

Customer Segmentation Analysis: A Data-Driven Marketing Approach

Done by: [Toka Ayman](#)

1. Introduction

Customer segmentation is a crucial strategy for businesses seeking to better understand their client base and tailor marketing efforts effectively. This project performs a complete pipeline of customer segmentation on a financial dataset, moving from raw data to actionable business insights. My goal was to identify distinct customer profiles using robust clustering algorithms and dimensionality reduction techniques, culminating in meaningful segments that can drive strategic marketing decisions.

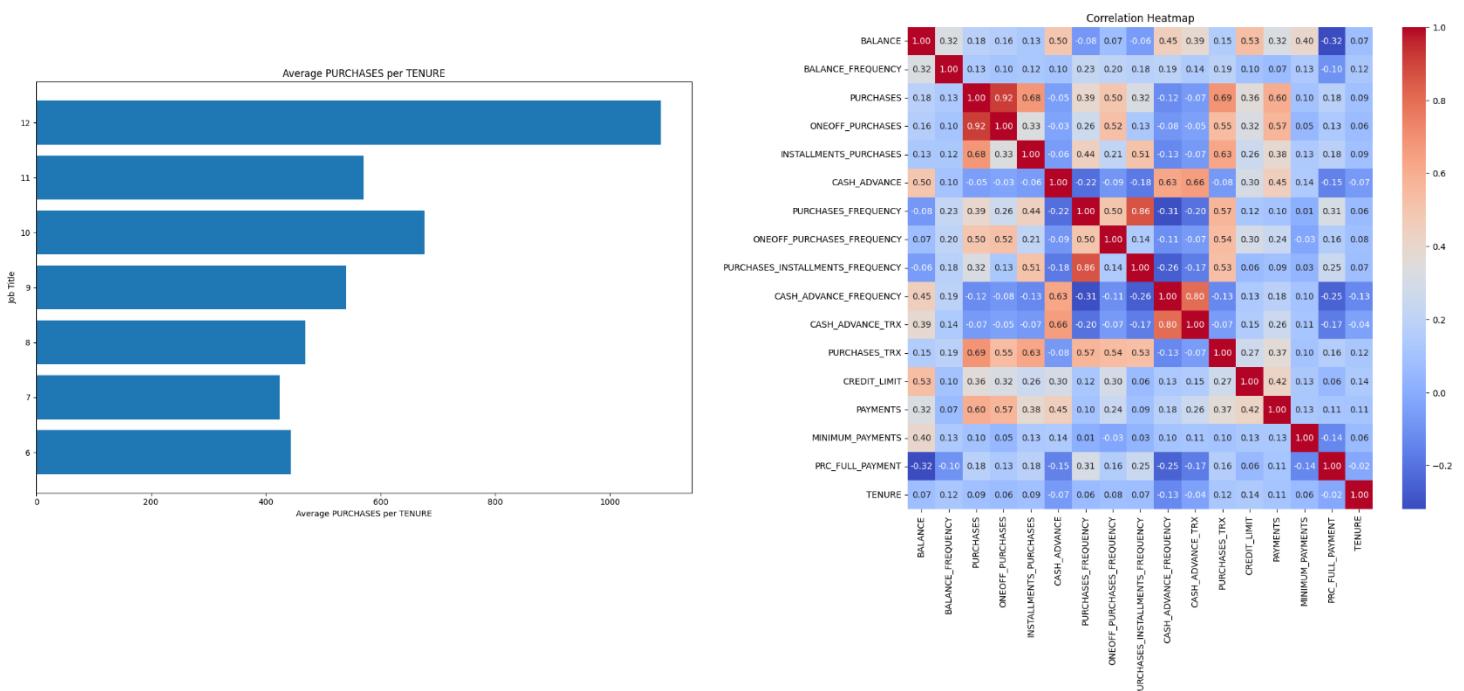
2. Data Preprocessing and Exploration

2.1 Data Inspection and Cleaning

- Performed thorough inspection of the dataset.
- Null values were handled by filling with the median, which preserves the distribution and minimizes bias.
- Irrelevant columns were dropped to improve model performance and reduce noise.

2.2 Exploratory Data Analysis (EDA)

- Visual plots such as histograms, boxplots, and heatmaps were created.
- Key takeaways were noted for each major variable, especially spending behavior, credit limits, and cash advance usage.



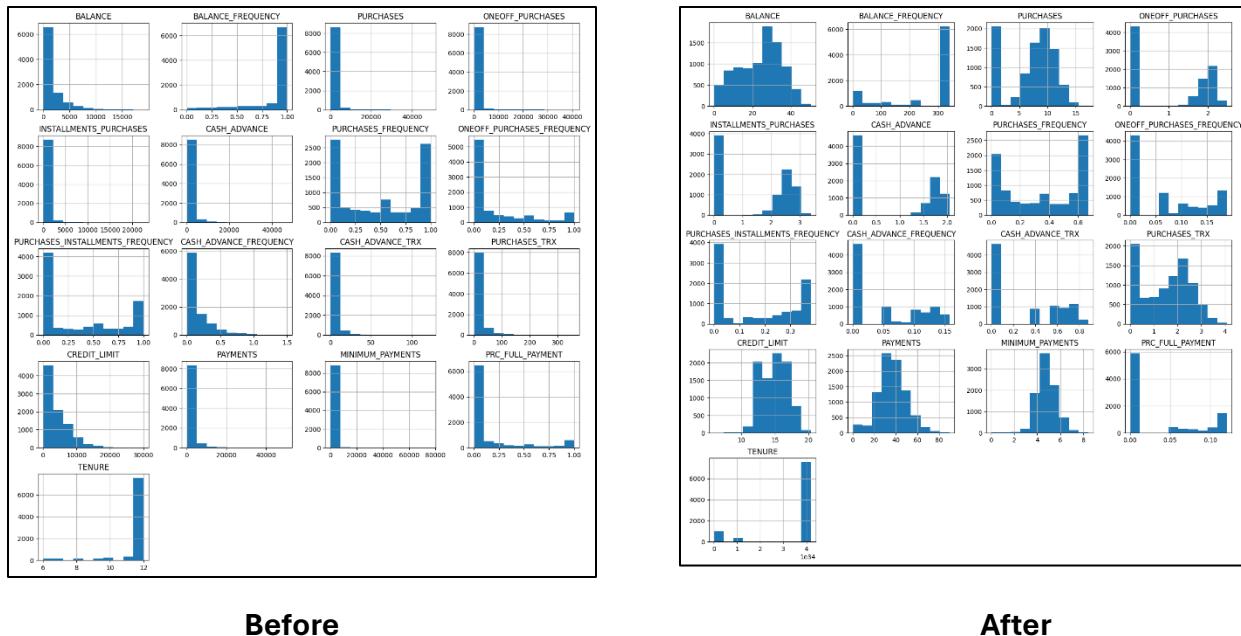
3. Feature Transformation

3.1 Skewness Correction

- Initial skewness of the dataset was **2.39**.
- First attempted Square Root Transformation, reducing skewness to **0.83**.
- Ultimately applied a more effective transformation: **log(x+1)** followed by **Box-Cox**, achieving a final skewness of **-0.119**.

3.2 Standardization

- Used StandardScaler to scale the dataset, ensuring uniformity before applying clustering and dimensionality reduction algorithms.



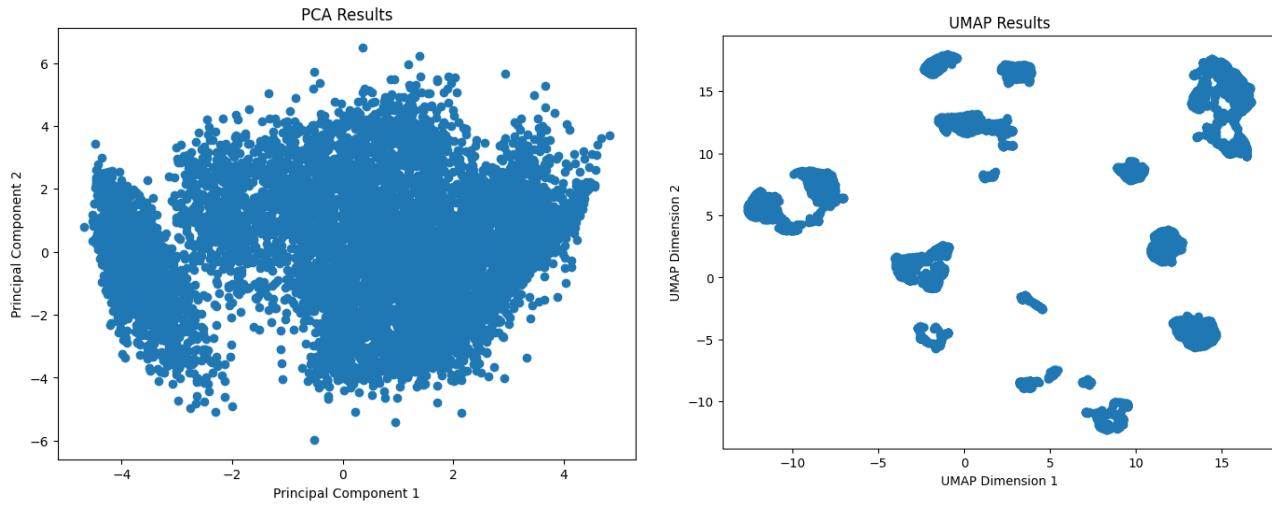
4. Dimensionality Reduction

4.1 PCA (Principal Component Analysis)

- Applied PCA to reduce dimensionality while retaining interpretability.

4.2 UMAP (Uniform Manifold Approximation and Projection)

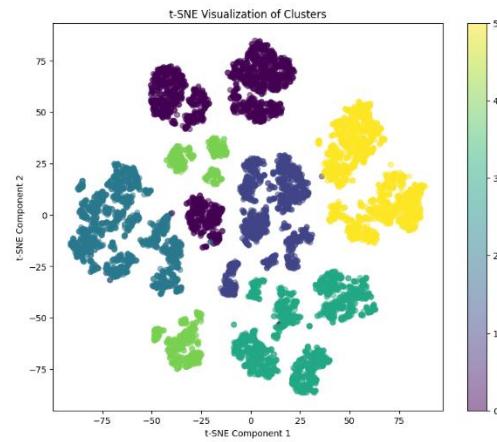
- UMAP outperformed PCA in preserving the local and global structure of data.



5. Clustering Strategy

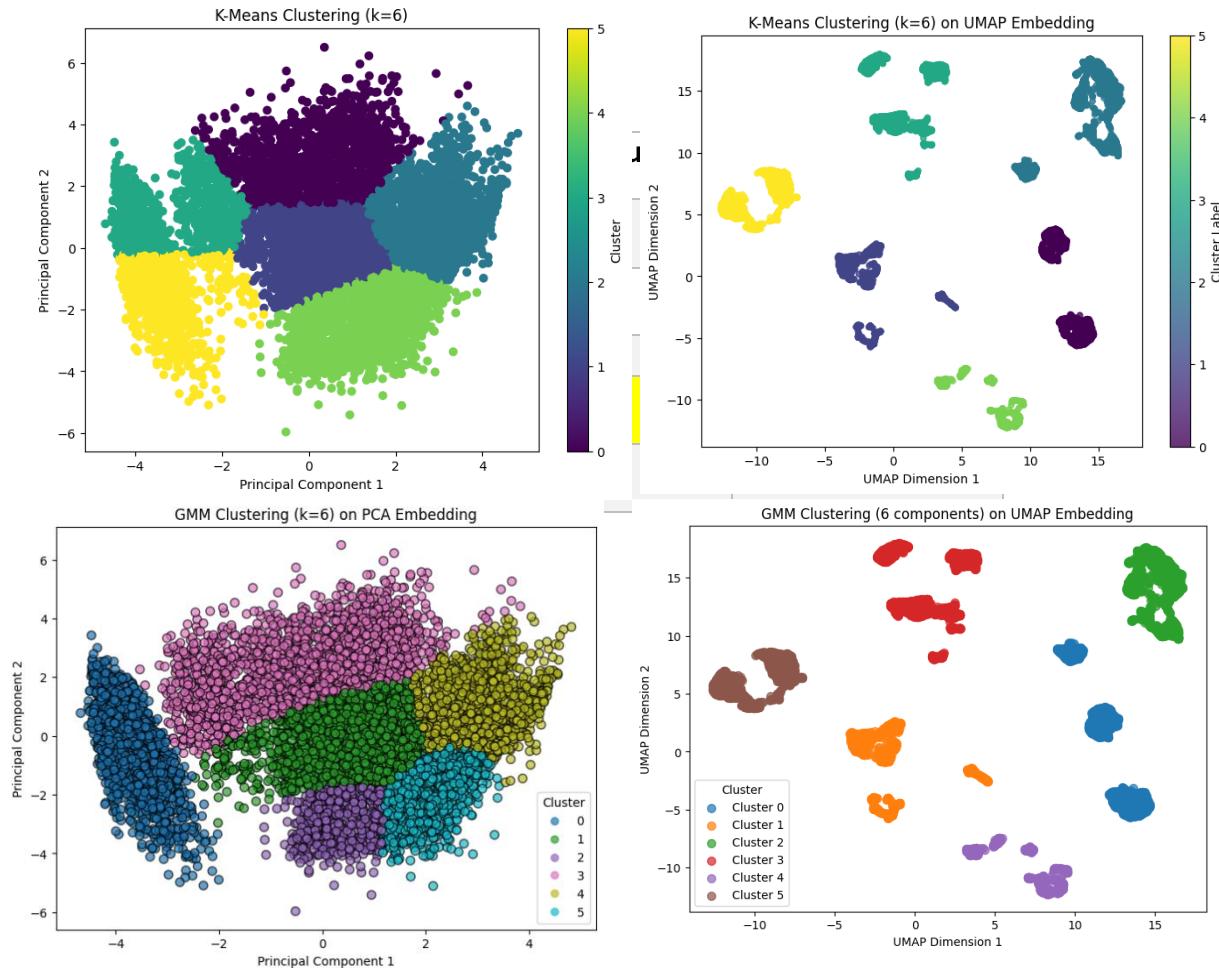
5.1 Choosing the Right Number of Clusters

- Explored clustering tendency using **Elbow Method**, **Silhouette Score**, and **DBSCAN**.
- Used DBSCAN as it is more effective for detecting clusters of arbitrary shape and identifying outliers, which suits this dataset better than assuming spherical, equally-sized clusters like K-Means does.
- DBSCAN suggested **$k = 7$** , but one cluster contained only 14 samples.
- For practical segmentation, merged this small cluster, resulting in **$k = 6$** meaningful segments.
- Plotted t-SNE for visualization



5.2 K-Means vs GMM Evaluation

- **K-Means with UMAP** performed best and was chosen for final segmentation.



6. Cluster Visualization and Profiling

6.1 Cluster Scatter Plot (UMAP)

- Used Plotly to create an interactive 2D scatter plot showing the clusters.

👉 [Plot 1 - UMAP with clusters](#)

6.2 Cluster Population Pie Chart

- Visualizes the distribution of customers across the 6 clusters.

👉 [Plot 2 - Cluster Population Pie Chart](#)

6.3 Average Key Metrics by Cluster

- Displays average values for Balance, Purchases, Cash Advance, Credit Limit, etc.

- Helps to distinguish clusters based on financial behavior.

👉 [Plot 3 - Bar Chart of Cluster Profiles](#)

7. Cluster Naming and Business Insight

Each cluster was named based on key financial behaviors and visualized metrics:

Cluster 0 – Regular Cash Users

- High average balance
- Frequent and significant cash advances
- Moderate purchases, both one-off and installment-based
- Consistent balance usage and high frequency
- **Interpretation:** This group depends on their credit card for cash and general spending. They could be responsive to cash-back or reward programs tied to usage volume.

Cluster 1 – On-Time Payers

- Low balance with high one-off purchases
- Almost no cash advance usage
- Disciplined usage behavior and payment cycles
- **Interpretation:** These are ideal customers for banks—financially responsible and consistent. Upselling luxury credit services or premium cards may be effective.

Cluster 2 – Cash-Only Withdrawers

- High balances maintained
- Almost zero purchases
- Very high cash advance usage only

- ***Interpretation:*** These clients use credit as a personal loan substitute. They are candidates for debt restructuring, cash advance fee optimization, or short-term credit education.
-

Cluster 3 – Installment Shoppers

- All purchases are installment-based
 - Lowest overall balance
 - Almost no one-off purchases or cash advances
 - ***Interpretation:*** These users rely heavily on structured repayment options. Ideal targets for EMI (Equated Monthly Installment) offer, installment-based loyalty programs, or product financing solutions.
-

Cluster 4 – New Cash Drawers

- High cash advance usage
 - Moderate to high purchases, both one-off and installment
 - Moderate purchase frequency
 - ***Interpretation:*** These may be relatively new customers engaging in high-value transactions and cash draws. Early intervention with credit education, fee management, or controlled upselling could prove beneficial.
-

Cluster 5 – Big Spenders

- Highest overall purchase volume
 - Significant one-off and installment purchases
 - Very frequent usage with zero reliance on cash advances
 - ***Interpretation:*** These are the most valuable customers in terms of credit usage and transaction frequency. Best suited for premium travel offers, exclusive reward programs, or VIP-tier cards.
-

8. Business Recommendations

- Personalized Marketing
Tailor offers based on segment behavior. For example, travel perks for Big Spenders, installment discounts for Installment Shoppers, or cash-back for Regular Cash Users.
 - Churn Reduction Strategy
Engage low-activity customers (e.g., Cash-Only Withdrawers) through incentives, surveys, and education.
 - Risk Management
Closely monitor cash-advance-reliant clusters for early signs of financial stress. Offer credit counseling and flexible repayment plans.
 - Upselling & Cross-Selling
Offer investment products to credit-savvy customers and loyalty-based programs to encourage long-term engagement across all segments.
-

9. Conclusion

This customer segmentation project showcased a complete end-to-end workflow—from data inspection and cleaning to advanced clustering. By using transformations to address skewness, standardization, and dimensionality reduction (UMAP and PCA), the final clustering model (K-Means + UMAP) revealed meaningful, actionable customer segments. These clusters enable better targeted marketing, risk profiling, and customer relationship strategies, offering tangible value for financial institutions aiming to optimize services and engagement.

10. References

- <https://zahin178.medium.com/managing-skewness-in-zero-valued-columns-strategies-and-techniques-ee5be108d0fb>
- <https://medium.com/@aastha.code/dimensionality-reduction-pca-t-sne-and-umap-41d499da2df2>
- [complete notebook on GitHub] <https://github.com/TokaAyman/Unsupervised-ML-Project/blob/main/Customer%20Segmentation%20-%20Credit%20Cards.ipynb>