

# Bank Customer Segmentation Report

 **Project Title: Understanding Bank Customers Using Clustering Techniques**

---

## 1. Project Objective

The goal of this project is to **group bank customers** based on their transaction behavior using **unsupervised machine learning**. By doing this, the bank can better understand customer needs, **offer personalized services**, and improve overall customer satisfaction. It also helps in identifying risky users and designing offers that reduce churn.

---

## 2. Dataset Overview

The dataset includes anonymous records of credit card usage for various bank customers. Key features are:

- **Spending:** Total money spent by a customer
- **Advance Payments:** Payments made in advance
- **Probability of Full Payment:** How likely the customer is to pay the full bill
- **Current Balance:** Current amount owed
- **Credit Limit:** Maximum allowed credit
- **Min Payment Amount:** Smallest amount paid
- **Max Spent in One Transaction:** Largest single shopping amount

We also created a new feature:

- **Balance Utilization Ratio** =  $\text{Current Balance} \div \text{Credit Limit}$

This helps us know if a customer is close to maxing out their credit.

---

### 3. Data Cleaning and Preprocessing

#### ✓ Missing Values & Outliers

- The dataset had **no missing values**.
- Outliers were removed using the **IQR method** to avoid misleading cluster formation.

#### ✓ Feature Scaling

- Used **StandardScaler** to bring all features to a similar scale.
  - This step is **important for clustering**, as techniques like K-Means and PCA are distance-based and get affected by unscaled features.
- 

### 4. Dimensionality Reduction

To reduce noise and simplify the dataset:

#### ◆ PCA (Principal Component Analysis)

- Reduced the dataset to fewer features while keeping **95% of original variance**.
- Helped in minimizing multicollinearity and improving model speed.


#### ◆ Autoencoder

- A **deep learning model** was used after PCA for further compression.
- We reduced features to just **5 hidden representations**, capturing deeper behavior patterns.

---

## 5. Clustering Algorithm Used

- ◆ **HDBSCAN (Hierarchical Density-Based Spatial Clustering)**
  - We projected the autoencoder output to **2D using UMAP**.
  - HDBSCAN was then applied to this projection.
  - It automatically detects the number of clusters and **ignores noisy data**.

 **Silhouette Score** (excluding noise): **0.6435**

Why HDBSCAN?

- No need to pre-set the number of clusters
  - Handles **clusters of different shapes and densities**
  - **Removes outliers** effectively
- 

## 6. Cluster Profiles

We found **4 main customer segments** (excluding noise). Here's a simple summary:

Cluster	Type	Traits
0	<b>High Spenders with Capacity</b>	High shopping spend, high credit limit, low balance utilization
1	<b>Over-Utilized Customers</b>	High balance utilization, medium payments – possibly risky
2	<b>Low Engagement Customers</b>	Low spending and advance payments – safe but not profitable
3	<b>Irregular High Spenders</b>	High single spends, low full payment probability – unpredictable

-1

Noise / Outliers

Do not follow any clear pattern – flagged for review

---

## 7. Suggested Promotional Strategies

Cluster	Recommended Strategy
0	Offer premium credit cards, cashback deals, and loyalty programs
1	Educate about credit usage, suggest EMI options or lower-limit cards
2	Engage them with tips, offer debit-based incentives, and budgeting tools
3	Give them alerts, limited-time offers, and reward-based behavior nudges
-1	Send to manual review or monitor behavior

---

## 8. Business Value and Impact

This project helps the bank in:

- **Targeting offers** to the right customers
  - **Retaining risky users** with early interventions
  - **Avoiding blanket marketing** – saving money and boosting efficiency
  - **Creating customer personas** for long-term CRM strategies
- 

## 9. Next Steps

- Try **K-Means and Hierarchical Clustering** on the same data to compare results.
- Add customer demographics like **age, income, and tenure** for deeper insights.
- Continuously update the model with **new customer data** for real-time segmentation.

---

## 10. Conclusion

With a structured ML approach using **PCA**, **Autoencoders**, **UMAP**, and **HDBSCAN**, we were able to find meaningful customer groups. These clusters reveal valuable patterns in user behavior, helping the bank improve services, **boost retention**, and **increase profitability** through intelligent decision-making.

---



### Note:

All visualizations and coding processes are included in the Jupyter Notebook file.