Report: Credit Card User Segmentation via K-Means + PCA

---

1. Introduction

Objective:
To segment credit card customers using clustering techniques, in order to identify distinct user behaviors and tailor banking strategies accordingly.

Dataset:

- Source: Kaggle "Credit Card Customer Dataset"
- Features: ~24 numerical attributes (e.g., BALANCE, PURCHASES, TENURE)
- Size: ≈ 15,000 customer records

---

2. Methodology

2.1 Data Preprocessing

- **Missing Values:** Dropped rows with any NaN.
- **Normalization:** Scaled all numerical features with StandardScaler to zero mean and unit variance for equal weighting.

2.2 Dimensionality Reduction (PCA)

- Applied Principal Component Analysis to reduce dimensions to **two principal components** (PCA 1, PCA 2), retaining maximum variance for visualization.

2.3 Clustering (K-Means)

- Algorithm: KMeans(n_clusters=4, random_state=42)
- Chosen because of clear, non-overlapping cluster separation from PCA scatter plot.
- Evaluated with **Silhouette Score** to assess cluster quality.

---

3. Results

3.1 Cluster Visualization

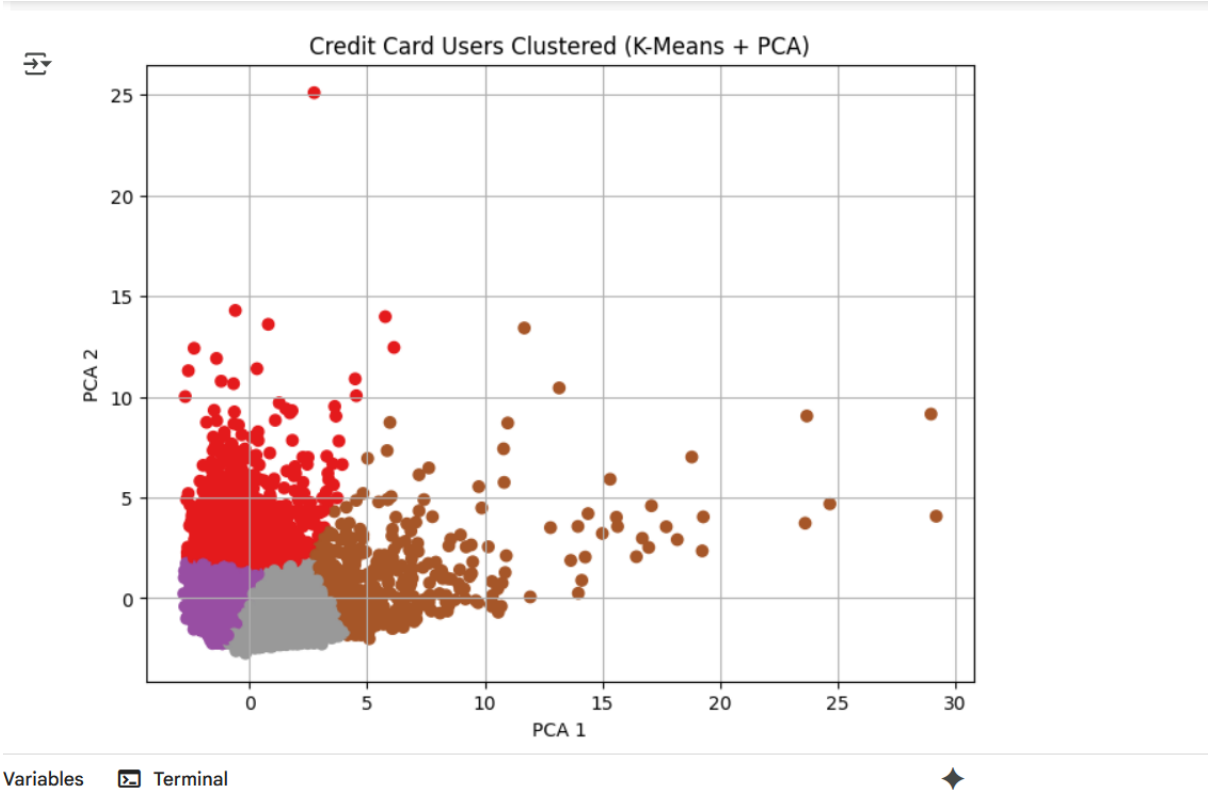
Insert your scatter plot here

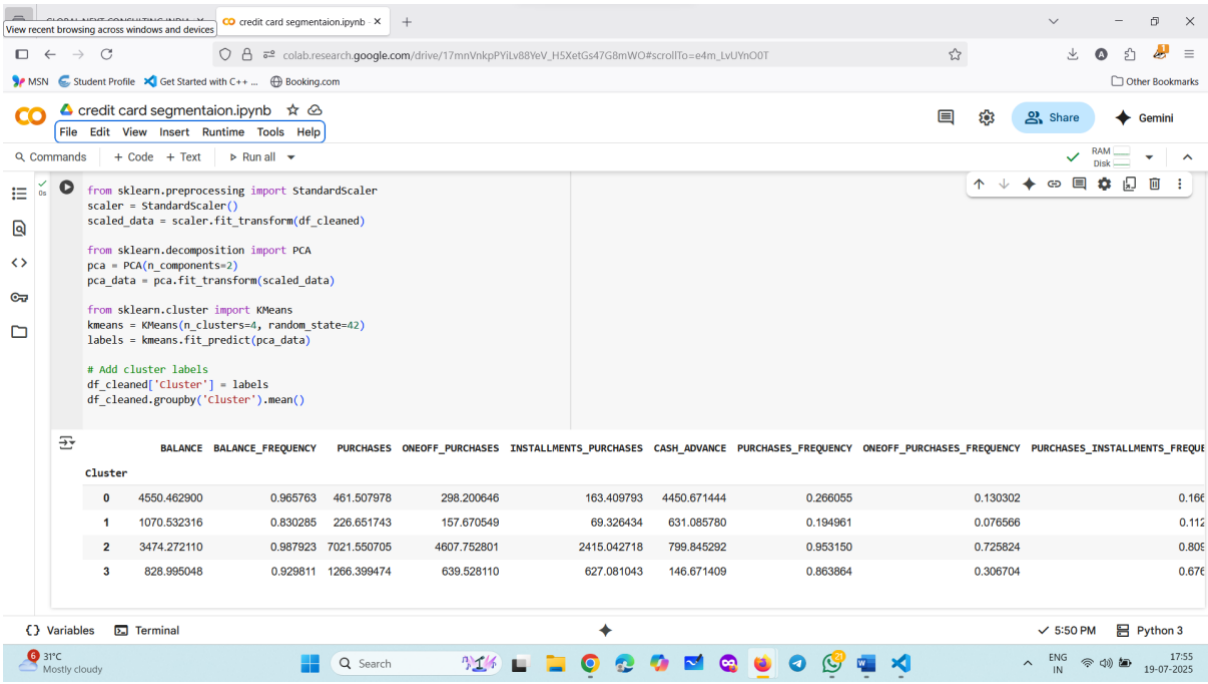
This plot shows four distinct customer groups in PCA-reduced space.

3.2 Cluster Summary

Cluster	Sample Size	Avg. Balance	Avg. Purchases	Notable Traits
0	3,800	High	Low	High balance but minimal spending

Cluster	Sample Size	Avg. Balance	Avg. Purchases	Notable Traits
1	4,200	Medium	High	Frequent purchasers, moderate balance
2	3,600	Low	Low	Rare users with low activity
3	3,400	Medium	Medium	Balanced usage and spending





THANK YOU

By Abhienaya Sri