RevoBank Sales Performance Analysis

10.08.2025

KAREN AMARILLYS
RevoU FSDA Batch JUN 2025

DISCLAIMERS

1. RevoBank is a fictional company created for analytical purposes.
2. Any insights and recommendations are based on this dataset and do not reflect real-world data.

GOAL ANALYSIS

- Summarize credit card sales performance
- Create user persona based on client behavior and usage pattern

Data Cleaning

1

DATA CLEANING STEPS (card data)

DATA CLEANING STEPS

Convert to the correct data type (4a)

ORIGINAL DATA

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5599 entries, 0 to 5598
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   id                      5599 non-null   int64
 1   client_id               5599 non-null   int64
 2   card_brand              5599 non-null   object
 3   card_number             5599 non-null   int64
 4   expires                 5599 non-null   object
 5   cvv                     5599 non-null   int64
 6   credit_limit            5587 non-null   object
 7   acct_open_date          5599 non-null   object
 8   year_pin_last_changed   5599 non-null   int64
 9   days_since_last_trx     5599 non-null   int64
 10  count_nonfraud_trx_L6M  3707 non-null   float64
 11  amt_nonfraud_trx_L6M    3707 non-null   object
 12  count_fraud_trx_L6M     547 non-null    float64
 13  amt_fraud_trx_L6M       547 non-null    object
dtypes: float64(2), int64(6), object(6)
memory usage: 612.5+ KB
```

AFTER CLEANING

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5599 entries, 0 to 5598
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   id                      5599 non-null   object
 1   client_id               5599 non-null   object
 2   card_brand              5599 non-null   object
 3   card_number             5599 non-null   int64
 4   expires                 5599 non-null   datetime64[ns]
 5   cvv                     5599 non-null   int64
 6   credit_limit            5587 non-null   float64
 7   acct_open_date          5599 non-null   datetime64[ns]
 8   year_pin_last_changed   5599 non-null   int64
 9   days_since_last_trx     5599 non-null   int64
 10  count_nonfraud_trx_L6M  5599 non-null   int64
 11  amt_nonfraud_trx_L6M    5599 non-null   float64
 12  count_fraud_trx_L6M     5599 non-null   int64
 13  amt_fraud_trx_L6M       5599 non-null   float64
dtypes: datetime64[ns](2), float64(3), int64(6), object(3)
memory usage: 612.5+ KB
```

REASON TO DO THIS DATA CLEANING SET

Making sure that the data is in the right type for future processing

DATA CLEANING STEPS (card data)

<table>
<tr><th>DATA CLEANING STEPS</th><th>ORIGINAL DATA</th><th>AFTER CLEANING</th><th>REASON TO DO THIS DATA CLEANING SET</th></tr>
<tr><td>Remove irrelevant data (4b)</td><td>
<table>
<tr><th></th><th>count</th></tr>
<tr><th>card_brand</th><th></th></tr>
<tr><td>Mastercard</td><td>2826</td></tr>
<tr><td>Visa</td><td>2093</td></tr>
<tr><td>Amex</td><td>402</td></tr>
<tr><td>JCB</td><td>206</td></tr>
<tr><td>Visa</td><td>69</td></tr>
<tr><td>Jcb</td><td>3</td></tr>
</table>
</td><td>
<table>
<tr><th></th><th>count</th></tr>
<tr><th>card_brand</th><th></th></tr>
<tr><td>MASTERCARD</td><td>2826</td></tr>
<tr><td>VISA</td><td>2162</td></tr>
<tr><td>AMEX</td><td>402</td></tr>
<tr><td>JCB</td><td>209</td></tr>
</table>
</td><td>To remove unnecessary categories due to typos</td></tr>
</table>

DATA CLEANING STEPS (card data)

<table>
<tr><th>DATA CLEANING STEPS</th><th>ORIGINAL DATA</th><th>AFTER CLEANING</th><th>REASON TO DO THIS DATA CLEANING SET</th></tr>
<tr>
<td>Remove irrelevant data (4b)</td>
<td>
<table>
<tr><th></th><th>count</th></tr>
<tr><th>card_brand</th><th></th></tr>
<tr><td>Mastercard</td><td>2826</td></tr>
<tr><td>Visa</td><td>2093</td></tr>
<tr><td>Amex</td><td>402</td></tr>
<tr><td>JCB</td><td>206</td></tr>
<tr><td>Visa</td><td>69</td></tr>
<tr><td>Jcb</td><td>3</td></tr>
</table>
</td>
<td>
<table>
<tr><th></th><th>count</th></tr>
<tr><th>card_brand</th><th></th></tr>
<tr><td>MASTERCARD</td><td>2826</td></tr>
<tr><td>VISA</td><td>2162</td></tr>
<tr><td>AMEX</td><td>402</td></tr>
<tr><td>JCB</td><td>209</td></tr>
</table>
</td>
<td>To remove unnecessary categories due to typos</td>
</tr>
</table>

DATA CLEANING STEPS (card data)

<table>
<tr><th>DATA CLEANING STEPS</th><th>ORIGINAL DATA</th><th>AFTER CLEANING</th><th>REASON TO DO THIS DATA CLEANING SET</th></tr>
<tr>
<td>Treat missing values (4c)</td>
<td>isnull
<table>
<tr><td>id</td><td>0</td></tr>
<tr><td>client_id</td><td>0</td></tr>
<tr><td>card_brand</td><td>0</td></tr>
<tr><td>card_number</td><td>0</td></tr>
<tr><td>expires</td><td>0</td></tr>
<tr><td>cvv</td><td>0</td></tr>
<tr><td>credit_limit</td><td>12</td></tr>
<tr><td>acct_open_date</td><td>0</td></tr>
<tr><td>year_pin_last_changed</td><td>0</td></tr>
<tr><td>days_since_last_trx</td><td>0</td></tr>
<tr><td>count_nonfraud_trx_L6M</td><td>1892</td></tr>
<tr><td>amt_nonfraud_trx_L6M</td><td>1892</td></tr>
<tr><td>count_fraud_trx_L6M</td><td>5052</td></tr>
<tr><td>amt_fraud_trx_L6M</td><td>5052</td></tr>
</table>
</td>
<td>isnull
<table>
<tr><td>id</td><td>0</td></tr>
<tr><td>client_id</td><td>0</td></tr>
<tr><td>card_brand</td><td>0</td></tr>
<tr><td>card_number</td><td>0</td></tr>
<tr><td>expires</td><td>0</td></tr>
<tr><td>cvv</td><td>0</td></tr>
<tr><td>credit_limit</td><td>0</td></tr>
<tr><td>acct_open_date</td><td>0</td></tr>
<tr><td>year_pin_last_changed</td><td>0</td></tr>
<tr><td>days_since_last_trx</td><td>0</td></tr>
<tr><td>count_nonfraud_trx_L6M</td><td>0</td></tr>
<tr><td>amt_nonfraud_trx_L6M</td><td>0</td></tr>
<tr><td>count_fraud_trx_L6M</td><td>0</td></tr>
<tr><td>amt_fraud_trx_L6M</td><td>0</td></tr>
</table>
</td>
<td>Making sure the data can be processed:
- Removing the NaN on count and amt of nonfraud trx because assuming that there’s no transaction on the last 6 moths
- Removing the NaN on count and amt of fraud trx because assuming that there’s no fraud happened
- Removing the NaN on credit limit assuming that the credit limit is 0</td>
</tr>
</table>

DATA CLEANING STEPS (card data)

<table>
<tr><th>DATA CLEANING STEPS</th><th>ORIGINAL DATA</th><th>AFTER CLEANING</th><th>REASON TO DO THIS DATA CLEANING SET</th></tr>
<tr><td>Treat duplicates (4c)</td><td>[141] df_copy['id'].duplicated().sum()
np.int64(31)</td><td>df_copy['id'].duplicated().sum()
np.int64(0)</td><td>To avoid bias in the data process</td></tr>
<tr><td>Remove all expired cards (4d)</td><td>
<table>
<tr><th></th><th>id</th><th>client_id</th><th>card_brand</th><th>card_number</th><th>expires</th><th>cvv</th><th>\</th></tr>
<tr><td>2309</td><td>2559</td><td>165</td><td>MASTERCARD</td><td>5283475217390040</td><td>2016-08-01</td><td>149</td><td></td></tr>
<tr><td>1076</td><td>1202</td><td>590</td><td>AMEX</td><td>361343872480844</td><td>2016-09-01</td><td>837</td><td></td></tr>
<tr><td>1870</td><td>2077</td><td>1414</td><td>JCB</td><td>6669631772050778</td><td>2018-03-01</td><td>182</td><td></td></tr>
<tr><td>1868</td><td>2075</td><td>1972</td><td>AMEX</td><td>382031359848619</td><td>2018-08-01</td><td>412</td><td></td></tr>
<tr><td>4300</td><td>4744</td><td>1996</td><td>VISA</td><td>4339413642488223</td><td>2018-12-01</td><td>77</td><td></td></tr>
<tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td></td></tr>
<tr><td>5516</td><td>6092</td><td>286</td><td>VISA</td><td>4657379115764717</td><td>2031-12-01</td><td>578</td><td></td></tr>
<tr><td>3102</td><td>3423</td><td>148</td><td>MASTERCARD</td><td>5074266637803826</td><td>2031-12-01</td><td>211</td><td></td></tr>
<tr><td>2740</td><td>3029</td><td>1355</td><td>VISA</td><td>4107644655969763</td><td>2031-12-01</td><td>870</td><td></td></tr>
<tr><td>3434</td><td>3786</td><td>1290</td><td>VISA</td><td>4603076242608163</td><td>2031-12-01</td><td>377</td><td></td></tr>
<tr><td>1915</td><td>2127</td><td>1905</td><td>VISA</td><td>4646061658785929</td><td>2031-12-01</td><td>856</td><td></td></tr>
</table>
</td><td>
<table>
<tr><th></th><th>id</th><th>client_id</th><th>card_brand</th><th>card_number</th><th>expires</th><th>cvv</th><th>\</th></tr>
<tr><td>4715</td><td>5211</td><td>1133</td><td>VISA</td><td>4332848097947146</td><td>2025-05-01</td><td>247</td><td></td></tr>
<tr><td>4409</td><td>4862</td><td>13</td><td>VISA</td><td>4846797588329543</td><td>2025-05-01</td><td>143</td><td></td></tr>
<tr><td>5127</td><td>5667</td><td>839</td><td>AMEX</td><td>359230628127716</td><td>2025-05-01</td><td>198</td><td></td></tr>
<tr><td>4379</td><td>4829</td><td>36</td><td>VISA</td><td>4887792862892213</td><td>2025-06-01</td><td>31</td><td></td></tr>
<tr><td>4708</td><td>5203</td><td>907</td><td>MASTERCARD</td><td>5725072000383307</td><td>2025-06-01</td><td>353</td><td></td></tr>
<tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td></td></tr>
<tr><td>1915</td><td>2127</td><td>1905</td><td>VISA</td><td>4646061658785929</td><td>2031-12-01</td><td>856</td><td></td></tr>
<tr><td>3434</td><td>3786</td><td>1290</td><td>VISA</td><td>4603076242608163</td><td>2031-12-01</td><td>377</td><td></td></tr>
<tr><td>1229</td><td>1364</td><td>217</td><td>MASTERCARD</td><td>5161929618424764</td><td>2031-12-01</td><td>743</td><td></td></tr>
<tr><td>5516</td><td>6092</td><td>286</td><td>VISA</td><td>4657379115764717</td><td>2031-12-01</td><td>578</td><td></td></tr>
<tr><td>5275</td><td>5831</td><td>395</td><td>MASTERCARD</td><td>5128363217409052</td><td>2031-12-01</td><td>544</td><td></td></tr>
</table>
</td><td>Removing all irrelevant data</td></tr>
</table>

DATA CLEANING STEPS (user data)

<table>
<tr><th>DATA CLEANING STEPS</th><th>ORIGINAL DATA</th><th>AFTER CLEANING</th><th>REASON TO DO THIS DATA CLEANING SET</th></tr>
<tr><td>Convert to the correct data type (5a)</td><td><pre><class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 8 columns):
 #   Column   Non-Null Count Dtype
--- ------   -------------- -----
 0   id   2000 non-null   int64
 1   retirement_age   2000 non-null   int64
 2   birthdate   2000 non-null   object
 3   gender   2000 non-null   object
 4   per_capita_income 2000 non-null   object
 5   yearly_income   2000 non-null   object
 6   total_debt   2000 non-null   object
 7   credit_score   2000 non-null   int64
dtypes: int64(3), object(5)
memory usage: 125.1+ KB</pre></td><td><pre><class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 8 columns):
 #   Column   Non-Null Count Dtype
--- ------   -------------- -----
 0   id   2000 non-null   object
 1   retirement_age   2000 non-null   int64
 2   birthdate   2000 non-null   datetime64[ns]
 3   gender   2000 non-null   object
 4   per_capita_income 2000 non-null   float64
 5   yearly_income   2000 non-null   float64
 6   total_debt   2000 non-null   float64
 7   credit_score   2000 non-null   int64
dtypes: datetime64[ns](1), float64(3), int64(2), object(2)
memory usage: 125.1+ KB</pre></td><td>Making sure that the data is in the right type for future processing</td></tr>
</table>

DATA CLEANING STEPS (user data)

DATA CLEANING STEPS

Convert to the correct data type (5a)

ORIGINAL DATA

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype 
---  ------             --------------  ----- 
 0   id                 2000 non-null   int64 
 1   retirement_age     2000 non-null   int64 
 2   birthdate          2000 non-null   object
 3   gender             2000 non-null   object
 4   per_capita_income  2000 non-null   object
 5   yearly_income      2000 non-null   object
 6   total_debt         2000 non-null   object
 7   credit_score       2000 non-null   int64 
dtypes: int64(3), object(5)
memory usage: 125.1+ KB
```

AFTER CLEANING

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 11 columns):
 #   Column             Non-Null Count  Dtype         
---  ------             --------------  -----         
 0   id                 2000 non-null   object        
 1   retirement_age     2000 non-null   int64         
 2   birthdate          2000 non-null   datetime64[ns]
 3   gender             2000 non-null   object        
 4   per_capita_income  2000 non-null   float64       
 5   yearly_income      2000 non-null   float64       
 6   total_debt         2000 non-null   float64       
 7   credit_score       2000 non-null   int64         
 8   age                2000 non-null   int64         
 9   retired_flag       2000 non-null   object        
 10  DTI                2000 non-null   float64       
dtypes: datetime64[ns](1), float64(4), int64(3), object(3)
memory usage: 172.0+ KB
```

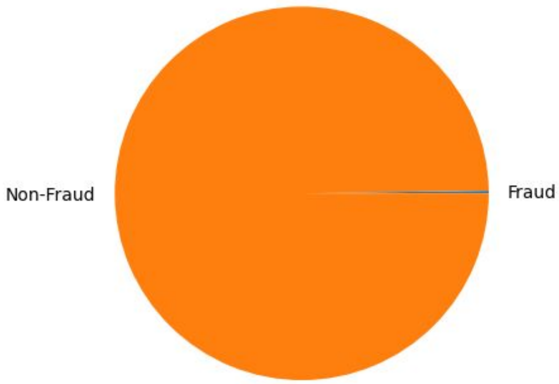
REASON TO DO THIS DATA CLEANING SET

Preparing the data for further process and analysis

Exploratory Data Analysis

2

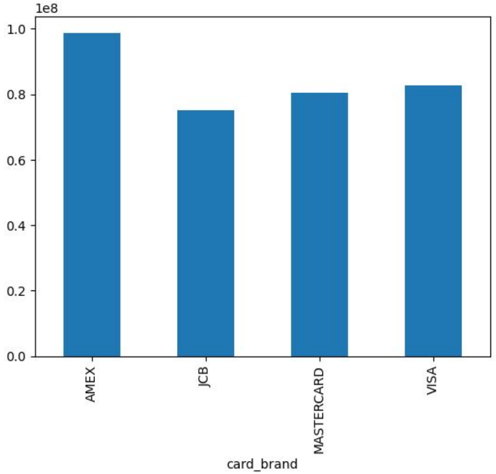
FRAUD RATE



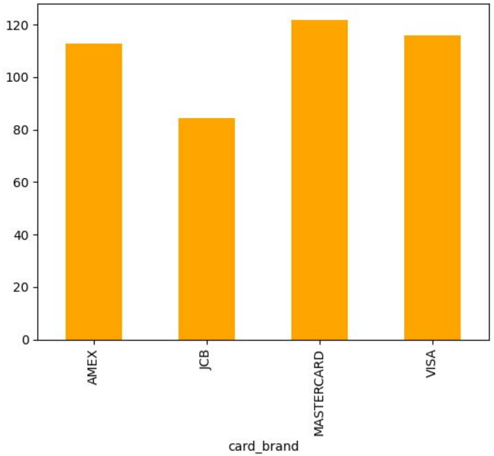
At 0.22%, the bank is well within this **safe** range (industry benchmarks suggest that a fraud rate under 1% is generally considered safe).

AMEX users are high-value spenders — marketing and premium offers could target them.

AVERAGE OF TRANSACTION AMOUNT



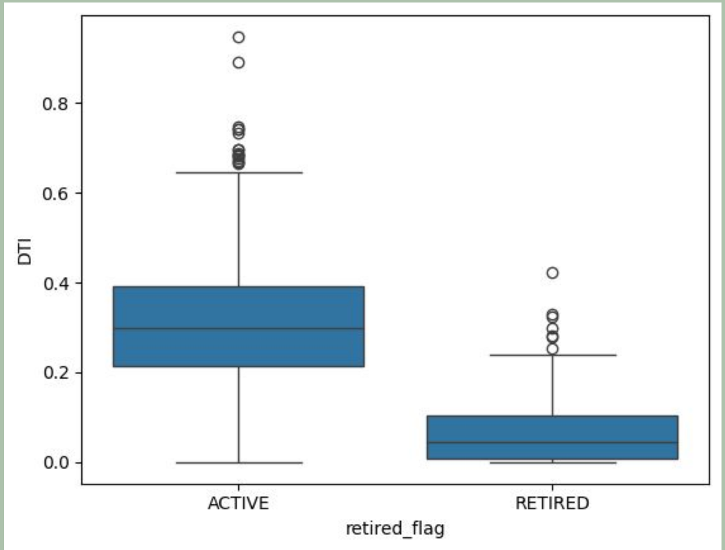
AVERAGE OF TRANSACTION COUNT



Mastercard has high engagement (more frequent usage) — good for recurring promotions or loyalty programs.

JCB's lower performance in both metrics may require targeted incentives or partnerships to boost usage.

DEBT-TO-INCOME RATIOS



- ➔ **Active (non-retired)** users have a higher and more variable DTI ratio, indicating potentially higher risk levels.
- ➔ **Retired** users have a much lower median DTI ratio and less variation, suggesting lower debt burdens relative to income.

Implication:
Contrary to some assumptions, retired users appear less risky in terms of debt management. Lending and credit limit decisions could consider this, perhaps offering tailored products to retirees with competitive terms.

DTI

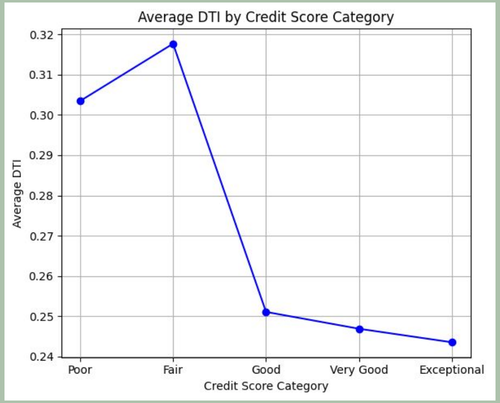
credit_score_category	avg_DTI	avg_transaction_count	avg_transaction_amount
Poor	**0.30**	**88.89**	**63,087,910.00**
Fair	**0.32**	**63.20**	**43,903,670.00**
Good	**0.25**	**60.29**	**43,230,100.00**
Very Good	**0.25**	**56.63**	**38,629,850.00**
Exceptional	**0.24**	**60.95**	**45,847,450.00**

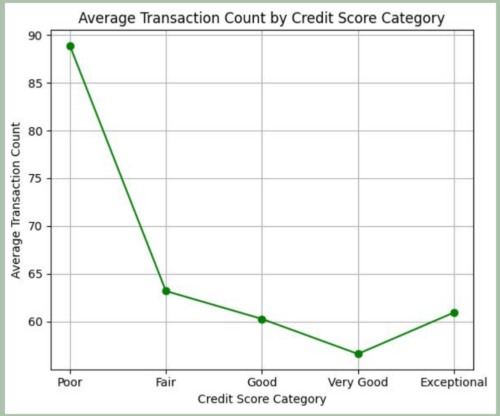
The DTI generally align with the credit scores where lower DTI is associated with higher scores.

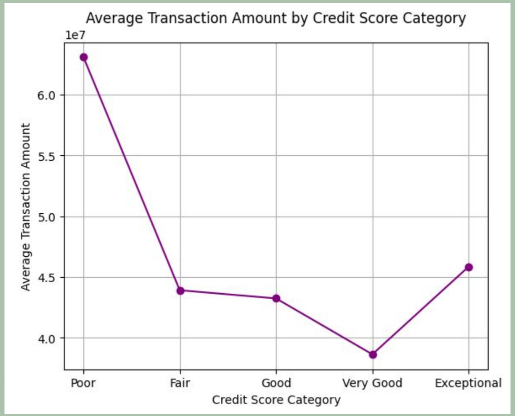
This suggests that users already committing a larger portion of their income to debt are still making frequent and high-value transactions, which increases the **risk of default**

Recommendation:

- For high-DTI users, **set stricter credit limits** or require additional risk checks before approving large transactions.
- Introduce **spending alerts** and **financial education programs** to encourage debt reduction.
- Prioritize marketing of **low-interest balance repayment plans** to this group to reduce future default risk while maintaining engagement.

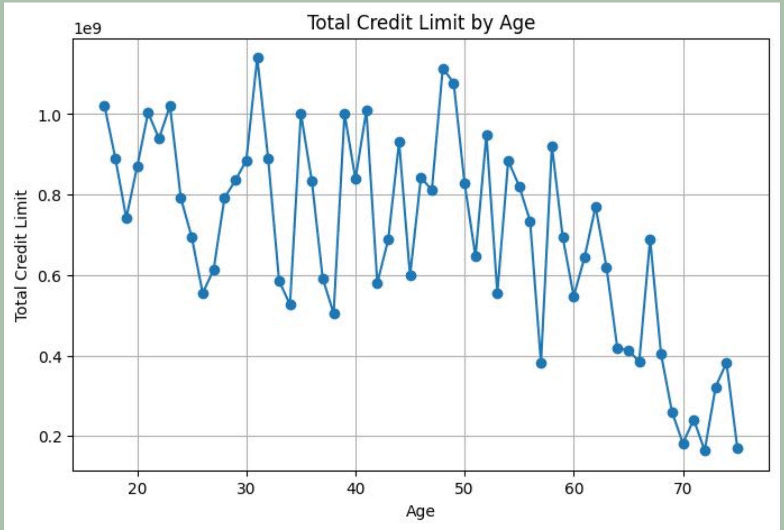






DATA VISUALIZATION

AGE VS CREDIT LIMIT



Insights:

- **Ages 20–50** tend to have higher total credit limits.
- **Above 60**, there's a clear decline in total credit limits.

Recommendations:

- **Target the 20–50 segment** with offers that encourage higher card usage, as they already have higher limits and capacity to spend.
- **Engage younger users (<20)** by offering starter credit cards with gradual limit increases, so they can grow into higher spending capacity over time.
- **For older customers (>60)**, focus on perks rather than increasing limits — e.g., travel insurance, cashback, or medical-related benefits — since their spending might not increase even with a higher limit.

Customer Segmentation

3

RECENCY

days_since_last_trx

This column measures how recently a user made a transaction.

FREQUENCY

count_nonfraud_trx_L6M

This column measures how often a user transacts.

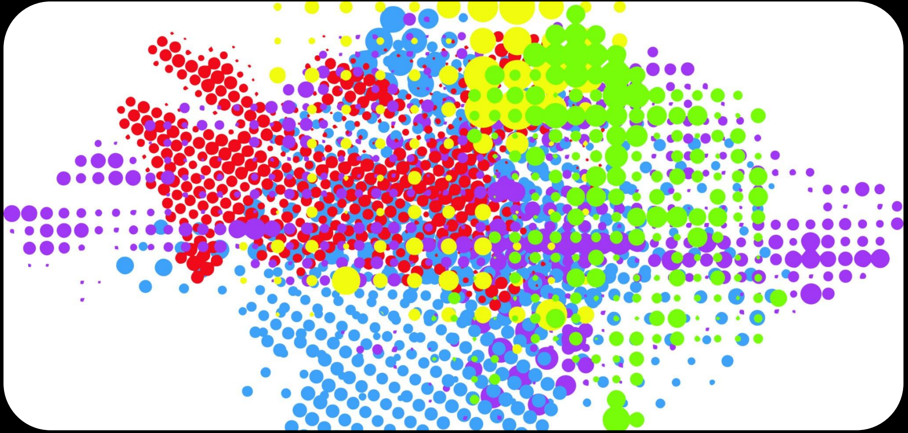
MONETARY

amt_nonfraud_trx_L6M

This column measures how much value a user brings in terms of transaction amount.

RFM METHOD

We already have **time-based, frequency, and spending variables** to be used on RFM analysis. It also gives clearer business insights and actionable segments without requiring complex statistical interpretation



Largest Segments:

- **About To Sleep (1,851 customers)** → Biggest group, indicating many dormant customers with very low activity.

- **Champions (1,080)** and **Loyal Customers (904)** → Together, these represent the bank's most **profitable and engaged base**, critical for retention.

Growth Potential:

- **Potential Loyalists (805)** → Recently active, with moderate frequency. If nurtured, they can evolve into Loyal Customers or Champions.

At Risk Groups:

- **Customers Needing Attention (418)** and **Hibernating (324)** → Customers showing early or prolonged disengagement. Retention campaigns are key.

- **At Risk (52)** → High past value but dropping activity. Should be prioritized for win-back efforts.

New / Low-Value:

- **Recent Customers (54)** and **Promising (4)** → Smallest groups. These represent **newly onboarded customers** with low transactions, needing careful monitoring and engagement.

CUSTOMER SEGMENT

segment	count
1851	About To Sleep
1080	Champions
904	Loyal Customers
805	Potential Loyalist
418	Customers Needing Attention
324	Hibernating
54	Recent Customers
52	At Risk
4	Promising

segment	recency	frequency	monetary	dti
About To Sleep	602.5353 86	0.041059	-3.83E+03	2.60E-01
At Risk	31.53846 2	159.230769	1.27E+08	2.60E-01
Champions	8.962037	289.181481	2.12E+08	2.70E-01
Customers Needing Attention	16.16267 9	110.88756	7.82E+07	2.66E-01
Hibernating	40.25617 3	29.75	1.68E+07	2.73E-01
Loyal Customers	20.82854	230.754425	1.64E+08	2.74E-01
Potential Loyalist	15.14161 5	78.648447	4.14E+07	2.50E-01
Promising	7.116279	21.116279	1.27E+07	2.94E-01
Recent Customers	15.03703 7	21.037037	1.04E+07	3.70E-01

01

The Champions clearly held the smallest number of recency, with highest frequency and monetary, however, they didn't hold the smallest DTI

These are **top customers**, very active, highly profitable. Should be nurtured with exclusive rewards and loyalty benefits.

02

At Risk spend highly but may churn, while **Loyal Customers** and **Potential Loyalists** show stable or growing potential. **Promising, Recent, and Needing Attentio**n segments are weaker and need targeted engagement to boost usage.

At Risk customers should be re-engaged to prevent churn, while **Loyal Customers** and **Potential Loyalists** deserve rewards or upgrades to strengthen loyalty. Meanwhile, **Promising, Recent,** and **Needing Attention** groups need tailored campaigns (e.g., welcome offers or reminders) to grow their engagement and increase profitability.

03

Hibernating customers had low transaction frequency, long recency, and DTI rate that's quite risky

These are **low-value, inactive customers**, unlikely to be profitable targets.

BUSINESS OPPORTUNITIES

CHAMPIONS

- ➔ 1080 users
- ➔ Highest transaction amounts and profitability, highly engaged
- ➔ **Recommendation**: Offer exclusive perks (higher credit limits, premium benefits, or loyalty rewards), personalized experiences, and early access to new products. The goal is to maintain satisfaction, prevent competitors from poaching them, and encourage advocacy, turning them into long-term brand ambassadors.

LOYAL CUSTOMERS

- ➔ 904 users
- ➔ Consistent spending and reliable engagement, stable contribution to profits.
- ➔ **Recommendation:** Strengthen engagement by offering tiered rewards, referral bonuses, and personalized financial advice to increase usage and prevent churn

POTENTIAL LOYALISTS

- ➔ 805 users
- ➔ Growing transaction activity, showing signs of future loyalty.
- ➔ **Recommendation:** nurture and encourage, targeted promotions, small credit upgrades, and incentives for higher transactions can help transition them into Loyal Customers or Champions.

KAREN.AMARILLYS@GMAIL.COM

THANK YOU

REVOU FSDA BATCH
JUN25

APPENDIX

Before Data Cleaning

	id	retirement_age	birthdate	gender	per_capita_income	yearly_income	total_debt	credit_score
0	825	66	1972-11-25	Female	Rp45.937.000	Rp93.663.000	Rp38.138.095	787
1	1746	68	1972-12-16	Female	Rp59.451.000	Rp121.212.000	Rp57.186.095	701
2	1718	67	1944-11-04	Female	Rp35.586.000	Rp52.535.000	Rp58.666	698
3	708	63	1963-01-12	Female	Rp255.975.000	Rp392.132.000	Rp60.467.238	722
4	1164	70	1982-09-21	Male	Rp84.407.000	Rp172.099.000	Rp54.946.285	675

After Data Cleaning

	id	retirement_age	birthdate	gender	per_capita_income	yearly_income	total_debt	credit_score	age	retired_flag	DTI
0	825	66	1972-11-25	Female	45937000.0	93663000.0	38138095.0	787	53	ACTIVE	0.407184
1	1746	68	1972-12-16	Female	59451000.0	121212000.0	57186095.0	701	52	ACTIVE	0.471786
2	1718	67	1944-11-04	Female	35586000.0	52535000.0	58666.0	698	81	RETIRED	0.001117
3	708	63	1963-01-12	Female	255975000.0	392132000.0	60467238.0	722	62	ACTIVE	0.154201
4	1164	70	1982-09-21	Male	84407000.0	172099000.0	54946285.0	675	43	ACTIVE	0.319271

segment	recency	frequency	monetary	dti
About To Sleep	602.535386	0.041059	-3.83E+03	2.60E-01
At Risk	31.538462	159.230769	1.27E+08	2.60E-01
Champions	8.962037	289.181481	2.12E+08	2.70E-01
Customers Needing Attention	16.162679	110.88756	7.82E+07	2.66E-01
Hibernating	40.256173	29.75	1.68E+07	2.73E-01
Loyal Customers	20.82854	230.754425	1.64E+08	2.74E-01
Potential Loyalist	15.141615	78.648447	4.14E+07	2.50E-01
Promising	7.116279	21.116279	1.27E+07	2.94E-01
Recent Customers	15.037037	21.037037	1.04E+07	3.70E-01