Customer Segmentation with Credit Card Customers

Please designate one team member to make a copy of this document and take notes to bring back to the class discussion

Overview

In this case we will be comparing clustering methods to identify customer grouping for a credit card company.

Credit card companies typically make money from three sources:

1. A small flat fee for each transaction, for example $0.25 per transaction
2. A percentage fee for each transaction, for example 1.5% - 3.5% per transaction
3. Interest on balances not paid off at the end of the month, for example 16-20% in interest for users who carry a balance.

Our client has given us data about their current customers, and has asked your team to develop data-driven customer segments using clustering techniques. The customer segments will be used for (1) marketing to potential customers, (2) retargeting existing customers with offers, and (3) developing new products (such as a "Super Platinum Card" or "College Starter Card").

Up until now the client has relied on heuristics such as "high income mothers" (Gender, Income Category, Dependent Count) and similar categorical features. Given the data dictionary below and summary statistics, develop an analysis plan for approaching this problem.

Part 1: Analytic Approach

Discuss the pros and cons of using demographic, psychographic, or behavioral clustering with your group. Remember, your approach should consider the data available for analysis, the needs of the customer, and the analysis approaches we may use (K-Means, Hierarchical, or DBSCAN).

Approach	**Pros**	**Cons**
Demographic (gender, age)	• Easy to interpret and explain to stakeholders. • General segments match standard store personas.	• Prone to bias (actual behavior often contradicts demographic stereotypes). • Low differentiation between customers with similar profiles but different habits.
Psychographic (lifestyle)	• Excellent for creating deeply contextualized advertising copy. • Humanizes the data.	• Data is often unavailable for newer users. • Difficult to validate without external surveys or third-party data enrichment.
Behavioral (detail activity)	• Highly relevant to profitability and risk. • Reflects actual usage rather than assumed usage. • Data capture is complete for existing clients.	• Cold Start Problem: Unavailable for new users. • Dormancy: Can become irrelevant if the user is inactive for a long period.

Discussion Notes:

- Group Discussion & Selected Approach: Our group advocates for a Behavioral approach as the primary method, supported by Demographic data as a secondary "contextual" layer.
- Reasoning: In the context of credit card products, particularly in developing markets, target users are often already onboarded as bank customers (holding savings accounts or debit cards). This ensures that detailed activity data is readily available to us. While behavioral data drives the clustering, demographic data remains essential for validating offers (e.g., ensuring a "Student" offer is restricted to the correct age group).

Part 2: Hypothesis

What hypotheses does your team have about what meaningful clusters may exist in the data? These can be about customer buying patterns, drivers of credit card spending, and what you may already know about how credit card companies market their products to customers.

Discussion Notes:

- Income and spending segment (income category)
- Lifestage segment (dependent, marital status)
- Frequency cluster (transaction freq, utilization rate)
- Unit price (barang yg di beli mahal atau engga)

Core Hypothesis: We hypothesize that the data will reveal clusters primarily defined by Financial Capacity and Life Stage.

Expected Clusters:

1. Income & Spending Power: We expect a clear split between high-income luxury spenders and budget-conscious users.
2. Life Stage Groups: Distinct clusters driven by Dependent_count and Marital_Status (e.g., Single Young Professionals vs. Families with Children).
3. Engagement Level: Segments defined by Total_Trans_Ct (frequency) and Avg_Utilization_Ratio.
4. Purchasing Behavior: A distinction between users who make a few expensive purchases (High Unit Price) versus those who make frequent small purchases

Part 3: Initial Measures

Create a short list of 2-3 measures from the dataset your team would start exploring with K-Means clustering. We want to get a broad sense of the clusters of customers rather than trying to find a small niche at this point. Over time we can expand our clusters to use several features, but we can only visually explore 2-3 dimensions at a time to start out with.

Discussion Notes:

Business Goal: Our primary objective is to increase the Utilization Rate, which drives sales volume and revenue.

Proposed Strategy: "Cluster based on Life Stage, then target with offers to increase Usage." By identifying the life stage first, we can find under-utilized segments (e.g., "Families with low credit usage") and design specific product offers to encourage higher utilization.

Selected Variables for K-Means: We will begin our exploration using a combination of Life Stage and Usage variables:

1. Dependent_count (Demographic/Life Stage): To differentiate family size and household needs.
2. Marital_Status (Demographic/Life Stage): To separate single vs. married spending profiles.
3. Avg_Utilization_Ratio (Behavioral/Usage): To identify current engagement levels and room for growth.
4. Total_Trans_Amt (Behavioral/Usage): To understand the monetary value of the customer.

In our debrief I will randomly be calling on teams to give a summary of their discussion, so please be prepared to share your insights with the class.

Data Dictionary

Column	**Description**
Customer_Age	Age of the customer
Gender	Gender of the customer
Dependent_count	Number of dependents of the customer
Education_Level	Educational level of the customer
Marital_Status	Marital status of the customer
Income_Category	Income category of the customer
Card_Category	Category of the credit card held by the customer
Months_on_book	Number of months the customer has been a bank client
Total_Relationship_Count	Total number of bank products held by the customer
Months_Inactive_12_mon	Number of months with inactivity in the last 12 months
Contacts_Count_12_mon	Number of contacts with the bank in the last 12 months
Credit_Limit	Credit limit on the credit card
Total_Revolving_Bal	Total revolving balance on the credit card
Avg_Open_To_Buy	Average open to buy credit line on the credit card
Total_Amt_Chng_Q4_Q1	Change in transaction amount over the last four quarters
Total_Trans_Amt	Total transaction amount in the last 12 months
Total_Trans_Ct	Total transaction count in the last 12 months
Total_Ct_Chng_Q4_Q1	Change in transaction count over the last four quarters
Avg_Utilization_Ratio	Average utilization ratio of the credit card

Data Distributions