Risk Analytics in Banking and Financial Services

A Python EDA + Power BI Project

Presented by: Priyasha Sinha Roy

ABOUT THE PROJECT

The project is a banking project where we have used a banking dataset to understand :

1. **How data can minimize financial risk in lending.**
2. **Identify high- vs. low-risk customer profiles.**
3. **Segment and score customers using banking, demographic, and financial behavior variables.**

We also created an automated Power BI dashboard where we are going to
show features like Loan analysis and Deposit Analysis and summary of the overall dataset.

Problem Statement

Financial institutions face potential losses from lending to risky customers.
Need a data-driven, explainable risk framework:
EDA + Scoring + Visualization = Actionable Risk Strategy.

About the Dataset

Contains: Client details, Financial metrics, Relationships, Demographics.
Multiple interlinked tables: Banking Relationship, Client-Banking, Gender, Investment Advisor, Period.
3,000 rows; 25 features.

Solution First

1. Developed a Risk Scoring Matrix to segment customers by risk.
2. High-risk clients identified through key attributes such as income, credit card balance, and asset ownership.
3. Implemented actions: pricing strategy, credit restrictions, and early warning systems.
4. Automated risk tier system with dynamic thresholds.

Visualisation Patterns by Nationality

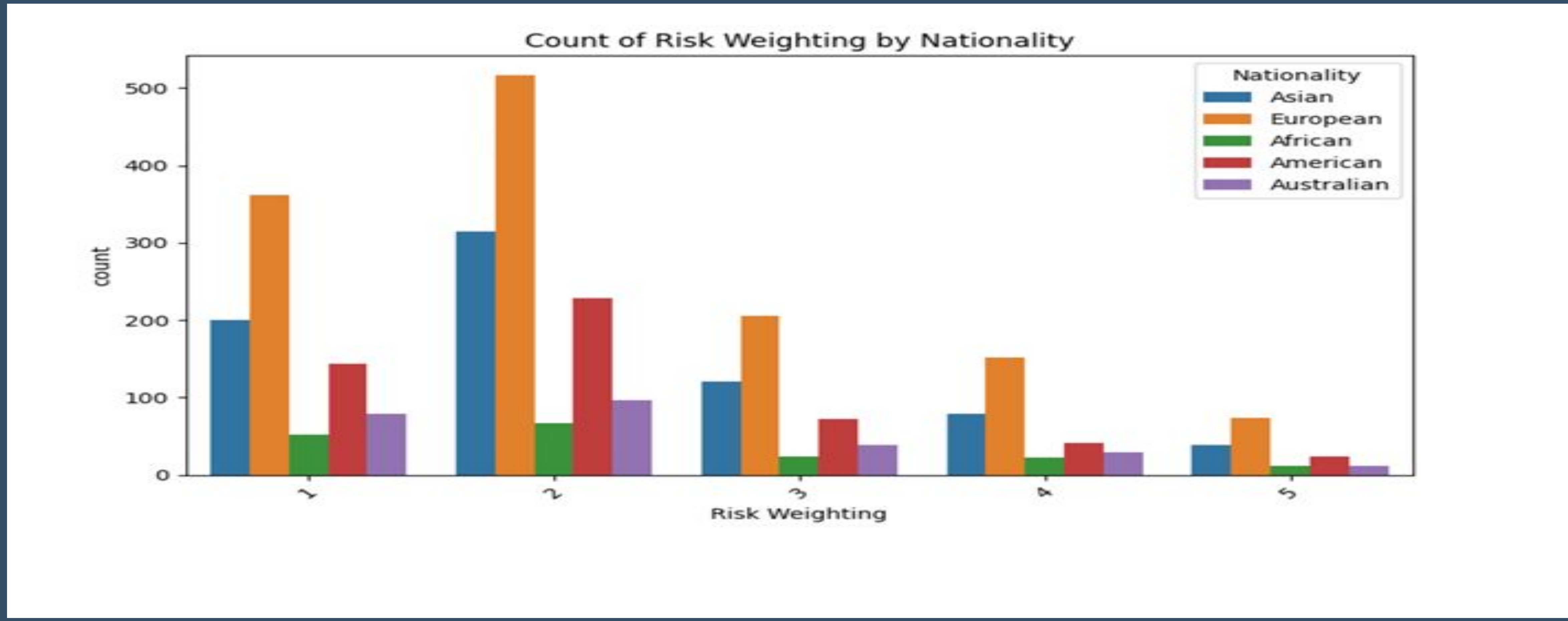
1. Risk Weighting By Nationality : Right Graph
   Observation:The graph shows Asians have the highest count in Risk Levels 4–5, followed by Europeans and Americans.

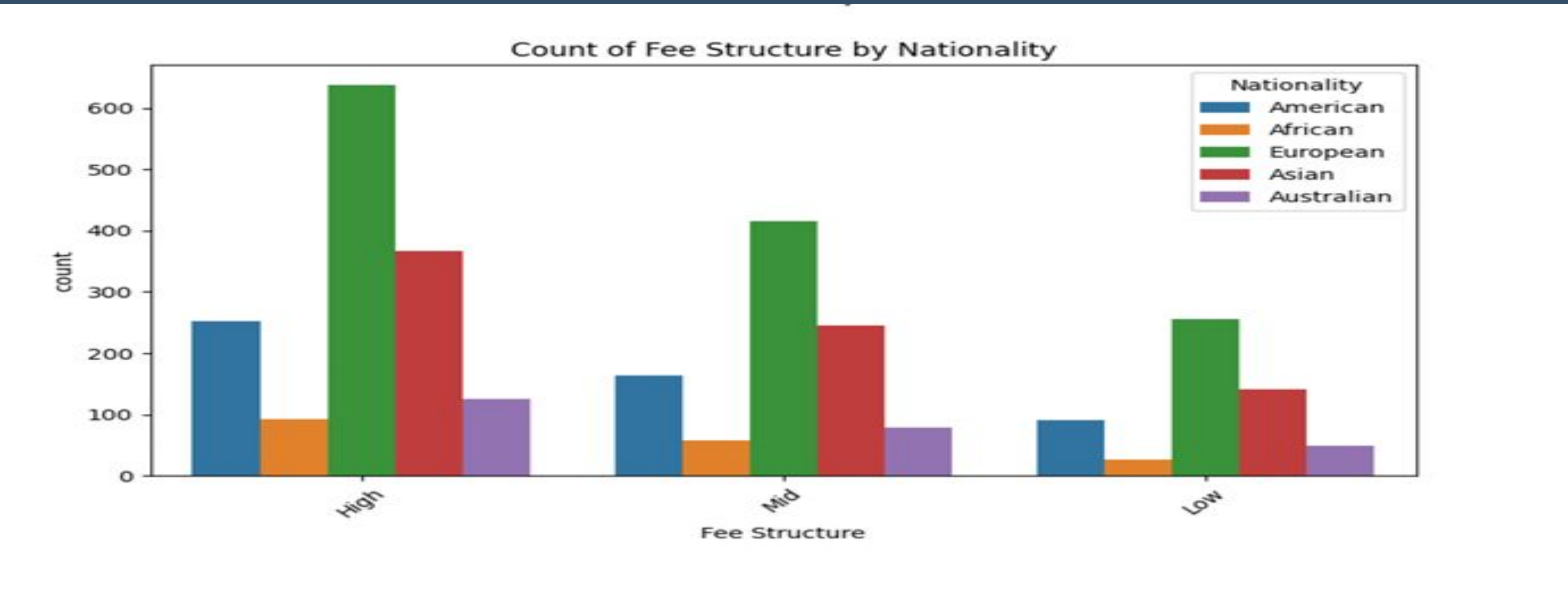
2. Fee Structure by Nationality: Left Graph
   Observation: Europeans dominate the High-Fee category, but right graph shows their risk is moderate (not the highest).

Higher-Risk Nationals: Asians > Americans > Europeans (based on left Graph).
High-Fee ≠ Highest Risk: Europeans pay high fees but are not the riskiest—Asians are.

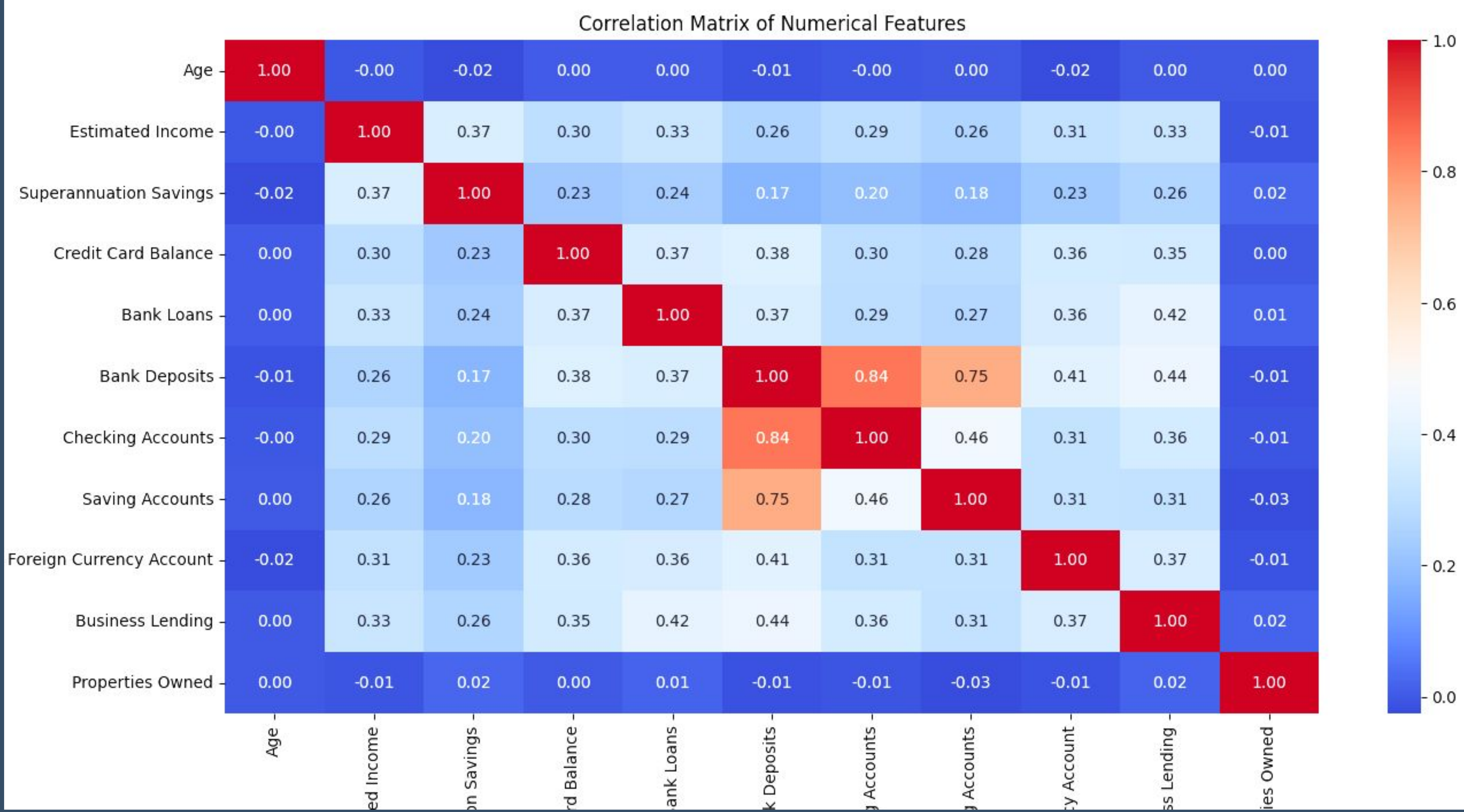
Why This Matters:
Fee Structure ≠ Risk Directly: High fees may indicate premium services (e.g., wealth management), not necessarily risk.
True High-Risk Groups: Asians & Americans (from Page 13) have higher-risk behavior (e.g., more credit cards, higher balances).

Actionable Insight: Scrutinize Asian/American applicants for debt patterns. Europeans may be low-risk but high-revenue customers (prioritize retention).





Correlation Matrix of Numerical Features: A HeatMap



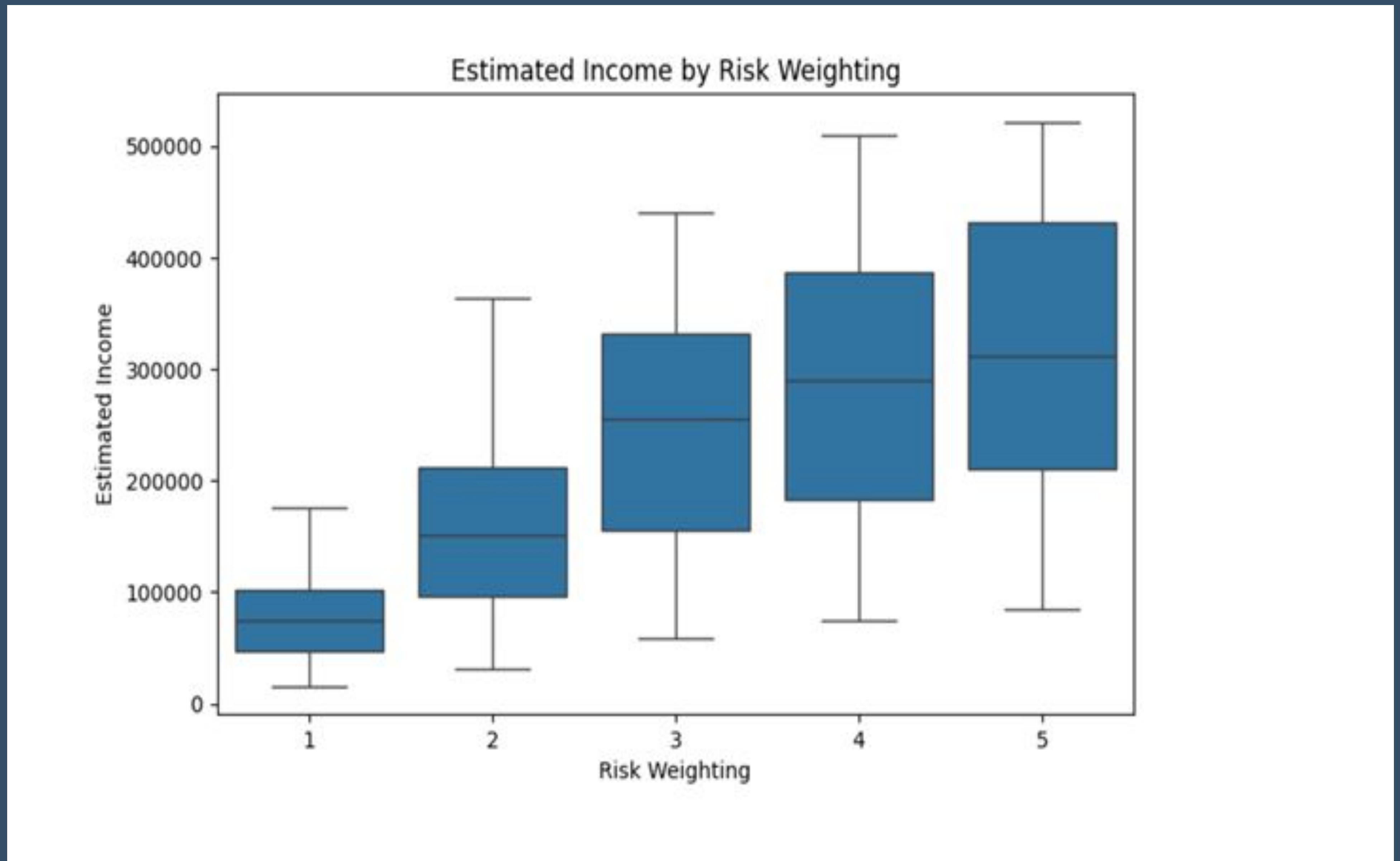
Bi Variate Analysis and Visualization Patterns from HeatMap

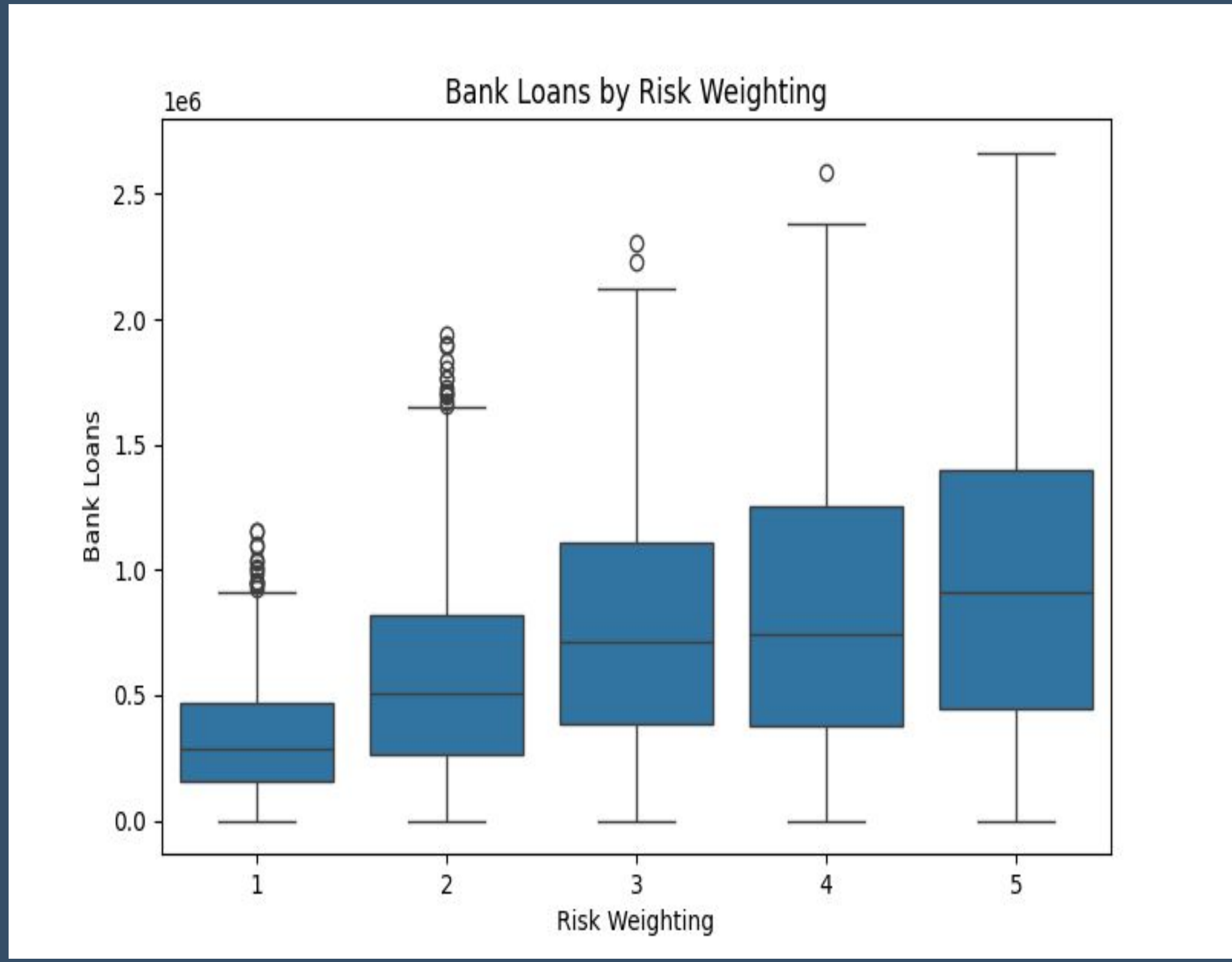
1. Checking vs. Saving Accounts: Strong positive correlation (0.75).
2. Age vs. Superannuation: Savings peak around middle age (40–60).
3. Income vs. Checking Accounts: Higher income links to more checking funds.
4. Bank Loans vs. Credit Card Balance: Positive relationship, suggesting leveraged spending.
5. Focus on: Deposit/savings stability (0.75), credit card/loan risks (0.37).
6. Ignore: Age, properties owned (minimal impact).
7. Act: Automate alerts for correlated risk pairs (e.g., credit card + loan spikes)

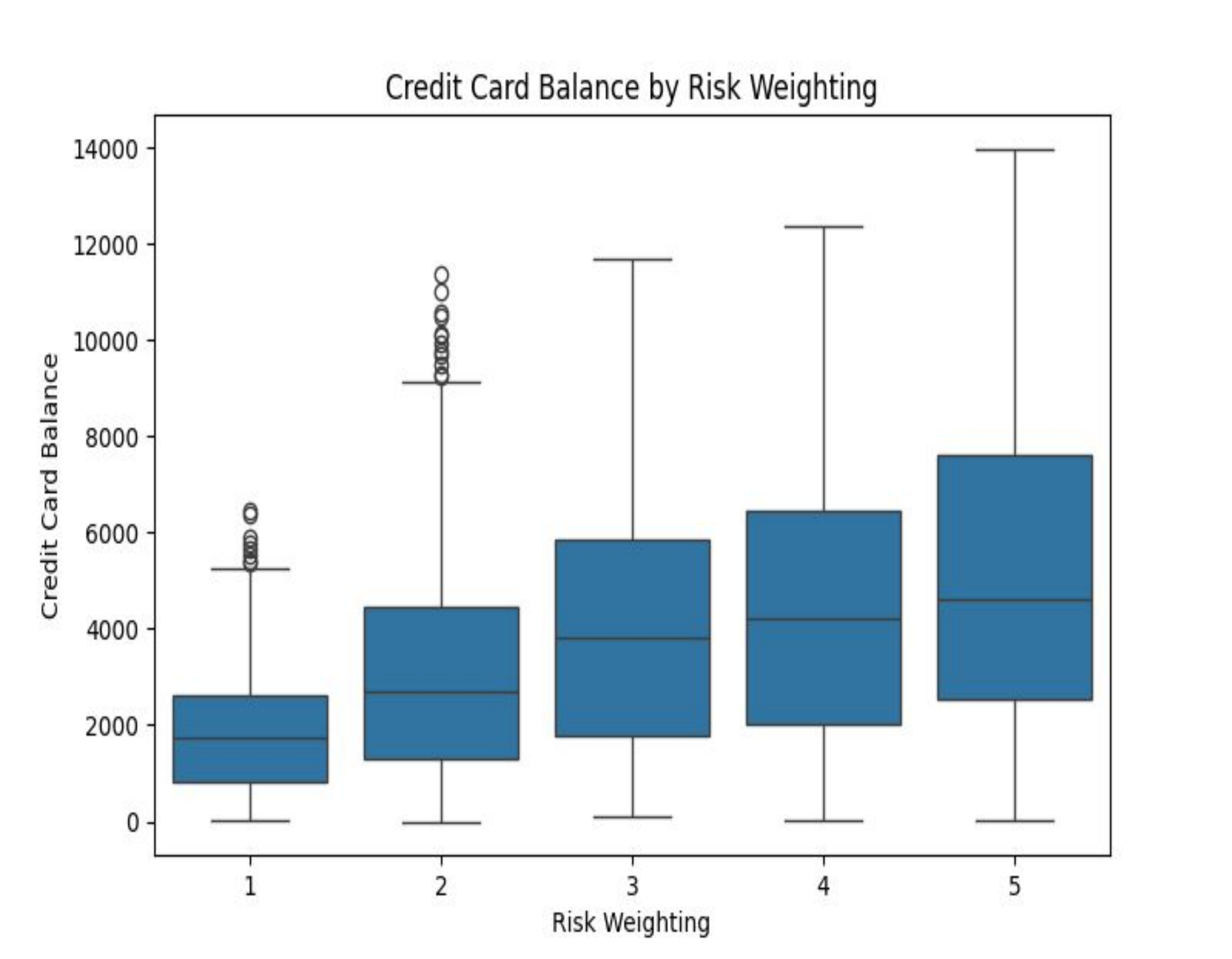
Visualization Patterns from : Boxplots of Financial Metrics by Risk Category

1. Across all financial metrics, higher risk customers (Risk 4–5) consistently show greater income, loan balances, and credit card debt, highlighting over-leverage despite high earnings.

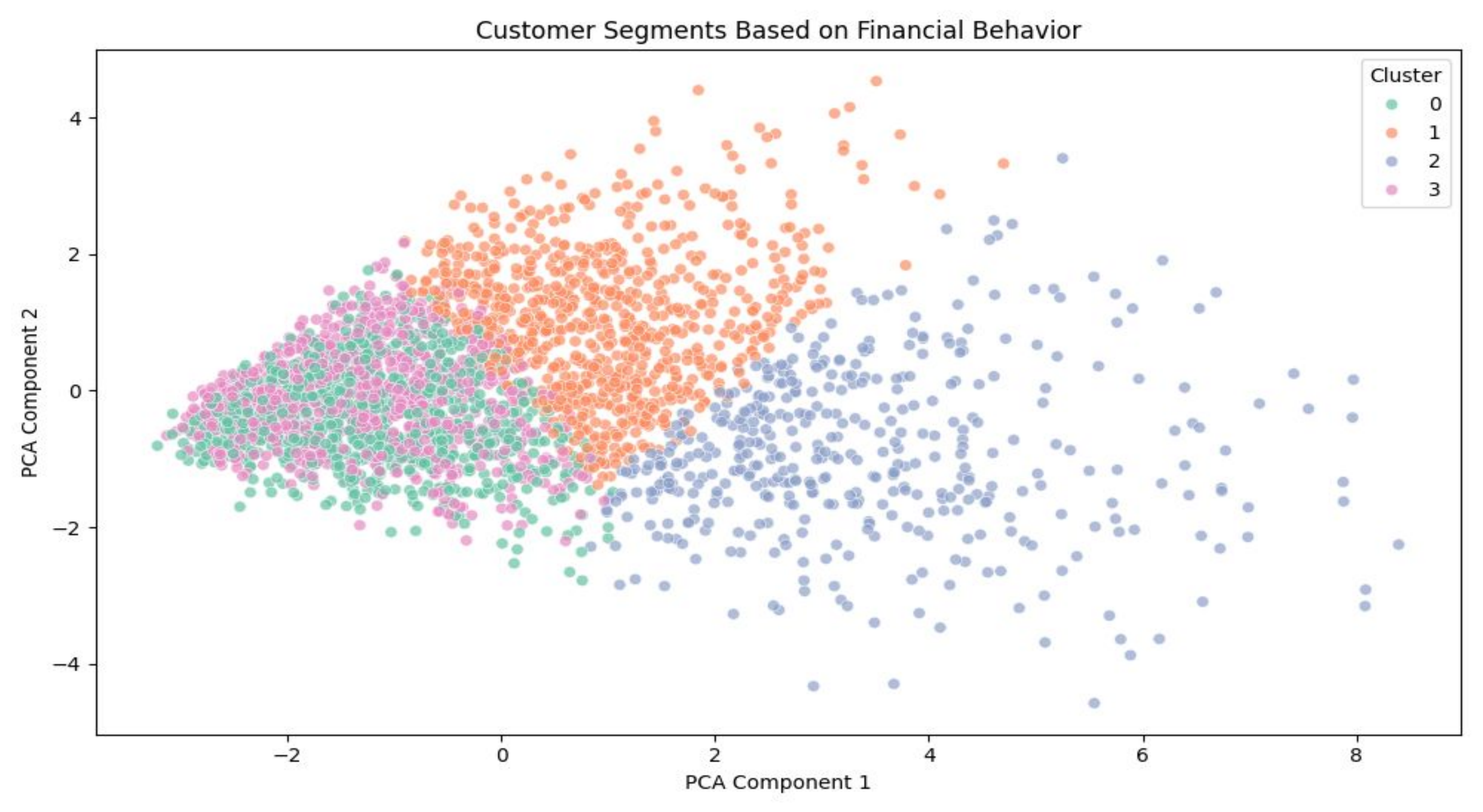
2. This indicates that absolute income or savings is not a reliable risk indicator—exposure and financial behavior matter more."







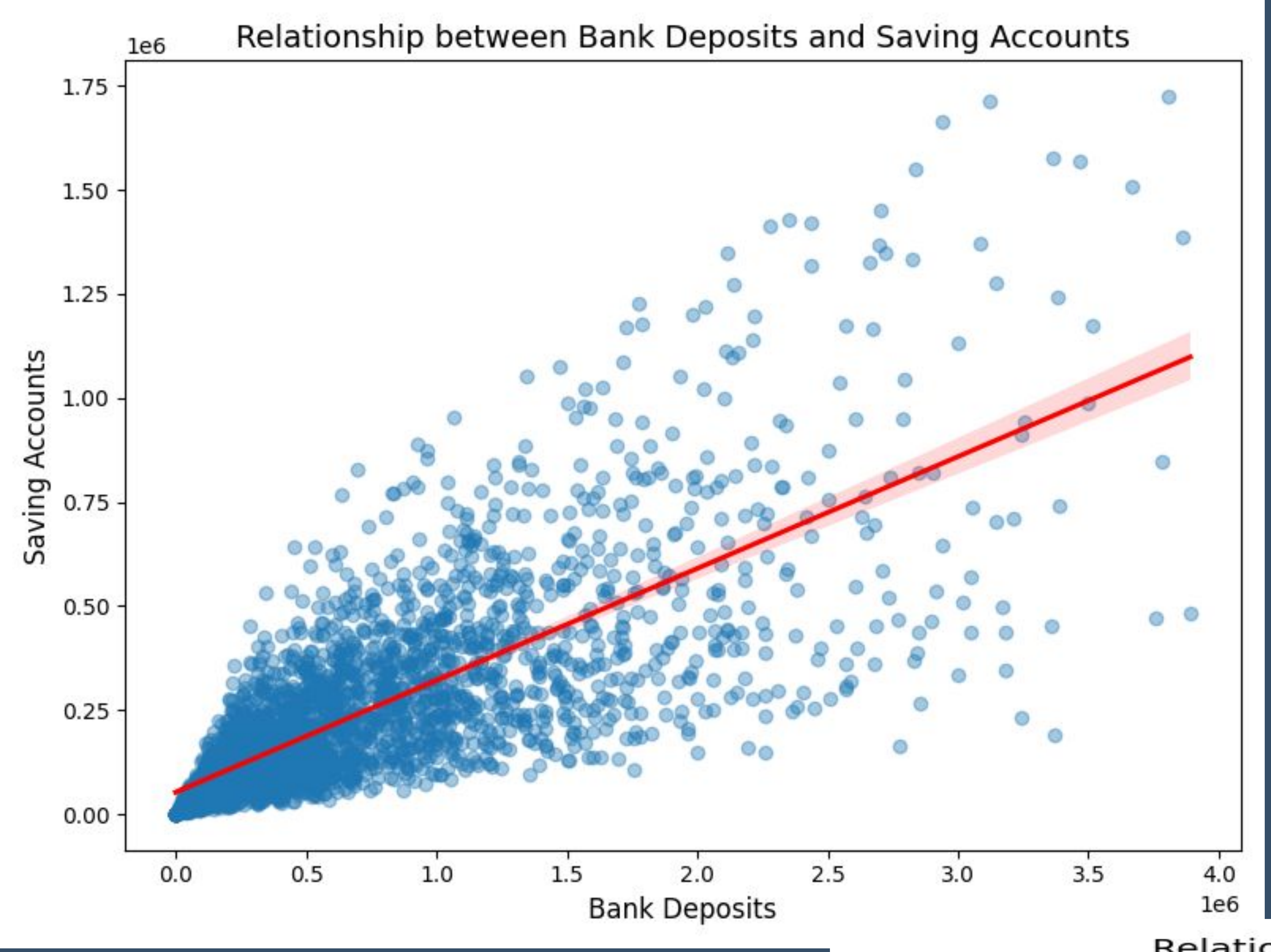
Clustering Analysis using K-means Machine Learning

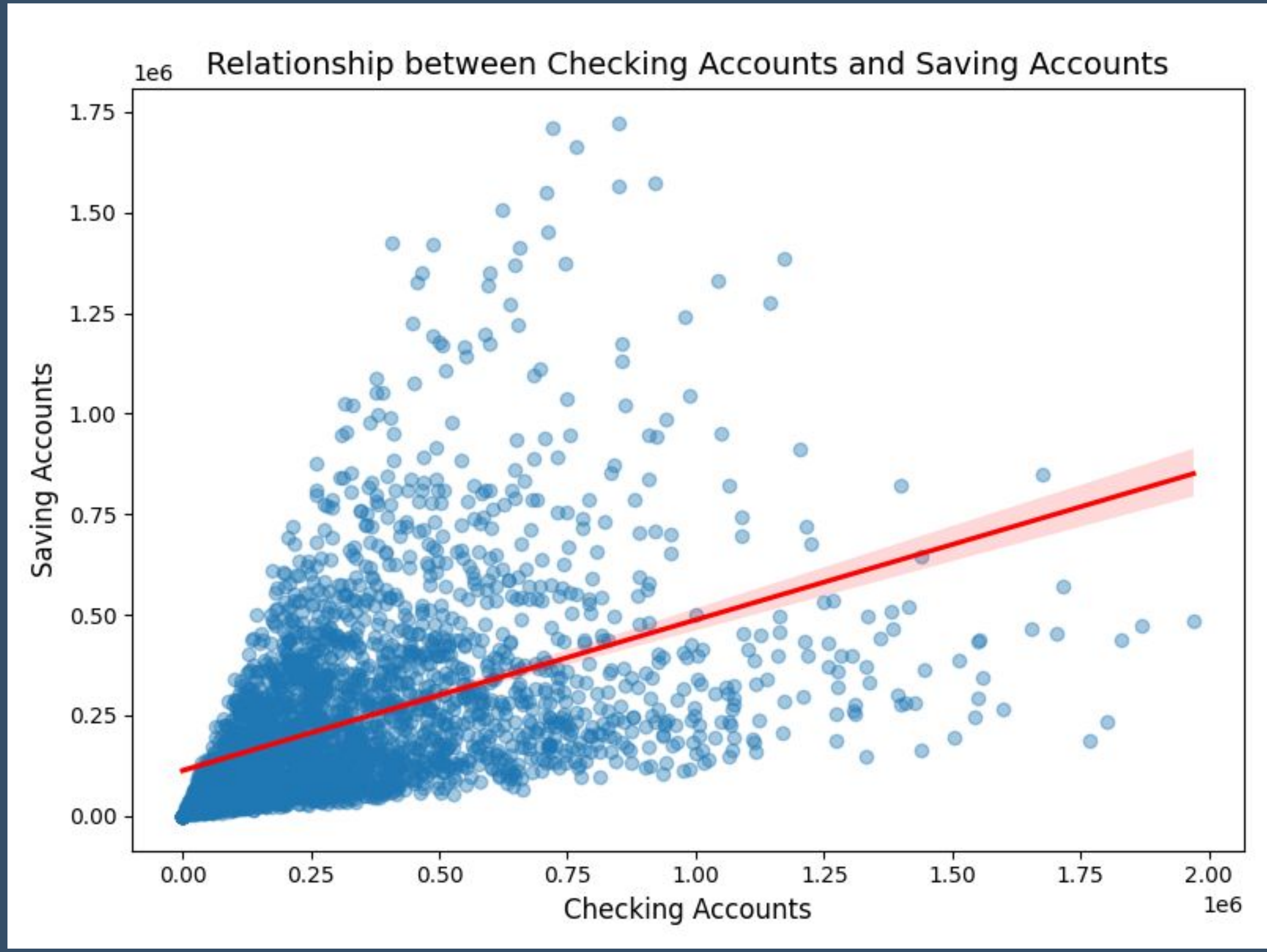


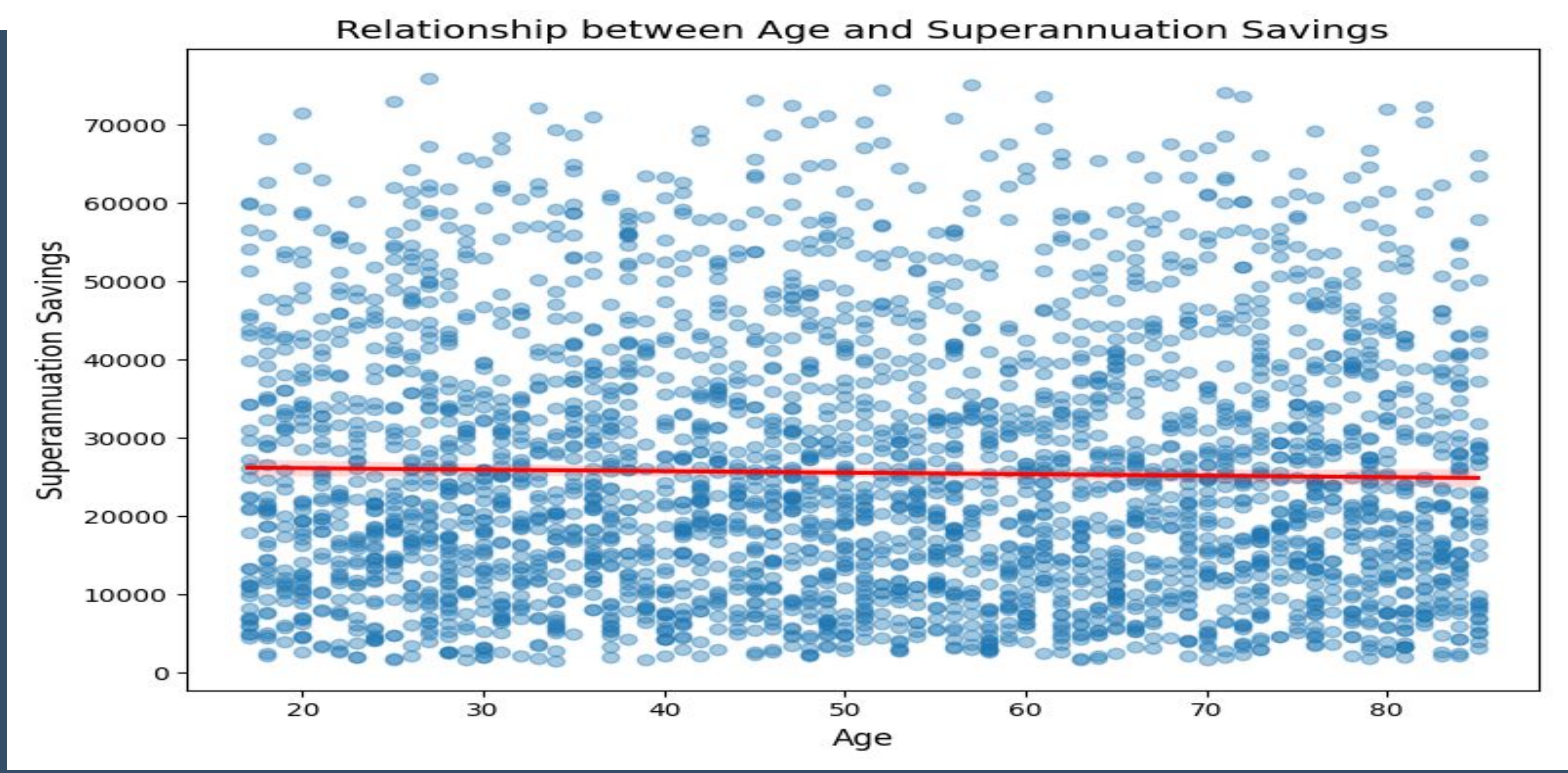
Key Insights from Customer Segmentation

1. One large customer group (Cluster 2) shows highly inconsistent financial behavior — these customers are unpredictable and present higher risk.
   → Action: Apply stricter credit checks and conservative lending policies for this group.
2. Two customer groups (Clusters 0 and 3) are financially stable and consistent.
   → Action: Offer better loan terms and cross-sell savings or investment products to these customers.
3. Segmenting customers using financial patterns helps the bank predict risk more accurately than using income or deposits alone.
   → Action: Integrate cluster segmentation into loan approval and marketing strategies.
4. The "spread" in financial behavior is a better indicator of risk than just account balances or loyalty tiers.
   → Action: Monitor behavioral changes within each cluster dynamically over time.

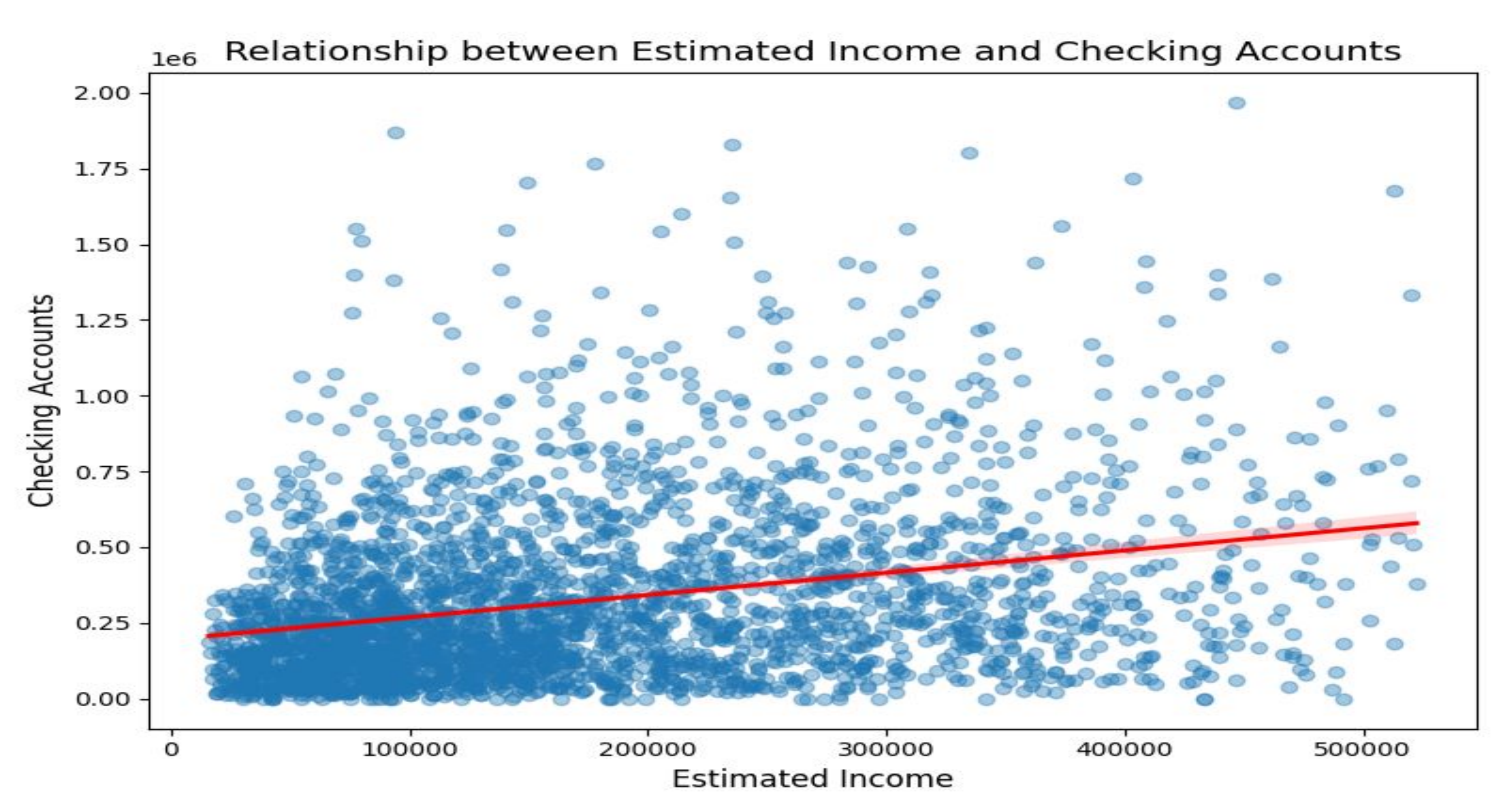
Bi-Variate Financial Relationships

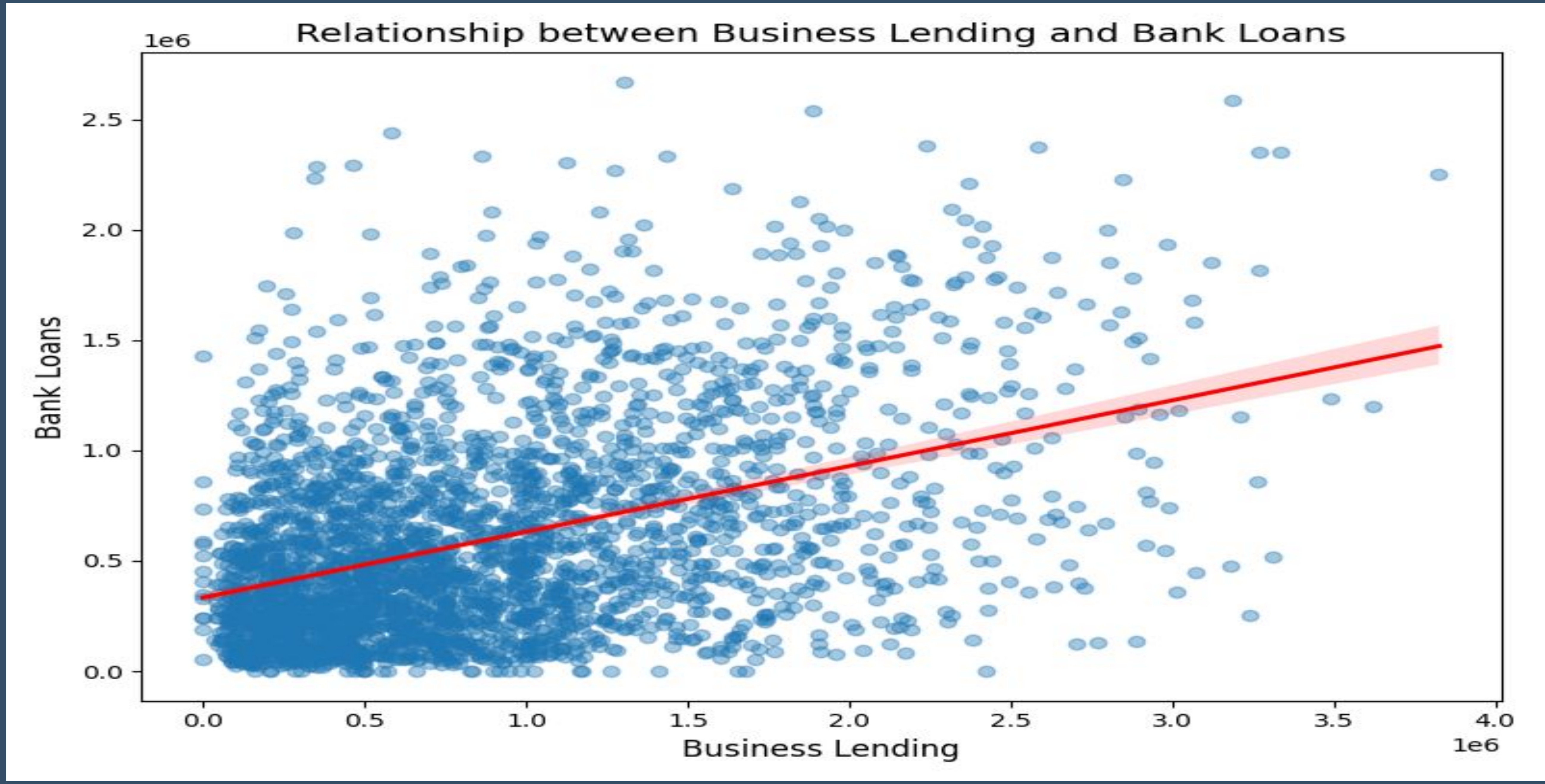


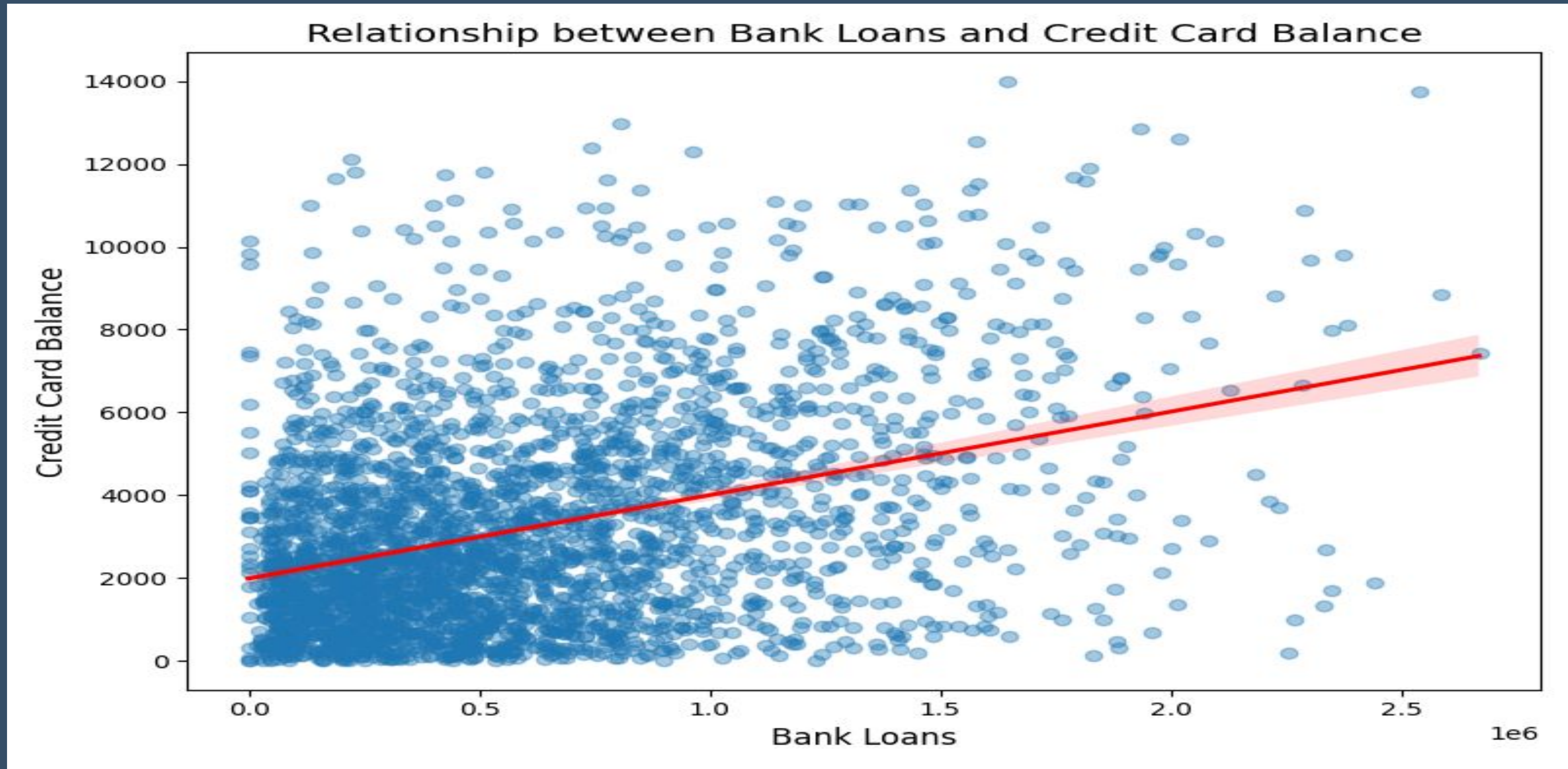




Bi-Variate Financial Relationships



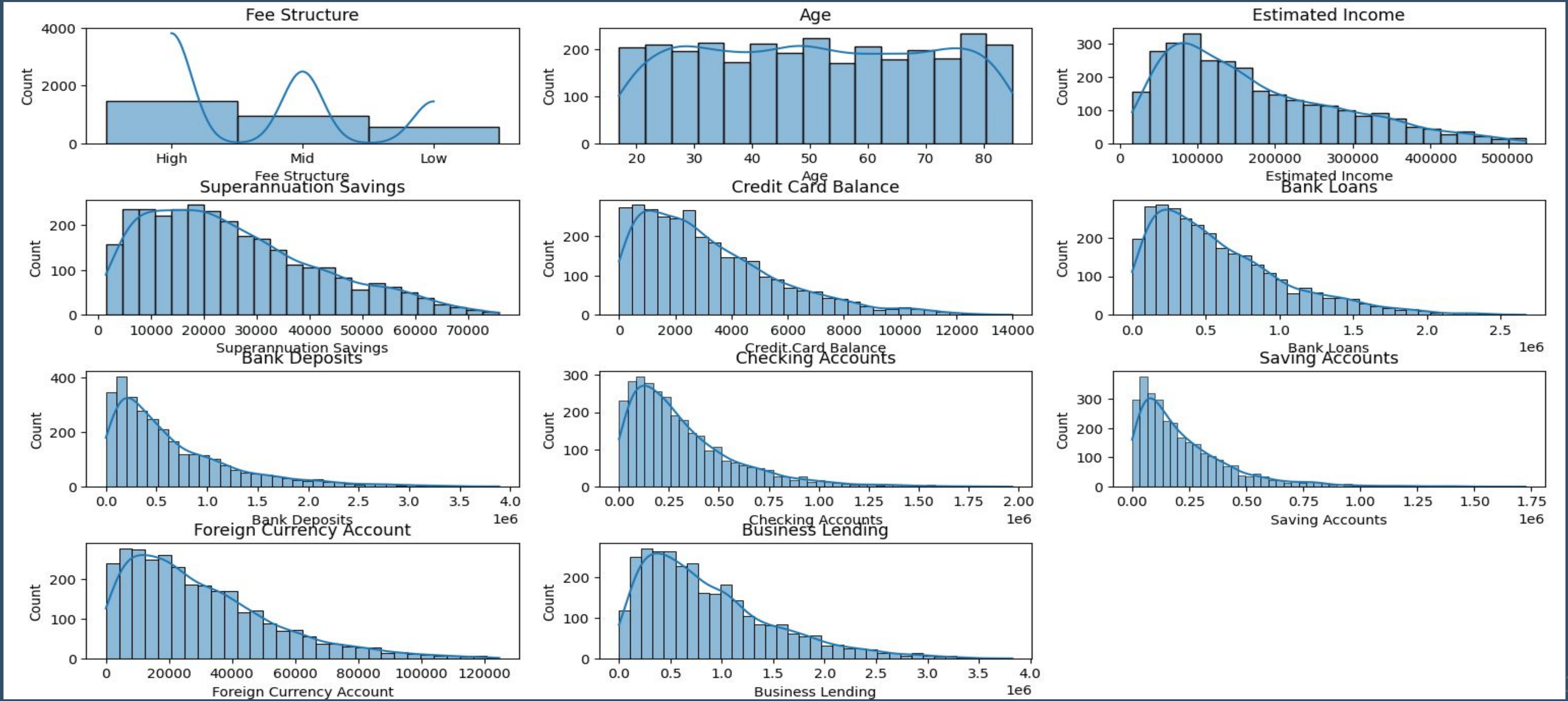




Paires of Financial Relationships – Insights and Actions

Financial Pair	Key Insight	Action to Take
Bank Deposits vs. Saving Accounts	Higher deposits usually align with higher savings — disciplined customers.	Prioritize for savings and investment upselling.
Checking vs. Saving/FC Accounts	Strong banking engagement across multiple accounts.	Offer bundled financial products.
Age vs. Superannuation Savings	Young customers with high retirement savings → liquidity risk.	Flag for income verification and cash flow assessment.
Estimated Income vs. Checking Accounts	Higher income often leads to higher checking balances.	Cross-sell premium banking services.
Bank Loans vs. Credit Card Balances	Heavy borrowers carry multiple debts → increased default risk.	Apply stricter loan approval and debt checks.
Business Lending vs. Bank Loans	Businesses stack multiple borrowings, raising risk.	Introduce collateral requirements or phased lending.

Univariate Analysis – Financial Metrics



Univariate Analysis – Financial Metrics Insights

Metric	Key Insight & Action
Age	Majority between 25–45 years → Focus lending and investment offers.
Estimated Income	Low-to-mid incomes dominate → Design middle-income products.
Superannuation Savings	Modest savings; few outliers → Promote retirement planning.
Credit Card Balance	Right-skewed; few high balances → Monitor and manage high-debt customers.
Bank Loans	Mostly small-to-moderate loans → Encourage controlled borrowing, limit big exposures.
Bank Deposits, Checking, Saving Accounts	Healthy savings behavior → Upsell investment and wealth products.
Foreign Currency Accounts	Low ownership → Target niche premium services.
Business Lending	Few heavy borrowers → Apply stricter evaluation for business loans.

Hidden Risk Patterns

Analysis of financial behaviors revealed two critical hidden risks:

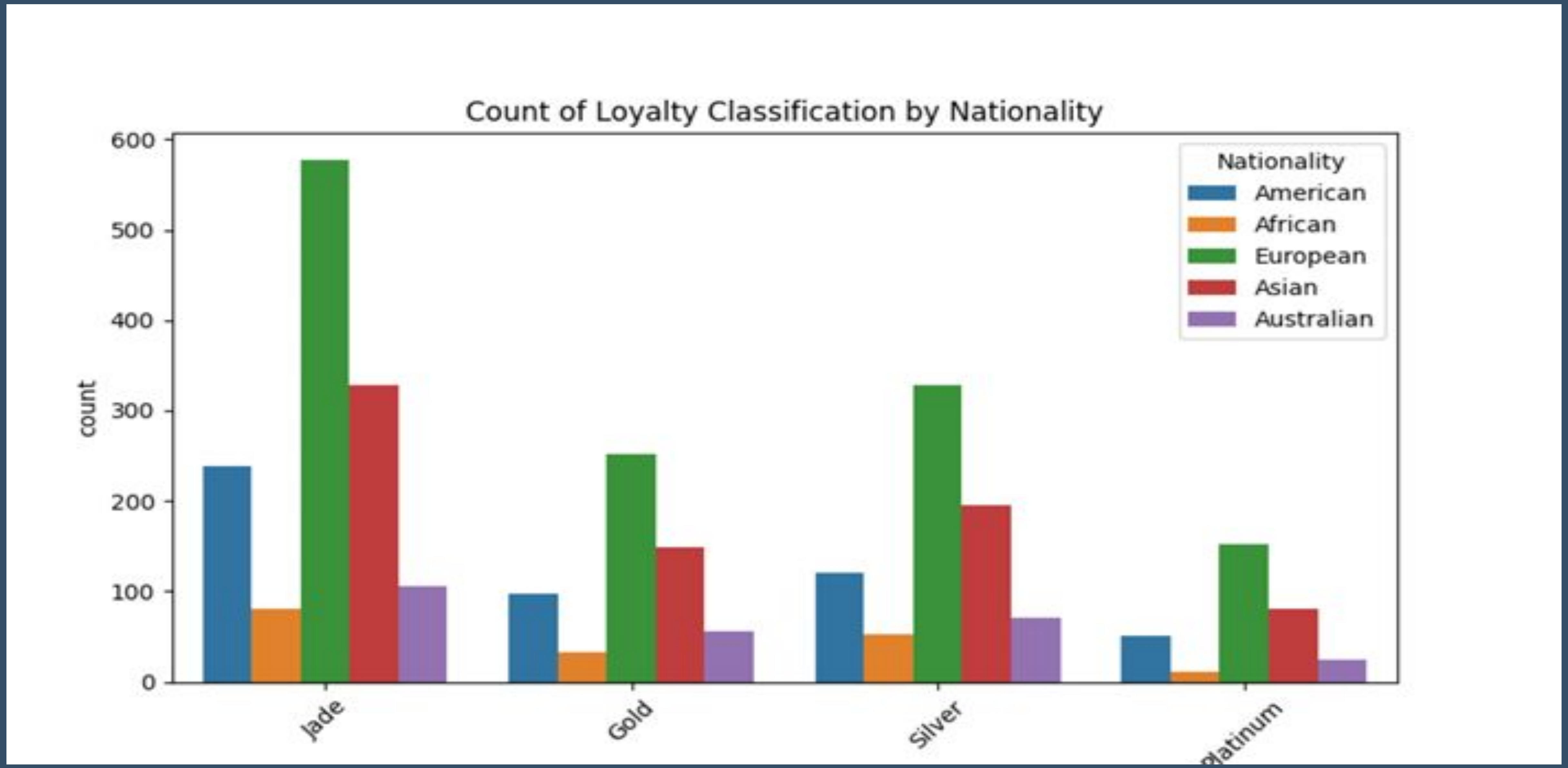
1. Young customers (<30) with abnormally high superannuation savings, indicating aggressive, illiquid investments that may lead to future cash flow constraints.

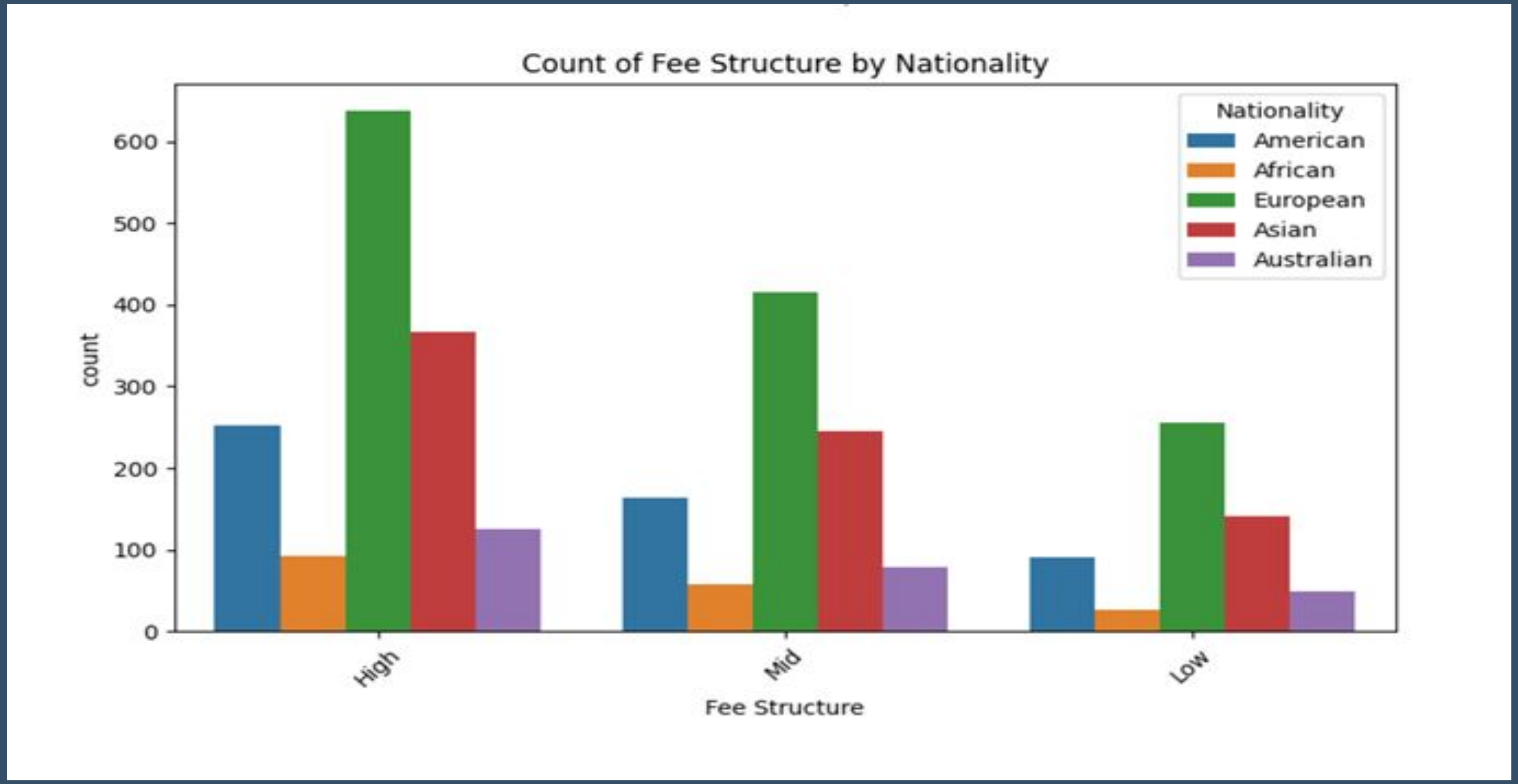
2. Customers with high bank deposits but disproportionately low savings, suggesting liquidity mismatch and higher risk under financial stress.

These patterns have been incorporated with moderate weight (5–10%) into the risk scoring framework to capture subtle but impactful risk behaviors.

Loyalty & Fee Structures ≠ Safety

1. Contrary to traditional belief, premium loyalty classifications (Platinum/Gold) and high fee structures do not correlate with lower risk.
2. Analysis shows high-risk customers are often represented within premium tiers, particularly among American and Asian nationalities.
3. Therefore, loyalty programs should not replace thorough risk evaluation and should be supplemented with behavioural and financial risk checks.





Risk Scoring Framework – Methodology

Step 1: Identify Key Risk Factors

Step 2: Assign Weights and Scoring:🧠 Formula:

Total Score = Weighted sum across all variables (0–100).

Variable	Weight	Scoring Highlights
Risk Weighting Level	30%	Levels 4–5 heavily penalized
Credit Card Balance	20%	>$10K = High Risk
Properties Owned	15%	3+ Properties = High Risk
Income Band	10%	High Income = Higher Points
Nationality	10%	Asian > American > European
Superannuation Savings	10%	2x+ normal = Higher Risk
Deposit/Savings Gap	5%	High deposits + low savings = Risk flag

Risk Scoring Framework – Execution and Actions

Step 3: Calculate and Classify

- Example: A high-risk customer scores 60/100 based on income, properties, credit behavior.
- Interpretation: Higher score = higher risk exposure.

Step 4: Apply Risk Tiers and Actions

Score Range	Risk Tier	Recommended Action
0–30	Low Risk	Approve at best interest rates
31–50	Moderate Risk	Approve with higher interest or limits
51–70	High Risk	Require strict collateral; short loan terms
71–100	Extreme Risk	Reject or escalate to manual underwriting

Risk Scoring Framework – Execution and Actions

1. Key Enhancements Dynamic Adjustments: If a customer's credit card balance spikes, increase their score by 10–20 points. If they pay off a property, reduce their score by 15 points.
2. Exceptions: Loyalty (Platinum/Gold): Subtract 10 points (reward trust but verify).
3. Young High-Savers: Add 20 points if liquidity ratios are poor.
4. Automation: Integrate with transaction data to update scores in real-time.
5. Why This Works Data-Driven:
   a. Uses the strongest correlations from the EDA (e.g., Asians > Americans in risk).
   b. Balanced: Weights reflect real-world impact (e.g., Risk Level matters more than nationality).
   c. Actionable: Clear thresholds for approval/rejection.

Bank Numbers

$1.77 bn

Bank Loan

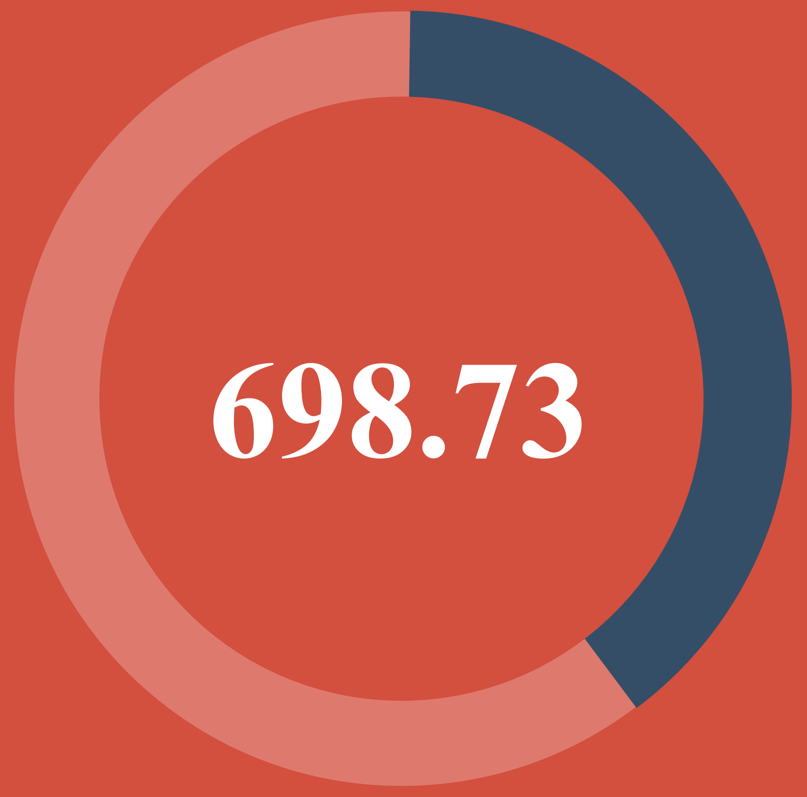
$2.01 bn

Bank Deposit

$2.60 bn

Business Lending

Saving amount and Checking amount



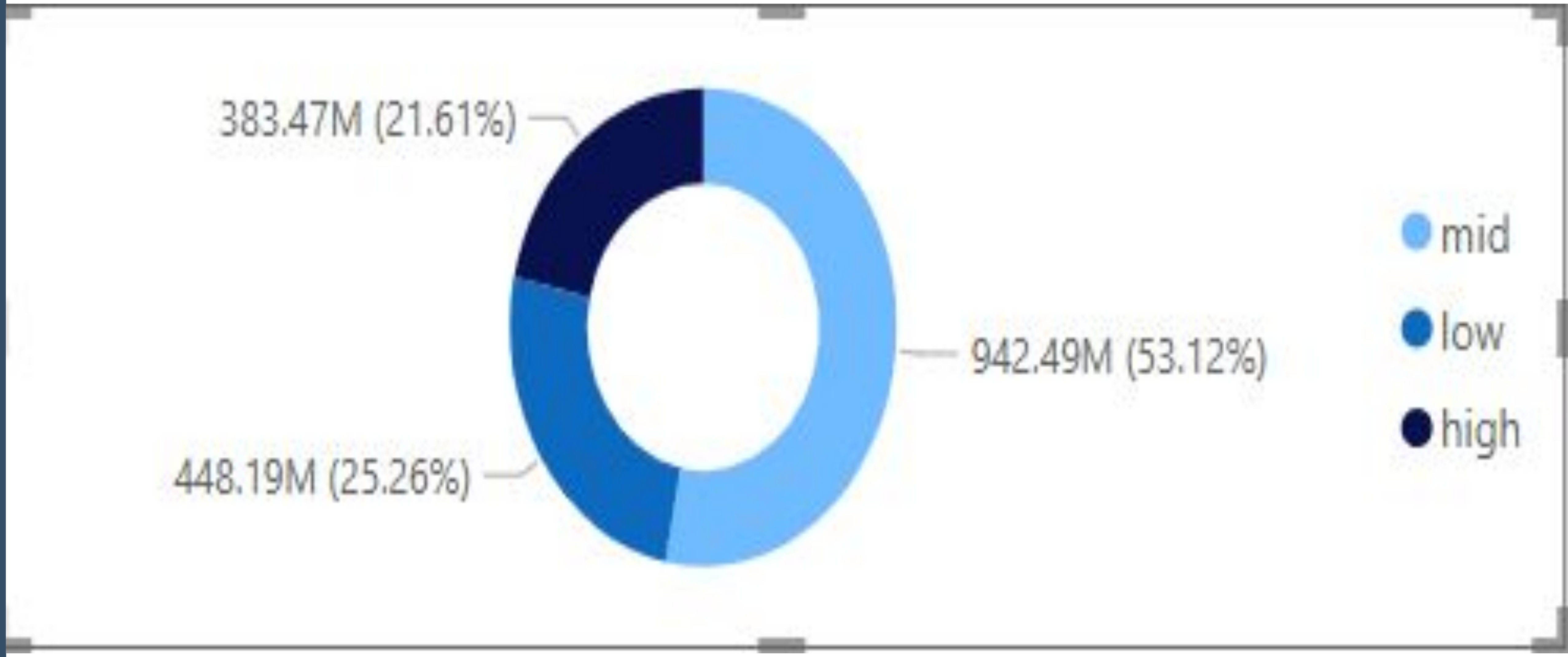
Saving Account Amount (in Millions)

Checking Account Amount (in millions)

$ 89.65 M

FOREIGN CURRENCY AMOUNT

Bank Loan by Income band



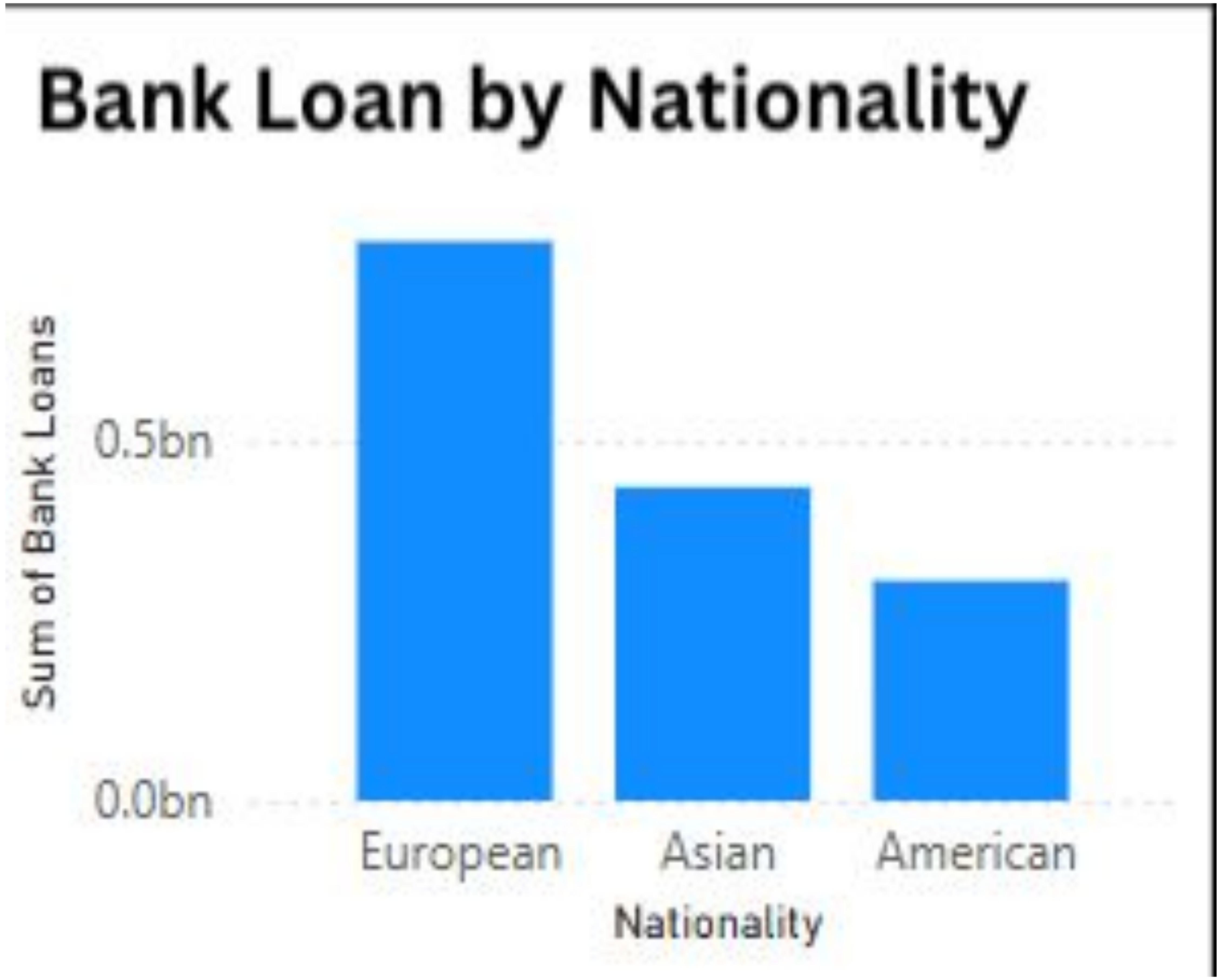
What we observe?

Bank loan is highest for Mid Income band and lowest for High Income band.

What we observe?

Bank loan is highest for European countries and lowest for Australian countries.

Bank Loan by Nationality



Conclusion

"We turned raw financial data into real, actionable intelligence — empowering the bank to lend smarter, protect assets, and grow stronger with every decision."

THANK YOU !

Any Questions?