Credit card customer segmentation

Constructor Academy - Data Science - February 2025

Kristina Liu, Marisa Davis & Natalia Neamtu

Challenge presentation

Finding the most useful customer segmentation to improve the marketing campaigns of the company.

Overview of the data

8500 customers and their activity over a 6-month period.

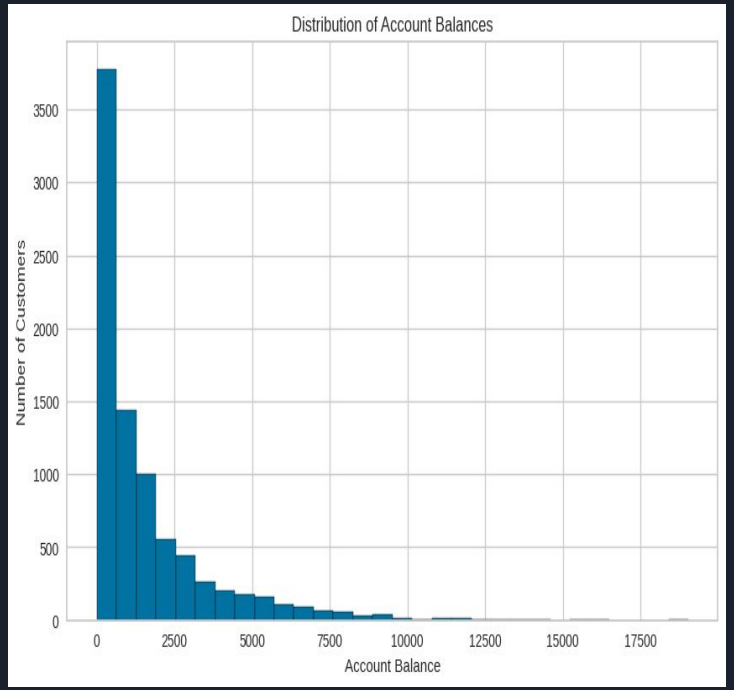
data source

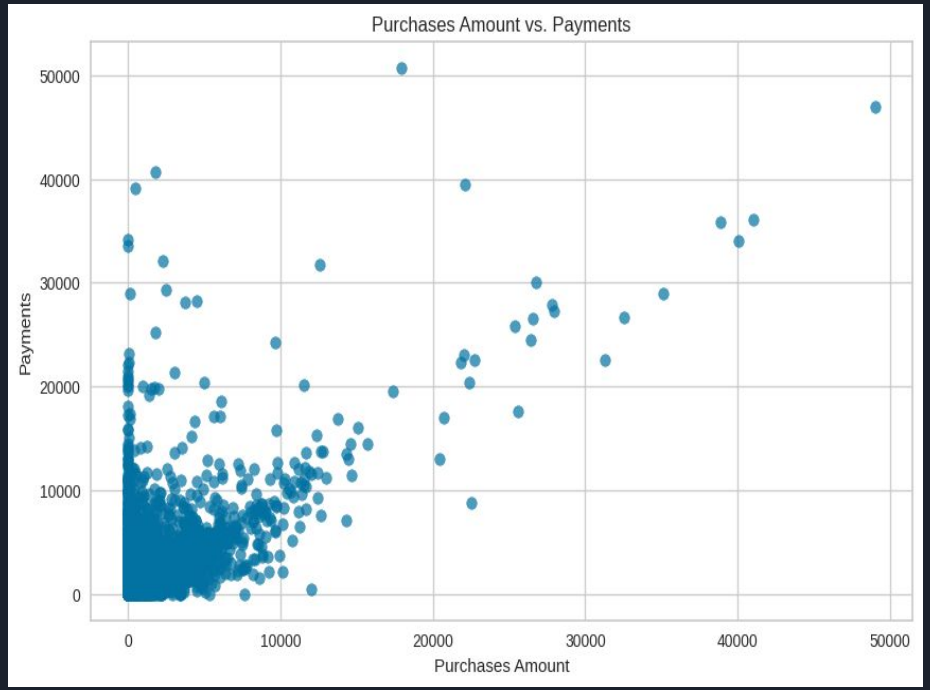
The dataset includes three types of features:

- amount-based (monetary values or quantities)
- frequency (recurring patterns or counts over time)
- number of times (total occurrences of specific events or actions).

Note: frequencies are coded with a score from 0 to 1, where 0 = not frequently and 1 = frequently.

A statistical overview





Customer Segmentation

Silhouette Method

2 segments

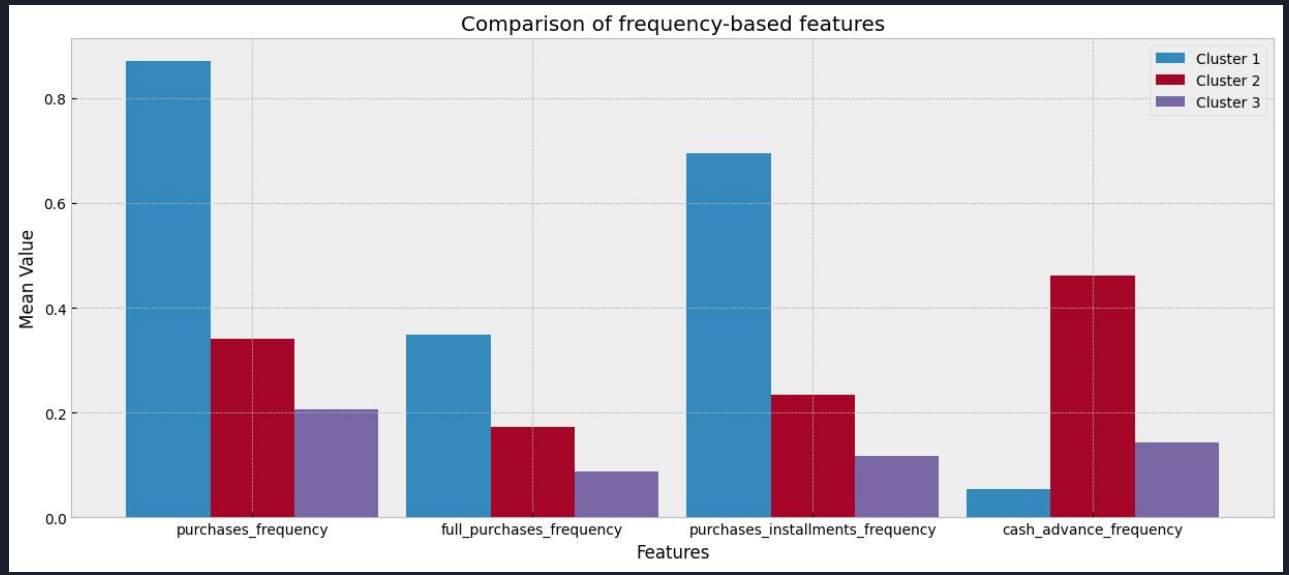
A closer look at K-means Implementation

Group	Nb of customers	Percentage of the sample
Group 1: “Affluent & Active Spenders” purchases amount 1725.74 USD balance in account: 2007.88 USD	6716	79%
Group 2: “Budget-Conscious Users” purchases amount: 292.93 USD balance in account: 1111.76USD	1784	21%

Customer segmentation in 3 groups

Cluster 1: **Frequent buyers with installments**

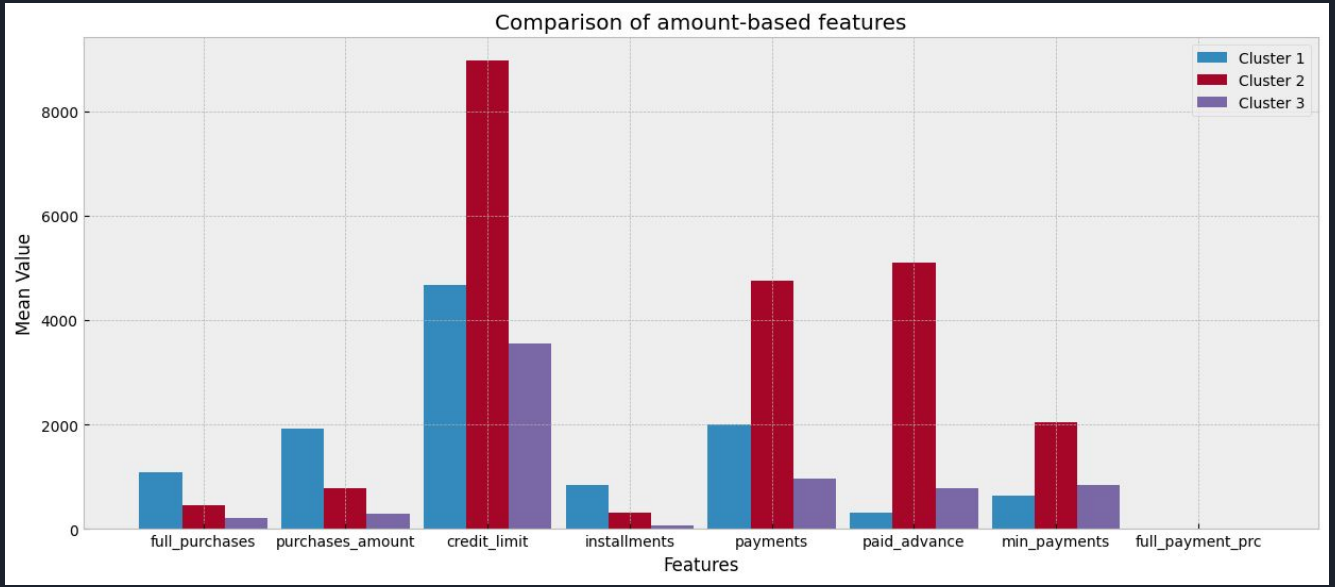
- Tend to make more purchases and prefer buying in installments spreading their payments over time rather than paying upfront.
- Less likely to use "Cash in Advance."
- Buy more frequently and in full compared to other clusters



Customer segmentation in 3 groups

Cluster 2: **Heavy “Cash-in-Advance “ users with higher payments**

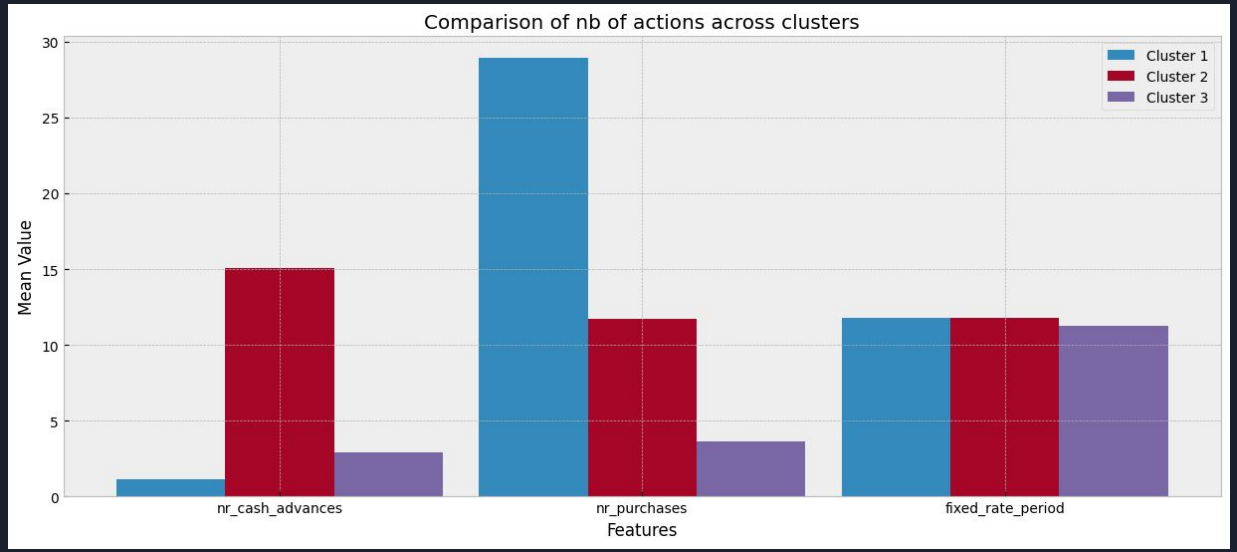
- Frequently rely on the "Cash in Advance" method.
- Have twice the credit limit of Cluster 1 and three times that of Cluster 3.
- Make higher payments overall, with a significantly higher minimum payment compared to other groups.



Customer segmentation in 3 groups

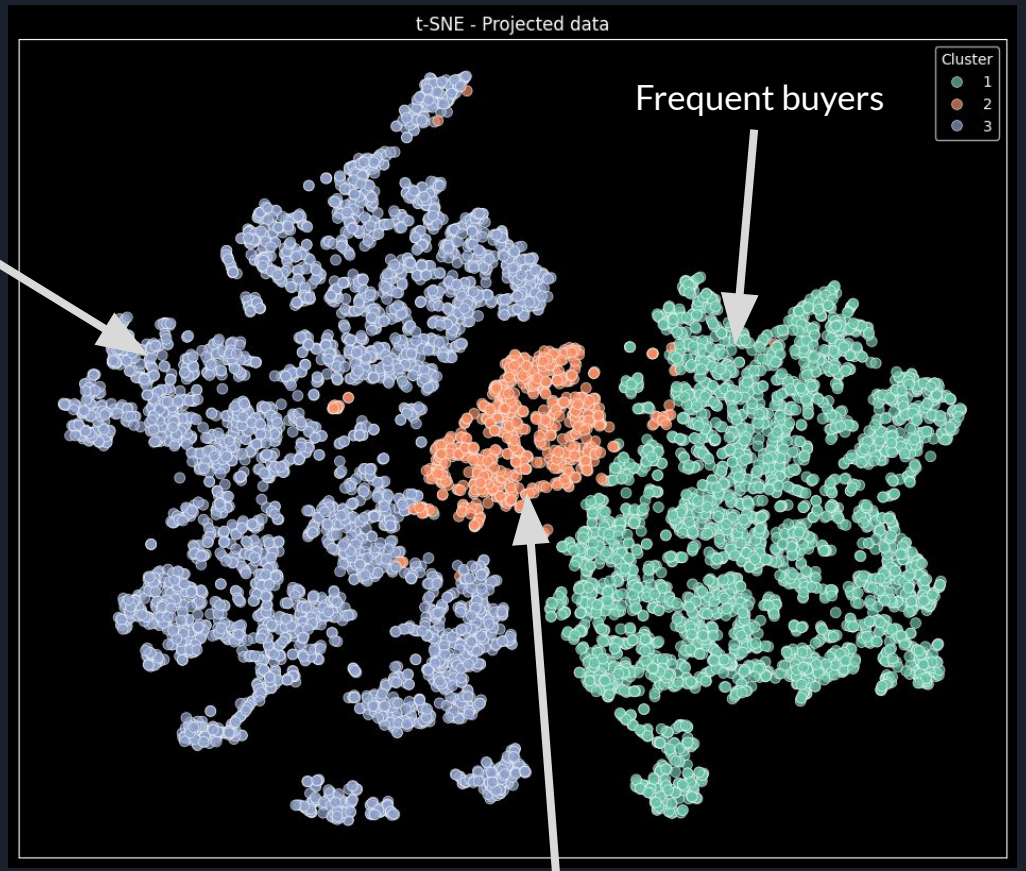
Cluster 3: **Casual, less active buyers**

- Make fewer purchases and don’t buy in large amounts.
- Less frequent use of installments or "Cash in Advance."
- No strong spending pattern—engage in a little bit of everything but without a dominant preference.



Who to target?

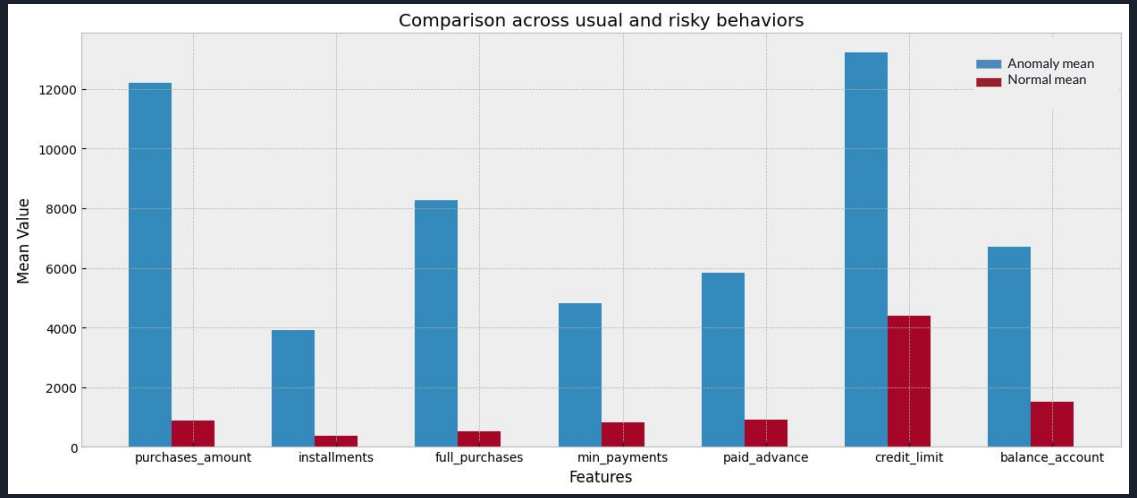
Group	Nb of customers	Percentage of customers
Group 1: Frequent buyers with installments	3474	41%
Group 2: “Cash-in-Advance” users	751	9%
Group 3: “Casual, less active buyers”	4275	50%



Why start with the casual users?

Percentage of customers that showed unusual, potentially risky, behaviors

Group 1: Frequent buyers with installments	0.56%
Group 2: “Cash-in-Advance” users	0.41%
Group 3: “Casual, less active buyers”	0.02%

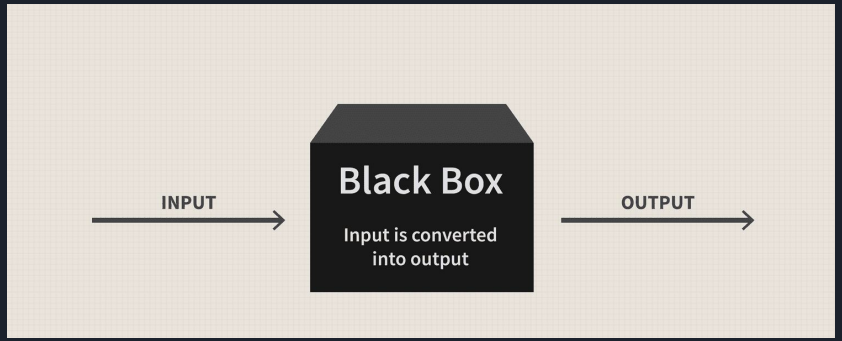


Validation of using 3 clusters

Goal: Investigate validity of clustering our data into 3 groups
Question: Unsupervised Learning can be a “black box”. We want to build trust in our clustering .
Method: DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Advantages:
Dynamic Clustering: Does not require specifying the number of clusters in advance. We can verify the best number of cluster.

*Other Advantages:
Handles Noise: Effectively identifies and separates outliers. Our data
Arbitrary Shapes: Can find clusters of any shape and size, not just spherical.
Scalable: Efficient with large datasets using spatial indexing.
Flexible Parameters: Uses density-based parameters for customization.



The Silhouette Value

The silhouette value is a metric used to measure how similar an object is to its own cluster (cohesion) compared to other clusters (separation). It ranges from -1 to +1:

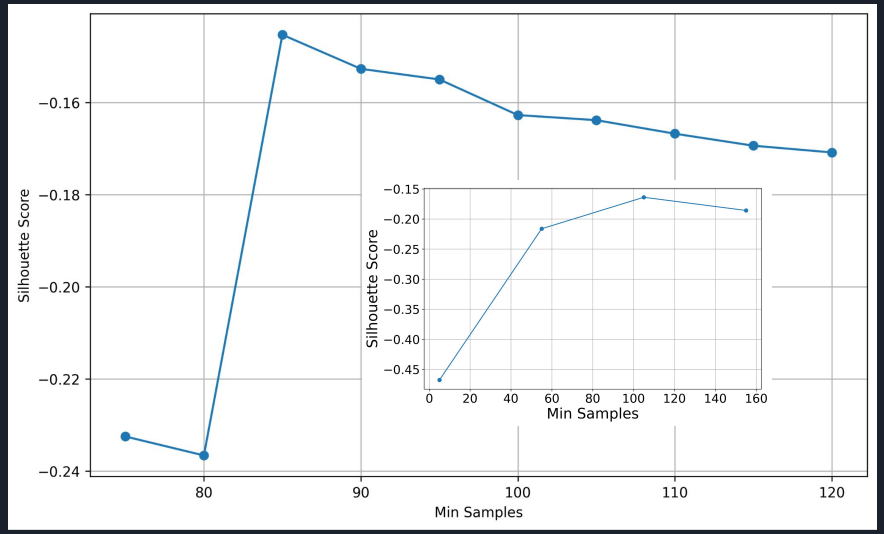
- +1: Well clustered (close to its own cluster, far from others).
- 0: On the boundary between clusters.
- -1: Misclassified (closer to a neighboring cluster).

Higher silhouette values indicate better-defined clusters, while values near 0 suggest overlapping clusters.

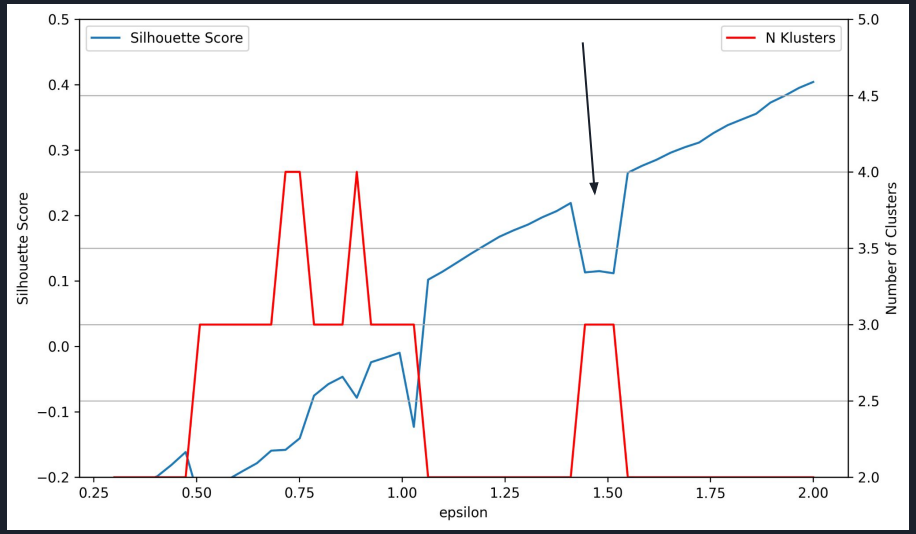
The silhouette value is useful for evaluating clustering quality and determining the optimal number of clusters. It is calculated by comparing the average distances within and between cluster

Scoring Number of Clusters

Min Samples = Density



Epsilon = Distance (the KEY parameter)



Conclusions

We identified three customer groups:

1. Frequent purchasers with installments
2. "Cash in Advance" users
3. Casual, less active users
    - representing the majority of the customer base, with lower financial risk and high market potential, making them an ideal target for a marketing campaign.

We applied three clustering methods, exploring different perspectives:

- Some approaches suggested two customer groups instead of three.
- However, validation confirmed the three-group segmentation.