Financial Transaction Report: A Multi-dimensional Analysis for Informed Business Decisions and Financial Management

Group 3: Ming Thet Paing, Aung Kyaw Win, Htet Myat Phone Naing, Thant Thaw Tun, Khair Hossin, Bhone Khant Theim

Parami University

CS 251-A: Statistical Programming with R

Dr. Mohamed Megheib

December 9, 2024

Final Project Report

1. Abstract

The development and widespread use of digital payment systems such as mobile wallets, QR scans, and card payments have made people purchase goods and services more conveniently. However, this increase in convenience also brings up a critical issue, which is 'impulsive buying behaviors', partly triggered by targeted advertisements and personalized recommendations. One of the negative consequences of impulsive purchases is unmanageable debt. To bring these consequences to a halt, a good depth of financial management knowledge and skills is necessary. Since financial institutions such as banks act as primary intermediaries between consumers (people) and money, these institutions can either educate consumers regarding healthy financial management or provide them with strong financial planning. Therefore, the authors collected a dataset from a U.S. financial institution that encompasses transactions, card data, users' card information data, and Merchant category codes (MCC). They did data analysis using R to extract customer spending patterns, behaviors, and debt situations. This study aims to reveal meaningful insights regarding customers' spending habits and their financial situations in that U.S financial institution. We aim to provide these insights back to that U.S. financial institution so that the institution can make more informed decisions and provide assistance to consumers for safer financial decisions and purchases.

2. Introduction

The rapid technological advancements in mobile wallets, online banking, and digital payment systems have changed the way people make purchases, from using the traditional cash-paying method to more innovative digital methods. The proliferation of these digital payment methods helps people efficiently make purchases. People can now use credit and debit cards, scan payments, and use mobile wallet applications to pay for bills, goods, and services instead of cash. This upgrades the way people live, raises their living standards, and provides them with a more efficient lifestyle. The digital payments market is growing at a rapid rate, and it is expected to expand in the future (Putrevu & Mertzanis, 2023). According to data from Statista (2021), the value of the digital payment market was around 4.7 trillion U.S. dollars in 2020 (Putrevu & Mertzanis, 2023). This indicates that digital payments such as credit cards and QR codes are increasingly used among people, although they have not fully replaced cash payments. For instance, credit cards have become an integral part of society since they provide convenience and flexibility to people's finances. According to the insight from the Clearly Payments Platform (2023), there are approximately 1.25 billion credit card holders globally in 2023. However, the increase in the level of convenience when making payments has changed the spending behaviors of people. According to Boopathy and Kanagaraj (2023), something called ‘impulsive purchases’ occurs when people can make payments more conveniently than ever before due to the development of digital payment systems. These impulsive purchasing behaviors have serious negative consequences, one of which is gaining unmanageable debt. The latest consumer debt data from the Federal Reserve Bank of the United States revealed that the total credit card debt of Americans in the third quarter of 2024 was 1.166 trillion dollars (Yee, 2024). Particularly, youths are vulnerable to impulsive purchasing behaviors due to a lack of strong self-control and high exposure to targeted online advertisements (Nyrhinen et al., 2024). Therefore, it is important to ensure that youths between 16 and 35 years old have financial management knowledge so that they will not get trapped in debt or impulsive buying behaviors. Since banks serve as the forefront institution that directly provides financial services to people, they have the capacity to inform and consult their customers regarding healthy financial management. Hence, the authors of this study work on analyzing the “Financial Transactions Data” dataset from a U.S. finance institution, which is uploaded on Kaggle. There are three primary objectives behind this study and data analysis. The first objective is to extract meaningful insights regarding the customers’ spending behaviors and debt differences based on income and age. These insights will be shared back to that financial institution so that the institution can make customized saving plans, credits, and loans suitable for different age and income groups. The end goal of this objective is to equip customers with healthy financial management plans, and the authors approach this goal by sharing revealing meaningful insights to that institution. The second objective is to extract information regarding spending and transaction trends across the United States to figure out potential business opportunities for business owners. The third objective is to extract customer insights regarding the correlations between credit scores and income, credit scores and spending, and the age group and average debt. These insights will also be shared with that financial institution so that the institution can make informed decisions for better financial management, customer acquisition, customer segmentation, and expansion.

3. Literature Review

Franco Modigliani and Richard Brumberg introduced something called "The Life Cycle Hypothesis" theory in 1954, which provides insights into how people make financial decisions throughout their lives (Chowdhury, 2023). The key takeaway of the theory is that the financial objectives and priorities of people are not static. Instead, these objectives and priorities change as people age. A good example of this concept is that people tend to prioritize investing in education when they are young, and they prioritize saving or doing investments for retirement plans when they reach their 30s or 40s. While the Life Cycle Hypothesis" provides the foundational framework that explains people's financial behavior, the rise of digital technology and its consequences challenge conventional financial strategies. People in the past might have had fewer and more deliberate priorities; however, the changes in priorities tend to ramp up due to digital technology developments and the prevalence of 'impulse purchases'. The 'impulse purchases', caused by the proliferation of digital payment systems, can make people sway from choosing concrete financial objectives or priorities. Impulsive purchasing behaviors refer to a sudden willingness to buy something without prior planning and consideration of post-purchase consequences (Nyrhinen et al., 2024). These behaviors are triggered by the targeted advertisements used by companies that are based on consumers' search behaviors, web browsing history, and other available information on the internet (Nyrhinen et al., 2024). Therefore, it is important to convey to young people that they are having impulse purchasing behaviors. Once it is found out, make them aware of the issue and educate them about financial management.

Credit cards are a common digital payment method. They are widely used around the world and are preferred by a lot of people worldwide. In terms of credit card users per country, Canada has the highest number of credit card holders, with a record indicating that 82.7 percent of the country's population own credit cards(Best, 2021). Meanwhile, 66.7% of the population have credit cards. The utilization of these credit cards brings both benefits and challenges. The bright side of credit cards is that they can be used to pay for unexpected emergency expenses such as repairing cars, accidents, and fixing computers. Additionally, since credit cards are not attached to the actual bank accounts, using credit cards can also reduce the risk of losing all the money when the card is lost or stolen. However, on the other hand, these cards can make people overspend and put them in debt with high interest rates. The average interest rate on people's credit cards in America was 24.61% in 2024, according to Credit Card Debt Statistics studies (Schulz & Shepard, 2024, November 13). With this, the total credit card debt of Americans is 1.166 trillion dollars, which is the all-time highest balance since the Federal Reserve Bank of New York started tracking the amounts (Schulz & Shepard, 2024, December 6, 2024). So, a lot of people are in debt when using digital payment systems such as credit cards. At that time, young people are, due to targeted advertisements and tech-driven personalized recommendation systems, influenced by impulsive purchasing behaviors. Therefore, examining whether these young people with credit card payment systems have debts has become important. To examine that, the authors analyzed a financial dataset from a U.S. financial institution to see the correlation between young people and the amount of debt.

Credit scores serve as a strong indicator for banks and credit card service providers in determining one's ability to repay the credit card debt. In general, credit card providers identify people with two categories: low-risk borrowers and high-risk borrowers. Low-risk borrowers are the ones who can repay the debt in time and possess a strong credit history. High-risk borrowers, on the other hand, are the ones who can possibly fail to repay their debt in time and do not possess a strong credit history. Many researchers argue that these credit card service providers, including banks and other financial institutions, consider income when calculating credit scores (Beer et al., 2018). However, many credit model-building agencies refuse this argument. Instead, these agencies indicate that the estimations of credit scores are based on length of credit history, credit limit utilization, level of indebtedness, and debt payment history (Beer et al., 2018). Although the agencies insist on not using income as a parameter for credit scores, the information they use, such as debt payment history, credit history, and level of indebtedness, are associated with income. Hence, whether income and credit scores are positively correlated is questionable. The authors of this report analyze the “Financial Transactions data” dataset from a U.S. finance institution to see if these two, income and credit scores, are positively correlated. Additionally, the correlations between credit scores and credit limits, and that of the number of credit cards and credit scores are examined to provide valuable customer insights into that U.S finance institution.

4. Data Description

The dataset that we are working on is called ‘Financial Transactions Dataset’, which is the combination of 4 sub-datasets: Transactions data, Cards_data and Users_data, and Merchant Category Codes. The dataset is from one of the most well-known data platforms, Kaggle, and the source of the dataset is a financial institution. The author who published this data on Kaggle has recommended using this dataset for Financial Analysis, Fraud Detection, and AI-power banking Solutions. The usability of this dataset is 10.0, which is the highest score in Kaggle. (The usability score is calculated by Kaggle, and the calculation is based on completeness, credibility, and compatibility. This particular dataset we are using has received 100% for all of these three parameters.) As mentioned above, three sub-datasets encompass the ‘Financial Transactions Dataset’, and they are in CSV file formats. The Transactions_data dataset includes information such as transaction amounts, timestamps, and the merchants' details, including merchant cities. The timeline of the data points is from 2010 to 2019. The Cards_data dataset includes credit and debit card details, with columns such as card limits, types, and activation dates. Users_data includes demographic information about customers and detailed account-related information.

5. Methodology

5.1 Data Collection: The dataset used to write this report is collected from the Kaggle Platform. The authors of this report did not do primary data collection but instead downloaded the data from reliable third-party organizations.

5.2 Exploratory Data Analysis (EDA) /Data Wrangling: The first step is reading the dataset files and merchant _categories JSON file. There are three datasets, namely transactions_data, users_data, and cards_data, plus one JSON file. Then, the authors use the summary to see and understand the overviews of each of the data files.

- Cleaning up the transaction_data (named as 'transactions' in R script) dataset:
    1. $ signs and minus signs from the "amount" column of the dataset are removed.
    2. The values in the "amount" column are changed to a numeric type.
    3. The "use_chip" column, which shows the transaction type, is changed to the factor type.
- Cleaning up the users_data (named as 'user_data' in R script) dataset:
    1. $ signs of the "yearly_income" column is removed.
    2. The values in the "yearly_income" column are changed to numeric.
    3. $ sign in the "total_debt" column is removed.
    4. The values in the "total_debt" column are changed to numeric.
- Cleaning up the merchant categories (named as mcc_df in R script) JSON file:
    1. The "mcc_code" column is converted into integers.
- Cleaning up cards_data (named as card_data) dataset:
    1. $ signs of the "credit_limit" column of the dataset are removed.
    2. The values in the "credit_limit" column are changed to numeric.
- Merging datasets:

1. The user_data and transactions dataset are merged by using a common identifier, which is 'client_id'. The merged dataset is named "data1".
2. The card_data and transactions datasets are merged by using a common identifier, which is 'client_id'. The merged dataset is named "data2".
3. Merchant Category is joined into "data1", which is the combination of user_data and transactions dataset. The joined dataset is named "data3".
4. The user_data and card_data datasets are merged by using a common identifier, which is 'client_id'. The merged dataset is named "data4".
5. The authors omitted null values in the 'data3' dataset by using the R code "na.omit(data3)'.

5.3 Linear Regression and Model Evaluation:

1. Feature Selection: Relevant predictors were chosen based on the domain knowledge and exploratory data analysis findings
2. Visual Analysis:
    a. Residual Plots to check for randomness in the residuals, confirming the linearity assumption
    b. Predicted vs. Actual Plots were used to visualize how closely predictions align with actual values
3. Training and Testing: The dataset was split into training (80%) and testing (20%) sets to evaluate the model's performance on unseen data. The linear regression model was trained on the training dataset using the Random Forest Regression model.
4. To assess the model's predictive power, the following metrics were used on the test dataset.
    a. Mean Absolute Error (MAE): Measuring the average magnitude of prediction errors in predictions provides an intuitive sense of prediction error.
    b. Root Mean Squared Error(RMSE): Captures the average magnitude of errors but penalizes larger errors more than MAE.

c. R^2 Score: The R^2 score ranges from 0 to 1 where values closer to 1 indicate a better fit.

6. Findings

6.1 Analysis of the incomes and geographical locations of the users using interactive maps

Understanding the geographical locations of the users and their impacts on income level can offer insights for both strategic market segmentations for businesses and addressing community needs for authorities.

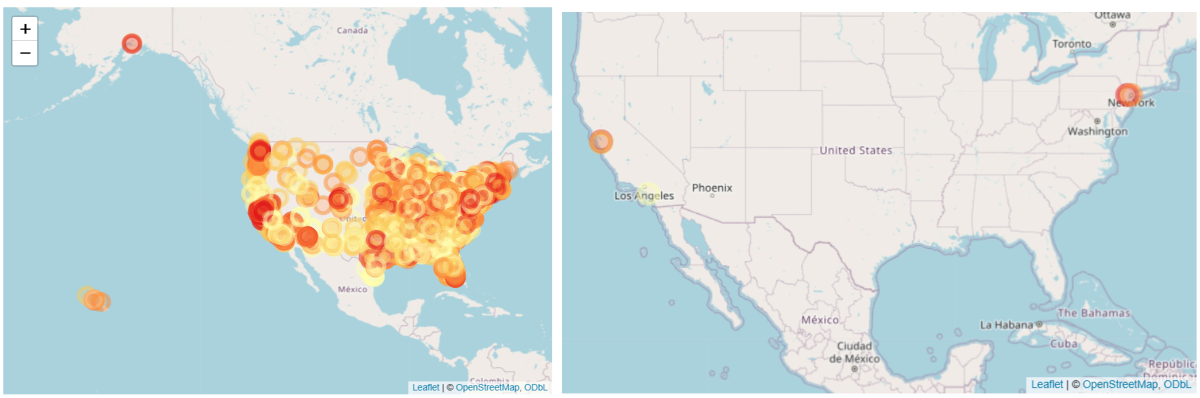

Figure 1: Interactive maps that show the exact geographical locations of the users of the bank

By using the Leaflet function, interactive maps are produced based on the longitudes and latitudes of the users provided by the user's data. The first map represents all the users from the dataset, offering a comprehensive overview of the geographical locations of the users. In contrast, the second one represents the top five users with the highest incomes. According to the first picture, it can be seen that the users highlight a wide representation of diverse states within the United States. Among them, the second picture highlights that the users with the highest incomes are from major financial states like New York and California. This could suggest the significance of geographical factors in income level since these regions with great economies can offer better financial opportunities.

6.2 Comprehensive analysis of payment transactions and spending

To understand users' or customers' behaviors including their spending patterns and market trends, it is crucial to conduct a comprehensive analysis of payment transactions and spending. Firstly, The above bar chart covers the total transaction payments for each of the top five transaction types using two payment methods: Online Transactions (red color) and Swipe Transactions (blue color). The transaction types include Grocery Stores and Supermarkets, Miscellaneous Food Stores, Service Stations, Eating Places and Restaurants, Drug Stores, and Pharmacies.

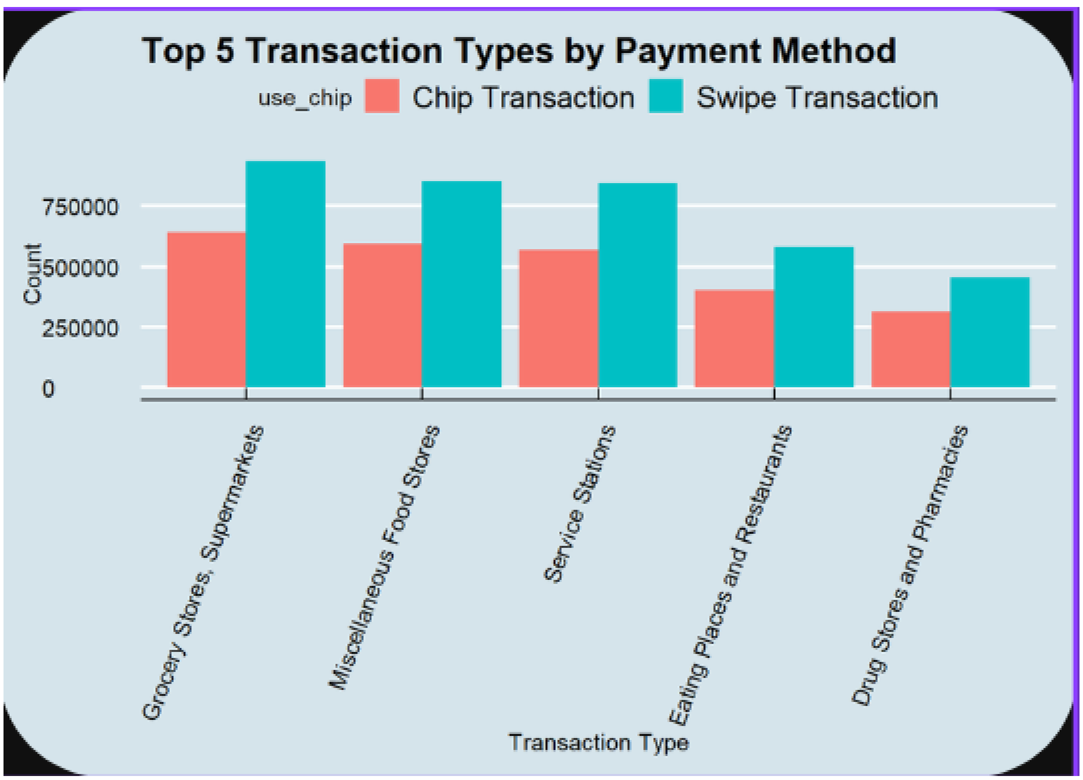

Figure 2: Top 5 Transaction Types by Payment Method

According to the bar chart, swipe transactions are more popular than online transactions in all categories. Thus, it can be said that individuals prefer to purchase items in stores generally. **Recommendation:** The usage of different payment methods for these businesses can help them make informed decisions on their financial management and optimize payment methods.

The following time series analysis covers from 2010 to 2020 for two payment methods: Online Transactions (represented in red) and Swipe Transactions (represented in blue).

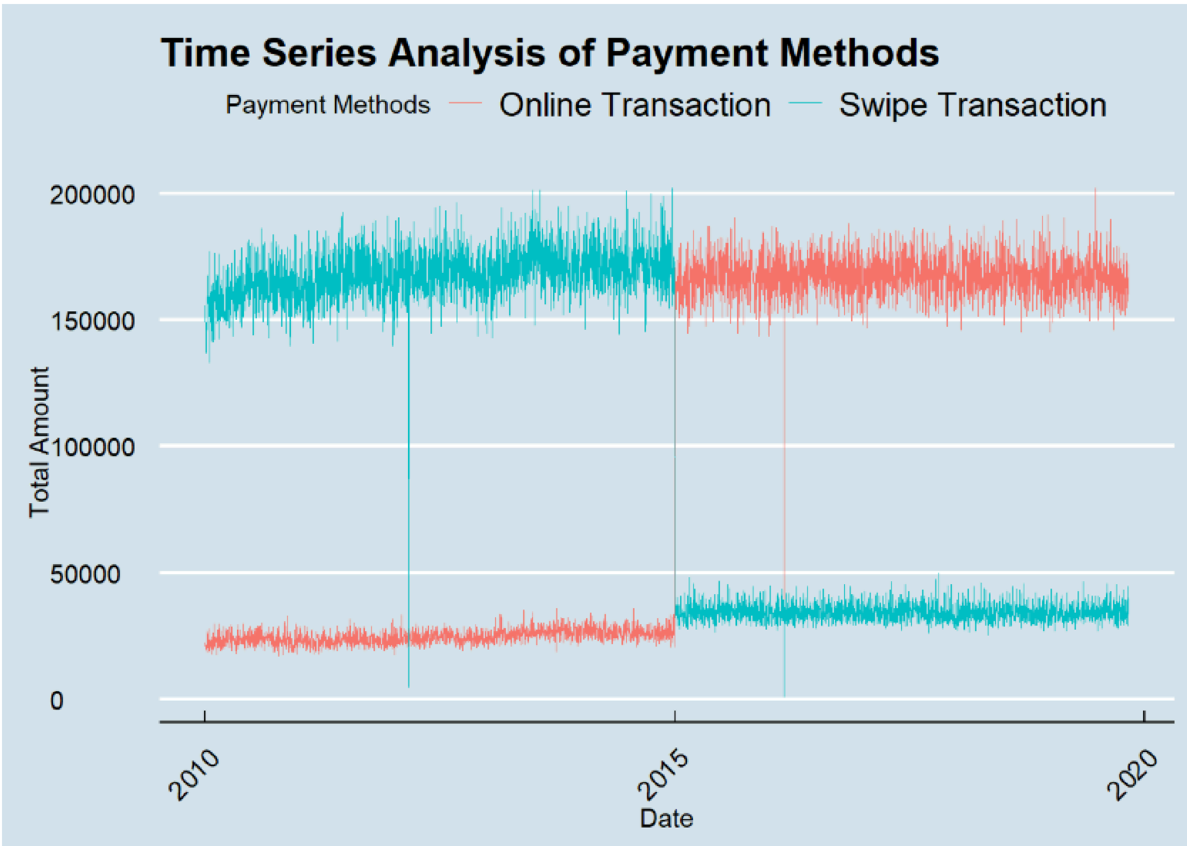

Figure 3: Time Series Analysis of Payment Methods

Before 2015, swipe transactions were at much higher levels than online transactions, meaning that people relied more on swipe transactions by using this method to transfer bigger amounts of money. However, their position changed significantly since the swipe transactions level dropped from around 150,000 to under 50,000, and the online transactions level increased from under 50,000 to around 150,000. Thus, the transition from physical swipe payment to online payment can be found. It might indicate the broader adoption of e-commerce and digital wallets. Moreover, there is a sudden drop to zero for both swipe and online transactions in a016. It might suggest a sudden interruption or anomaly in transactions, possibly a technical error from the banking institution. From this result, businesses can understand the change in customer preferences on payment methods and make informed decisions.

The following horizontal bar plot revealed the top 10 merchant cities by transaction count, excluding transactions categorized as “ONLINE”. Online transactions dominate the entire plot, preventing the reader from clearly identifying the significance of physical merchant locations. Thus, the exclusion of online transactions ensures that the focus remains on physical merchant locations, offering a clear perspective on geographically driven consumer behavior.

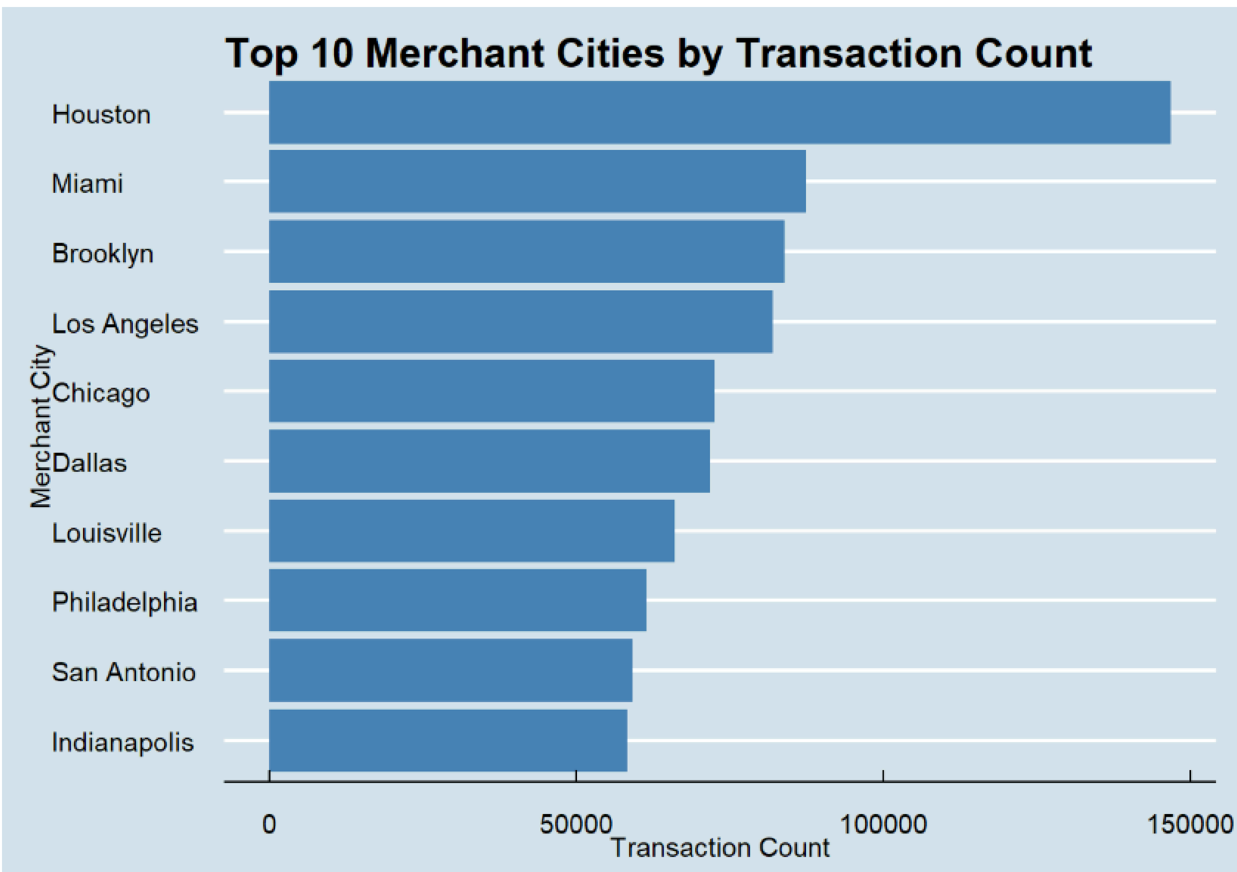

Figure 4: Top 10 merchant cities by transaction count

This bar plot ranks the top 10 merchant cities by transaction counts from the highest to lowest. It is obvious that Huston stands out for the significantly highest transaction volumes with around 150,000 dollars. Thus, Miami starts at a bit under 50000, and from Miami to Indianapolis, there is a gradual decrease with a relatively small difference in transaction counts. It highlights the significant role of Huston as the biggest financial hub for individuals from users_data. This analysis provides valuable insights into regional economic trends and suggests opportunities for businesses to allocate resources, expand, and target the relevant locations strategically.

The following bar chart features the spending level low based on four income categories: low income (0–30,000 USD), middle income (30,000–70,000 USD), upper middle income (70,000–150,000 USD), and high income (above 150,000 USD).

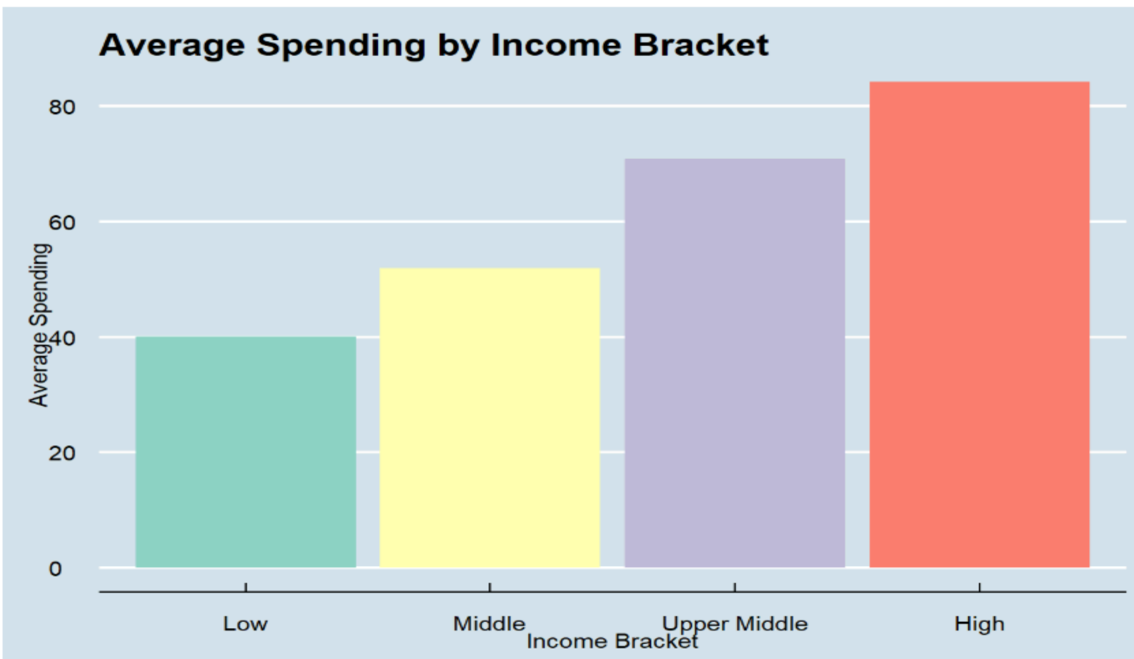

Figure 5: Average Spending by Income Bracket

A bar chart comparing average spending across these brackets revealed a positive trend, with low-income individuals spending around $40 and high-income individuals spending over $80 on average. This trend highlights a clear positive relationship between income levels and spending behavior, and the average spending of the highest group approximately doubles the average spending of the lowest income group. Hence, it can be concluded that the higher the incomes are, the higher the spending will be. This analysis can help banks design customized savings plans, loans, or credits suitable for the spending capacities of different income groups.

6.3 Analysis of the Relationship between Credit_score and Yearly_income

While some search suggests the positive impacts of yearly income on credit score, it is important to check if it is true for this users_data to help individuals make better decisions to improve their creditworthiness.

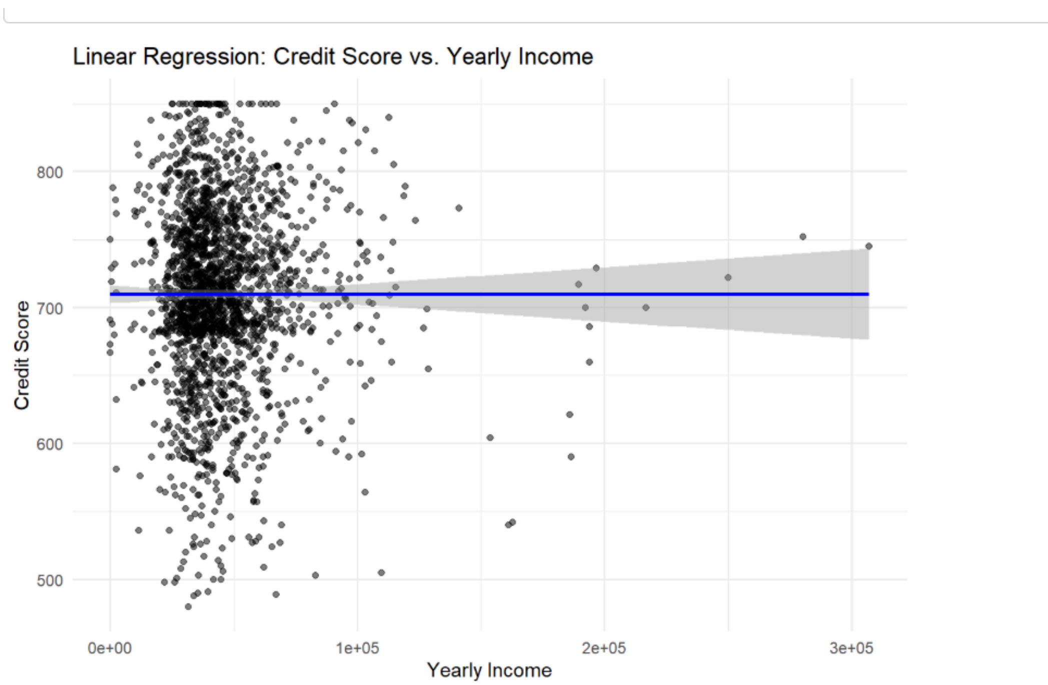

Figure 6: The linear regression model for credit score and yearly income

Based on the model indicating the relationship between credit score and yearly income, the linear regression includes **yearly income** as an independent variable and **credit score** as a dependent variable. According to this model, the intercept is estimated to be at 709.7, which implies the expected credit score of roughly 709.7 when the yearly income is zero. The coefficient for yearly income is 0.0000004872 (4.872e-07). Since it is a very small number and the graph shows an almost straight line, this coefficient is not statistically significant. Moreover, a very large p-value (0.994) suggests that yearly income does not have a statistically significant influence on credit score. An extremely low R-squared value indicates that yearly income does not explain the variations in credit score effectively.

Although multiple sources expect that yearly income has an impact on credit score, this dataset fails to show any significant correlation between yearly income and credit score of the users. It is because some users with higher incomes may not manage their finances wisely, leading to lower credit scores but some users with lower incomes have more responsible financial management, leading to higher credit scores. However, it is necessary to consider other factors that influence the credit score, such as credit history, debt payment, and so on.

For American youths, especially those with low income, the lack of a significant positive relationship between income and credit scores informs them that better income is possibly not enough to improve their credit scores. Thus, addressing American youth debt issues can require more comprehensive approaches and relevant financial education.

6.4 Analysis of Findings on Average Spending According to Credit Score Range

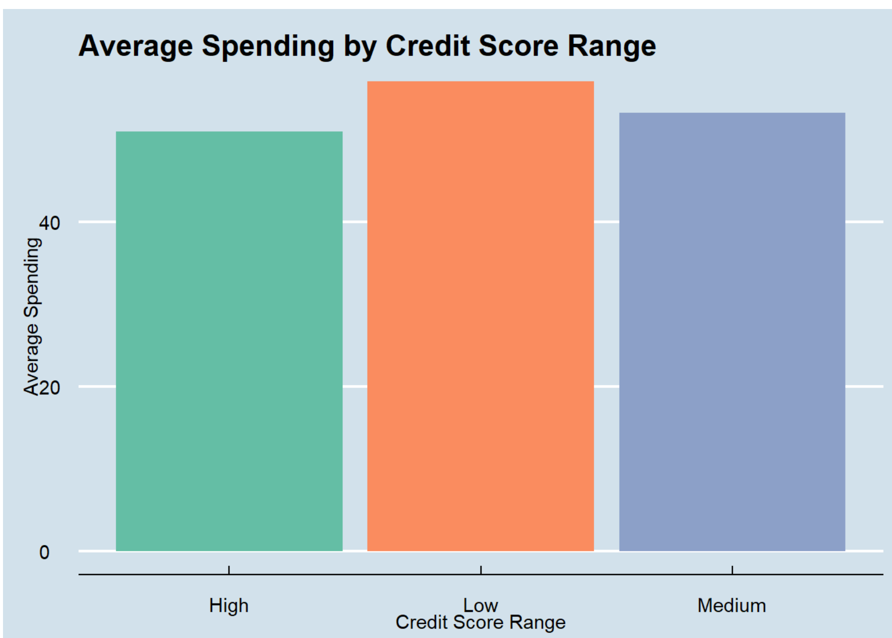

Figure 7: Data Visualization of Average spending by credit score range

Based on the figure mentioned above, the credit score range is divided into three(3) categories: low (in which the credit scores are below 600), medium (credit scores between 600 and 699), and high (credit scores at 700 and above). The figure includes the **Credit Score Range** as an independent variable and **Average Spending** as a dependent variable. By following the results of finding the average spending of individuals in those three brackets, it states that all of the average scores are above the line of 40(forty) units. The graph clearly shows that individuals with low credit score brackets had the highest amount of average spending, followed by those in the medium and high credit score brackets with a slightly lower amount of average spending. Thus, this analysis reveals that individuals with lower credit scores exhibit higher average spending compared to those with medium and high credit scores, despite financial constraints. This can contribute to the other possible factors of individual interests, loans, and debts. In the case of income not being a key driver of credit scores, those with low credit scores may have to rely on and are more likely to take more loans and interest.

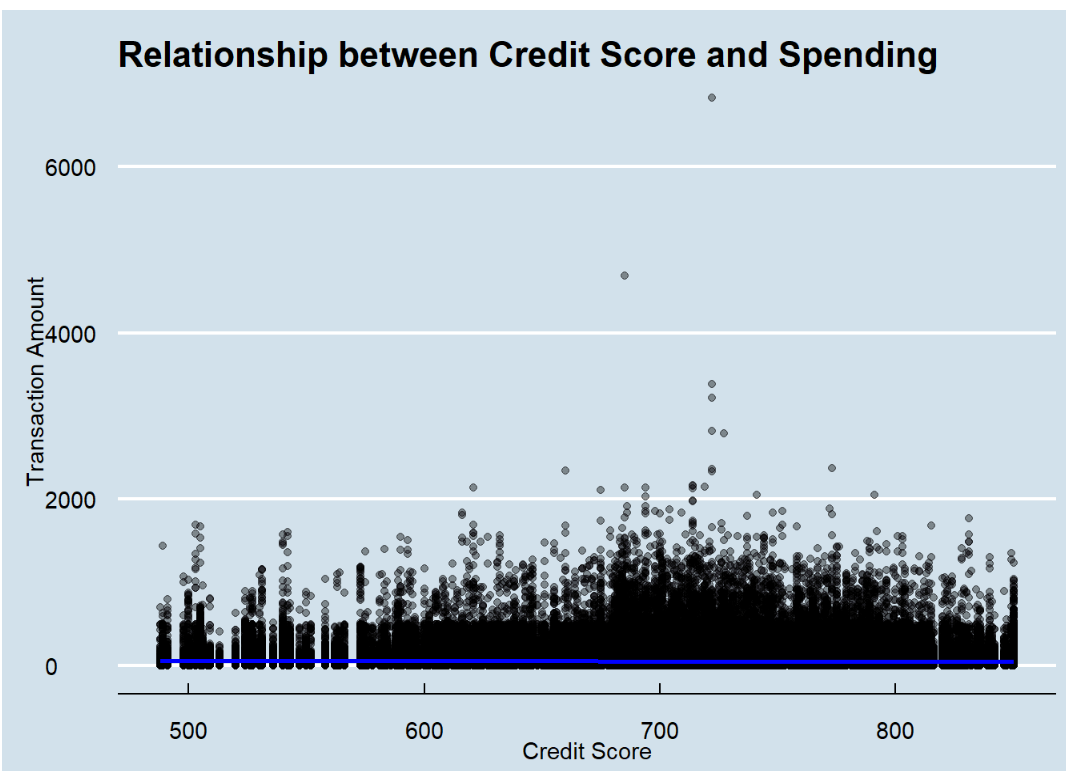

Figure 8:Analyzing the relation between credit scores and spending

The figure shown above is the linear regression visualization of analyzing the relationship between credit scores and the spending amount of individual users to determine if there is any significant effect of an individual's credit score on their transaction spending amount. The linear regression model above includes Credit Score as an independent variable, Transaction Amount as a dependent variable, scatter plots to show an individual's credit score and spending amount, and the blue regression line representing the relationship between two variables. As the result shows that the coefficient of credit score is negative(-) 0.0308, it means that the transaction amount of spending is decreasing by 0.0308 for every 1-point increase in credit score. For the residuals, the results show that the majority (75%) of the residuals exist between -41.2 (1st Quartile) and 20.0 (3rd Quartile) with the median at 17.4. However, there is the highest peak of the residual of 6768.5, indicating the presence of extreme individual credit scores.

6.5 Analysis of Findings on the Relationship Between Age and Average Debt

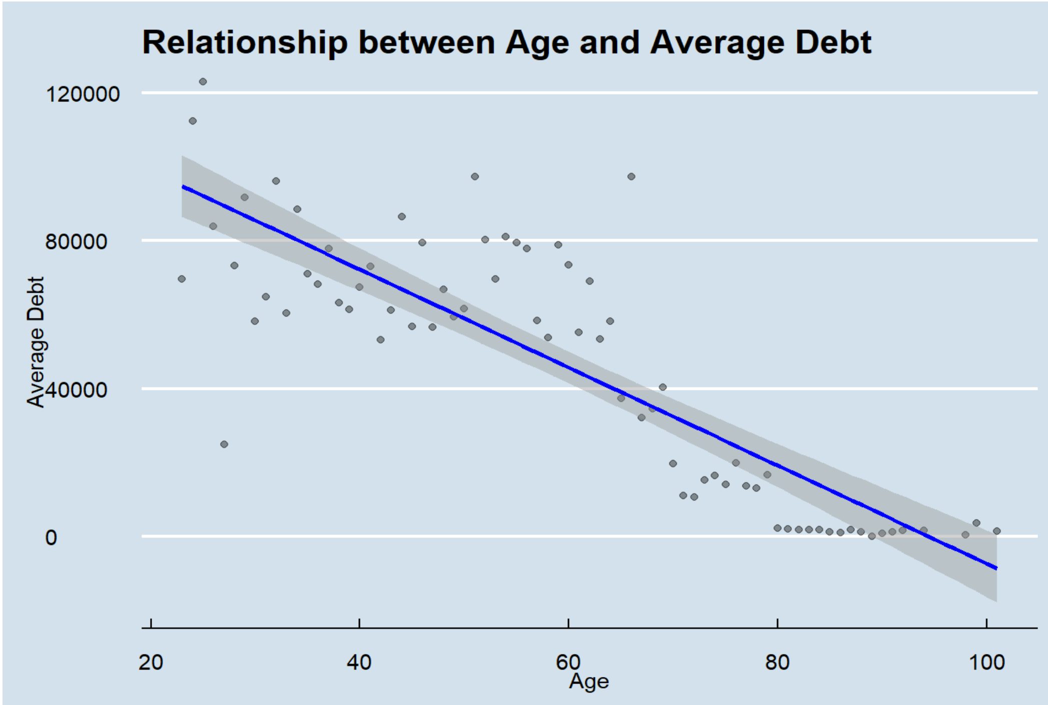

Figure 9: The linear regression model for Age and Average Debt

Based on the linear regression model mentioned above, by using the visualization of ggplot2, includes **Age** as an independent variable, **Average Debt** as the dependent variable**,** and **a blue linear regression line** crossing the plot in descending order. This analysis identifies a strong negative relationship between age and average debt because of the clear visualization of the linear regression line slope, indicating that the average debt gradually goes down as age increases. Even though scatter plots that stand for the average debt for a specific age show some fluctuations at certain ages, the trend is going downward. Thus, it can be concluded that younger individuals have a higher average debt than older individuals. This is because the older individuals have already paid off or have done repayment on debt on buying houses, things, or other financial issues and obligations to be spent on while the younger individuals are at the start of taking loans for their education, personal loans, and using credit debit card. This finding reflects the user's behaviors in terms of financial habits, responsibilities, or debt repayment capacity over time according to the age difference.

6.6 Analysis of the Relationship between Average Debt and Age Group

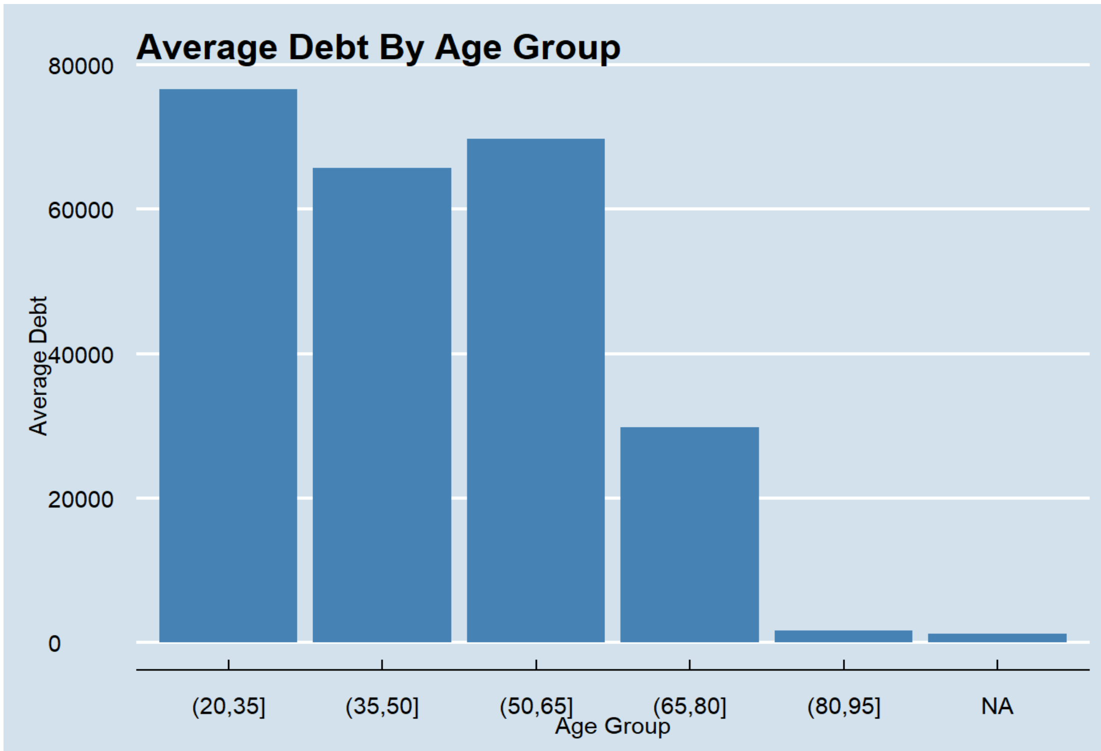

Figure 10: Bar Graph showing the relationship between Age Group and Average Debt

In this bar graph of data analysis, the individual users are categorized into different age groups: 20-35, 35-50, 50-65, 65-80, 80-95, and unavailable data for age groups. According to the results, the younger age group (between 20 and 35) has the highest average debt, nearing about $80,000, followed by the 35-50 and 50-65 groups. The older group of 65-80 has significantly lower debt levels than the younger group, with the amount nearly around, followed by the 80-95 age group exhibiting the lowest average debt. As a result, young individuals are likely to have the burden of debt. Thus, this graph also shows the negative trend of debt levels regarding age groups, which means that the lower the age of individuals, the higher the average debt amount of the individuals. It indicates a potential life-cycle effect on debt accumulation and repayment.

6.7 Analysis of the Relationship between Credit Limit and Credit Score

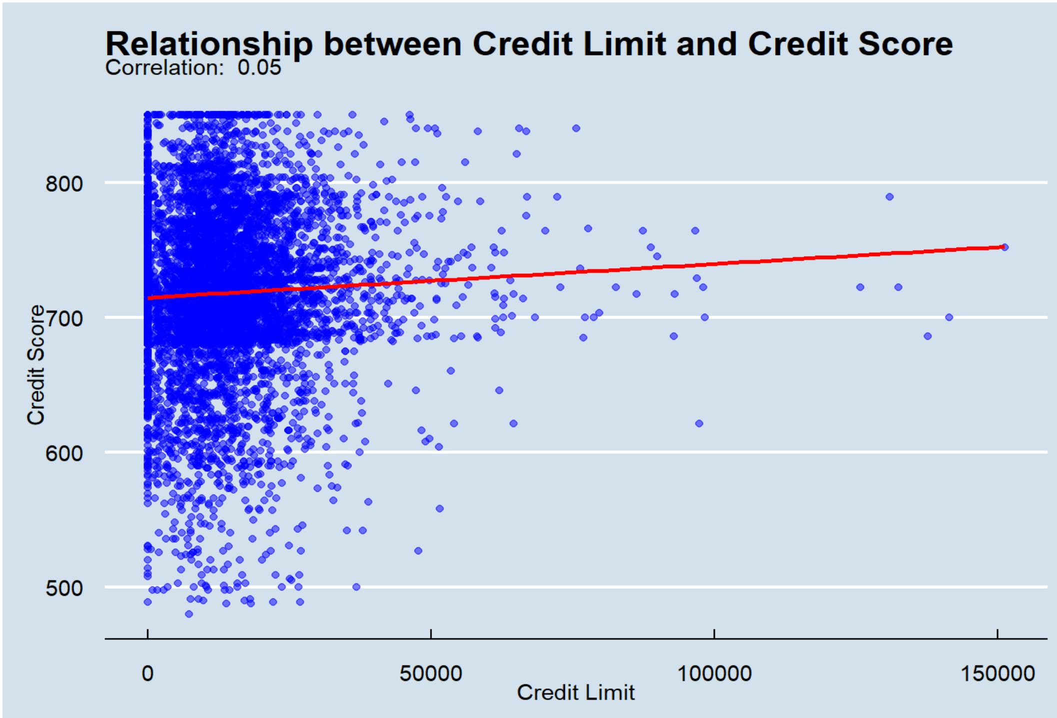

Figure 11: Data Visualization of Relationship between Credit Limit and Credit Score

Figure 11 includes **Credit Limit** as an independent variable, Credit Score as a dependent variable, a red line indicating weak positive linear correlations, and blue scatter points representing individuals' credit scores and limits. There is a very weak positive correlation (5%) between credit score and credit limit. According to the data in the scatter plot, it indicates that individuals with higher credit scores (above 700) occasionally had credit limits exceeding $100K. Most of the credit limits are clustered within a lower range, suggesting that other factors beyond credit score influence credit limit. Moreover, the red linear regression line (fitted) shows a slightly positive slope, highlighting that credit score alone does not fully determine credit limit allocations, reflecting a weak positive correlation.

6.8 Exploratory Analysis of the Relationship between Yearly Income and Credit Limit

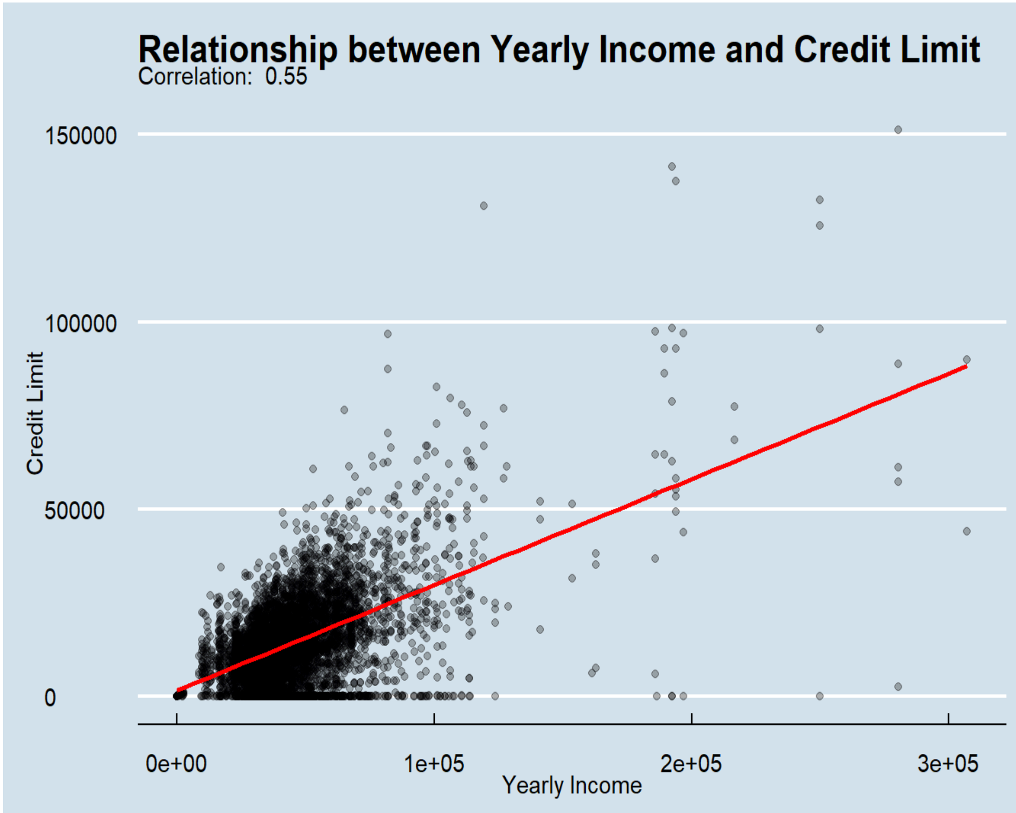

Figure 12: Scatter Plot indicating the relation between yearly income and credit limit

This data result indicated a moderate correlation (55%) between yearly income and credit limit of individuals, which means that credit limits will tend to increase as yearly income increases. However, as the relation is significant, the moderate correlation (0.55) indicates that there may be other factors beyond yearly income influencing the credit limit allocations. With that being said, individuals with higher yearly incomes are more likely to have larger credit limits that can vary depending on their financial situation. Thus, if a customer can increase their yearly income, they are eligible for potentially higher chances of accessing financial loans and credit debts. Not only from the customer’s side but also from the side of the institution, they can conclude and decide to have a range of credit limits by predicting the individual’s ability.

6.9 Exploratory Analysis of Relationship between Total Debt and Number of Cards

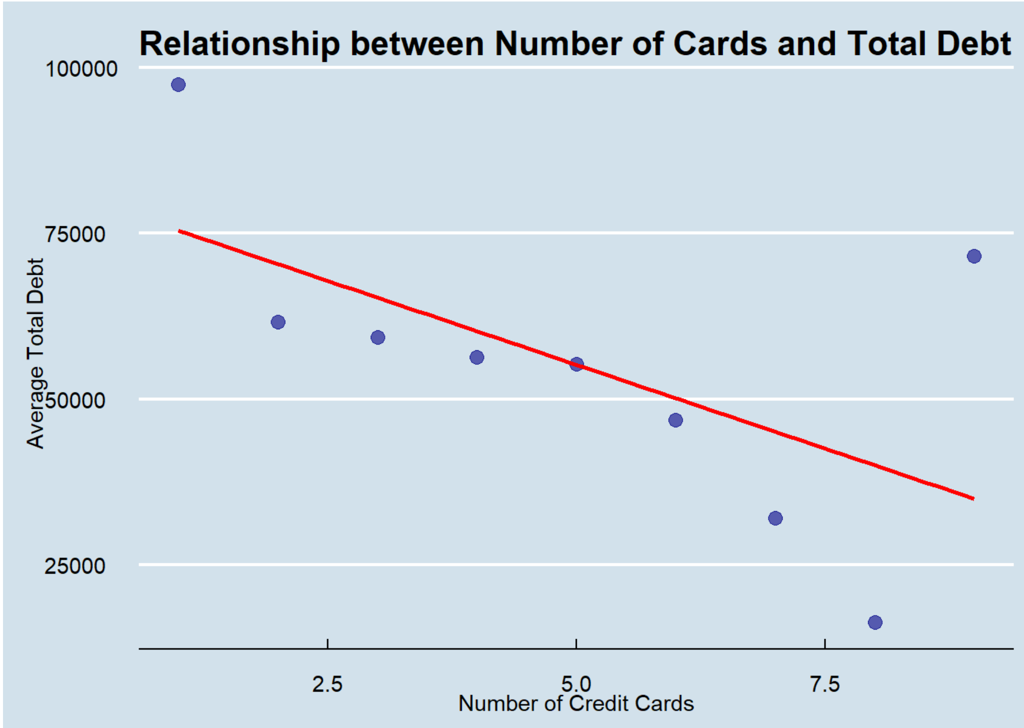

Figure 13: Linear Regression Data Analysis of Number of Cards and Total Debt

The purpose of using a linear regression plot is to search for the findings and if there are significant points in the relationship between the average total debt and the number of credit cards. This linear regression model includes the Number of Credit Cards as an independent variable and the Average Total Debt as the dependent variable. There is a negative coefficient for the number of credit cards, suggesting that the average total debt gradually decreases as the number of credit cards increases. The regression line also shows a downward slope trend, ensuring the negative correlation between the number of credit cards and the average total debt. As a result, it can be concluded that those who own more credit cards are more likely to handle their credit cards and debts. So, the bank institution should call those individuals with many credit cards to subscribe to the premium or higher credit products because of having less risk assessment with debt risk.

6.10 Analysis of Findings in Credit Limit Prediction Model

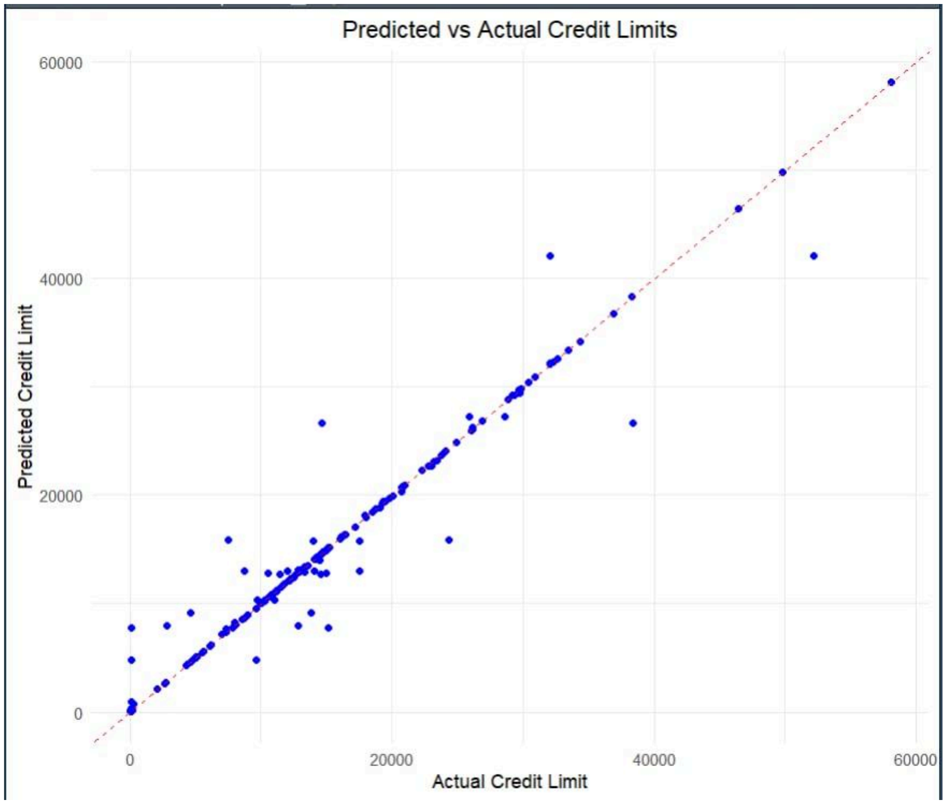

Figure 14: Data Visualization of Credit Limit Prediction Model

The model was trained to predict the credit limit of clients using a Random Forest regression model. The model was trained and evaluated using key financial and demographic features such as yearly income, credit score, current age, total debt, and average spending. The RFR model is trained with **80%** of the (train_data) but tested with the remaining 20% (test_data), and it consists of 100 decision trees. For the evaluation process, the model is evaluated by the following **three** metrics: **Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score.**

Metric	Value
MAE	780.64
RMSE	2264.05

R^2 Score	0.95

By this result, **780.64 of MAE** metric means the average deviation of magnitude of errors between two variables of actual credit limit and predicted credit limit. At that same time, the **RMSE value (2264.05)** provides the insight of having larger errors while the **R^2 score** of **0.95 (95%)** of the variance in the credit limit indicates the model's ability to make accurate predictions. The blue scattered dots in this analysis of the prediction model are aligned with the red dashed line (representing perfect predictions), indicating the high accuracy of the model. Moreover, outliers are observed, suggesting areas for potential model improvement, and further improvements can be made by featuring expansion and handling outliers.

However, the model is effective for predicting credit limits and can support decision-making processes related to credit allocation for the financial institution.

7. Limitations

Despite the above comprehensive analysis based on three dataset-users_data, cards_data, and transactions_data- this study is subject to several limitations and challenges. Firstly, the analysis was done in a short period, limiting further explorations that can take more time. Secondly, there is a limitation related to analytical tools since datasets were analyzed by using only R studio with limited expertise. Moreover, one of the major challenges is that the size of transactions_data is too enormous to be processed with standard laptops, leading to slow processing time and difficulty in identifying patterns and anomalies in some visualizations. Since the use of the combination of three datasets can create data complexity, it is important to handle them carefully to avoid errors and create accurate results. While the insights can explain the events between 2010 and 2019 effectively, they might not reflect the current trends. Furthermore, it is crucial to consider that the datasets only represent the users of a banking institution and transactions using its service rather than the whole population in the US. Lastly, findings drawn from the analysis may be influenced by the researchers' perspectives, which could cause potential bias. Based on the limitations of this research, future studies should aim to address these issues by using various approaches such as using more advanced analytical tools and libraries, using better devices, additional data, and so on.

8. Conclusion

This study provides valuable insights into comprehending information about the customers of a financial institution, including their financial status, spending, transaction trends, cash flows, and financial management across the United States. Firstly, high-income users being mainly from financial hubs like New York and California indicate that geographical factors influence income levels. Secondly, a significant transition from swipe to online payment after 2015 highlights a shift in customer preferences regarding payment methods. Regarding their spending patterns, the study suggests that the higher the incomes are, the higher the spending would be. Unexpectedly, while there are several claims that yearly income influences credit score, there is a lack of any significant correlation between yearly income and credit score for this dataset. Furthermore, regarding debt, it is found that younger individuals are more likely to have higher debt levels. A negative correlation between the number of credit cards and total debt shows that people with more credit cards manage their debt more effectively. Lastly, in the case of credit limits, there is a strong positive relationship between credit limit and yearly income. However, credit limits and credit scores have a very weak positive correlation, with only 5 percent. Therefore, this study plays an important role in understanding a bank’s customers’ spending trends, financial information, and management.

9. Discussion

This study can be beneficial for not only the bank and businesses but also the users themselves. For example, by understanding the geographical impacts on the income level, banks and businesses can make better targeting and customer segmentation. Additionally, changing customers' preferences to online transactions can help businesses adapt to the new trends. For better financial health, younger people who tend to have higher debt should have a more effective financial management system. Regarding financial management, it can also be found that owning and managing multiple credit cards can increase credit scores. Lastly, it is pointed out that the study needs to explore other factors influencing credit scores like marital status and credit history due to the lack of its correlation with yearly income and credit limits. Hence, this study can help banks, businesses, and individuals make more informed decisions for their financial well-being.

(Please see the references on the t page)

References

Beer, R., Ionescu, F., & Li, G. (2018, August 13). Are Income and Credit Scores Highly Correlated? FEDS Notes. https://www.federalreserve.gov/econres/notes/feds-notes/are-income-and-credit-scores-highly-correlated-20180813.html

Boopathy, G., & Kanagaraj, P. (2023). The Impact of Digital Payments on Consumer Spending Habits. Tuijin Jishu/ Journal of Propulsion Technology, ISSN: 1001 - 4055, 44(4), 5373. DOI: https://doi.org/10.52783/tjjpt.v44.i4.1896

Best, R. D. (2024, May 22). Credit card penetration in 161 countries worldwide 2011, 2014, 2017, 2021. https://www.statista.com/statistics/675371/ownership-of-credit-cards-globally-by-country/

Clearly Payments (2023). The Growth of the Credit Card Industry in 2023. https://www.clearlypayments.com/blog/growth-of-credit-card-industry-in-2023/

Chowdhury, M. H. (2023). The Financial Management Practices of Young Adults in Helsinki. A Comparative Study. Arcada University of Applied Sciences: International Business, 2023. https://www.theseus.fi/bitstream/handle/10024/817389/Manzar_Chowdhury.pdf?sequence=2&isAllowed=y

Nyrhinen, J., Sirola, A., Koskelainen, T., Munnukka, J., & Wilska, T. (2024, April). Online antecedents for young consumers' impulse buying behavior. Computers in Human Behavior. Vol.153. https://doi.org/10.1016/j.chb.2023.108129

Putrevu, J., & Mertzanis, C. (2023). The adoption of digital payments in emerging economies: challenges and policy responses. Digital Policy Regulation and Governance. DOI: 10.1108/DPRG-06-2023-0077

Schulz, M., & Shepard, D. (2024, December 6). 2024 Credit Card Debt Statistics. https://www.lendingtree.com/credit-cards/study/credit-card-debt-statistics/

Yee, A. (2024, February 10). Credit card debt among young people rising, Fed report shows. Chicago Sun Times. https://chicago.suntimes.com/2024/2/9/24067906/credit-card-debt-auto-loans-mortgage-interest-rates-millennials-genz-nyfed-report

end