering**

Created new features to better capture customer behavior:

* PURCHASES\_RATIO
* ONEOFF\_PURCHASE\_RATIO
* INSTALLMENTS\_PURCHASES\_RATIO
* CASH\_ADVANCE\_RATIO
* CREDIT\_UTILIZATION
* AVG\_PURCHASE\_VALUE
* PURCHASES\_PER\_TENURE
* PAYMENT\_RATIO
* AVG\_CASH\_ADVANCE\_TRX

Used **log1p** transformation to reduce skewness.

### **2. Determining Optimal Clusters**

* Applied PCA to reduce dimensionality for clustering.
* Calculate **K-Means** score using **Silhouette Score**.
* Tested K-Means with k values from 2 to 10.
  + **Optimal K** : 5
  + **Silhouette score** : 0.498
* Calculate **GMM** score using **Silhouette Score**.
  + **Silhouette Score** (Hard Clusters): 0.216
* Selected **k=5** due to highest silhouette score (0.498).

### **3. Customer Segmentation**

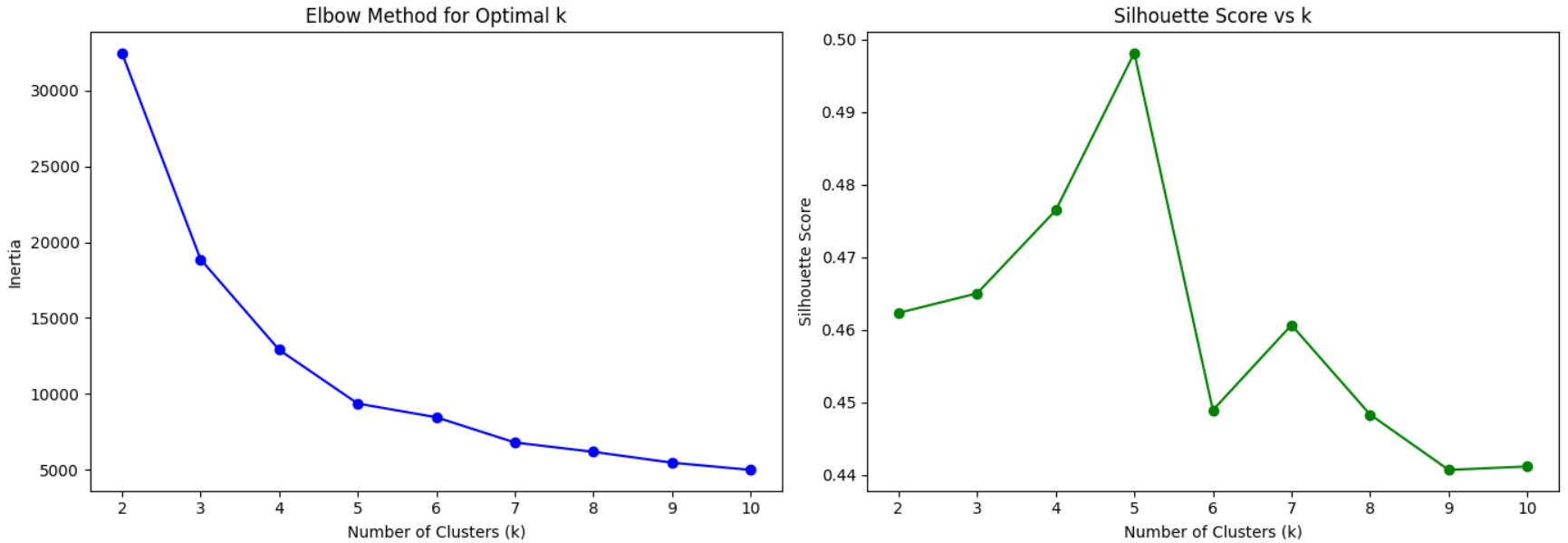
* Used K-Means clustering with k=5 to segment customers.
* Assigned cluster labels and analyzed behavioral traits.

#### **Cluster Profiles:**

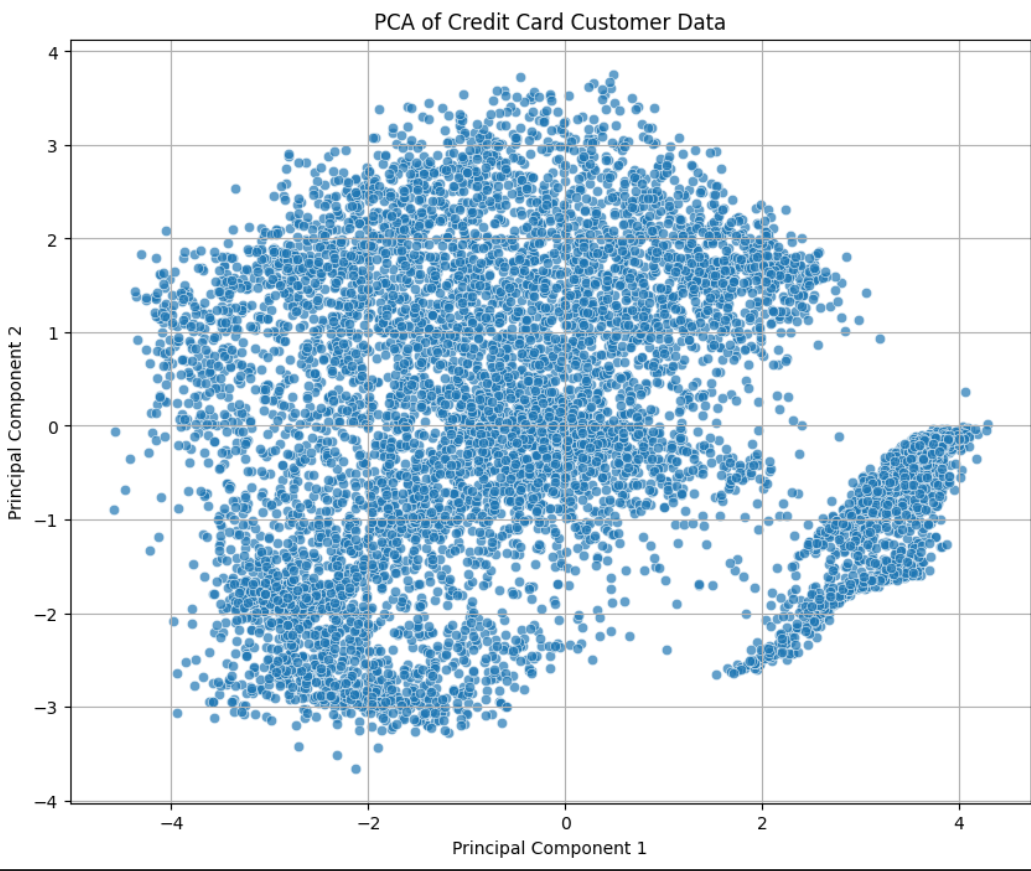
| **Cluster** | **Behavior Summary** | **Financial Traits** | **Count** |
| --- | --- | --- | --- |
| 0 | One-off Spenders | Moderate utilization, low full payment rate | 2121 |
| 1 | Disciplined Full Payers | High repayment, low utilization, avoids cash | 1675 |
| 2 | Frequent Installment Shoppers | Active but low repayment, avoids one-off/cash | 1634 |
| 3 | Heavy Cash Advance Users | Financially stressed, high utilization | 2098 |
| 4 | Balanced, High-Value Engaged Users | Strong all-round usage and repayment behavior | 1422 |

### **4. Visualization and Analysis**

* **Silhouette Score Plot**: Used to evaluate clustering quality.

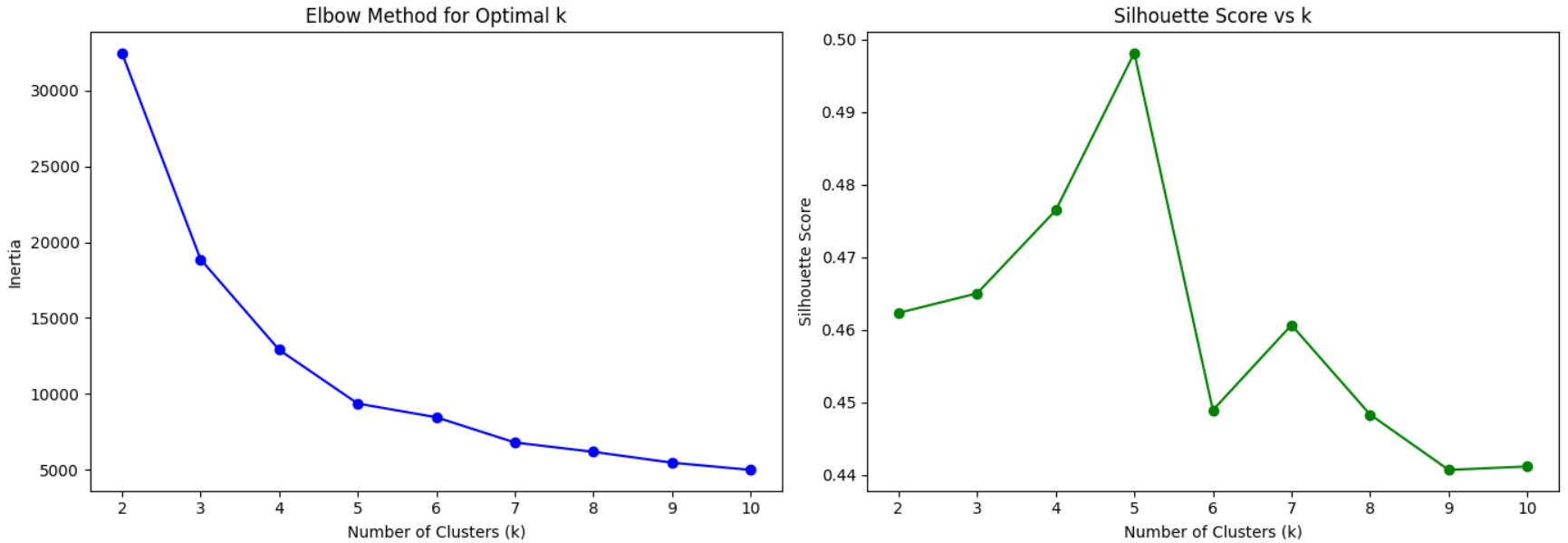


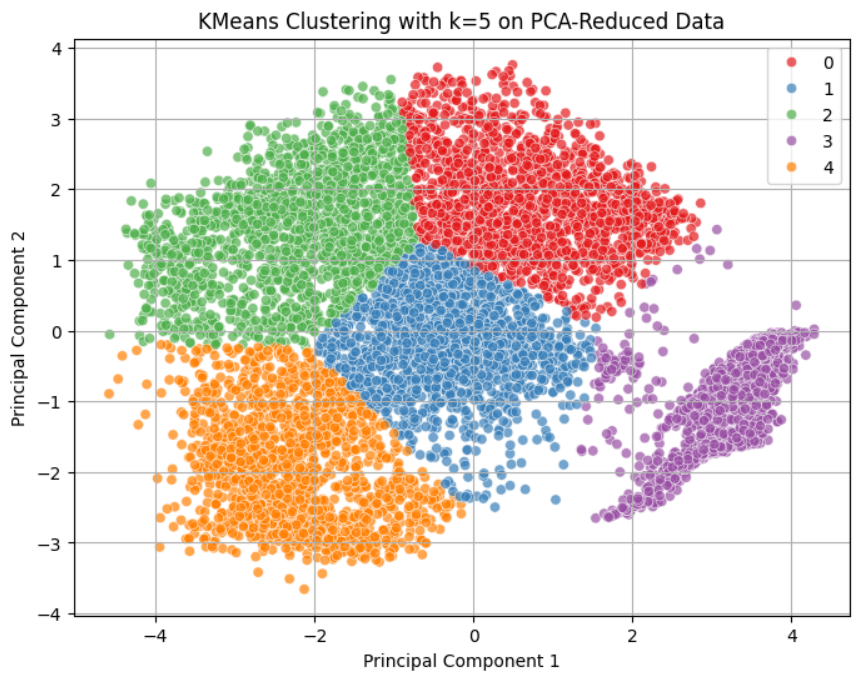
* **PCA-Reduced Cluster Plot**: Showed separation in 2D space.



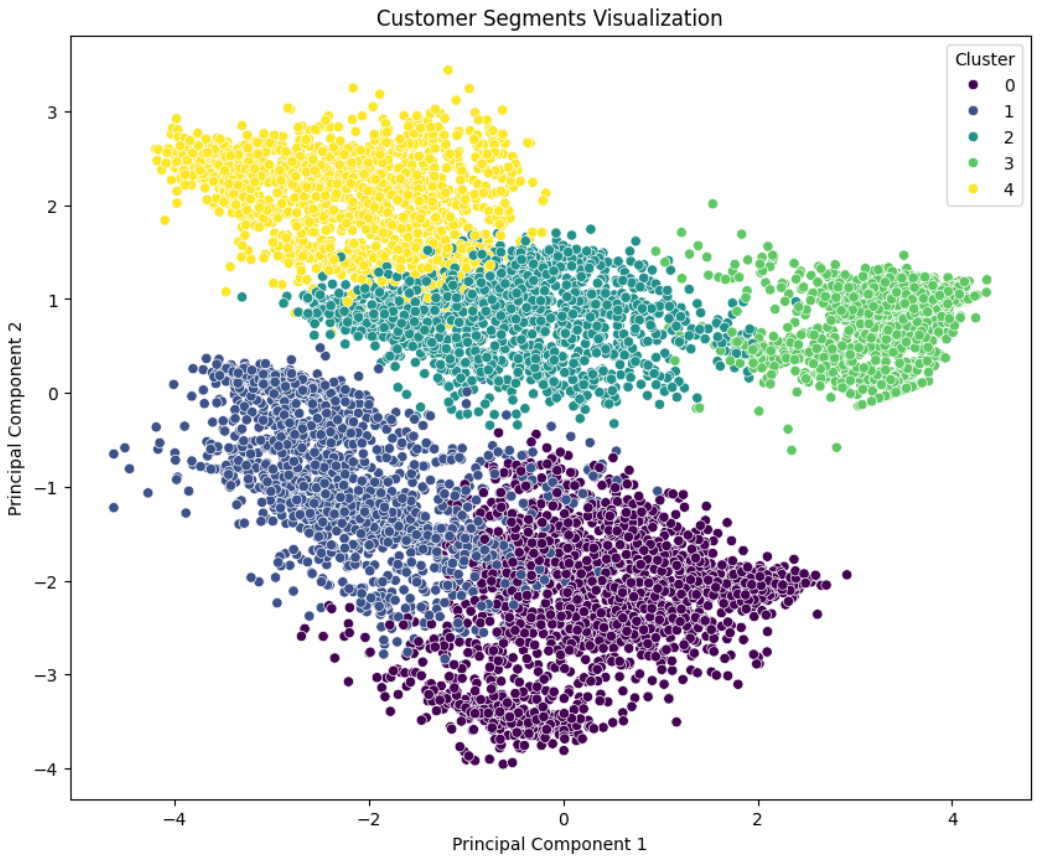
#### **Business Insight (for each visualization):**

* Silhouette plot confirmed k=5 as most stable.

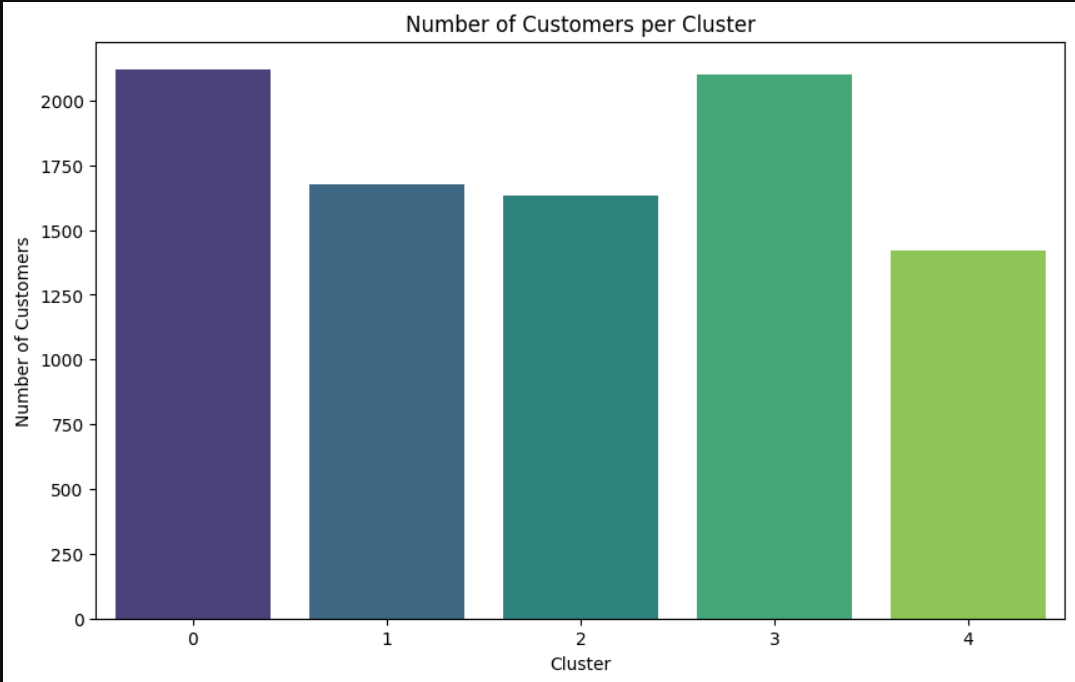




* PCA plot revealed good visual separation of clusters.



* Number of Customers per each Cluster



### **5. Business Insights and Recommendations**

#### **Customer Segment Characteristics**

1. **Cluster 0: Mixed Behavior – One-off Spenders**
   * **High:** ONEOFF\_PURCHASE\_RATIO, CREDIT\_UTILIZATION (slightly), PURCHASES\_PER\_TENURE
   * **Low:** PURCHASES\_FREQUENCY, INSTALLMENTS\_PURCHASES\_RATIO
   * **Insight:** This group tends to make larger one-off purchases rather than installment purchases, and they use a moderate portion of their credit limit.
2. **Cluster 1: Disciplined Full Payers**
   * **High:** PRC\_FULL\_PAYMENT, PURCHASES\_RATIO, INSTALLMENTS\_PURCHASES\_RATIO, PURCHASES\_FREQUENCY, PAYMENT\_RATIO
   * **Low:** CREDIT\_UTILIZATION, CASH\_ADVANCE\_RATIO, AVG\_CASH\_ADVANCE\_TRX
   * **Insight:** This group uses installment-based purchasing, pays their balances in full, and avoids cash advances, suggesting financial discipline.
3. **Cluster 2: Active Installment Shoppers**
   * **High:** PURCHASES\_FREQUENCY, INSTALLMENTS\_PURCHASES\_RATIO, PURCHASES\_PER\_TENURE
   * **Low:** PRC\_FULL\_PAYMENT, ONEOFF\_PURCHASE\_RATIO, CASH\_ADVANCE\_RATIO
   * **Insight:** These users make frequent purchases via installments but rarely pay in full. They tend to avoid one-off purchases and cash advances.
4. **Cluster 3: Heavy Cash Advance Users**
   * **High:** CASH\_ADVANCE\_RATIO, AVG\_CASH\_ADVANCE\_TRX, CREDIT\_UTILIZATION
   * **Low:** Almost all purchase-related features, especially PURCHASES\_PER\_TENURE, PURCHASES\_FREQUENCY, PRC\_FULL\_PAYMENT
   * **Insight:** This group relies heavily on cash advances, seldom makes purchases, and has low repayment behavior—potentially financially stressed users.
5. **Cluster 4: Engaged Multi-Mode Users**
   * **High:** ONEOFF\_PURCHASES\_FREQUENCY, PRC\_FULL\_PAYMENT, PURCHASES\_RATIO, PURCHASES\_PER\_TENURE, PAYMENT\_RATIO
   * **Low:** CREDIT\_UTILIZATION, AVG\_CASH\_ADVANCE\_TRX
   * **Insight:** These users engage in both installment and one-off purchases, frequently use their cards, and pay well. They maintain healthy usage and repayment habits with minimal reliance on cash advances.

#### **Segment-Specific Behavior Patterns**

* High PRC\_FULL\_PAYMENT + Low CASH\_ADVANCE\_RATIO = low risk.
* Each cluster shows distinct purchase and credit usage profiles.

#### **Opportunities and Challenges**

**Opportunities:**

* Reward loyal payers (Clusters 1, 4) with exclusive benefits.
* Convert Cluster 0 to regular users via installment promotions.

**Challenges:**

* Cluster 3 is high-risk due to heavy reliance on cash advances.

#### **Strategic Recommendations**

**Marketing:**

* **Cluster 0**: Offer usage-based discounts, promote installment plans.
* **Cluster 1**: Premium memberships and cashback.
* **Cluster 2**: Loyalty points for timely full payments.
* **Cluster 3**: Financial advice, reduce advance limits.
* **Cluster 4**: Retain with exclusive upgrades and rewards.

### 

### **Conclusion**

This project successfully segmented credit card customers using K-Means clustering with **k=5**, achieving a **silhouette score of 0.498**, which indicates moderately strong cluster separation. The analysis uncovered **five distinct customer segments**, each with unique credit usage, payment behaviors, and spending patterns. These segments can guide personalized marketing, improve customer retention, and inform product development strategies.