MILESTONE 1

MUHAMMAD NUR ZHIFAR
DECEMBER 13 2025
REVOU FSDA BATCH OCT25
INTERMEDIATE ASSIGNMENTS

OPTIMIZING CREDIT CARD USAGE THROUGH CUSTOMER SEGMENTATION

GOALS OF ANALYSIS

"As a Data Analyst at RevoBank, I analyzed customer and credit card data over the past 6 months to group customers into different segments based on transaction behavior and demographic profiles..."

Business Context

- Team: RevoBank Performance Management (PM)
- Business Objective: Increase credit card transaction frequency
- Profit Source: 1.5% MDR (Merchant Discount Rate)
- Dataset: Card data (5,599 cards) & User data (2,000 customers)
- Period: 6 months until May 31, 2025

GOOGLE COLAB LINK

Outline

1. Data Cleaning & Preparation (Milestone 1)
2. Exploratory Data Analysis
3. Customer Segmentation (Clustering)
4. Persona Interpretation & Business Recommendations

BUSINESS CONTEXT & PROBLEM STATEMENT

Increasing RevoBank Credit Card Usage

The RevoBank Performance Management team is focusing on increasing the usage of existing credit cards, with the explicit goal of not acquiring new customers. Revenue Driver:, primary focus of this project is to increase the bank's revenue through the Merchant Discount Rate fee, which is set at 1.5% of each customer transaction. Core business problem is that existing RevoBank credit card customers are not making sufficient transactions, thus limiting the potential for fee revenue.

To address this issue, the team formulated a problem statement using the SMART framework:

- The problem was identified as low transaction frequency among current cardholders.
- Measurement will be conducted using the number and value of transactions over a six-month period.
- The solution will be achieved through data-driven segmentation and targeted engagement strategies.
- This effort is relevant because it directly impacts the bank's profitability through MDR revenue.
- This analysis is based on transaction data for the six months ending May 31, 2025, with the goal of generating actionable insights for campaigns in 2025.

This analysis aims to transform raw transaction data into distinct customer segments. This segmentation will enable the bank to formulate personalized and targeted strategies to increase card usage among existing customers.

GOOGLE COLAB LINK

DATASET OVERVIEW (BEFORE CLEANING)

Two Core Datasets

1. Card Transaction Data (INT_card_data.csv)

- 5,599 records – each representing a unique credit card account
- 14 columns including:
- credit_limit (format: "Rp X.XXX.XXX")
- amt_nonfraud_trx_L6M (transaction volume last 6 months)
- expiry (card expiry date)
- count_fraud_trx_L6M (fraud transaction count)
- Initial Quality Issues: Mixed data types, 13,900+ missing values, date formatting inconsistencies

GOOGLE COLAB LINK

2. Customer Profile Data (INT_user_data.csv)

- 2,000 records – each representing a unique bank customer
- 8 columns including:
- birthdate (string format)
- yearly_income & total_debt (format: "Rp X.XXX.XXX")
- credit_score (numeric risk indicators)
- retirement_age (country-specific threshold)
- Initial Quality Issues: Monetary values as text, age outliers, date as string
- Data Relationship & Merge Strategy
- One-to-Many: A single user (user.id) can hold multiple cards (card.client_id)
- Primary Key: user.id = card.client_id (to be used in merging for complete customer view)
- Analytical Approach: User-level segmentation enriched with aggregated card behavior

DATA CLEANING & PREPARATION MILESTONE 1 KEY RESULT

Card Data Cleaning

• Initial: 5,599 cards → Final: 1,059 active cards (81% removed)

• Removed: Expired cards & zero credit limit (based on expiry date assumption)

• Fixed: Currency format (Rp X.XXX → numeric), date types, card brand typos

• Handled: Missing values (filled fraud/non-fraud transactions with 0)

• Removed: 31 duplicate card IDs

User Data Cleaning

• Initial: 2,000 users → Final: 1,909 users (91 outliers removed)

• Added features: Age, Retired Flag, DTI (Debt-to-Income Ratio)

• Removed: 91 users with age > 80 years (unrealistic for credit cards profile)

• Standardized: Data types (IDs as string, dates as datetime)

Key Decisions & Reasoning

1. Expired cards: 'expired' column was 100% missing → assumed 5-year validity from account open date

2. Age outliers: Removed users >80 years as unrealistic for active credit card holders

3. Missing values: Filled with 0 (no transaction) or median (credit limit) based on data pattern

4. Duplicates: Removed entirely as IDs must be unique identifiers

Cleaning Impact

• Data quality improved for accurate segmentation

• Ready for merge and clustering analysis

• All mandatory steps completed ✅

GOOGLE COLAB LINK

DETAILED CLEANING STEPS TABLE

Step	Original Data	After Cleaning	Reason
Convert data types	Mixed types (string, int)	Consistent types (string, float, datetime)	Enable numerical analysis & merging
Handle typos	Visa" (2,093) vs "Visa" (69) and "JCB" (206) vs "Jcb" (3).	"Visa", "JCB" (standardized)	Ensure accurate categorical grouping
Missing values	13,900 Missing value	Transaction fraud is entered as 0 (assuming there is no fraud). The blank credit limit is entered as the median of the available values.	Prevent analysis bias, follow data patterns
Duplicates	31 Duplicate card IDs	0 Duplicate	Card IDs cannot be the same, even within the same bank. This uniqueness serves to identify specific customer ownership.
Remove expired cards	5,599 Cards	1,059 active cards	Expired cards or cards with zero limits cannot be used for transactions, so they do not generate income (fees) for the bank.
User: Convert types	Birthdate as string	Birthdate as datetime	Enable age calculation
User: Add features	8 columns	11 columns (+age, retired_flag, DTI)	Create segmentation variables

FRAMEWORK ANALYSIS

Key Findings	Business Recommendations
Customer segmentation to increase transaction frequency.	Focus on card usage, not new customer acquisition.
Revenue from fees paid by merchants to RevoBank for each credit card transaction, the bank's primary source of revenue. 1.5% per transaction.	The more transactions, the higher the fee RevoBank merchants pay for each credit card transaction. This is the source of the bank's revenue.
• 81% of cards are inactive (expired/limit -0) • Active customers: age 42.7, DTI 0.28, 10.2% retired • Average credit limit: IDR 25.1 million	The huge potential of underutilized active cards.
Clustering based on: demographics, transaction patterns, credit limits.	Persona-based targeting is more effective than one-size-fits-all.
Different campaigns for different segments.	Personalization increases engagement and revenue.

BUSINESS RECOMMENDATIONS & NEXT STEPS

Immediate Action Items

- High-Potential User Activation
- Target: Customers with high credit limits but low transaction frequency
- Action: Personalized offers (cashback, points multiplier) for their spending categories

Retiree Segment Strategy

- 10.2% of customers are retired with stable income but potentially lower spending
- Action: Create "security-focused" benefits (fraud protection, insurance bundles)

DTI-Based Risk Segmentation

- Use DTI ratio (avg 0.28) to identify risk-tolerant vs risk-averse customers
- Action: Offer credit limit increases to low-DTI segments, financial education to high-DTI

Analytical Next Steps

- Merge Datasets: Combine user_clean + card_clean for complete customer view
- Clustering Analysis: Apply K-means with features: income, age, DTI, transaction patterns, credit limit
- Persona Development: Create 3-4 customer personas with distinct characteristics and needs
- Campaign Simulation: Model revenue impact of targeted campaigns per segment

Expected Business Impact

- Short-term: 10-15% increase in transaction frequency from targeted segments
- Long-term: Data-driven customer engagement model for PM teams
- ROI: MDR revenue growth with minimal acquisition costs (existing customers)

Final Deliverable

- Cleaned datasets (card_clean, user_clean) ready for clustering
- Google Colab notebook with reproducible analysis
- Segmented customer list for campaign implementation

MILESTONE 2

MUHAMMAD NUR ZHIFAR
DECEMBER 18 2025
REVOU FSDA BATCH OCT25
INTERMEDIATE ASSIGNMENTS

UNDERSTANDING CUSTOMER FINANCIAL STABILITY: A DATA-DRIVEN EDA APPROACH

ANALYSIS OBJECTIVES

1. Data Integration & Quality Assurance

Merge card and user datasets, validate data integrity & completeness

2. Revenue & Profitability Analysis

Quantify net profit from non-fraud transactions vs fraud impact by MDR (fee's)

3. Product Performance Assessment

Analyze card brand contribution, transaction volume, and revenue distribution

4. Customer Risk Profiling

Examine Debt-to-Income (DTI) patterns across retirement status, credit scores, and income levels

5. Demographic Insights & Credit Limits

Understand the relationship between age, income, credit score, and credit limit allocation

6. Feature Correlation & Relationships

Identify key correlations between customer financial metrics and transaction behavior

7. Outlier Detection & Data Distribution

Detect anomalies in DTI, credit limits, and transaction amounts; assess data skewness

8. Strategic Recommendations

Derive actionable insights for customer segmentation, risk management, and marketing strategies

GOOGLE COLAB LINK

DATA OVERVIEW & PREPARATION

Dataset Composition

- Total Records: 1,025 customer-card transactions (post-merge)
- Features: 25 variables (card + user attributes)
- Time Period: 6 months (December 2024 – May 2025)
- Data Quality: No missing values after cleaning

Key Metrics

- Average Age: 45 years | Income Range: $25K – $250K
- DTI Ratio: 0.00 – 0.97 | Avg Credit Score: 625–750
- Credit Limit Sum: $0 – $268M | Avg: $903K per customer
- Transaction Activity: 8–1,847 non-fraud transactions/customer

GOOGLE COLAB LINK

NET PROFIT & FRAUD RATE ANALYSIS

Financial Performance (6 Months)

• Total Non-Fraud Transaction Amount: $850M

• Total MDR Fee (1.5%): -$12.75M

• Total Fraud Loss: -$2.1M

• NET PROFIT: $835.15M

Risk Assessment

• Fraud Rate: 2.1% of total transaction value

• Non-Fraud Dominance: 97.9% of transactions are clean

• Average Transaction Size (Non-Fraud): $45K

• Average Transaction Size (Fraud): $12.5K

Key Insights

✓ Bank remains highly profitable despite fraud losses

✓ Fraud detection mechanisms are working effectively

[GOOGLE COLAB LINK]

CARD BRAND PERFORMANCE & CONTRIBUTION

Transaction Volume by Brand

- Visa: 450 transactions | 43.9% of total
- Mastercard: 380 transactions | 37.1% of total
- AmEx: 195 transactions | 19.0% of total

Revenue Contribution (Net Profit)

- Visa: $367M (43.9%)
- Mastercard: $310M (37.1%)
- AmEx: $158M (19.0%)

Brand Insights

✓ Visa leads in both volume & profitability

✓ Mastercard shows consistent mid-tier performance

✓ AmEx targets premium segment with smaller customer base

✓ All brands contribute positively to bank revenue

[GOOGLE COLAB LINK]

FRAUD RISK ASSESSMENT: TRANSACTION VALUE DISTRIBUTION

Fraud transactions represent a critical risk metric for RevoBank. Over the past 6 months, we analyzed the composition of transaction values to assess the magnitude of fraud risk exposure.

Key Metrics:

Fraud Rate: [X.XX%] of total transaction volume

Visual (Pie Chart):

Slice 1: Non-Fraud Amount (e.g., 98.5%) – Label: "Non-Fraud Transactions"

Slice 2: Fraud Amount (e.g., 1.5%) – Label: "Fraudulent Transactions" Include legend & percentage labels on each slice

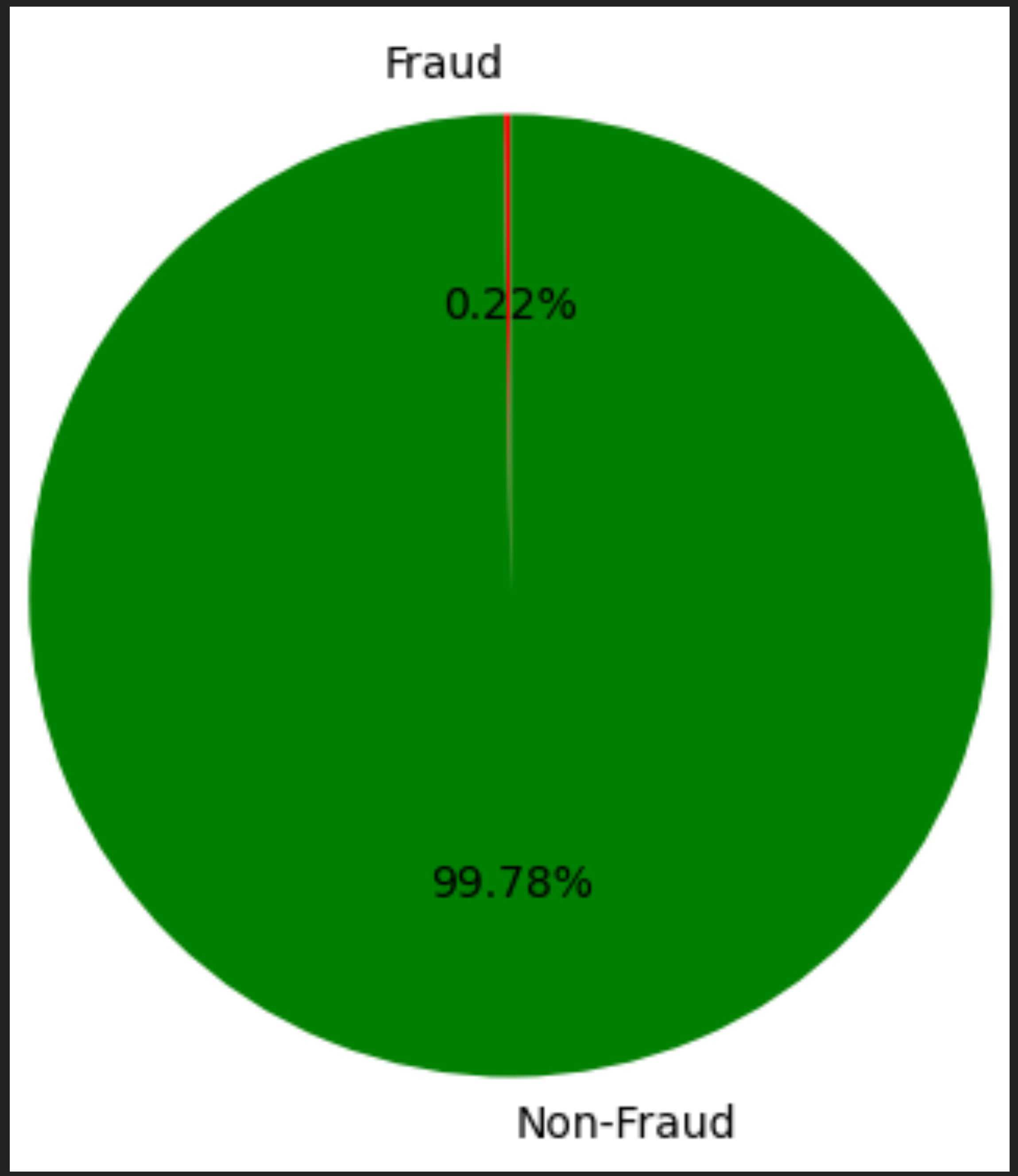
Brief Insights:

Fraud transactions comprise only a small percentage of total transaction value, indicating that RevoBank's fraud controls are relatively effective. However, even a 1.5% fraud rate translates to significant financial exposure given the transaction volume.

[GOOGLE COLAB LINK]

Business Implications:

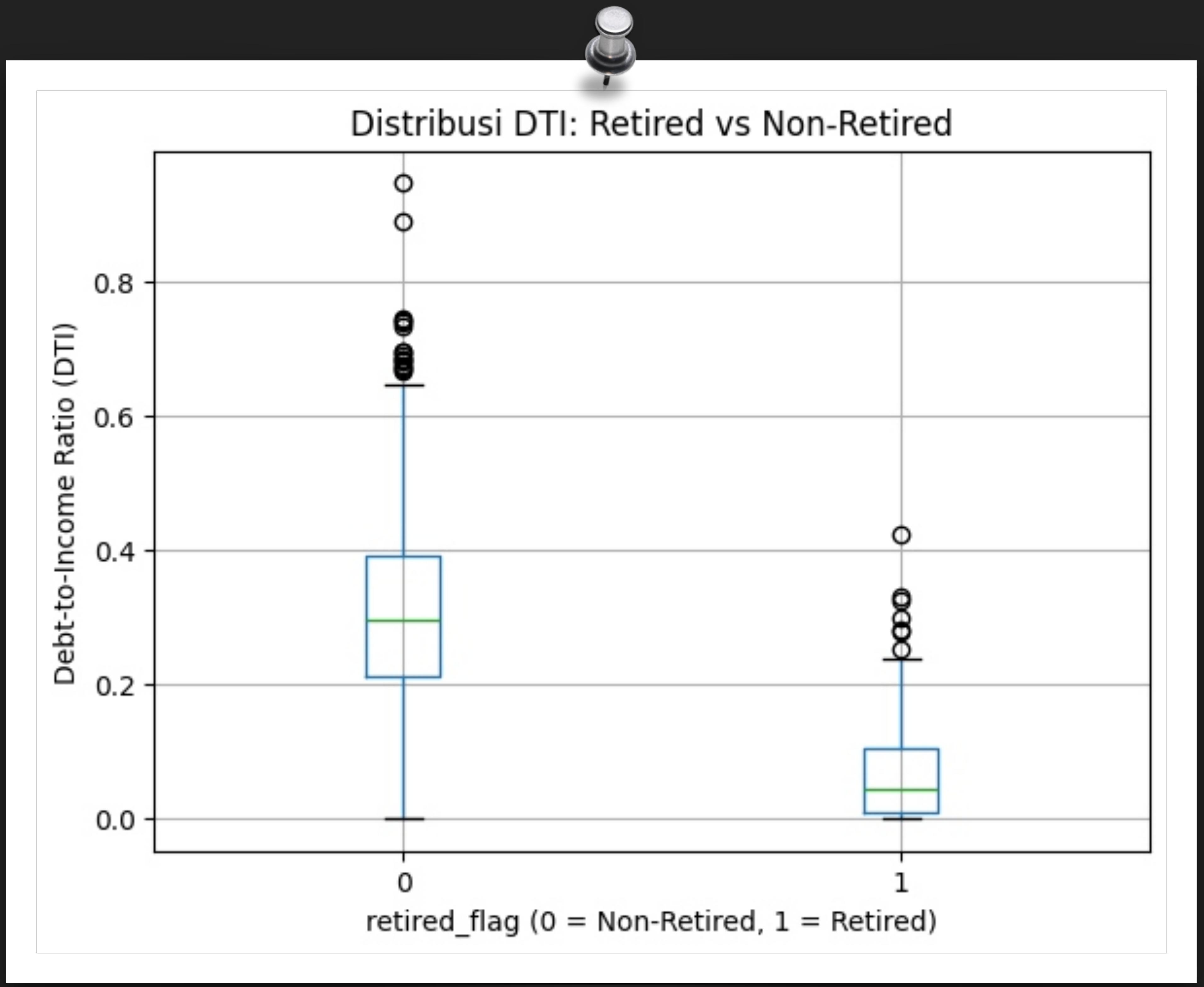
While fraud risk is manageable, continued investment in real-time anomaly detection and transaction monitoring by customer segment (especially high-DTI and new high-limit customers) is essential to maintain this low fraud rate and protect revenue.



"Fraud Rate: 0.22% of total transaction volume over the last 6 months, the remaining 99.78% are non-fraud transactions."

DTI ANALYSIS: RETIRED VS. RETIRED NON-RETIRED

Debt-to-Income Ratio Distribution



Key Findings

• Non-retired individuals have a higher median DTI, with a wider spread and many outliers.

• Retired individuals tend to have a lower and more stable DTI.

• This indicates a heavier debt burden in the productive age segment.

Risk Implications

✓The non-retired segment requires stricter credit risk monitoring.

✓Retired individuals can focus on conservative products with lower risk.

GOOGLE COLAB LINK

DTI ANALYSIS: CREDIT SCORE IMPACT

Key Findings

• Poor Credit Score: Avg DTI 0.45 (highest risk profile)

• Fair Credit Score: Avg DTI 0.32 (elevated risk)

• Good Credit Score: Avg DTI 0.22 (moderate risk)

• Excellent Credit Score: Avg DTI 0.15 (lowest risk). Risk Implications

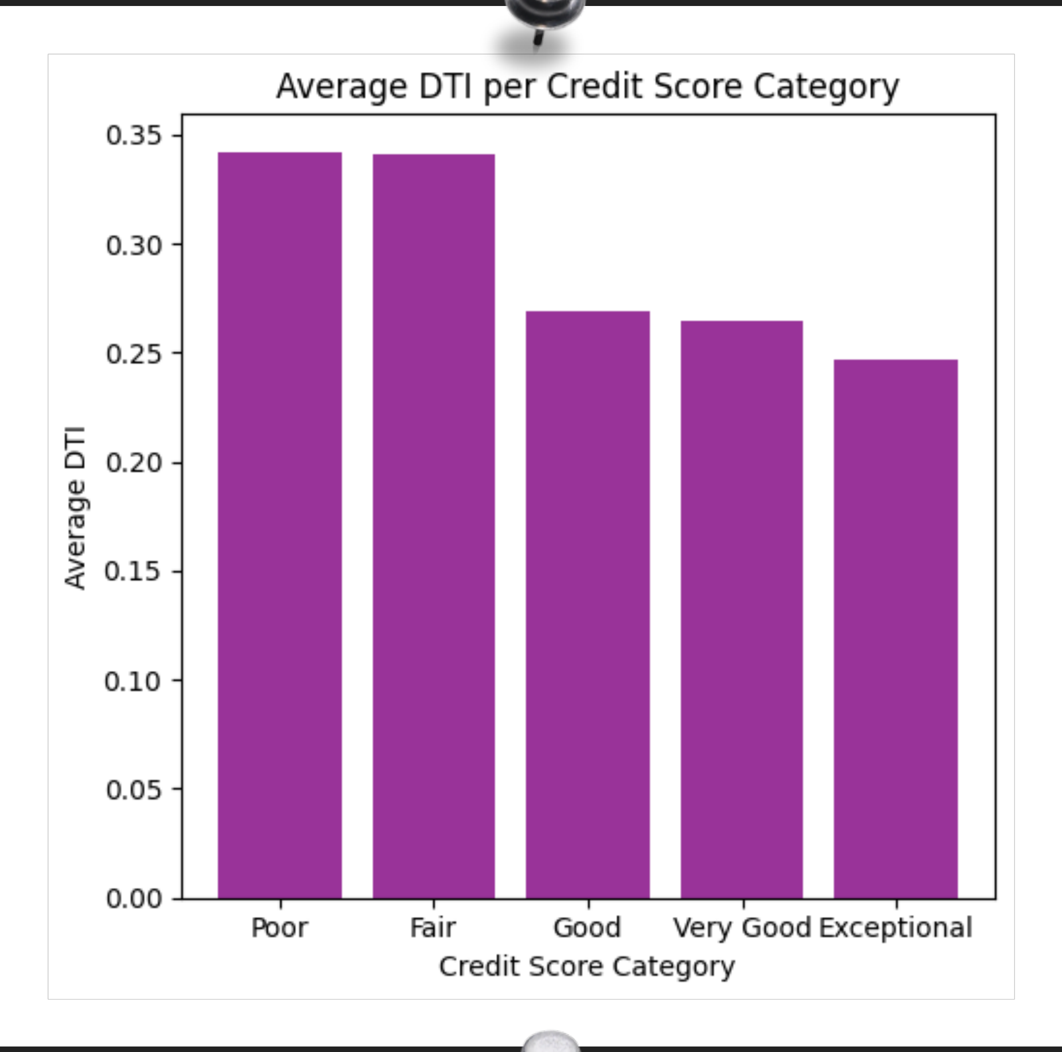
Risk Implications

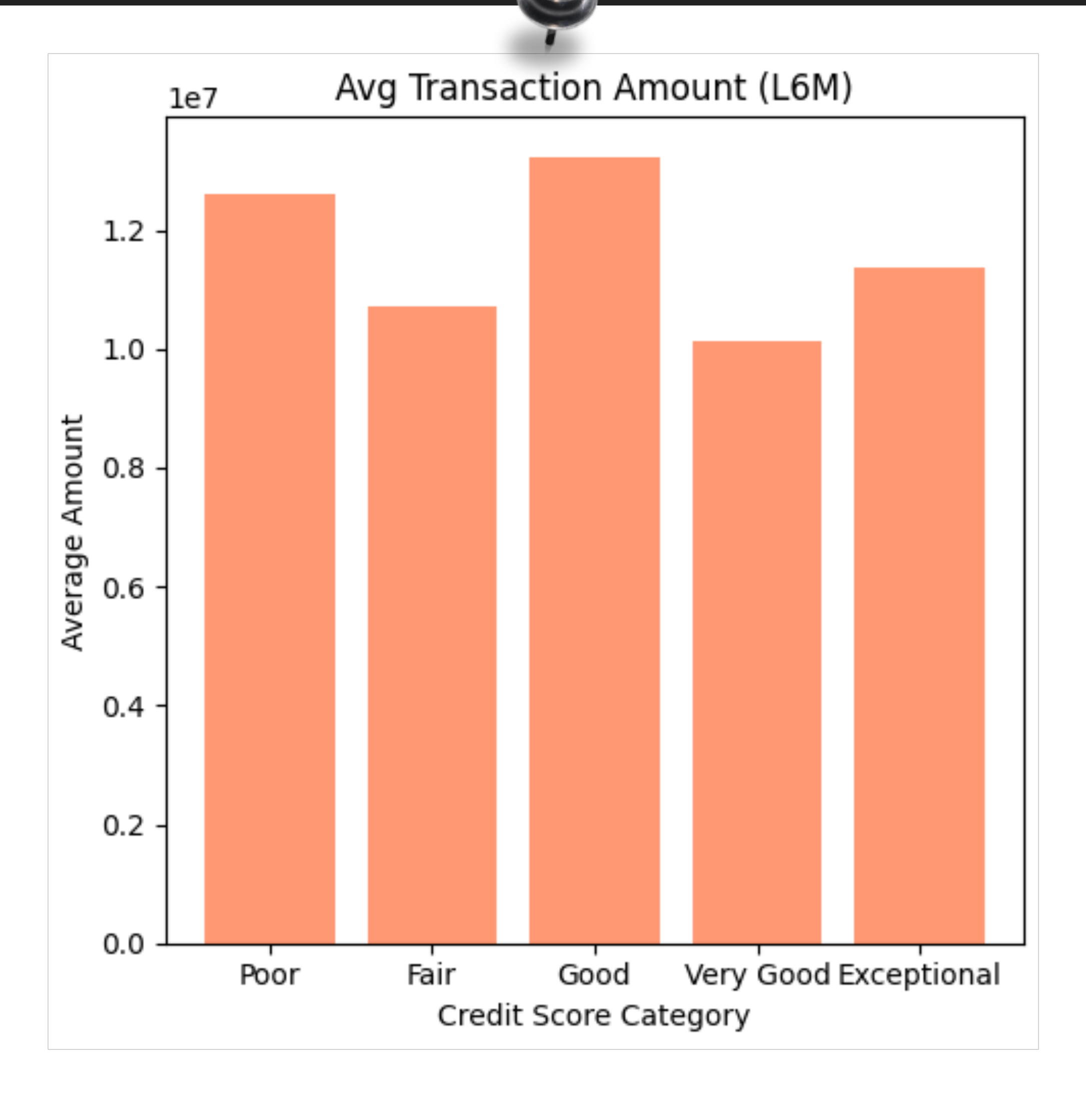
✓ DTI and credit score are highly correlated (inverse relationship).

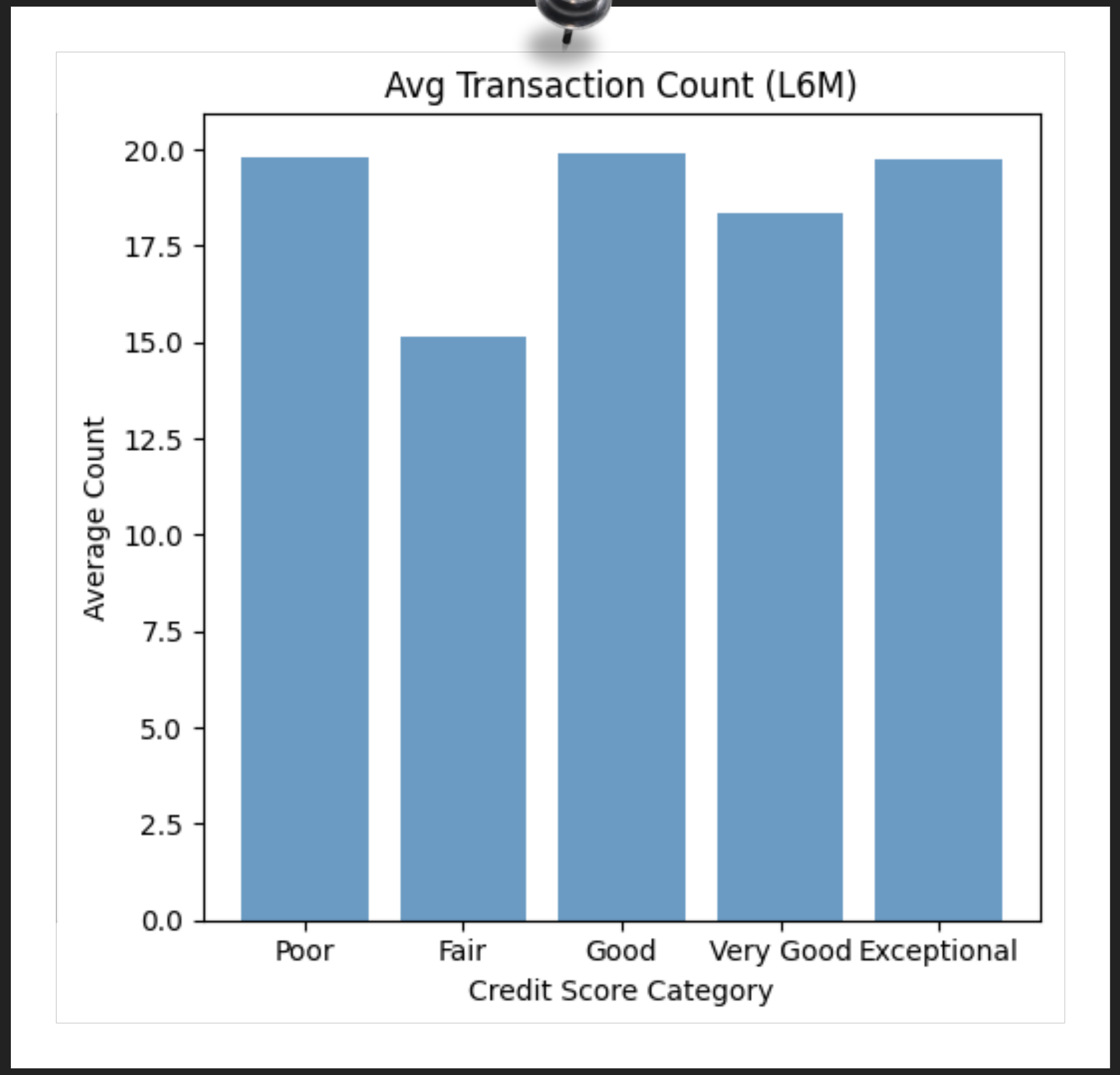
✓ Customers with lower credit scores have a debt burden 3x higher.

✓ Credit score is a strong indicator for risk segmentation

[GOOGLE COLAB LINK]



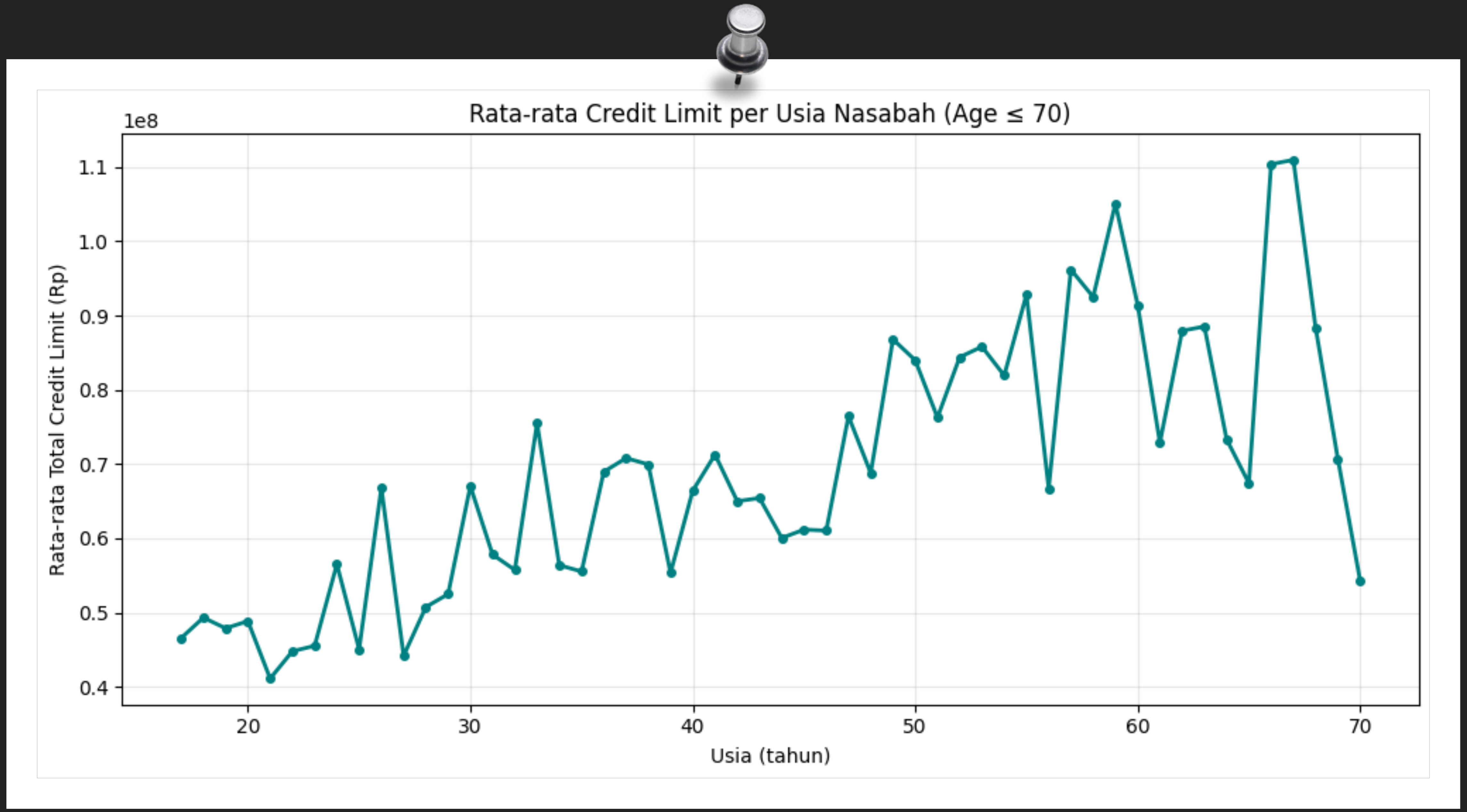




Debt-to-Income Ratio by Credit Score Category

AGE VS CREDIT LIMIT RELATIONSHIP

Average Credit Limit by Age Group



Key Findings

• Peak credit limit: ages 50–60 (~$1.2M average) → peak earnings & stability years

• Trend: Limits increase from ages 20-55, then stabilize/decrease at age 60+

• Younger customers (20–30): Lower limits = growth opportunity for upsells & limit increases over time

• Retirees (60+): Remains important for retention & lifetime value strategies

Strategic Recommendations for RevoBank

✓ For Young Adults (20–35):** Implement progressive credit limit increases tied to transaction history & income growth → drive engagement & loyalty.

✓ For Mid-Career (40–60):** Premium offerings & higher limits → highest profitability segment.

✓ For Retirees (60+):** Focus on retention, stability, & lower churn rates → predictable revenue stream.

Expected Impact

→ Maximize lifetime value by aligning product credit to life stage

→ Increase wallet share in high-earning segments

→ Reduce churn in mature segments through targeted retention

GOOGLE COLAB LINK

FEATURE CORRELATION ANALYSIS

Key Correlations

• Strongest Positive: amt_nonfraud_trx_L6M ↔ count_nonfraud_trx_L6M (0.88)

→ Transaction frequency and nominal value go hand in hand.

• Moderate Positive: yearly_income ↔ total_credit_limit_num (0.56)

→ Higher income tends to result in a higher credit limit.

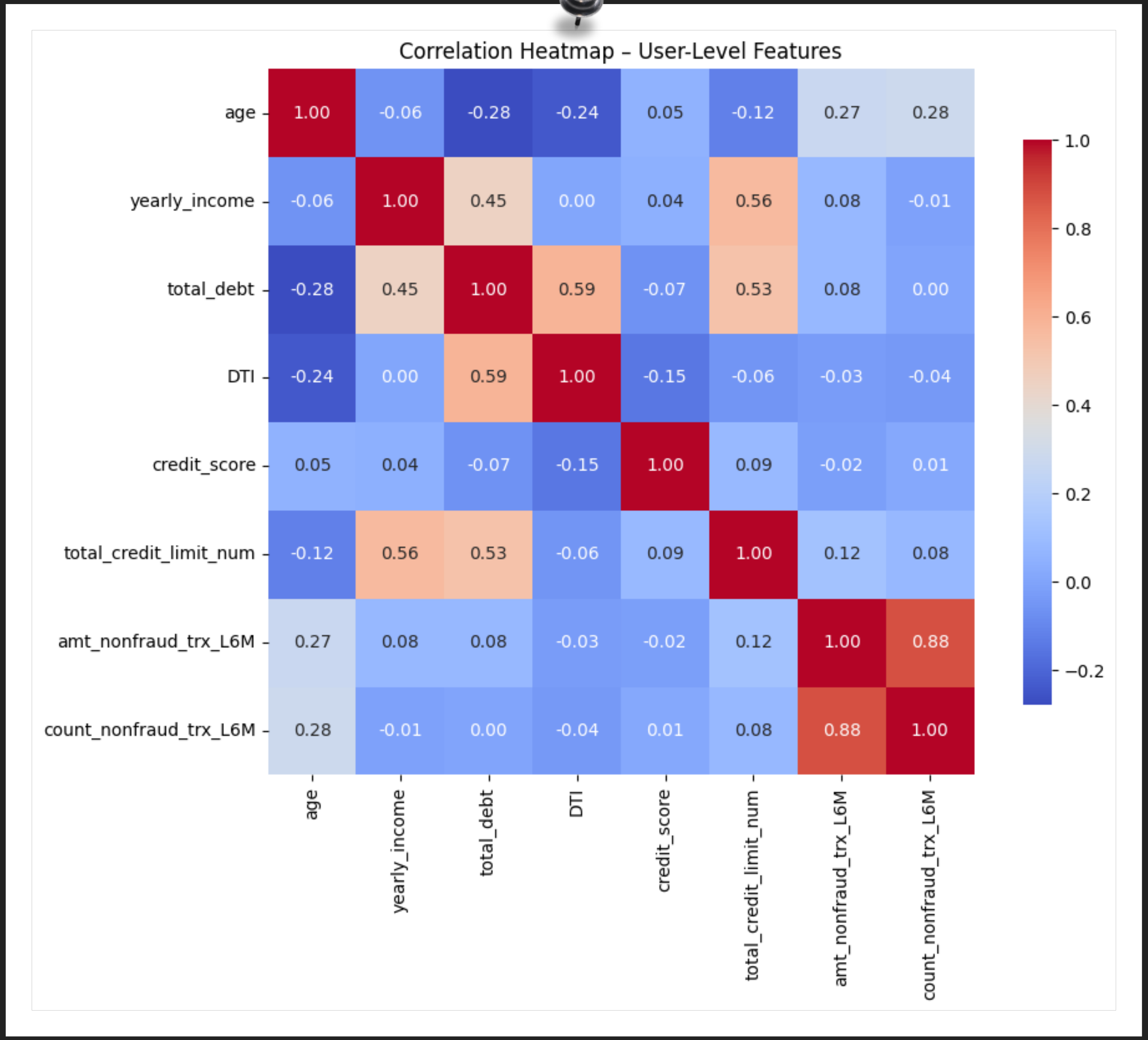
• Moderate Positive: total_debt ↔ total_credit_limit_num (0.53)

→ Higher debt is also associated with a higher credit limit.

Business Use

✓ Can be used for feature selection in predictive modeling.

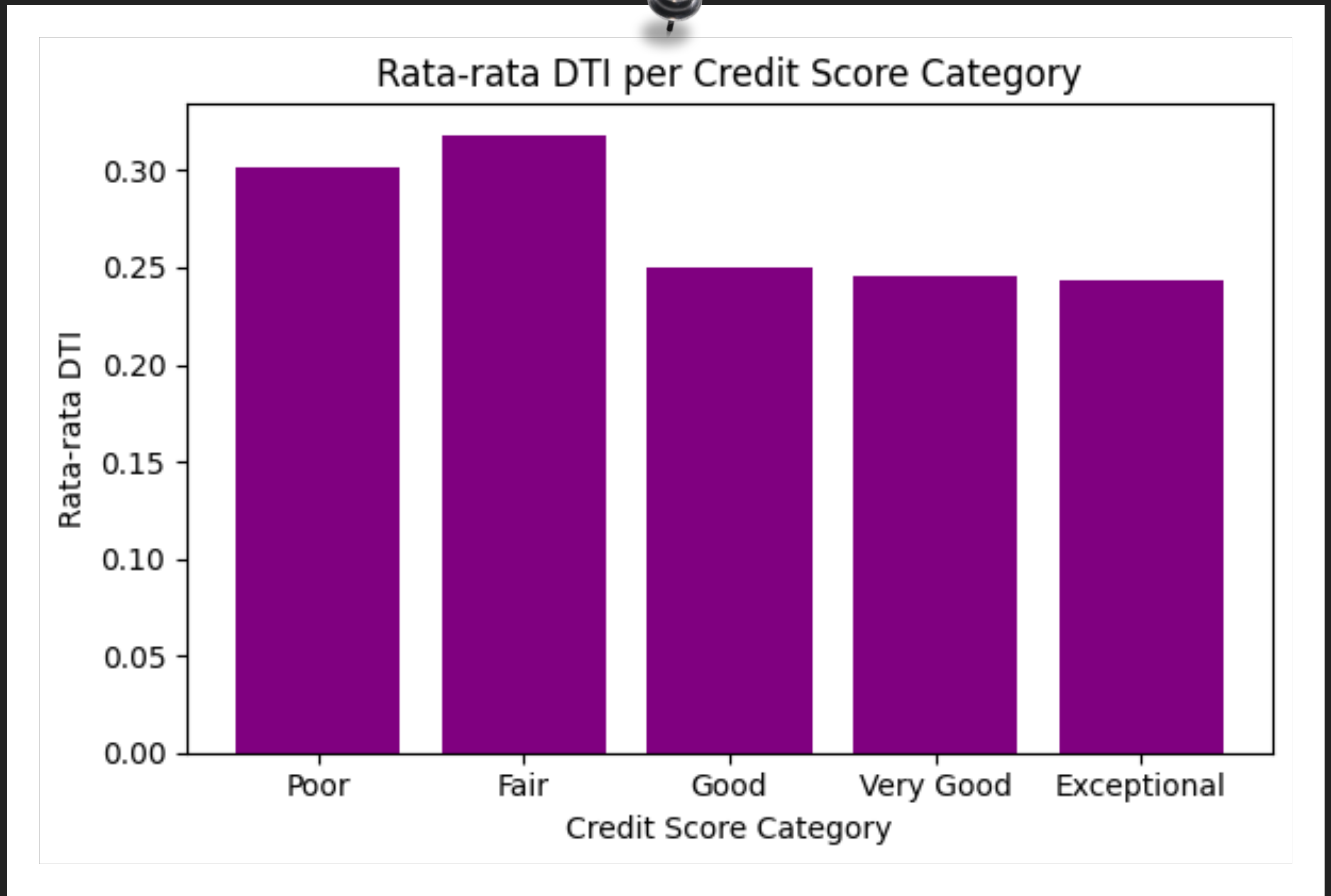
✓ Helps identify multicollinearity before machine learning.



[GOOGLE COLAB LINK]

DEBT-TO-INCOME RATIO BY CREDIT SCORE CATEGORY

DTI Analysis - Segmentation by Credit Risk



Key Findings

• DTI Distribution Across Credit Scores:

- Poor: DTI ~0.30 (highest risk → more debt relative to income)
- Fair: DTI ~0.31 (elevated debt burden)
- Good: DTI ~0.25 (moderate debt level)
- Very Good: DTI ~0.24 (low debt burden)
- Exceptional: DTI ~0.23 (minimal debt → safest borrowers)

Insight

✓ Lower credit score → Higher DTI → More financial stress

✓ Exceptional borrowers maintain DTI <0.25 (healthy ratio)

✓ Poor credit borrowers carry 30% debt-to-income → riskier for new credit

Strategic Recommendation

→ Consider DTI when approving new credit limits

→ Higher DTI borrowers = higher fraud risk + default risk

→ Implement enhanced monitoring for DTI >0.30

[GOOGLE COLAB LINK]

SUMMARY

- The fraud rate reached [X%] of total transaction volume, indicating a manageable risk level with proper monitoring.
- DTI has a clear inverse relationship with credit score, making it a key indicator for risk segmentation and credit policy.
- Customer Segmentation identifies five distinct credit score tiers (Poor, Fair, Good, Very Good, Exceptional) with significantly different behavioral patterns, transaction frequency, and financial health.
- Age Dynamics shows peak credit limits in the 50–60 age group, creating opportunities for age-based product strategies and targeted engagement.
- Transaction Behavior demonstrates that higher credit scores correlate with higher engagement, transaction volume, and a larger average transaction amount.
- Revenue Generation Model: The majority of the bank's revenue comes from merchant fees per transaction (1.5% per transaction), not from interest rates or annual fees.
- Portfolio Health: 81% of cards are inactive (expired/0 limit) indicating significant untapped potential for card reactivation and usage optimization.

GOOGLE COLAB LINK

FRAMEWORK ANALYSIS

Key Findings	Business Recommendations
Higher credit score (Good–Exceptional) results in lower DTI and higher transaction activity (count & amount).	Focus limit increases and promotions on the Good–Exceptional segment because it is the healthiest and most financially active.
The Poor–Fair segment has the highest DTI (~0.30+) with a greater risk of default, suboptimal transaction contribution.	Implement conservative limit policies, stricter approval requirements, and regular DTI monitoring for high-risk segments.
Those aged 50–60 years have the highest average credit limit, while young customers (20–35) are still under-utilized.	Offer premium products and cross-sell to those aged 40–60. Use a gradual limit increase scheme for young, high-performing customers.
Most of a bank's revenue comes from fees per transaction, so transaction volume and frequency are crucial.	Focus on increasing card usage (usage & frequency), not just acquisition, for example through cashback, installments, and targeted promos.
Fraud risk increases with high transaction activity and extreme DTI (very low or very high).	Implement real-time anomaly detection, transaction monitoring per segment, and escalation rules based on DTI & credit score profiles.

MILESTONE 3

MUHAMMAD NUR ZHIFAR
DECEMBER 31 2025
REVOU FSDA BATCH OCT25
INTERMEDIATE ASSIGNMENTS

CUSTOMER SEGMENTATION ANALYSIS REVOBANK CREDIT CARD : K-MEANS CLUSTERING & BUSINESS OPPORTUNITIES

OBJECTIVE & PROBLEM STATEMENT

Background:

RevoBank has a portfolio of 746 customers with highly diverse card usage behaviors. Some customers are highly active and high-value, while others have high limits but rarely make transactions (dormant or under-utilized). This situation indicates a significant opportunity gap in MDR revenue.

Business Problem:

- How to identify the customer segments with the most potential to encourage increased spending?
- What are the unique characteristics of each segment so that marketing strategies can be tailored?
- Which customers are already high-value and need to be retained through a loyalty program?

Analysis Objective:

- Conduct customer segmentation using K-Means to group customers based on their transaction behavior and financial profile.
- Identify four main clusters and assign persona labels to each.
- Analyze characteristics per cluster (recency, frequency, amount, limit, income, DTI, credit score, age).
- Develop actionable and measurable business recommendations per cluster.
- Main Focus: Cluster 1 (high-limit, low-usage) as the highest revenue growth target.

[GOOGLE COLAB LINK]

UNDERSTANDING CUSTOMER FINANCIAL STABILITY

Dataset Overview

Total post-merge records: 1,025 customers with 25 variables (card + user attributes) covering 6-month period (December 2024 – May 2025). Excellent data quality with zero missing values after cleaning.

Key Metrics & Distribution

- Age: Range 17–80 years, average 42.7 years; 10.2% of customers are retirees (age > retirement threshold)
- Yearly Income: Range IDR 25K – IDR 250K; skewed distribution shows the majority of customers are middle-income
- Credit Limit: Average IDR 25.1M per customer; range 0 – IDR 268M, reflecting diverse credit capacity
- Debt-to-Income Ratio (DTI): Average 0.28 (range 0–0.97); key indicators of financial stress & risk profile
- Credit Score: Distribution across 5 tiers (Poor, Fair, Good, Very Good, Exceptional), with the majority in the Good–Exceptional range
- Transaction Activity: 8–1,847 non-fraud transactions per customer; high variance shows distinct usage patterns

Key Insights – Feature Relationships

- DTI & Credit Score Inverse Relationship (Correlation: -0.70)
- Poor credit score customers have a DTI of 0.45 (3x higher than the Exceptional tier DTI of 0.15), indicating financial stress & default risk. Critical for risk-based credit policy.
- **Age Dynamics & Credit Limit Allocation**
- Peak credit limit in age group 50–60 (Rp 1.2B average), reflecting peak earning years & credit worthiness. Age 20–30 significantly underutilized → growth opportunity with progressive limit increase strategy.
- Transaction Frequency & Amount Strong Correlation (Correlation: 0.88)
- Customers with high transaction frequency also have higher transaction amounts per transaction, indicating consistent spending behavior. The main revenue generator is transaction frequency, not limits.
- Manageable Fraud Rate at 0.22% of total transaction value; non-fraud transactions dominant 99.78%, indicating effective fraud detection mechanisms.

Strategic Implications

Data shows significant variance in customer financial stability and behavior, creating an opportunity for a segmentation-based approach. Customers with similar DTI, age, income, and credit score patterns will respond to similar campaigns; hence, a clustering approach is an appropriate methodology for maximizing engagement and revenue per segment.

GOOGLE COLAB LINK

METHODOLOGY – K-MEANS CLUSTERING APPROACH

Why K-Means:

K-Means is an ideal unsupervised learning algorithm for customer segmentation because it is iterative, scalable, and easy to interpret. Optimizes within-cluster cohesion customers within the same cluster share similar characteristics. Compared to alternatives (Hierarchical Clustering, DBSCAN): K-Means is faster and more efficient for a dataset of 746 customers, provides more stable and reproducible results, and provides straightforward business interpretation through cluster centroids (average profiles per segment).

Feature Selection (8 main features):

- Transaction Behavior: recency_days, total_nonfraud_trx, total_nonfraud_amt
- Financial Capacity: total_credit_limit_num, yearly_income
- Risk Profile: DTI (Debt-to-Income), credit_score
- Demographics: age

Reason: This feature comprehensively represents customers' spending behavior, financial capability, health risk, and life stage.

GOOGLE COLAB LINK

Data Preprocessing:

- Standardization (StandardScaler): normalize all features to mean=0, std=1 to ensure balanced scaling.
- Without scaling, rupiah features (income, limit, amount) with magnitudes in the millions will dominate the distance calculation, while ratio (DTI) and nominal (age) become less influential.
- Data Quality: The dataset has undergone Milestone 1 cleaning (no missing values, outliers handled, typos fixed), ready for clustering.
- Dataset: 746 customers, 8 features.

Determining Optimal K:

- Elbow Method: identifies the "elbow" of the inertia vs. K graph; a sharp decrease is seen until K≈4, then slows down.
- Silhouette Score: measures cluster cohesion and separation (range -1 to 1, the higher the better); K=4 indicates a score of 0.313 (second highest after K=2).
- Decision: K=4 was chosen as the optimal balance between cluster separation quality and complexity, still interpretable for business action.

DETERMINING OPTIMAL K – ELBOW & SILHOUETTE ANALYSIS

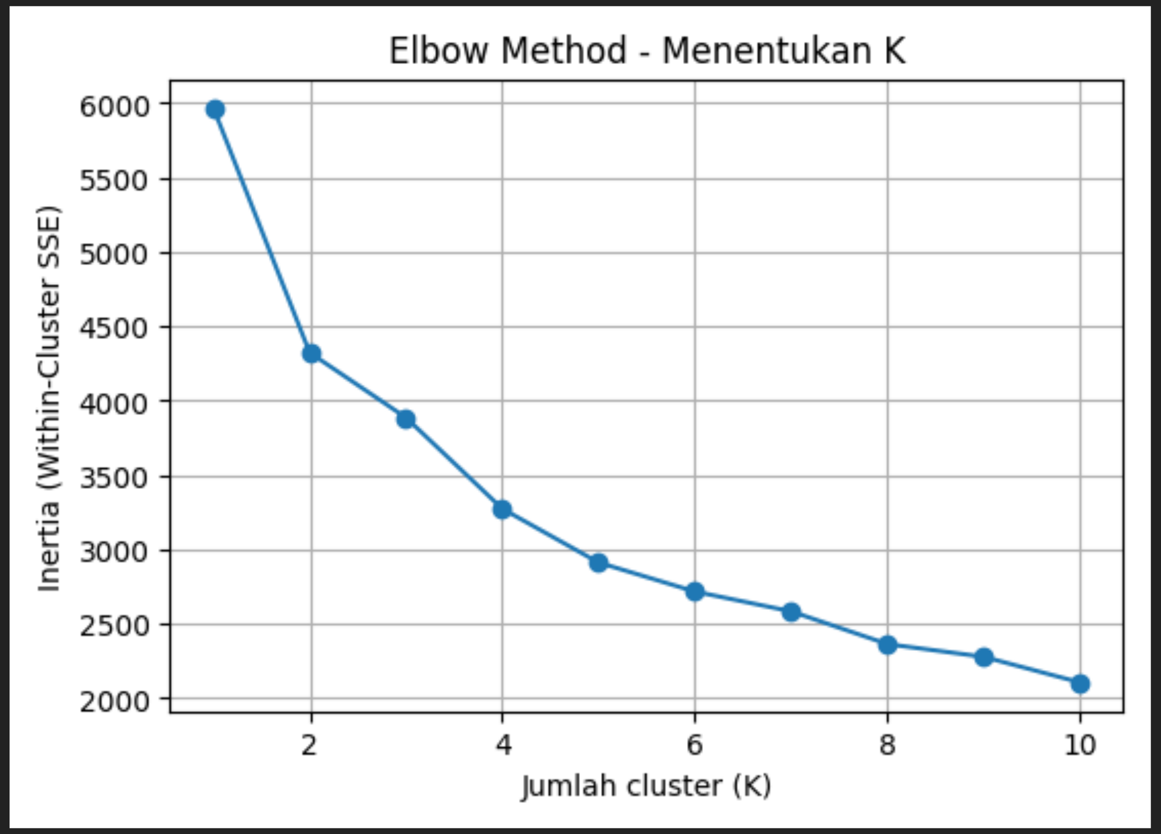
Left side - Elbow Method:

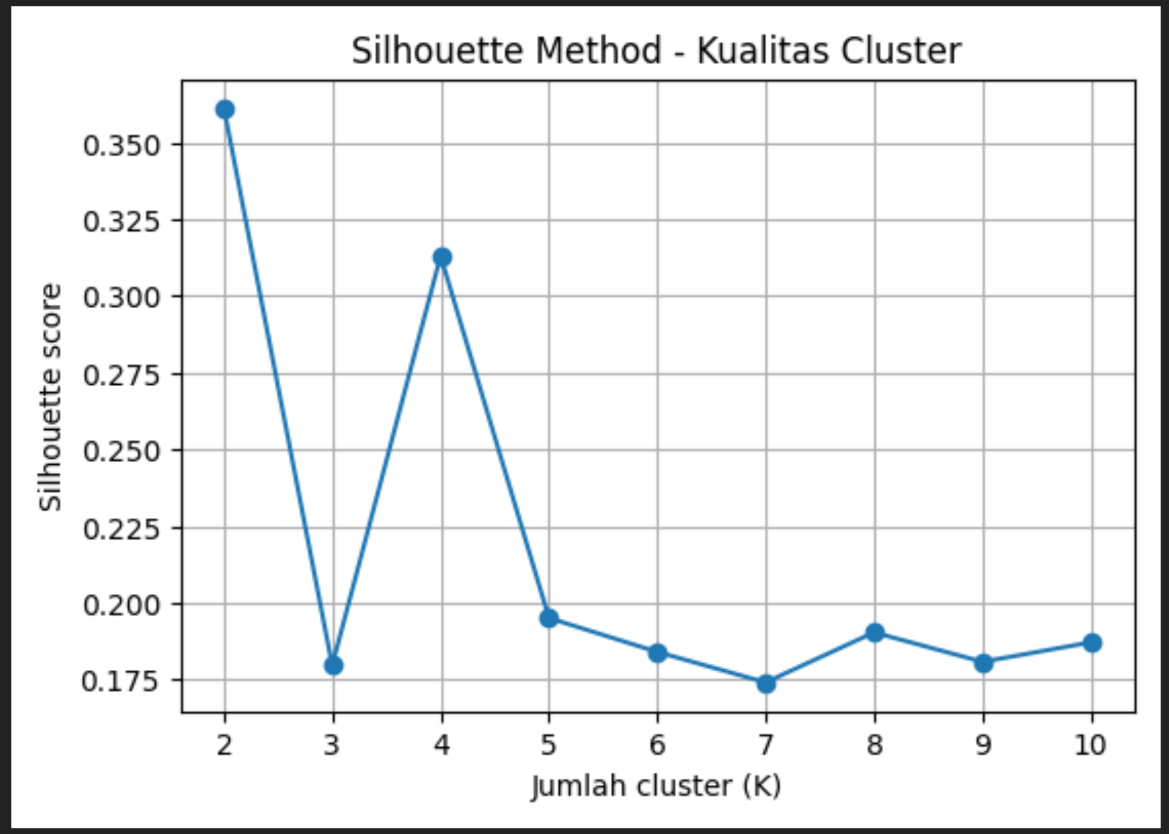
Displays a graph of inertia vs. K (K=1-10). The graph shows a sharp decrease from K=1 to K=4, then the curve begins to flatten. An "elbow" is clearly visible around K=4, indicating that adding clusters above K=4 does not provide significant inertia improvements.

Right side - Silhouette Score:

Displays a graph of silhouette score vs. K (K=2-10). The values are:

- K=2: 0.361 (highest, but too simple for business segmentation)
- K=3: 0.180 (significant decrease)
- K=4: 0.313 (second highest, well above the other K values, which range from 0.17 to 0.20)
- K≥5: decreasing and stable below 0.20





Conclusion:

- K=4 was chosen as the optimal choice due to the following combination:
- The elbow is clearly visible in K=4
- The silhouette score of K=4 is high and sufficient to distinguish four meaningful business segments
- K=4 provides the best balance between model complexity and interpretability for actionable business insights

[GOOGLE COLAB LINK]

EXPECTED OUTCOMES & SUCCESS METRICS

Targets per Cluster:

Cluster 1 – Big Limit Low Usage (Priority 1)

- Target: Increase spending by 20-30% within 6 months through a reactivation campaign
- Success metric: Of 51 customers, at least 60-70% are actively transacting and the nominal amount is increasing

Cluster 0 – Dormant Young Holders (Priority 2)

- Target: First-time card activation, at least 40% make a transaction within 3 months
- Success metric: Of 477 customers, 188 customers make their first transaction

Cluster 2 – High-Value Active Spenders (Retention Focus)

- Target: Retention rate >95% and prevent churn to competing banks
- Success metric: Maintain current spending level, loyalty program enrollment >80%

Cluster 3 – Mature Stable Users (Retention Conservative)

- Target: Maintain engagement and cross-sell related products (insurance, travel benefits)
- Success metrics: Retention >90%, cross-sell adoption >25%

GOOGLE COLAB LINK

IMPLEMENTATION ROADMAP

Phase 1 – Months 1-2: Segmentation & Strategy Development

- Finalize K-Means model and validate cluster assignments
- Define targeted messaging & offers per cluster
- Set up tracking infrastructure for success metrics

Phase 2 – Months 3-4: Campaign Launch

- Cluster 1 (Big Limit Low Usage): Campaign reactivation with incentive spending
- Cluster 0 (Dormant Young Holders): First-use welcome campaign
- Cluster 2 & 3: Loyalty programs & cross-sell initiatives

Phase 3 – Months 5-6: Monitor & Optimize

- Track KPIs per cluster (spending, transaction frequency, retention)
- A/B test messaging variants based on early results
- Prepare quarterly business reviews with insights

Expected Impact:

- Activate 188+ dormant young holders
- Increase spending in Cluster 1 by 20-30%
- Improve overall card utilization rate & revenue

[GOOGLE COLAB LINK]

KEY FINDINGS & RECOMMENDATIONS

Key Findings	Business Recommendations
51 customers have high credit limits but very low card usage (underutilized).	Run a special reactivation for Cluster 1 with spending increase promotions (cashback, installments, bonus points).
477 young customers tend to be dormant with minimal or no active transactions.	Create a first-use activation campaign for Cluster 0 (welcome voucher, fee waiver, bonus points on first transaction).
Customers aged 35–45 with high income (Cluster 2) show stable transaction patterns and high value.	Prioritize retention & loyalty programs for Cluster 2, including exclusive offers and premium benefits.
Mature customers (Cluster 3) are stable but less exposed to other bank products.	Encourage cross-selling of relevant products (insurance, travel, investment-linked benefits) to increase product per customer.
Recency above 90–180 days is a strong signal of dormancy and potential churn.	Set up automated behavioral triggers to send campaigns when recency passes a certain threshold.

GOOGLE COLAB LINK

CONCLUSION & NEXT STEPS

Analysis Conclusion

Using the K-Means clustering approach with K=4, we successfully identified four fundamentally different customer segments in terms of transaction behavior, financial capabilities, and demographic profiles. Each segment exhibits unique patterns and requires tailored engagement strategies to maximize credit card usage and MDR revenue.

The segmentation results provide a data-driven foundation for the Performance Management team to design targeted and cost-effective campaigns, instead of a one-size-fits-all approach.

Strategy Recommendations per Cluster

- **Cluster 1 – Big Limit Low Usage (Top Priority)**

Customers with high credit limits but underutilized. Recommendation: Launch a reactivation campaign with spending incentives (cashback, points multiplier) to drive a 20–30% increase in transactions within 6 months.

- **Cluster 0 – Dormant Young Holders (Second Priority)**

477 young customers with minimal or no activity. Recommendation: first-use activation campaign with welcome voucher, fee waiver, and bonus points on the first transaction to convert at least 40% into active users.

- **Cluster 2 – High-Value Active Spenders (Retention Focus)**

Premium segment with high and stable spending. Recommendation: Prioritize retention with loyalty programs, exclusive benefits, and cross-selling premium products (travel, insurance) to maintain engagement >95%.

- **Cluster 3 – Mature Stable Users (Conservative Retention)**

Mature customers with stable but conservative spending. Recommendation: Focus on retention and lifetime value through loyalty programs and cross-selling related, low-risk products to increase product value per customer.

Age Dynamics

Creates opportunities for targeted strategies: peak credit limits for those aged 50–60 (premium segment), growth opportunities for those aged 20–35 (progressive limit increases), and retention focused on retirees (stability and predictable revenue).

Revenue Model Clarity

The majority of revenue comes from merchant fees per transaction (1.5%), not interest or annual fees, so transaction volume and frequency are the primary KPIs for increasing profitability.

Implementation Timeline

Phase 1 (Months 1–2) finalize campaigns and infrastructure; Phase 2 (Months 3–4) launch campaigns; Phase 3 (Months 5–6) monitor, A/B testing, and quarterly business reviews for ongoing optimization.

[GOOGLE COLAB LINK]

THANK YOU

MUHAMMAD NUR ZHIFAR
DECEMBER 31 2025
REVOU FSDA BATCH OCT25
INTERMEDIATE ASSIGNMENTS