2025

  TAN, James Anthroi A. - 240350922

  TAN, Xiuhao - 240244326

  ZHUANG, Beibei - 240383013

IT114116 - DSAI

2025/7/10



Credit Card Customer in DSA Bank

Table of Content

[Introduction 2](#_Toc202813805)

[Data 3](#_Toc202813806)

[Setting and Utilizing GitHub 4](#_Toc202813807)

[Roles 4](#_Toc202813808)

[Product Backlog items 4](#_Toc202813809)

[Sprint Backlog items 4](#_Toc202813810)

[Evaluation 5](#_Toc202813811)

[Conclusion 7](#_Toc202813812)

# Introduction

This project was initiated by the DSAI bank, and the project is aimed at consulting the credit card utilization circumstance on customers and provide management suggestions or strategy to seek characteristics of potential customers and have innovative insight for strengthening the market share. The suggestion and strategy are based on the following considerations:

1. They like to know the expected income of customers who are not stated their income information.
2. Overview of Existing and Attrition Customer circumstance
3. The pattern of customers' credit card utilization behaviours
4. New customer labelling and credit limit prediction

# Data

The only one dataset is from DSA bank and stored in an open dataset platform Kaggle. The dataset includes customer information and transaction records. The following is the metadata table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Field | Type | Description | range | Sample |
| CLIENTNUM | String | Customer Identification | - | 768805383 |
| Customer\_Age | Integer | Customer's Age in Years | 0 to 120 | 45 |
| Dependent\_count | Integer | Number of dependents of the customer's family | - | 4 |
| Credit\_Limit | Integer | Credit Limit on the Credit Card | - | 21341 |
| Total\_Revolving\_Bal | Integer | The money has not been paid off in full by the end of the billing cycle. | - | 233 |
| Avg\_Open\_To\_Buy | Integer | the difference between the total credit limit on the card and the current balance | - | 15412 |
| Total\_Trans\_Amt | Integer | Total Transaction Amount (Last 12 months) | - | 25 |
| Total\_Trans\_Ct | Integer | Total Transaction Count (Last 12 months) | - | 5 |
| Avg\_Utilization\_Ratio | Double | The % of credit consumed in terms of credit limit | 0 to 1 | 0.1547 |

# Setting and Utilizing GitHub

Provide the link of your GitHub project and some screen capture, including treeviews and document changing capture.

**GitHub Link**: <https://github.com/XiuhaoTan/TAN_4475Project.git>

# Roles

|  |  |  |
| --- | --- | --- |
| Name | Role | Responsibilities |
| TAN James Anthroi | Serving Product Owner |  |
| TAN Xiuhao | Product Owner |  |
| ZHUANG Beibei | Development Team |  |

# Product Backlog items

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sprint** | **Task** | **Category** | **priority** | **status** |
| 1 | Uploading data to R | DATA | LOW | Under Review |
| 1 | Data Cleansing | DATA | HIGH | Under Review |
| 2 | Relationship between price and bedroom numbers | Modelling | Middle | Committed |

# Sprint Backlog items

Use the template provided and make a screen capture for all sprint backlog item record

# Evaluation

1. Provide one insight of the customer’s information with respect to different education background

From the data, most customers have a graduate education (3,128 people), followed by high school (2,013) and uneducated (1,487). This shows that many customers are well-educated, which might affect how they use credit cards.

This is useful for the bank because those people with higher education may have more stable jobs and income, so they could be potential high-value customers. While people with lower education might need different credit strategies or support.

1. Guess the Income of Unknow Income Category customers

Customers whose income category is listed as “Unknown” have an average open-to-buy amount of **8,401.52**. That value falls between the "$40K - $60K" group (**4,290.27**) and the "$60K - $80K" group (**9,603.78**), indicating these individuals likely earn somewhere in that mid-range bracket.

1. Observe relation between transaction amount and it counts and provide one insight regards to customer consumption pattern

Customers who use their card more often (higher transaction count) generally spend more overall, with transaction frequency explaining 65% of spending patterns. This means customers' total spending is driven more by how often they buy things than by how much they spend each time. For the bank, this suggests focusing on encouraging frequent card use—like rewards for small daily purchases—will likely grow revenue more than pushing bigger individual transactions. Also, tracking how often a customer uses their card can help predict their spending, spot those at risk of leaving, and set better credit limits.

1. Observe one distinguish utilization behavior in the customer dataset and raise one insight regards to customer behavior.

The most distinct pattern shows Blue cardholders utilize 29% of their credit limit—over six times more than Platinum users (4.4%)—despite having much lower credit limits. This demonstrates entry-level users heavily depend on credit for daily expenses, while premium cardholders preserve almost their entire limit for emergencies, reflecting fundamentally different financial approaches.

1. Extract and name the four observable types of customers and introduce the characteristics of each customer

We found four clear customer types:

1. **Everyday credit users** (Blue cardholders) who regularly use 25-30% of their credit limit for daily expenses and need higher limits to avoid overutilization.
2. **Emergency savers** (Platinum/Gold users) who keep over 95% of their high credit limits untouched, treating credit as safety nets for unexpected costs.
3. **Frequent spenders** who drive most revenue through high transaction volume (50+ monthly purchases), even when individual amounts are small.
4. **Ultra-careful balancers** who maintain near-zero revolving balances (<$305) and exceptional utilization under 10%, showing strong financial discipline.
5. By comparing the similar cases with existing customers, evaluate the quality of the classification model, and you may suggest some advice(s) the improvement of the model quality

The current model uses customer demographics (like age, education, and income) to predict credit utilization behavior but shows mixed results. It successfully identifies some customers with excellent utilization patterns, especially those with stable banking relationships. However, it frequently mislabels customers with poor utilization, likely because it ignores key financial behaviors like transaction frequency and balance levels—which we know heavily influence utilization. This limitation makes it unreliable for spotting high-risk customers who need intervention.

To improve the model, add key financial features (transaction counts, revolving balances) and consider transaction patterns that better reflect real spending behaviors. Also, testing different modeling approaches might yield more accurate predictions for all utilization levels, not just the excellent ones.

1. By comparing the similar cases with existing customers, evaluate the quality of the Prediction model, and you may suggest some advice(s) to improve the model quality.

The random forest model for credit limits gives weirdly inconsistent results - some new customers get super low limits ($3k) while others get very high ($24k) without clear reasons. It's because the model only considers basic demographics (age/income) and misses important spending behavior. Real banks decide limits based on how you actually use credit - your payment history, balances, and spending frequency. To improve this, we should add: 1) transaction history, 2) card type (Platinum users get higher limits), and 3) better income estimates for unknown-income customers. The error plot shows the model stabilizes after 400 trees, so the algorithm itself works fine - it just needs better input data.

# Conclusion

DSAI Bank’s customer analysis shows several useful patterns that can help improve its business. First, most of the customers have a high level of education—with over 3,000 having graduate degrees. This suggests that many users are likely to have stable incomes and could be good candidates for premium banking services. Meanwhile, those with lower education levels (like high school or no formal education) may rely more on credit, which means they might benefit from extra financial support or personalized banking help.

Another insight comes from customers whose income was marked as “Unknown.” Based on their credit usage, these customers likely earn between $60K and $80K, which puts them in the middle-income range. Understanding this helps the bank treat these users fairly when deciding on credit limits or promotional offers.

A key spending pattern revealed that how often a customer uses their card has a stronger effect on total spending than how much they spend each time. Frequent, small purchases—like coffee or groceries—add up over time. This means the bank should consider giving rewards for regular card use instead of focusing only on large transactions.

There’s also a big difference in how customers use their credit limits. For example, **Blue cardholders** use around 29% of their available credit, while **Platinum cardholders** only use 4.4%. This suggests that Blue users depend on their credit more for daily life, while premium users keep their limits for emergencies. With that in mind, the bank should consider raising limits for Blue users to reduce financial pressure and improve customer experience.

From this behavior, we can identify **four main types of customers**:

1. **Everyday Credit Users** – They often use a large part of their credit limit and may benefit from higher limits or personalized budget tools.
2. **Emergency Savers** – They keep their credit unused unless they really need it and may value extra security features or premium rewards.
3. **Frequent Spenders** – These customers make many small purchases and would respond well to cashback, points, or loyalty programs.
4. **Ultra-Careful Balancers** – They manage their credit very cautiously, keeping their balances low and showing strong financial discipline.

Looking at the bank’s classification model, it works well in some areas—especially for spotting customers with strong financial habits—but struggles to correctly identify users who may misuse credit. This seems to happen because the model relies too much on basic data like age and income, and not enough on actual behaviors like spending patterns or credit balance history. To fix this, the model should include features like how often a customer makes purchases, their card type, and how much credit they use regularly.

Similarly, the prediction model that decides credit limits has a major flaw. It gives some new users very low limits (around $3K) and others very high ones (up to $24K) without any clear reason. The issue is that it doesn’t include important data like transaction history, payment habits, or accurate income estimates. While the random forest algorithm works well on a technical level, the input data needs to be improved for better results. Adding more detailed information will make the credit limit predictions more fair and realistic.

In summary, if DSAI Bank uses these insights and takes action—like helping customers who rely heavily on credit, rewarding frequent card use, improving prediction models, and customizing offers based on customer types—it can reduce financial risk, grow customer loyalty, and see stronger business growth over the next year.