UNIVERSITI MALAYSIA PAHANG
AL-SULTAN ABDULLAH

CENTRE OF MATHEMATICAL SCIENCES

GROUP PROJECT

DATA SCIENCE PROGRAMMING II (BSD2223)

SEMESTER II 2023/2024

GROUP NAME: DOLLAR $IGN

TITLE: CREDIT CARD CUSTOMER SEGMENTATIONS

PREPARED BY:

MATRIC ID	NAME	SECTION
SD22061	ANG MEI YING	01G
SD22049	THAM ZHI YUN	01G
SD22007	MIZA SYAZWANA BINTI MOHD SAFIAN	01G
SD22035	NURUL NAJWA BINTI NORHISHAM	01G

ABSTRACT

This study explores the segmentation of credit card customers to understand diverse user behaviours and preferences. By employing advanced data analytics techniques, we categorise customers into distinct segments based on their customer demographic, credit limit, card category, income category, dependent count, revolving balance and overall relationship between all of the segments. Financial organisations can more efficiently adapt their services and marketing campaigns by using the segmentation model, which uses clustering algorithms to find patterns. High spenders, revolving credit users, low balance transactors, and inactive users are among the key segments that have been identified. Each has distinct demands and characteristics related to money. The process of segmenting data yields significant insights that credit card issuers may utilise to improve customer satisfaction and profitability through targeted marketing, risk management, and personalised customer support.

TABLE OF CONTENTS

1.0 Introduction

Within the financial services industry, credit cards are essential tools that allow users to easily transact, manage their finances effectively, and access many benefits customized to suit their needs. Credit card companies on the other side are always looking for new ways to attract new customers and maintain the loyalty of the existing ones while increasing profits to thrive in this harsh financial market of today. Credit card segmentation is a process that strategically divides a large customer into a discrete one based on shared attributes and traits and it's one of the effective ways to get into a company’s goal of profits. Companies can improve customer satisfaction and loyalty by customizing their marketing strategies, product offerings, and customer service techniques to better cater to the unique needs of each group by having a thorough understanding of these segments.

This study will identify important subsets of credit card users by examining a large dataset that includes customer demographics, transaction history, credit limits, and more. These segments also will offer useful information that can be used to create a focused marketing campaign and improve credit risk management. To find hidden patterns and correlations between variables, advanced analytics and data-driven insights must be utilized. It also calls for a comprehensive strategy that goes beyond conventional measurements and considers psychographic factors, technology preferences, and changing market trends. In a market that is constantly changing, financial institutions can establish themselves as dependable allies in the financial journeys of their clients while promoting inclusivity and innovation.

The objective of this study is:

- To identify and analyse customer segments based on customer demographics, eventually understand the key factors influencing long-term loyalty and attrition for the banks

- To explore the variation of credit limits across several factors

- To examine correlations between revolving balances and other variables in order to understand their influence on spending habits and payment behaviour

2.0 Project Description

This project's goal is to delve into the complexity of credit card customer behaviour by analysing the Bank Churners dataset, which provides a comprehensive view of anonymized credit card user data. The dataset includes a variety of demographic information, consumer behaviour indicator and transactional data, offering a large source of detailed analysis. The focus is on understanding what influences customer behaviour, including customer demographic, exploratory data analysis which focus on credit limit, revolving balance analysis, and relationship analysis. The project seeks to uncover the pattern and meaningful insights that can benefit financial institutions by leveraging this data.

The main goal is to analyse the characteristics of credit card customers to identify segments based on their demographic details, credit limit, revolving balances, and the relationship of these attributes. Financial institutions can customise their services and marketing strategies by segmenting customers into different groups to better meet the needs of each segment. With the use of this study, it will be possible to better understand the unique characteristics of both loyal and at-risk customers, leading to the development of more effective customer retention plans.

At the beginning stage, this analysis will focus on the demographics of the customer base. We will look at age, gender, marital status, and educational attainment to see how these demographic characteristics affect the way that customers behave. We can determine which customer categories are more likely to sustain long-term relationships with the credit card provider and which are more likely to experience attrition by examining the distribution of these demographic parameters among various customer segments. To effectively present the results of this demographic analysis, visualisations such as pie charts, donut charts, bar plots, scatter plots, boxplot, and histograms will be used as support.

Another important aspect of the project is exploratory data analysis, with focus on credit limits. Understanding how credit limits are distributed among different customer segments and how they relate to the card category, income category and dependent count to be able to provide valuable insights for optimising credit limit allocations. Various visualisations will be used including pie chart, donut chart, bar chart and histogram. By uncovering the pattern and relationship from that, we can find the most significant factors that influence individual’s credit limit.

Finally, the project will delve into revolving balance analysis and relationship analysis. By examining the total revolving balance, we aim to identify potential financial distress among customers and understand how this influences their behaviour. Moreover, we will conduct relationship analysis to explore correlations between different variables. Scatter plots will be the main visual tool to highlight the finding relationships, providing a comprehensive view of the factors driving customer behaviour. Overall, the insights gained from this analysis will be instrumental in helping financial institutions design targeted interventions to enhance customer satisfaction and loyalty.

3.0 Data Description

A dataset named "Predicting Credit Card Customer Segmentation" is found at Kaggle and will be used in this project to achieve the objectives in this project. There are a total of 10128 rows and 23 columns inside the data set. **Table 3.1** shows the characteristics and descriptions of data in the dataset.

No.	Column name	Characteristic	Description
1.	CLIENTNUM	Qualitative	Unique identifier (ID number) for each customer
2.	Attrition_Flag	Qualitative	Flag indicating whether the customer has churned out or not
3.	Customer_Age	Quantitative	Age of customer
4.	Gender	Qualitative	Gender of customer
5.	Dependent_count	Quantitative	Number of dependents that customer has
6.	Education_Level	Qualitative	Education level of customers
7.	Marital_Status	Qualitative	Marital status of customer
8.	Income_Category	Qualitative	Income category of customer
9.	Card_Category	Qualitative	Type of card held by customer
10.	Months_on_book	Quantitative	How long the customer has been recorded on the books
11.	Total_Relationship_Count	Quantitative	Total number of relationships the customer has with the credit card provider
12.	Months_Inactive_12_mon	Quantitative	Number of customers has been inactive in the last twelve months
13.	Contacts_Count_12_mon	Quantitative	Number of contacts customer has had in the last twelve months
14.	Credit_Limit	Quantitative	Credit limit of customer
15.	Total_Revolving_Bal	Quantitative	Total revolving balance of customer
16.	Avg_Open_To_Buy	Quantitative	Average financial budget of the

			customer
17.	Total_Amt_Chng_Q4_Q1	Quantitative	Total amount changed from Quarter 4 to Quarter 1
18.	Total_Trans_Amt	Quantitative	Total transaction amount
19.	Total_Trans_Ct	Quantitative	Total transaction count
20.	Total_Ct_Chng_Q4_Q1	Quantitative	Total count changed from Quarter 4 to Quarter 1
21.	Avg_Utilization_Ratio	Quantitative	Ratio of revolving balance to the credit limit
22.	Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_1	Quantitative	Naive Bayes classifier for predicting whether someone will churn or not based on characteristics (Result 1)
23	Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_2	Quantitative	Naive Bayes classifier for predicting whether or not someone will churn based on characteristics (Result 2)

Table 3.1: Table of Data Description

4.0 Data preparation

Once the dataset is obtained from Kaggle, it is necessary for data preparation before further processing and analysis. Firstly, we need to know the structure of data, data types and understand all the variables inside the data. We have a total number of 10128 rows and 22 columns inside the original dataset before data cleaning is processed. We have removed the unused column including "CLIENTNUM" which indicates the customer id, "Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_1" and "Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_2" which indicate the Naive Baiyes classifier for predicting whether or not someone will churn based on selected characteristics. After that, we rename the column to ensure that its name accurately represents the data it holds and provide the audience an early description of the column's content. We rename the column "Customer_Age" to "Age" which provides a clearer and more direct label for this attribute.

Next, we check for missing values or null values inside the data. We found that there are 1519 null values in column "Education_Level" , 749 null values in column "Marital_Status" and 1112 null values in column "Income_Category". Since the missing data is exclusively of string type, we cannot utilize mean or median imputation methods to address the null values. Therefore, we handle the missing values by deleting the rows that contain null values. Overall, we have deleted 3046 rows with null values. It is important to clean all the null values inside the dataset to ensure the result of analysis is accurate and reliable.

Additionally, since the yearly income is represented in categorical type like $40k-$60k, therefore we had added a new column named "Mean_Income" to find the mean income for each income category. At the end of data preparation, we have a total number of 7082 rows and 21 columns in the dataset.

5.0 Data Analysis, Results and Discussion

Our GUI dashboard consists of one home page and four different content including Customer Demographic, Exploratory Data Analysis, Revolving Balance Analysis and Relationship Analysis. Each content section serves a distinct purpose in providing insights and understanding about various aspects of credit card customer segmentation.

5.1 Customer Demographic

Figure 5.1 show the first content of our GUI dashboard which is Customer Demographic. It provides a comprehensive analysis of the different characteristics of the credit card customer base. We can explore key demographic indicators such as age distribution, gender distribution, marital status distribution and education level distribution in this section. By using the slider input, we have the flexibility to customize our analysis to explore demographic trends and patterns in a targeted manner by selecting specific age ranges of interest. This section helps users to understand the customer base better and make it easier to tailor their services to targeted marketing.

Figure 5.1 Dashboard Customer Demographic

Based on **Figure 5.1**, the total number of customers across all age groups, ranging from 18 to 90 years old is 7081. Among these customers, 1113 of customers have churned with an attrition rate of 15.72%. From the results shown, individuals over the age of 65 have the lowest number of customers. This indicates that older individuals are less likely to hold credit cards therefore contribute less to attrition rates compared to other age groups. The attrition rates can be caused by several factors including dissatisfaction with current services or being attracted to offers from competing banks.

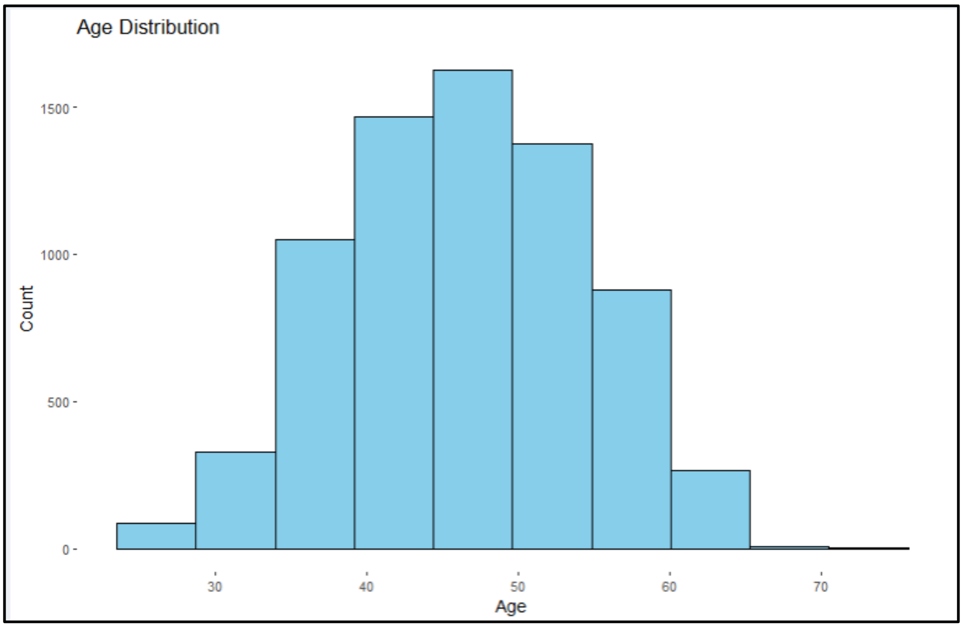


Figure 5.2 Count of Customer by Age

The histogram analysis reveals interesting insights about the distribution of credit card holders by age. Based on **Figure 5.2**, the largest segment of credit card holders falls within the age range of 45-50 with the total number of customers is 1627. This age group represents the highest concentration of customers, indicating that individuals in their early 45s are more likely to hold credit cards. Additionally, the age range 40 to 45 exhibits a considerable proportion of credit card holders with a total number of 1468 customers. These findings suggest that middle-aged individuals form a significant customer base for credit card services. The higher concentration of credit card holders in the 40-50 age range may be attributed to the fact that individuals within this age group are typically in their prime working years, therefore they have stable income compared to other age groups. Thus, they are more likely to qualify for a credit card and have the financial resources to manage credit responsibly. A study from Wei et al. (2018) has shown that credit card usage level is generally higher among individuals with age of 30 years and above with monthly income more than RM4, 000 and white-collar occupation. People tend to spend more when they are financially stable and own several credit cards. Based on Figure 5.2, we can see that the distribution is right skewed due to the presence of outlier in the 65-75 age range. The age group of 70-75 represents the smallest portion of credit card holders with a total number of 2 customers. This indicates that older individuals may have a relatively lower propensity for credit card usage. This may be due to older individuals more prefer to use cash for any transaction as they are not proficient in using credit cards. In addition, some older individuals may have concerns about the complexity of credit card usage and worry about security risks of credit cards compared to cash. Overall, understanding the age distribution of credit card holders is crucial for tailoring marketing strategies and designing targeted financial products that align with the preferences and needs of specific age segments.

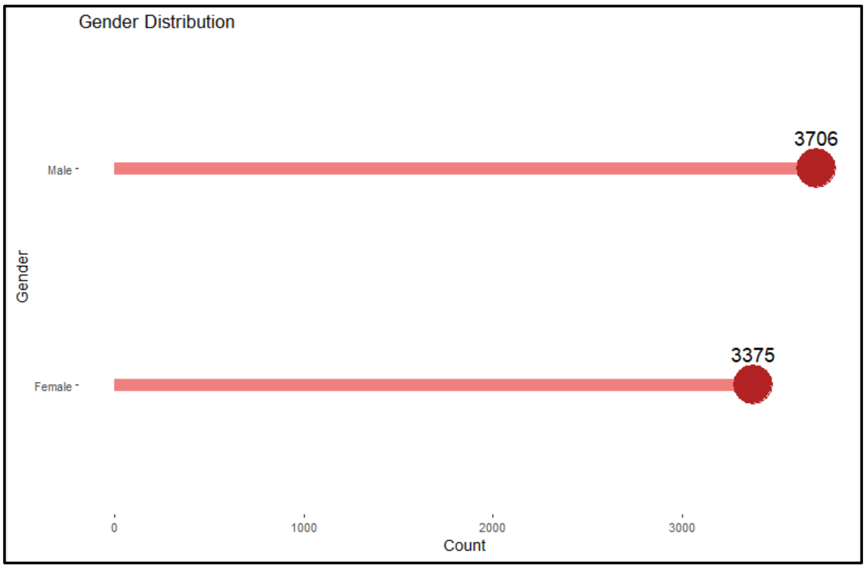


Figure 5.3 Count of Customer by Gender

The Lollipop chart provides insightful analysis of gender distribution among credit card holders. Based on **Figure 5.3**, we can see that males have a total number of 3706 customers which is higher than females. This suggests that males are more active in utilizing credit card services offered by the institution compared to females. However, there is only a slight difference of 331 customers between males and females. These findings indicate that both male and female customers have similar levels of engagement with the credit card services. The differences are due to the different spending habits between males and females in which males may tend to use credit cards for larger transactions such as vehicles and electronics while females purchase more on clothes, groceries or other shopping items. In addition, male tend to spend more than females, therefore resulting in higher credit card usage than females (*How Men and Women Manage Money Differently* , 2023). Understanding the gender distribution is instrumental in segmenting customers effectively. Financial institutions can use this information to tailor products, services and marketing campaigns to cater to specific needs and preferences of male and female customers.

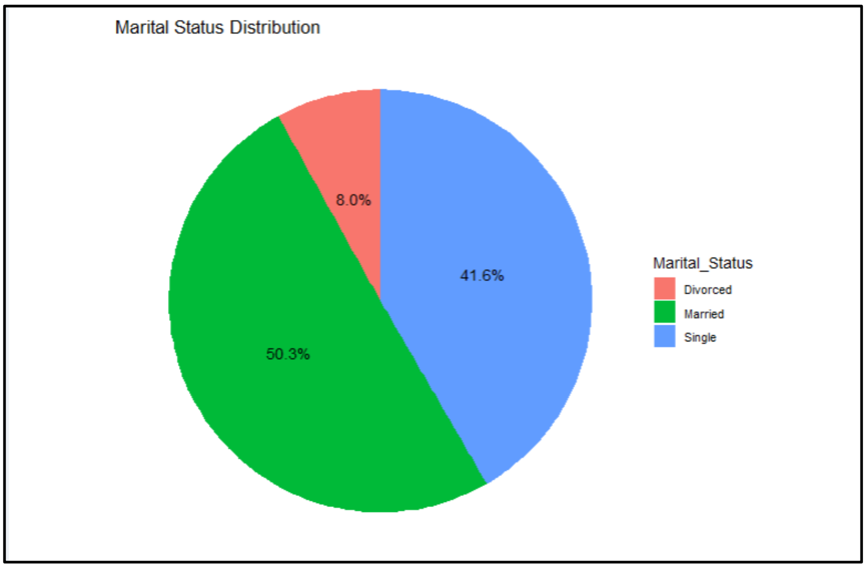


Figure 5.4 Pie Chart Marital Status Distribution

The pie chart analysis provides valuable insights into the distribution of credit card holders by marital status. Based on **Figure 5.4**, the largest segment of credit card holders is married individuals, accounting for a significant percentage of 50.3%. This indicates that married individuals are more likely to hold credit cards compared to other marital status categories. This pattern can be attributed to the fact that married individuals have more responsibility compared to other marital status to manage household finances such as rent payment, utility bills, groceries, childcare costs and other expenditures. The next prominent category is single, comprising 41.7% of credit card holders. This suggests that unmarried individuals form a substantial portion of the customer base. Unlike married individuals who need to share expenses, single individuals often rely on credit cards to manage their personal expenses including groceries, transportation, entertainment, rent cost and others. Divorced individuals represent a smaller percentage of 8.0%, indicating the lowest prevalence of credit card usage among this demographic. Following a divorce, individuals are likely facing a change in income. They may be going from two incomes to one, and adjusting to making alimony, child support or other payments. This change in monthly income may affect their ability to make payments on time and in full for a period of time, therefore reducing the credit card usage. In addition, Jasmin Suknanan (2021) also shows that divorce can impact an individual's credit score, therefore making it difficult for divorce individuals to manage their finances. Understanding the distribution of credit card holders by marital status is crucial for tailoring marketing strategies, developing targeted financial products, and offering appropriate benefits that cater to the needs and preferences of specific marital status segments.

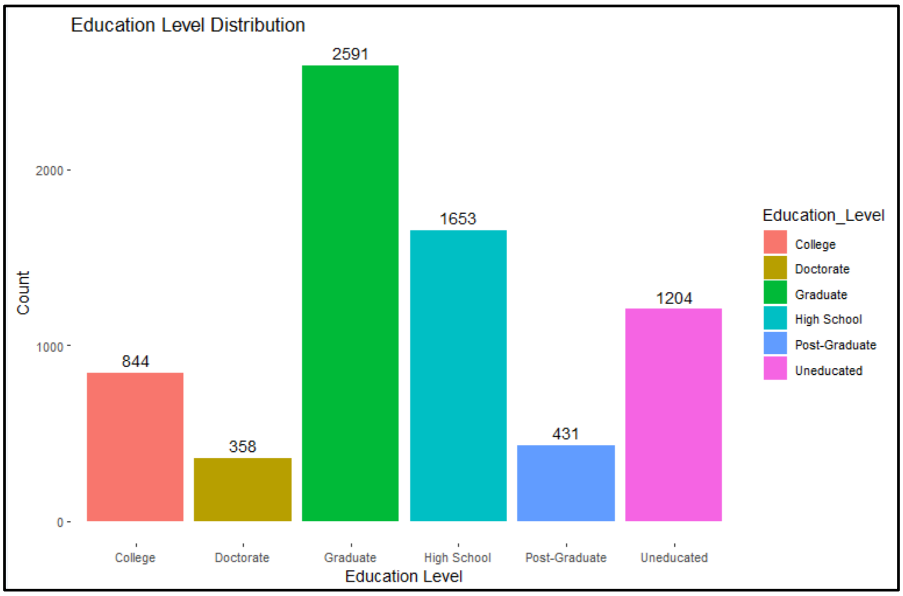


Figure 5.5 Bar Chart of Counts by Education level

The bar chart analysis provides valuable insights into the distribution of credit card holders by education level. Based on **Figure 5.5**, the largest segment belongs to those with a Graduate level of education, constituting a number of 2591 from total 7082 cardholders. This may suggest that individuals with graduate degrees are the most likely demographic to own and use credit cards, potentially due to higher income levels, better financial literacy, or a combination of both. Following Graduate level holders, the second-largest group is the High School educated cardholders with a number of 1653 customers. The Uneducated group represents the third largest with a number of 1204 customers which may reflect factors such as lower income levels or a lack of access or awareness of credit facilities. Furthermore, College educated individuals account for 844 cardholders, Post-Graduate individuals represent 431 cardholders, and Doctorate holders are the smallest group, with only 358 cardholders. It's interesting to note the significant drop from the Graduate to Post-Graduate and Doctorate levels. This may indicate that those with higher educational qualifications beyond a graduate degree may not prioritise credit card usage or prefer alternative financial management methods. These insights could be valuable for credit card companies to strategize their marketing efforts based on education levels.

5.2 Exploratory Data Analysis

In this subchapter, we aim to explore the variation of credit limits across several factors including card category and income level. Besides that, we also want to investigate the relationship between mean income and card category.

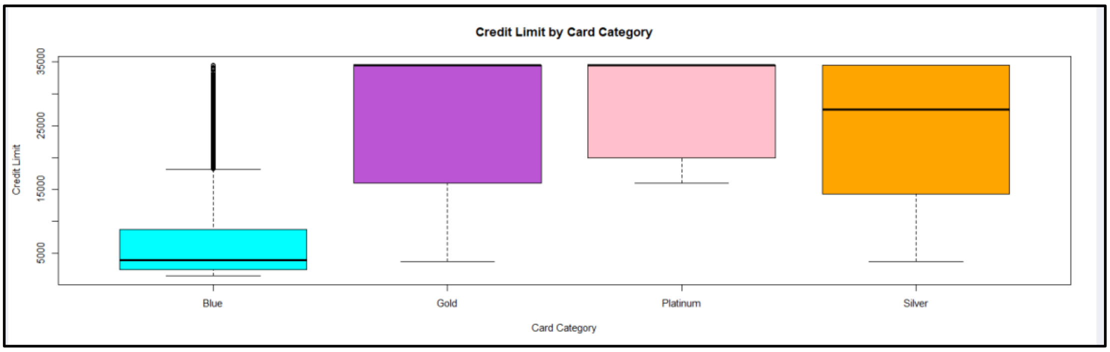


Figure 5.6 Boxplot of Credit Card Limit by Category

The boxplot analysis of credit limits by card category reveals several key insights. Based on Figure 5.6, the Blue card category has a relatively narrow boxplot with the median credit limit closer to the lower quartile, indicating lower variability within the middle 50% of the data. In addition, the presence of several outliers above the upper whisker suggests that while most Blue cardholders have lower credit limits, only a few have significantly higher limits, likely due to higher creditworthiness or spending patterns. In contrast, the Gold card category displays a much wider and more symmetrical boxplot with a higher median credit limit, a larger interquartile range (IQR), and no outliers, indicating a more uniform distribution of credit limits among Gold cardholders.

The Platinum card's box plot is similar to that of the Gold card but with a slightly lower range, a high median close to the upper quartile, and substantial variability without any outliers. This suggests that Platinum cardholders generally have high credit limits with a consistent distribution. The Silver card category exhibits the widest boxplot with the largest IQR, indicating the highest variability in credit limits. The central position of the median and the absence of outliers show a diverse set of credit limits within this category. The outliers in the

Blue card category likely result from a broader range of credit limits, with most customers having lower limits but a few with much higher limits. Moreover, from **Figure 5.6**, we can see that blue card category has the minimum credit limit compared to other card category, followed by silver, gold and platinum. This result in card benefit for each category.

In conclusion, there is a clear progression in median credit limits from Blue to Gold to Platinum to Silver cards, with increasing variability. The Blue card category stands out with its outliers, indicating a subset of customers with significantly higher credit limits. Overall, as we move from Blue to Silver card categories, both the credit limits and their variability increase, reflecting the diverse creditworthiness and spending patterns of customers within these groups.

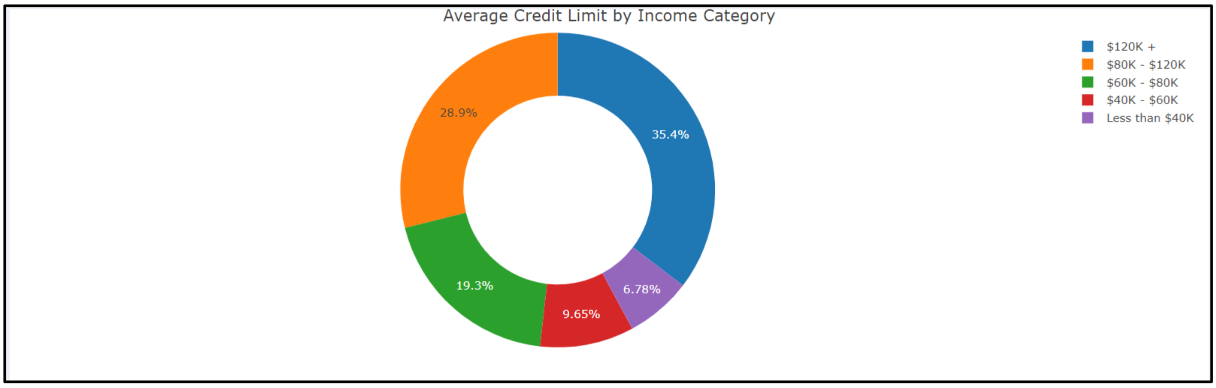


Figure 5.7 Donut Chart of Average Credit Card Limit by Income Category

The donut chart illustrating average credit card limits by income category reveals significant insights into how income levels influence credit access. Based on **Figure 5.7**, the chart shows that individuals earning over $120K have the highest average credit limit, representing 35.4% of the total. This high limit reflects their strong creditworthiness and financial stability, which make lenders confident in offering them substantial credit. The second highest is recorded by individuals with income category between $80K and $120K, represent for 28.9% of the total. It shows that they enjoy high credit limits, though slightly lower than the top earners.

Moving down the income scale, those earning between $60K and $80K hold 19.3% of the total credit limits, showing a moderate level of credit access. Their credit limits are lower than the top two categories due to comparatively less income but still reflect a decent level of financial reliability. Individuals with incomes between $40K and $60K, making up 9.65% of the total, have the second lowest average credit limits. Lenders offer them more conservative limits due to perceived higher financial risk. Lastly, those earning less than $40K have the lowest average credit limits at 6.78%, reflecting the significant risk lenders associate with lower income levels and their limited capacity to handle large credit amounts.

In summary, the chart underscores a clear correlation between higher income and higher credit limits. Those with higher incomes benefit from greater credit access due to their perceived financial stability and ability to repay, while lower-income individuals face stricter limits due to higher associated risks. This distribution highlights the role of income in shaping financial opportunities and access to credit.

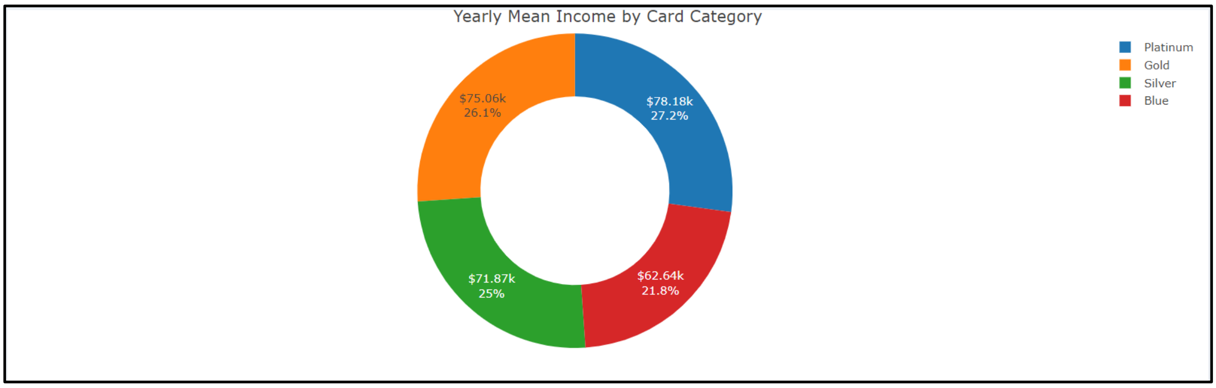


Figure 5.8 Donut Chart of Yearly Mean Income by Card Category

The donut chart illustrates the yearly mean income associated with different card categories. The blue segment, representing the highest average income of $78.18k, indicates that this card category is favoured by higher earners, likely due to its premium benefits and exclusive features. In contrast, the red segment, with an average income of $62.64k, shows that this category caters to lower-income individuals, possibly offering more accessible or entry-level benefits. The orange and green segments, with average incomes of $75.06k and $71.87k respectively, fall in between, appealing to middle-income earners who seek a balance of value and benefits without the high requirements of the top-tier cards. This distribution highlights how different card categories are tailored to distinct income levels, guiding issuers in targeting their products effectively. Higher earners are drawn to the perks of premium cards, while those with lower incomes tend to use more accessible options. This segmentation helps financial institutions design and market their cards to meet the specific needs and preferences of various income groups.

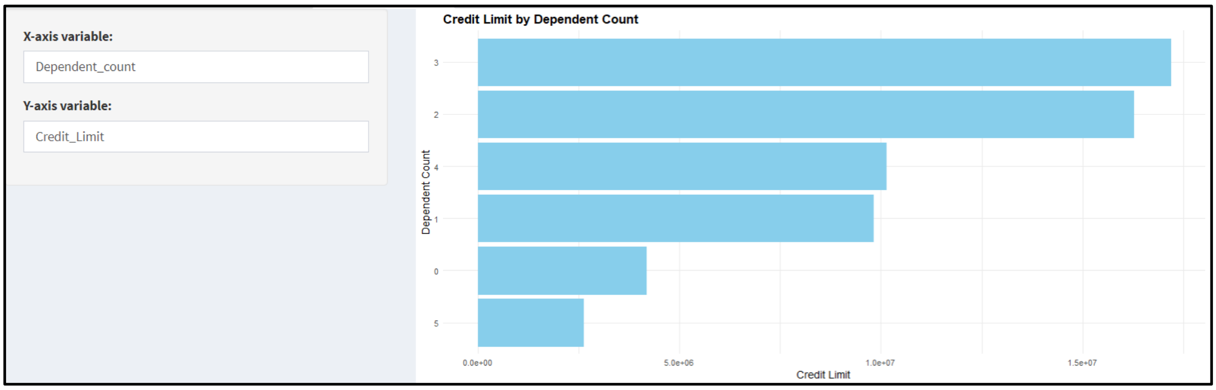


Figure 5.9 Horizontal Bar Chart of Credit Limit by Dependent Count

The bar chart reveals the relationship between credit limits and the number of dependents a person has. **Figure 5.9** shows that individuals with three dependents enjoy the highest credit limits, around $15 million. This suggests that having three dependents is often seen as a sign of stability and responsibility, possibly reflecting a steady income and a stable family life, which makes them more trustworthy to lenders. Following closely is those with two dependents, also enjoying high credit limits, likely due to similar reasons of perceived financial stability and responsibility.

People with four dependents have moderate credit limits, slightly lower than those with fewer dependents, possibly because supporting a larger family can strain finances, which might affect their perceived creditworthiness. Those with just one dependent have lower credit limits, suggesting they might have fewer financial obligations, but also possibly less financial stability compared to larger families. Individuals with no dependents have the lowest credit limits, indicating that lenders might see them as having less financial responsibility or a younger demographic that hasn't yet established financial reliability.

Finally, those with five dependents, while having a slightly higher credit limit than those with no dependents, still fall on the lower end of the spectrum. This group might face significant financial responsibilities, which could make lenders cautious. Overall, the chart underscores how credit issuers assess financial stability and responsibility, granting higher credit limits to those perceived as more financially stable due to their family situations.

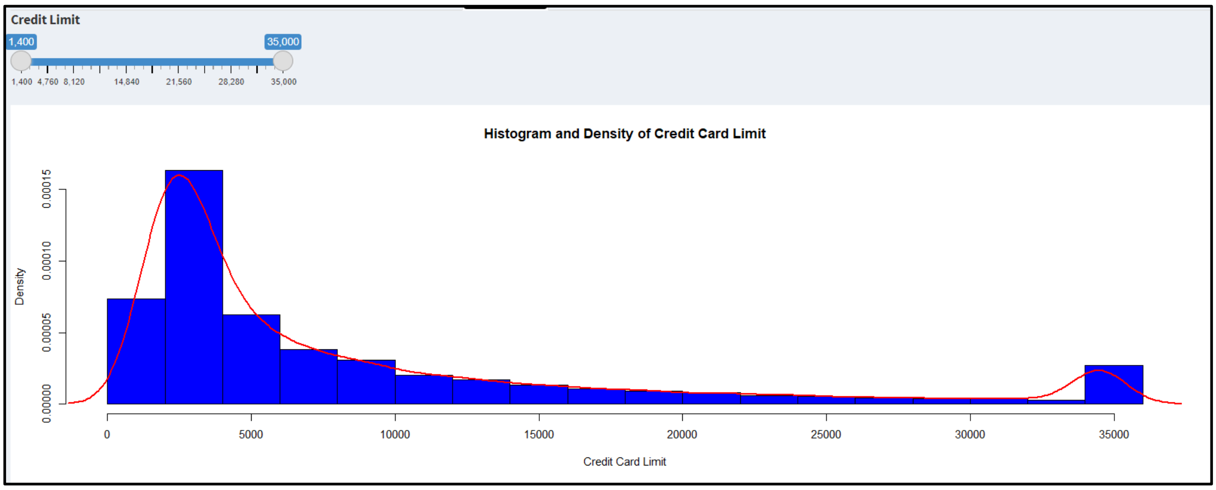


Figure 5.10 Histogram and Density of Credit Card Limit

The histogram and density plot of credit card limits reveal an interesting distribution among cardholders. **Figure 5.10** shows that most individuals have credit limits between $3,000 and $5,000, indicating that this is a common range for standard or entry-level credit cards. This likely reflects the cautious approach of lenders who set moderate limits for most consumers, based on typical income levels and creditworthiness. However, there's a smaller yet noticeable peak around $35,000, suggesting that a select group of individuals have access to significantly higher credit limits. These high limits are likely given to people with excellent credit histories, higher incomes, or those who hold premium credit cards. This bimodal distribution highlights how credit issuers differentiate between standard consumers and those deemed more financially stable, offering higher credit limits to the latter. Overall, the chart illustrates a strategic approach in credit limit allocation, balancing risk management with meeting the needs of various consumer segments.

5.3 Revolving Balance Analysis

In this subchapter, we aim to explore which customer segments mostly having zero-revolving balances. By identifying the characteristics of customers with zero-revolving balances, we can understand their financial behaviour and spending habit.

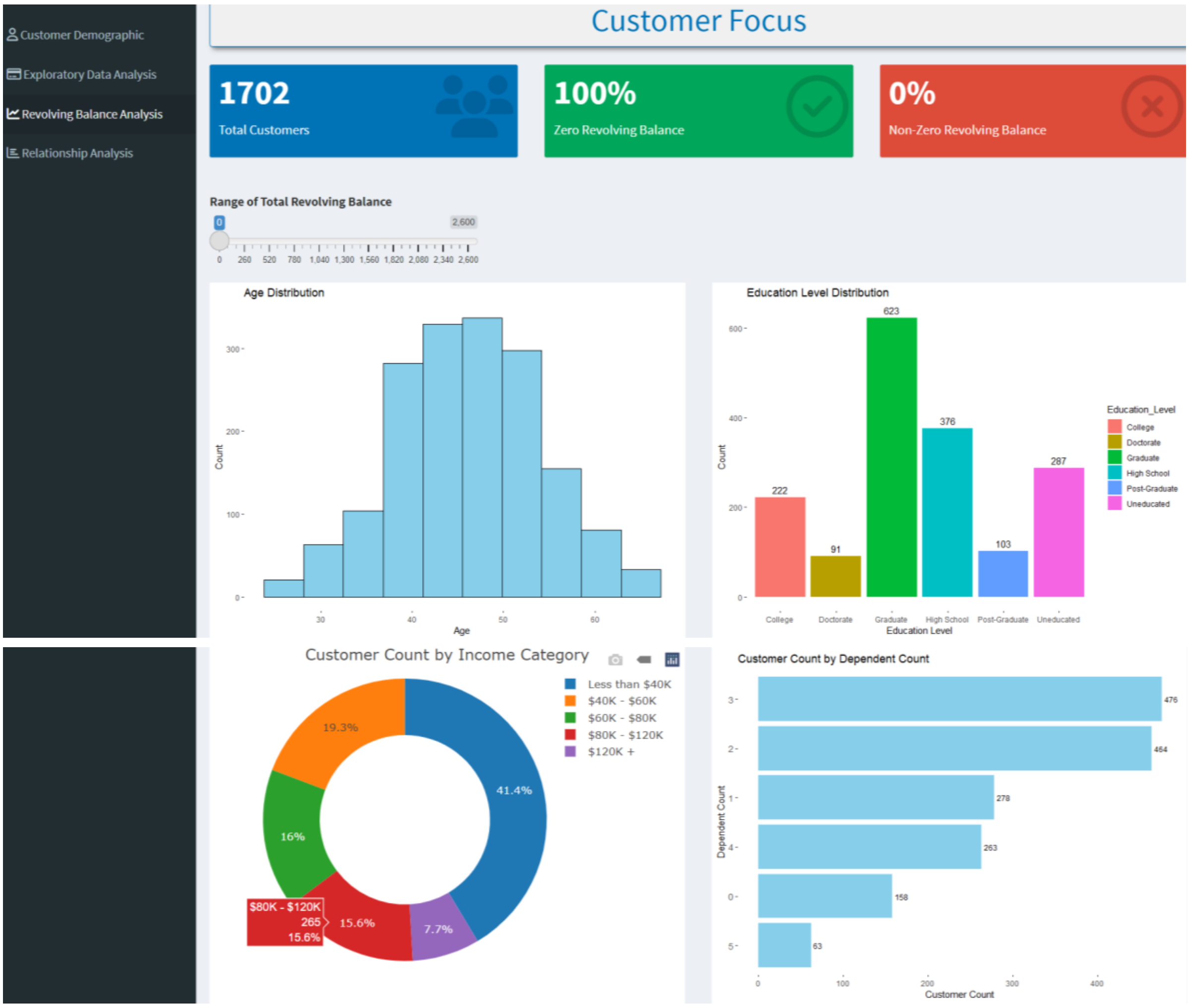


Figure 5.11 Dashboard Focus on Revolving Balance

Figure 5.11 shows the dashboard which focuses on the customers with zero revolving balance by a slider to filter the revolving balance. To identify the customer pattern based on characteristics of customers, a variance of customers characteristic will be observed in this dashboard. Based on **Figure 5.11**, the focus will be done by visualisation with customer's characteristics such as age, education level, income category and number of dependents. By comparison of **Figure 5.11** with **Figure 5.1**, we may observe that there are 1701 customers who have zero revolving balance out of a total of 7081 customers, which equals 24.04%. The revolving balance credit card is the amount that is carried over to the following month, if the balance is low, it may bring negative impact to the credit health. Hence, the zero revolving balance records show that the customers are fully using the recent card limit and show zero in the utilization ratio. Thus, a focus on the zero revolving balance customers will be done to identify the pattern behind it and eventually facilitate in modifying the credit card plans toward the customers.

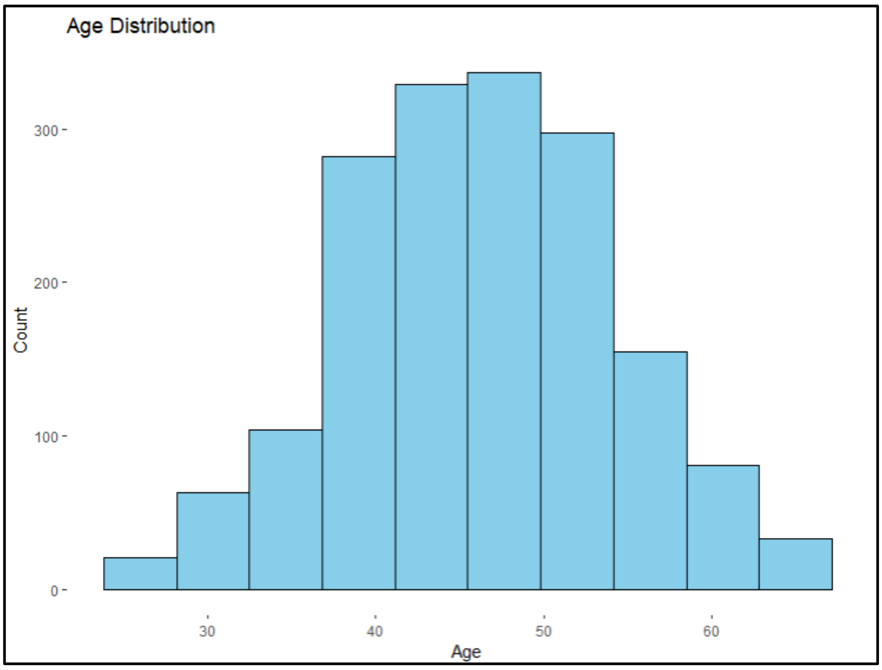


Figure 5.12 Histogram of Age Distribution

Figure 5.12 illustrates the age distribution histogram focusing on customers with zero revolving balance. The distribution takes the shape of a normal distribution, with the median value being equal to the mode value. The data indicates that the majority of customers with zero revolving balance fall within the age range of 45 to 50, while the smallest number of customers below the age of 30 fully utilise their credit limits. This particular age group shows the highest concentration of customers with zero revolving balance, suggesting that individuals in their mid-40s are more inclined to max out their credit card limit. This trend could be attributed to the fact that a significant portion of credit card holders are also within the age range of 45 to 50, as depicted in **Figure 5.1**. Additionally, individuals in this age group may have substantial financial obligations, such as mortgage payments, educational expenses for their children, or healthcare costs, which may prompt them to utilise their full credit limit without considering the negative impact on their credit score. They may use the entire credit limit for daily expenses for themselves and their families. In summary, the majority of customers aged 45 to 50 are financially responsible for their families, and most credit card users fall within this age bracket. Conversely, the least customers with zero revolving balances are those under the age of 30. This could be attributed to the fact that younger generations have access to various alternative credit options like digital wallets and buy-now-pay-later services, therefore reducing their reliance on traditional credit cards. For instance, teenagers in Malaysia prefer using E-Wallet services such as TouchNGo E-Wallet, ShopeePay, and Boost over credit cards. In addition, young consumers may view credit card usage as irresponsible and linked to overspending, leading them to be cautious about accumulating debt and thus limiting their credit card usage. In short, fewer teenagers opt for credit cards due to these perceptions. In conclusion, a focus can be done toward the 45 until 50 years old customer, to increase their credit cards limit as their financial weight may be heavier than the other age category customer.

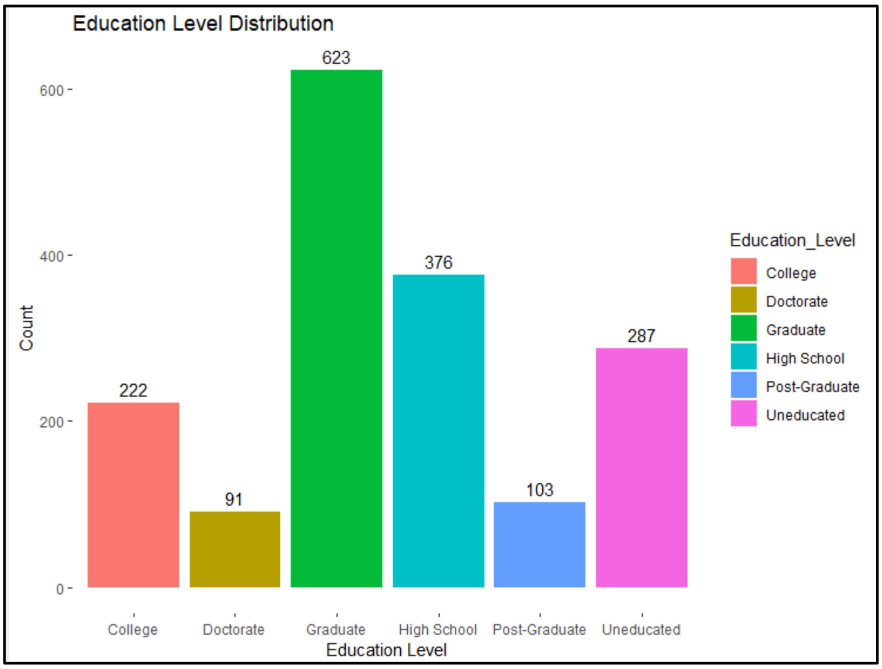


Figure 5.13 Bar Chart of Education Level Distribution

Figure 5.13 illustrates the distribution of education levels among customers who have no revolving balance. From the majority of these customers, a total number of 623 customers have a graduate education level. Following this are high school level with a total of 376 customers, uneducated level with a total of 287 customers, college level with a total of 222 customers, post-graduate with a total of 103 customers, and doctorate with a total of 91 customers. The highest number of graduate customers with zero revolving balance may be attributed to the fact that most credit card users are graduates, as shown in **Figure 5.5**. Additionally, graduates often have jobs and regular income, which can result in higher credit limits on their credit cards. This higher credit limit encourages spending among customers. On the other hand, doctorate customers tend to have more stable jobs and regular income compared to customers with other education levels. This financial stability and higher education level may lead to less reliance on credit and a better understanding of financial management. Therefore, doctorate customers are less likely to fully utilise their credit limits. In conclusion, a focus should be done toward the graduated customer to help them have better understanding on correct ways of credit card usage. For example, organising “Sharing Sessions” or talks in high schools to educate students about basic financial management and responsible credit card usage may help in reducing the number of customers with zero revolving balance as majority of no revolving customers are graduated and high school levels.

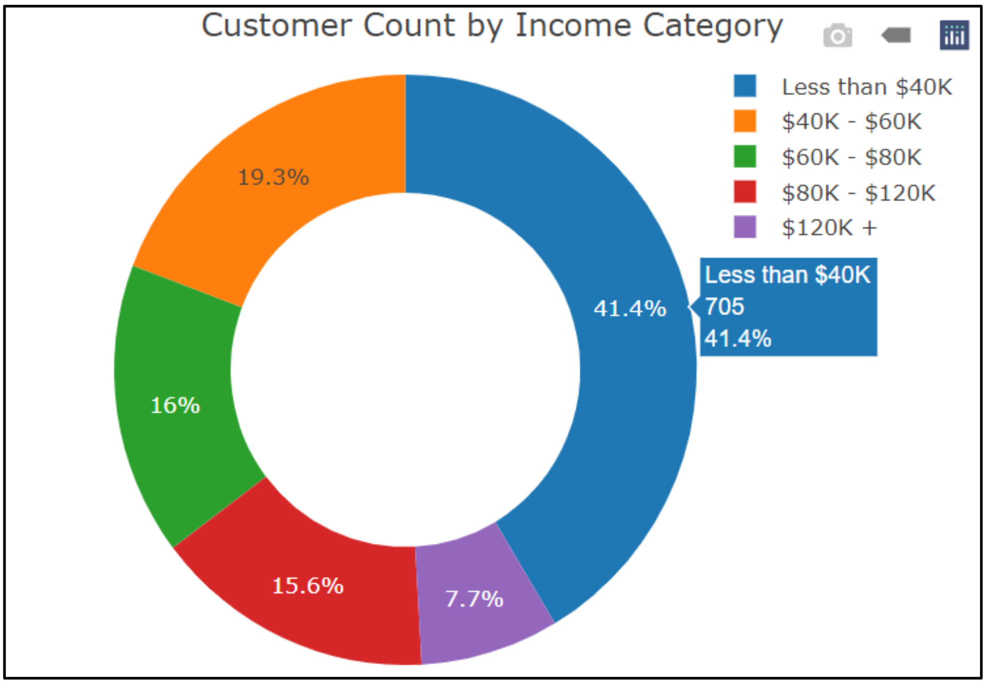


Figure 5.14 Donut Chart of Income Category

Figure 5.14 shows a donut chart which represent the percentage of customers who have no revolving balance by yearly income category. By observation toward the chart, we may observe that 41.40% of customers with no revolving balance has the income less than $40k, which equal to a total of 705 customers, follow by 19.30% or 328 customers with income level between $40k and $60k, 16% or 273 customers with income level between $60k and $80k, 15.60% or 265 customers with income level between $80k and $120k. Lastly, the lowest customer count is recorded by customer with yearly income more than $120k which represents by 7.7% or 131 customers. The result show that customers with lower income levels tend to utilise their credit card limits fully due to the absence of a financial safety net. Their limited income often results in insufficient savings, therefore leading to the need to rely on credit card for unexpected expenses, eventually resulting in a low revolving balance. Additionally, lower income customers may utilise credit cards to establish a credit history. In summary, low-income customers rely on credit cards for unexpected expenses and to build their credit history, consequently leading to a lack of revolving balance. Moreover, customers with higher yearly incomes may have access to alternative financial products. Therefore, individuals with higher incomes often have a wider range of financial options including making loans with lower interest rates, which is more appealing than accumulating credit card debt. Therefore, it can be concluded that there is a negative correlation between the yearly income level of customers and the number of customers without a revolving balance. As the yearly income level increases, the number of customers without a revolving balance will decreases. In conclusion, it is crucial to focus on customers with lower income levels as they are more likely to rely on credit cards compared to those with higher incomes. Higher credit card limits may be planned for the less income customer, and hence they will use credit cards more frequently for their spending. For example, in Malaysia, the credit limit for individuals earning RM36,000 or less per year is two times their monthly salary for each bank, to increase the frequency of customers using credit cards.

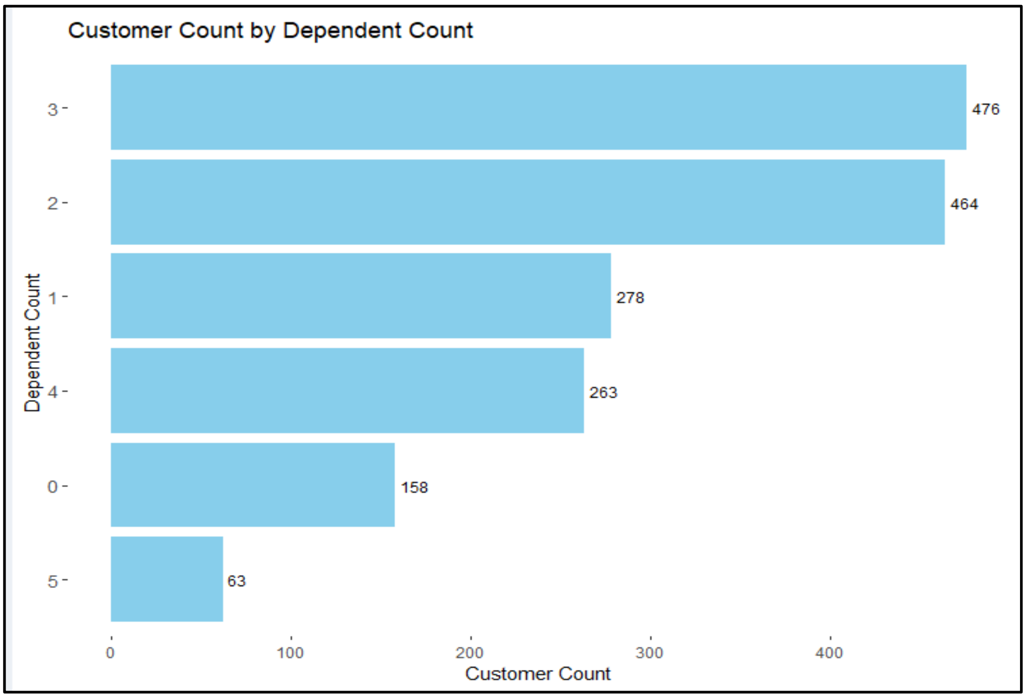


Figure 5.15 Horizontal Bar Chart of Customer Count by Dependent Count

Figure 5.15 show the ranking of customer counts for customer for zero-revolving balance by dependent counts. We may observe that the highest customer count recorded by customers with 3 dependents, followed by customers with 2 dependents, customers with only 1 dependent, customers with total 4 dependents, customers without dependents and lastly customer with a total 5 dependents. Based on **Figure 5.15**, customers who with 3 dependents have a total number of 476 customers, followed by 464 customers who with 2 dependents, 278 customers who with only one dependent, 263 customers who with total 4 dependents, 158 customers customers who without dependent, and lastly 63 customers who with 5 dependents. As we observed that the top of ranking record by customers with three dependents, this may be due to the reasons that majority customers are having three dependents. According to the U.S. Census Bureau, there were approximately 83.9 million families in the United States in 2021, with an average family size of 3.13 people per household (Average Family Size in the U.S. 1960-2022 | Statista, 2024). Additionally, financial concerns are also one of the significant reasons that have led to smaller family sizes. People want to have just one kid and offer the child the best within their means after realising in 2007–2008 that raising children is a costly and significant commitment. This may result in the majority of three dependent customers and lead to the fact that most of the three dependent customers have zero revolving balance. Furthermore, there is a question as to why the number of zero-revolving balances is lower for customers without dependents. This can be explained by the fact that the majority of them are fresh graduates who are still economically unstable and get financial support from home. According to a Jobstreet salary report, the minimum income level for fresh graduates before was between RM1,949 to RM2,8361. This income level might not be sufficient to cover all expenses, suggesting that some fresh graduates might receive financial support from home. Lastly, we focus on customers who have a total of five dependents but less zero revolving balance. This may be due to the reasons that customers with five dependents might have more financial responsibilities, leading to lower credit card usage. They might be more cautious about their spending due to the larger number of dependents. Hence, a focus should be done towards the customers who have 3 dependents, by increasing the credit limit, so they can more frequently use credit cards, and thus their credit history will be created too.

5.4 Relationship Analysis

In this last subchapter, we aim to explore the relationship between the variables in the dataset. It is critical to examine how variables are related to each other in order to identify distinct customer segments within the dataset and help banker to tailor targeted marketing strategies.

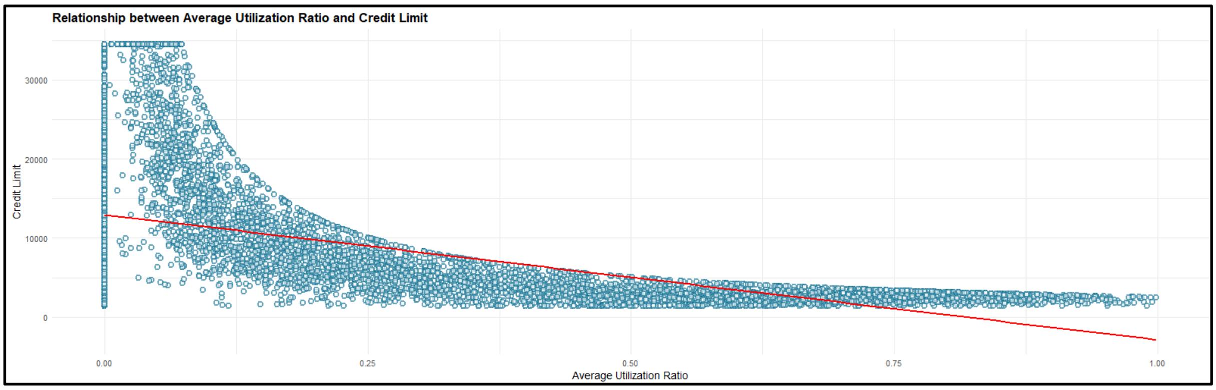


Figure 5.16 Scatter plot and trend line of the relationship between Average Utilization Ratio and Credit Limit

Figure 5.16 shows the relationship between the credit limit and average utilization ratio. The credit limit is represented by the vertical axis, which ranges from 0 to about 35,000, and the average utilization ratio, which ranges from 0 to 1, shows how much of the cardholder's available credit is being used. Basically, average utilization ratio is calculated by total revolving balance divided by credit limit. Based on **Figure 5.16,** there is a negative correlation between credit limit and average utilization ration. This indicates that as credit limit decrease, average utilization ratio will increase. People with lower credit limit tend to have a higher utilization ratio which indicates that they are using a larger percentage of their available credit. This may be due to financial institutions may reduce credit limits for borrowers who use a larger percentage of their credit to reduce risk, which might be responsible for this trend. Furthermore, the data point distribution indicates that most cardholders have lower credit limits, with the majority of them clustered at the lower end of the credit limit spectrum. This pattern may also indicate that people with lower credit limits use cash rather than credit cards more often, which leads to higher utilization ratios when those people do use their cards. As a result, at lower utilization ratios and higher credit limits, the density of scatter points is higher, suggesting that cardholders with higher credit limits typically utilize a smaller percentage of their available credit. In summary, the graph illustrates a negative correlation between the average utilization ratio and the credit limit, highlighting how higher usage of available credit is associated with lower credit limits, possibly due to financial institutions' risk mitigation strategies and the usage patterns of cardholders.

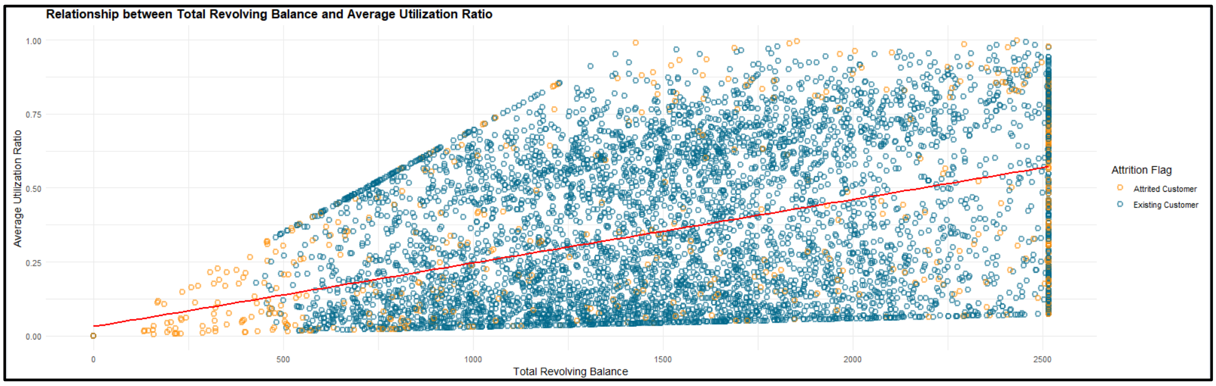


Figure 5.17 Scatter plot and trend line of the relationship between Total Revolving Balance and Average Utilization Ratio by Attrition Flag

Figure 5.17 illustrates the correlation between financial metrics, focusing on existing customers (blue) and attrited customers (orange). The revolving balance represents the total amount owed without being paid in full, while the average utilization ratio indicates how much of the customer's available credit is being used. The graph's dense blue point population indicates that most data come from existing customers, while the reduced number of orange dots represent attrited customers. The positive correlation between the average utilization ratio and total revolving balance indicates that an increase in the total revolving balance is likely to result in an increase in the average utilization ratio. This suggests that customers with larger revolving balances use more credit, which raises their utilization ratios. Existing customers' credit behaviour varies widely, while attrited customers show more cautious credit behaviour, with lower utilization ratios and balances. The graph concludes that greater revolving balances are linked to higher utilization ratios, suggesting that customers tend to use a greater percentage of their available credit when they owe more. This information can help financial institutions better manage risk and customize credit products.

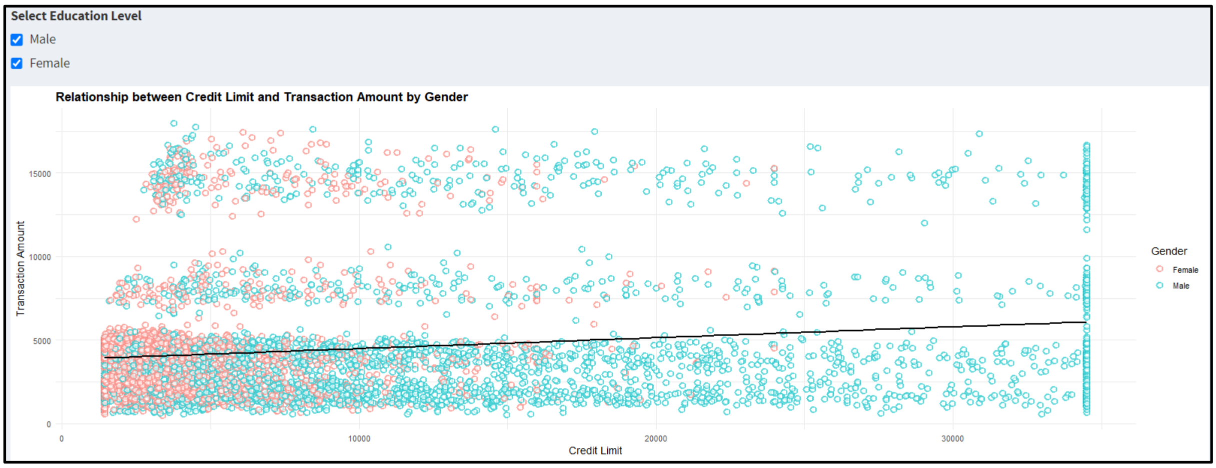


Figure 5.18 Scatter plot and trend line of the relationship between Credit Limit and Transaction Amount by Gender

By using scatter plot and trend line, we tend to analyse and understand how gender affects data points. Based on **Figure 5.18**, data points for both genders are distributed across the graph, with males typically having higher credit limits than females. This may be due to higher employment rates of males compared to females. Female data points are similarly dispersed, but some exceptions exist, such as those with high transaction amounts compared to their credit limits.

The graph's black trend line shows a minimal or weak relationship between the credit limit and transaction amount, suggesting that a rise in credit limit does not always result in a corresponding rise in transaction volume. However, the positive correlation demonstrates that independent of credit limits, men and women display a variety of spending habits, with men typically having higher credit limits. Most people have lower credit limits and make smaller transactions, according to the dense population of data points at lower credit limits and transaction amounts for both genders.

In conclusion, the graph shows a positive correlation between the credit limit and transaction amount, with both genders displaying a variety of spending behaviours that are not exclusively reliant on the credit limit. Men typically have higher credit limits, possibly due to higher employment rates, suggesting that the amount of transactions is more heavily influenced by personal spending patterns and money management.

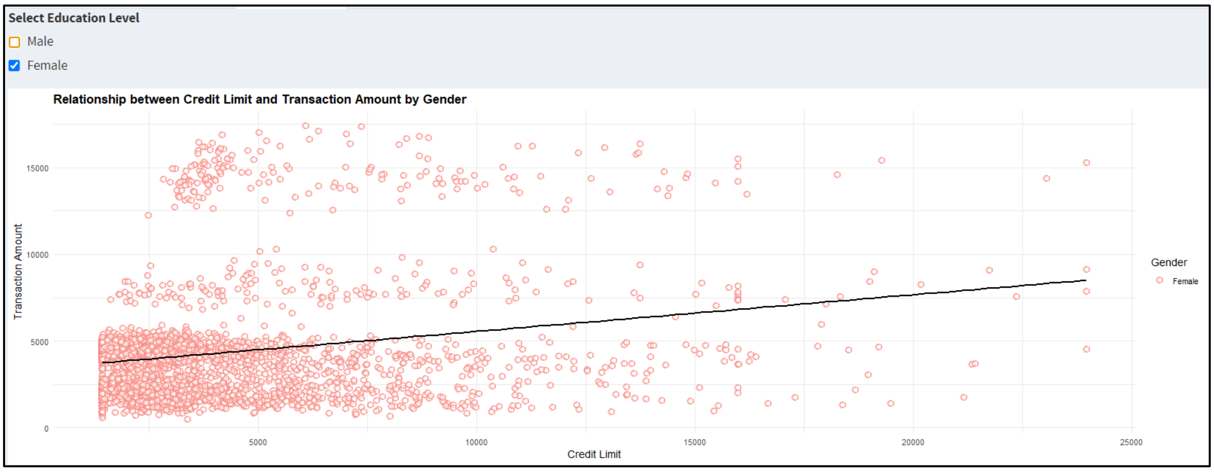


Figure 5.19 Scatter plot and trend line of the relationship between Credit Limit and Transaction Amount by Female

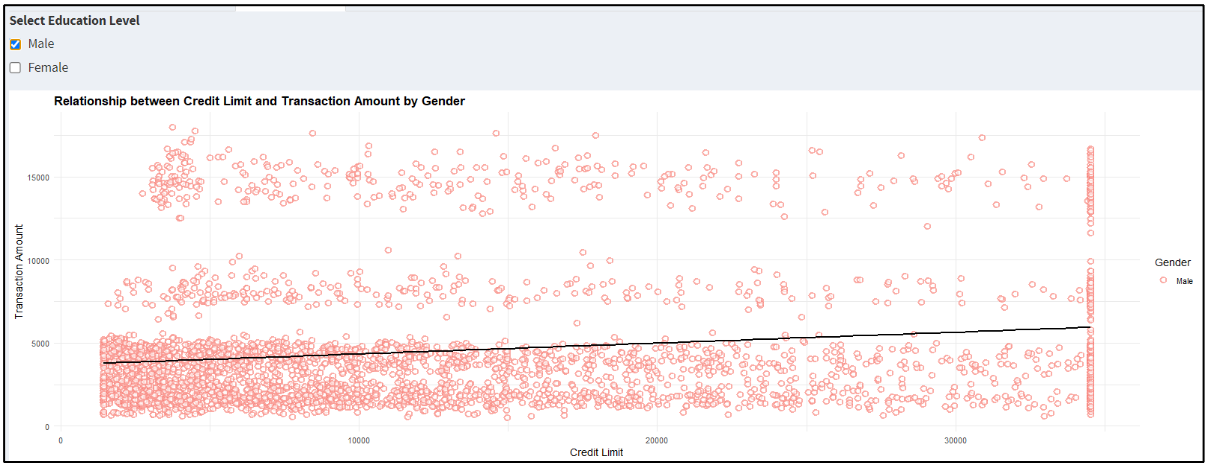


Figure 5.20 Scatter plot and trend line of the relationship between Credit Limit and Transaction Amount by Male

By using the check box input, **Figure 5.19** and **Figure 5.20** shows a weak positive correlation for males and females, suggesting that transaction amounts slightly increase as credit limits increase. However, there are variations in the distribution and clustering of data points. Female scatter plots are sparser as credit limits rise, with most females making smaller transactions. On the other hand, male scatter plots show a vertical cluster at higher credit limits, suggesting that some males with high credit limits have a wider range of transaction amounts. This vertical clustering suggests more variability may be present among males with high credit limits. Males also exhibit a wider range of spending behaviours, with more outliers at the high end of the credit limit spectrum, exhibiting both low and high transaction amounts. These disparities suggest that men exhibit higher levels of variability and a wider range of transaction amounts at higher credit limits, while women tend to have a more consistent spending pattern with fewer extreme outliers.

6.0 Conclusion

The objectives of this study are first, to identify and analyse customer segments according to age, gender, marital status, and level of education to find out which are more likely to remain loyal to the credit card company over the long run and which are more likely to experience attrition. Second, to investigate the relationship between credit limits and factors such as card category, income category, and so on. Last, to identify patterns and create a suitable marketing strategy and to examine correlations between revolving balances and other variables to learn how these factors affect spending and payment behaviour. To achieve all the objectives, data are analysed and visualised for better understandings towards the credit card customer segmentations as done in Chapter 5.0 in this study.

The first objective in this study is to identify and analyse customer segments according to age, gender, marital status, and level of education to find out which are more likely to remain loyal to the credit card company over the long run and which are more likely to experience attrition. To achieve this objective, a dashboard named "Customer Demographic" is created as shown in **Figure 5.1** and each characteristic of customers and its effect on attritions are discussed in Chapter 5.1 of the study. In short, we may conclude that there are a total of 7081 customers with age range from 18 to 90, and 15.72% of the customers are attrited. By observation towards the analyses on customers' behaviours, we can conclude that the graduate married male customers who aged between 45 until 50 are more likely to remain loyal to the credit card company over the long run whereas the divorced females who aged over 65 and have doctorate education levels are at risk of attrition.

Next, the second objective of the study is to investigate the relationship between credit limits and factors such as card category, income category, and so on to identify patterns and create a suitable marketing strategy. To achieve these goals, various visualisations have been done to find the pattern inside, such as boxplot, donut chart, horizontal bar, and histogram as shown in Chapter 5.2. By observations toward the data analyses in Chapter 5.2, we may conclude that most of the holders have the credit limits between $3000 and $5000, and there is a positive relationship between card category level and yearly income level with credit card limit. The higher the card category level or the higher the early income category, the higher the credit limit of credit cards. This is because higher earners are drawn to the perks of premium cards, while those with lower incomes tend to use more accessible options. In addition, the numbers of dependents do not show an obvious relationship with the credit card limit, the higher credit card limits are set for the customers who have 3 dependents. In short, high income level customers tend to have high category level credit cards, to get the higher credit limits and get extra benefits.

Also, the last objective is to examine correlations between revolving balances and other variables to learn how these factors affect spending and payment behaviour. To achieve the objective, a dashboard named “Revolving Balance Analysis” is created as shown in **Figure 5.11** which focuses on the customers with zero revolving balance by a slider to filter the revolving balance. By focusing on the characteristics of customers without revolving balances, we may conclude that there are a total of 1702 customers who are facing problems without revolving balance. The graduated customers who are between 45 and 50 years old, having yearly income less than 40 thousand dollars and have 3 dependents are the main characteristics of customers who are facing problems of zero revolving balance. To help them to solve the problem, a plan modifying will be done for those customers in order to increase their credit card limit or help them to understand the correct ways of using credit cards smartly. A higher revolving balance will result in higher utilization ratio and increase the customers’ performance in their credit history.

Moreover, to achieve the third objectives, a number of scatter plots has been done to identify the relationship inside the data as discussed in Chapter 5.4. In this chapter, three relationships are determined: the relationship between average utilization ratio and credit limit, relationship between total revolving balance and average utilization ratio, and relationship between credit limit and transaction amount. There is a strong negative correlation between utilization ratio with credit limit, the higher the utilization ratio, the lower the credit limit. Also, there is a strong positive correlation between total revolving balance and the average utilization ratio, the higher the revolving balance, the higher the utilization ratio. Lastly, there is a weak positive correlation between the credit limit and transaction amount, the higher the credit limit, the higher the transaction amount.

In conclusion, as credit card plans management should design more attractive plans which may attract more graduate married male customers aged between 45 until 5, as they are the main customer population who are always active in using cards. Also, the credit limits of each customer may be adjusted carefully by handling a variety of factors. Lastly, focus on no revolving balance customers to modify their card planning, may help to improve their financial performance, as well as improve the profit by the bank. As the higher the revolving balance, the higher the utilization ratio, as well as the higher the performance in financial stability.

7.0 Limitation of the Project

After conducting the study, we found that there are some limitations in our study. To improve the understanding in credit card customer segmentations and improve the performance of relevant study in the future, some suggestions will be provided to overcome the limitation. **Table 7.1** shows the limitations of the study and the suggestions to overcome the limitations.

Limitations	Description	Solutions
Dashboard Functionalities	Lack of tooltips button to guide user to understanding the ways to exploring dashboard	Notes may be added to the dashboard to give the hints on steps to explore the dashboard. Also, the suggestions on exploring the dashboard can be shown in the figure.
Limited Ways of Data Interactivity	Only available for a few interactions, like slider and tick box	Add the other interaction ways to increase the data interaction and get different patterns by the filter applied.
	Cannot zoom in to focus on specific values	Mouse interactions can be used to implement zooming in the dashboard, so focus on certain parts of any visualisation.
Difficult to understand the values	No labels for values in histogram and bar chart	Adding value labels for easier understanding and deeper analysis into each specific value.
	Card category and education level does not show ascending in visualisation	Arrange the categorical data ascending according to the order by using factors and levels.
Limited data analyses	Too many data in the data set, difficult to analyse all the data	Build a subset of dataset to just focus on certain specific data to get deeper understanding towards the dataset.

Table 7.1 Table of Limitations and Solutions

References

Average family size in the U.S. 1960-2022 | Statista. (2024, May 22). Statista. *https://www.statista.com/statistics/183657/average-size-of-a-family-in-the-us/*

Bank, A. (n.d.). Credit card limit: Learn what is a credit card limit | Axis Bank. *https://www.axisbank.com/progress-with-us-articles/how-to-increase-your-credit-card-limit-and-why-it-is-useful*

Finwealth. (n.d.). *https://www.finwealth.com.my/articles/financial-planning/fresh-graduate-financial-planning-understanding-your-assets-and-liabilities/*

ggplot2 barplots : Quick start guide - R software and data visualization . (n.d.). STHDA. Retrieved June 2, 2024, from http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization

ggplot2 scatter plots : Quick start guide - R software and data visualization . (n.d.). STHDA. Retrieved June 2, 2024, from http://www.sthda.com/english/wiki/ggplot2-scatter-plots-quick-start-guide-r-software-and-data-visualization

How Men and Women Manage Money Differently . (2023, January 19). Directions Credit Union. https://directionscu.org/2023/01/19/how-men-and-women-manage-money-differently/

iMoney Editorial. (2015, December 28). *How credit card limit works in Malaysia*. iMoney Malaysia. https://www.imoney.my/articles/credit-card-limit

Jasmin Suknanan. (2021, August 31). *Does getting a divorce affect your credit score?* . https://www.cnbc.com/select/does-divorce-affect-credit-score/

Plotly R Open Source Graphing Library. (n.d.). Retrieved June 2, 2024, from https://plotly.com/r/

Rajpurohit, V. (2024, April 4). Why credit card essential for PhD / research scholar in 2024. *Research Voyage*. https://www.researchvoyage.com/credit-card-phd-research-scholar/

Wei, C.-Y., Chin, L.-K., Kuah, Y.-C., & Chia, M.-S. (2018). ATTRIBUTES OF ACADEMICIANS' CREDIT CARD USAGE BEHAVIORS IN MALAYSIA. *Journal of Contemporary Issues and Thought*, *8*. https://doi.org/10.37134/jcit.vol8.8.2018

Appendix

The dataset and R source codes is provided in the link below:
https://drive.google.com/drive/folders/1bUTJaA2V1-LEZ9CaLokyOgiZDr8K1stp?usp=sharing

The link for kaggle:
https://www.kaggle.com/datasets/thedevastator/predicting-credit-card-customer-attrition-with-m

Figure below is the screenshots of the dashboard in this project.

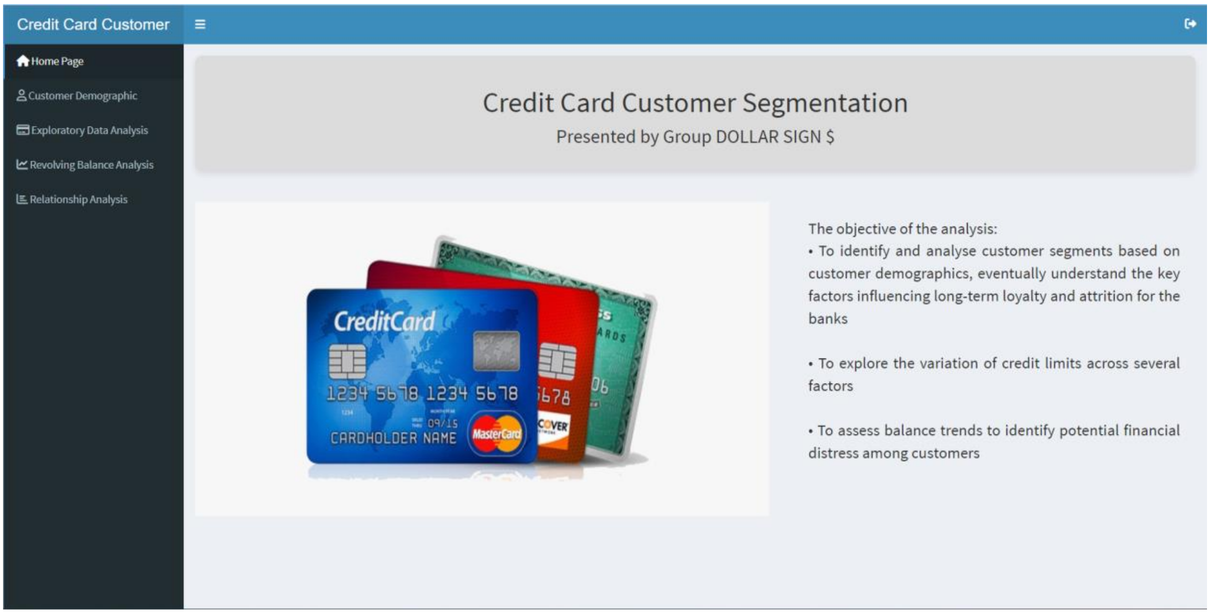
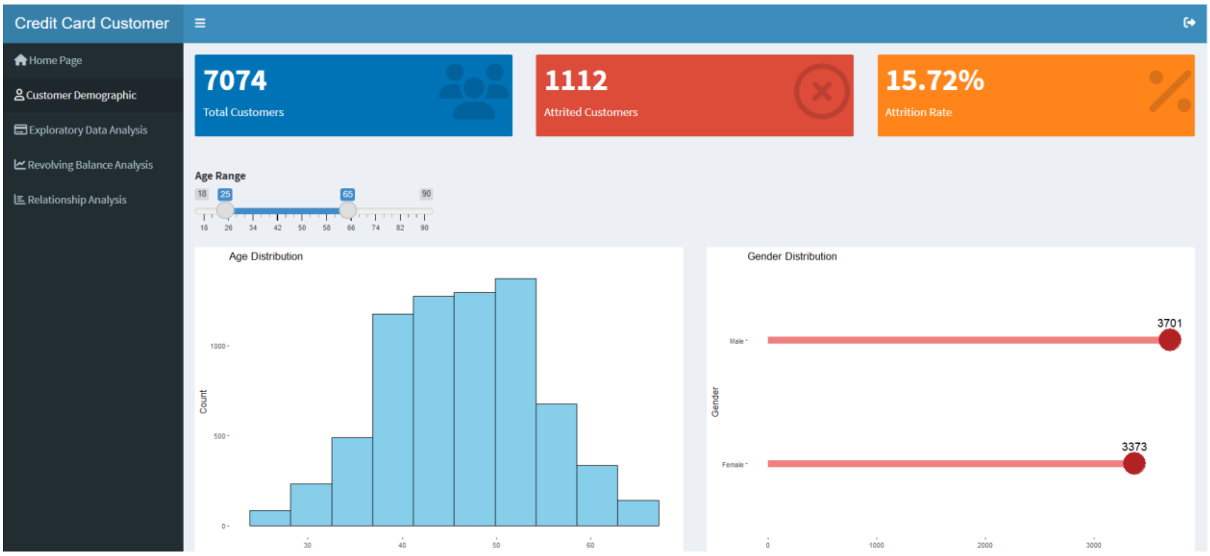


Figure 1: Home Page



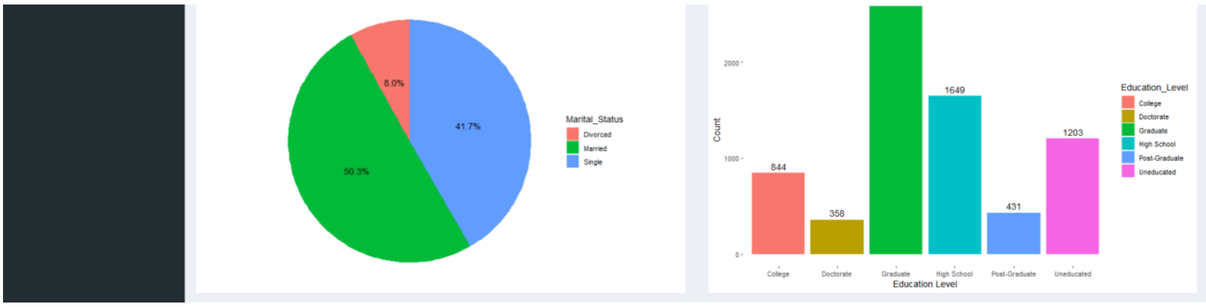


Figure 2: Customer Demographic

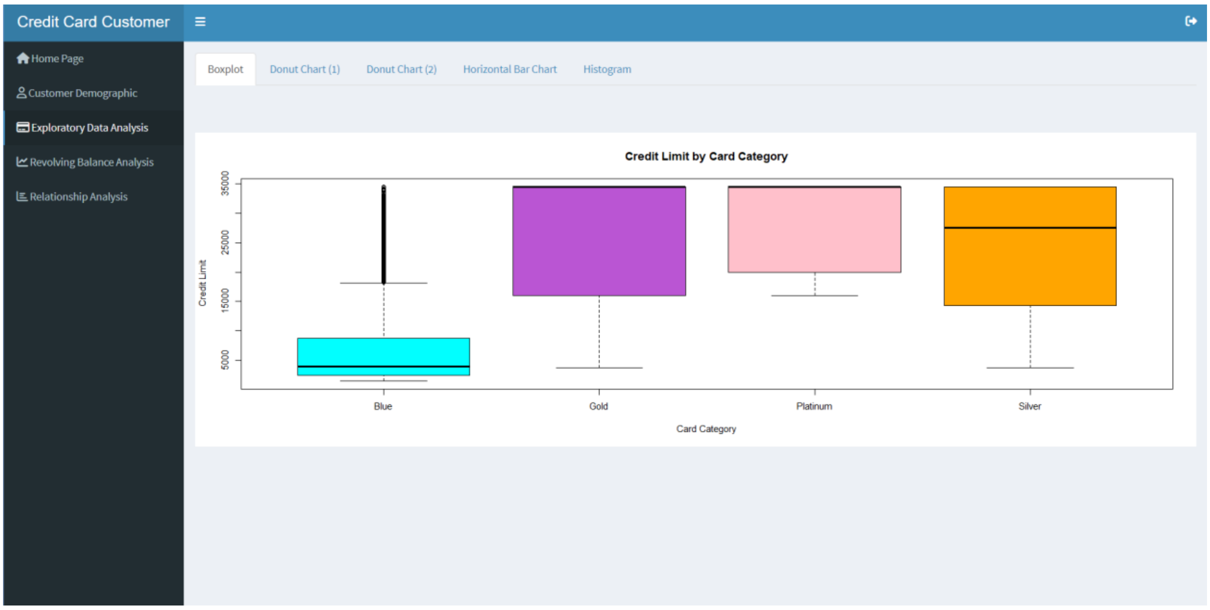


Figure 3: Exploratory Data Analysis (First Tab)

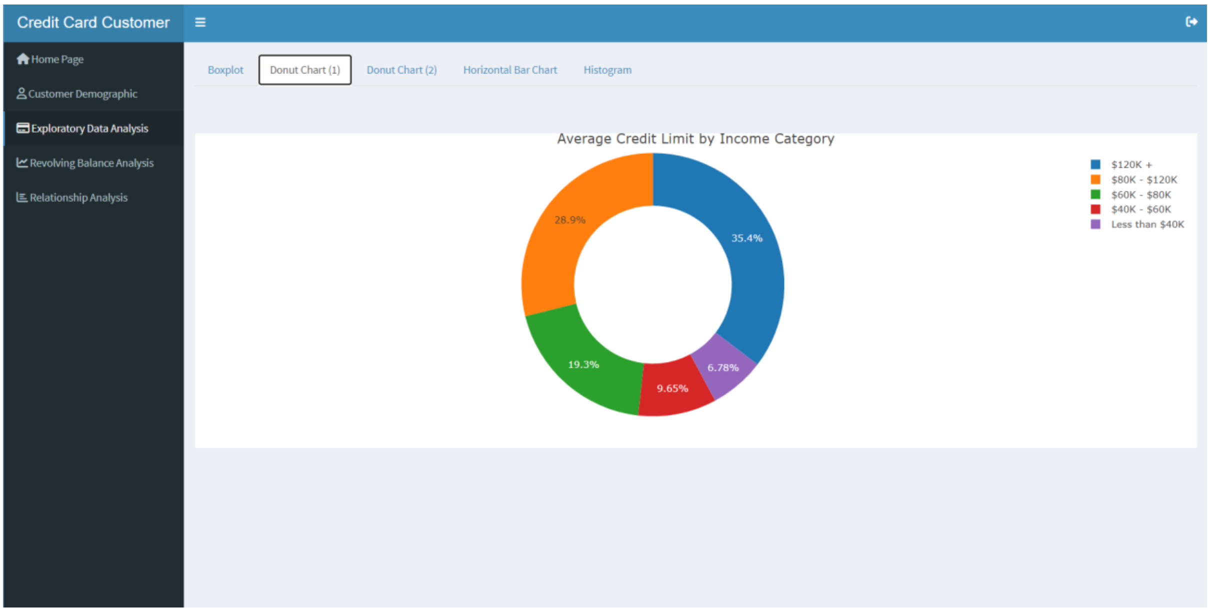


Figure 4: Exploratory Data Analysis (Second Tab)

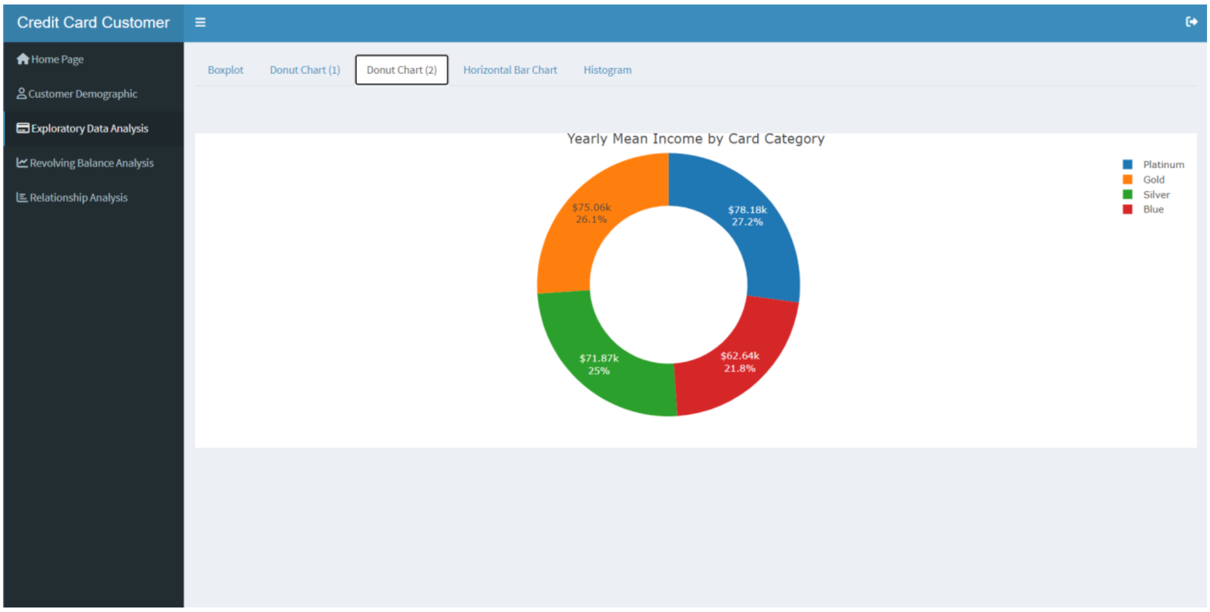


Figure 5: Exploratory Data Analysis (Third Tab)

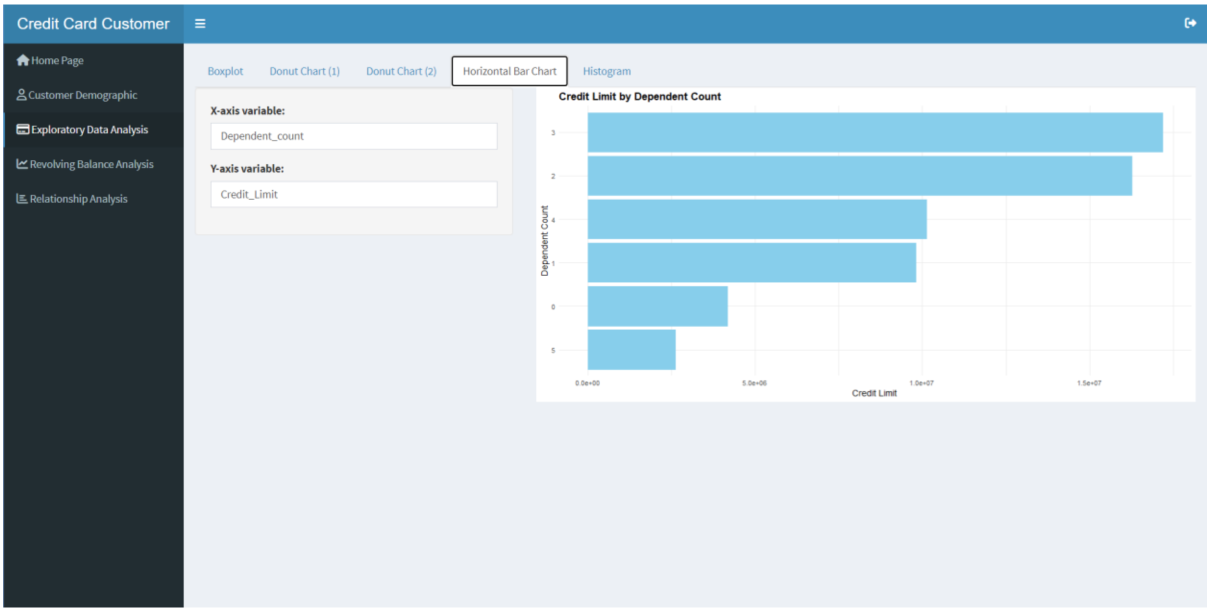


Figure 6: Exploratory Data Analysis (Fourth Tab)

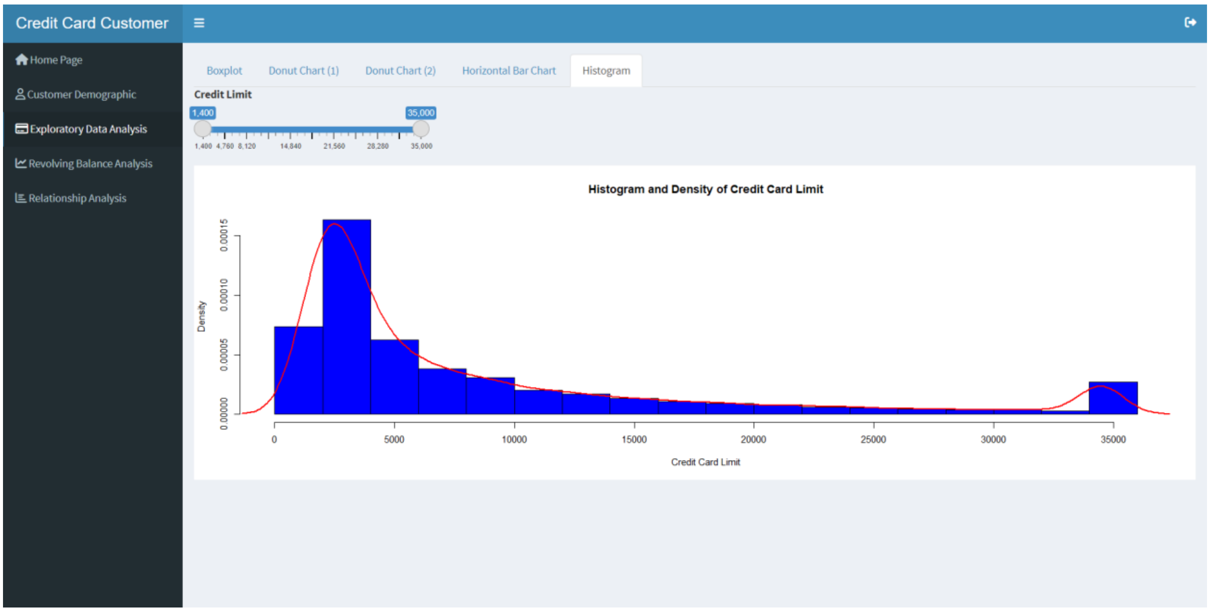


Figure 7: Exploratory Data Analysis (Fifth Tab)

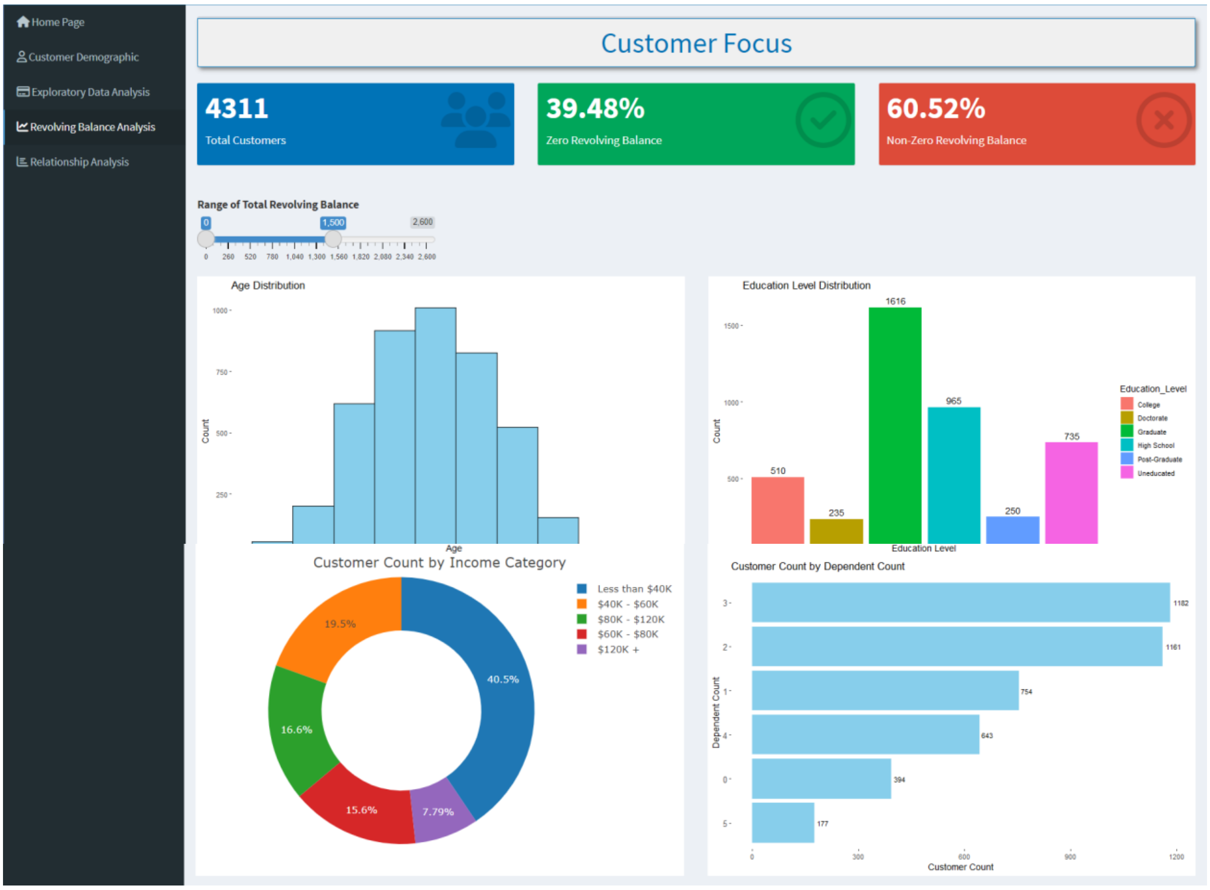


Figure 8: Revolving Balance Analysis

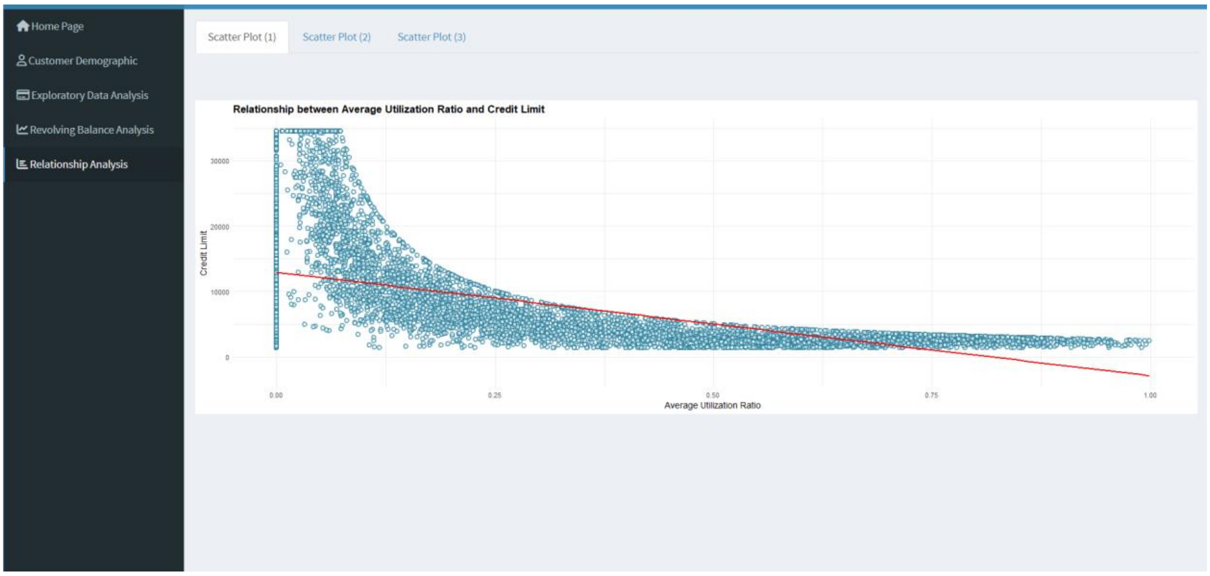


Figure 9: Relationship Analysis (First Tab)

Figure 10: Relationship Analysis (Second Tab)

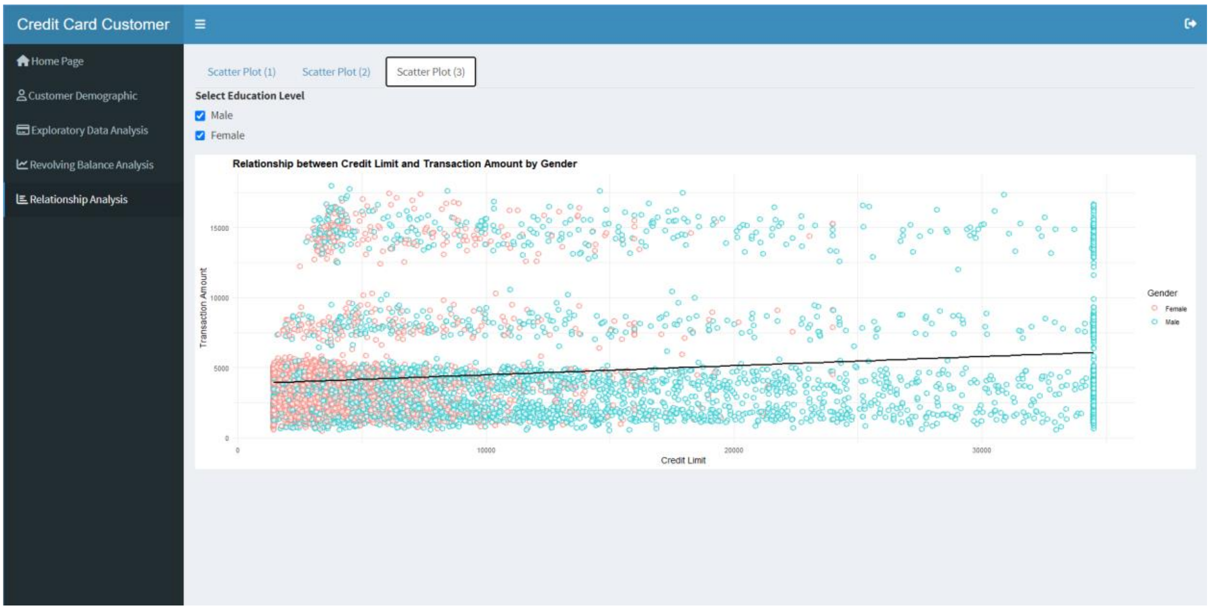


Figure 11 : Relationship Analysis (Third Tab)

	COURSE: DATA SCIENCE PROGRAMMING II (BSD2223)		MARKS: 120 (30%)
UNIVERSITI MALAYSIA PAHANG PUSAT SAINS MATEMATIK	NAME: ANG MEI YING	ID NO: SD22061	
	GROUP PROJECT	SECTION NO: 01G DUE DATE: 22/6/2023	/120

CLO	PLO	Marks	%
CLO2: Analyse and summarise data using appropriate programming tools	PLO2: Cognitive Skills and Functional work skills with focus on Numeracy skills. **C4: Analysis**	/40	/10
CLO3: Develop programming codes to solve problems	PLO3: Functional work skills with focus on Practical, and Digital skills. **P4: Mechanism**	/40	/10
CLO4: Demonstrate verbal and written communication skills	PLO5:Functional work skills with focus on communication skills. **A3: Valuing**	/20	/5
CLO5: Relate entrepreneur skills in assigned task	PLO8:Entrepreneural skills **A4: Organising values**	/20	/5

OBJECTIVES

Programming skill is vital to solve practical problems in various disciplines including data science, engineering and technology. This group project in Data Science Programming II requires students to develop a user-friendly dashboard or Graphical User Interface (GUI) using R language as the final output. An effective dashboard serves as a starting point and provides at-a-glance views of the data. From this project, the students should be able to apply all the fundamental of programming skills they have learned in this course. The project weights 30% of the total marks which composed of four (4) CLOs.

INSTRUCTIONS

Imagine you plan to become an entrepreneur/ business man/women in the near future. You are required to explore the types of business that you are interested to venture in order to prepare yourself for all the challenges that you may be faced. The decision to choose a particular type of business is based on data-driven decision. Go through the following steps as a guide to complete the group project:

1. You are assigned in a group of three (3) or four (4) members. Select a group leader.
2. The leader should coordinates and distributes the tasks among the group members. Each member must be responsible to ensure that all the given tasks are completed in time and well presented
3. Select a case study that is related to entrepreneurship or business e.g. ehailing services, ecommerce, food delivery services, etc. The suggested projects are given in the Appendix.
4. **Project Approval**
   Choose a project title that has available data. Free data can be accessed from various platforms such as github, kaggle and etc. Obtain an approval on the chosen topic/ project from your lecturer before or by **31st May 2023, 5.00 PM**.
5. **Project Scope**
   Data: Use the chosen data of sample size at least 1000 with minimum three (3) quantitative and qualitative data for each, respectively.
   Programming language: Use R language to analyse the data by following the data science process which was learned in the previous course such as Introduction to Data Science to obtain the data insights.
   Dashboard (GUI): Contruct a dashboard (GUI) in order to summarise and display the data using Rshiny package. The dashborad should have the following criterias/properties.
   - Clear and concise: Present information of the relevant data in a clear and concise manner.
   - User-friendly: Contain user-friendly elements with intuitive navigation and easy-to-use features where the user can interact with the dashboard and easily access the information they need.

Data Visualization: Use effective data visualization techniques, such as charts, graphs, and tables, to make the data more accessible and easy to understand.

Use navigation/information components eg. message boxes, pull-down menu and others.

Include other attractive elements including the use of colour, dashboard design and arrangement of information.

A catchy dashboard (GUI) name.

6. **Project Report Content**

Prepare a report consists of **10-15 pages** including appendices. Your project report must be in the given format and include the following information or components:

Cover page – Project title, name of group leader and members with ID no.

Content – List of Chapters and page numbers

Introduction – What is the study about? The objectives of the study?

Project description

The analysis of an existing case study which motivates this project.

Why this project is chosen?

Present the importance of this project and its innovative values.

State the ideas on how this project can be extended or improved.

Data description

Choose a set of data that can support the addressed issue with a sample size of at least 1000. The variables of the data should composed of quantitative and qualitative data.

Data Preparation

Conduct an appropriate exploratory data analysis on the chosen data to provide insightful interpretation of the descriptive statistics and graphical representations.

Data Analysis, Results and Discussion

Use statistical analysis to analyse the data and present the results and discussion on the findings.

Conclusion

Provide the main conclusion based on the statistical analysis and findings. The conclusion should be consistent with the objective of the addressed issue. In addition, provide the significance of the proposed solution related to entrepreneurship.

Limitations of the Study

Describe any limitations of your study and suggest how they might be overcome in a future project. Your description should be related to the data analysis.

Appendix

Includes all attachments of any relevant evidences (data, project approval form and etc.) Provide all the R source codes in completing this group project.

Complete GUI Screenshot

7. **Project Report Submission**

Submit the hardcopy report to the lecturer's office before or by 16th June 2023, 4:00 PM, AND The softcopy file in .pdf version through KALAM before or by 17th June 2023, 5:00 PM.

8. **Project presentation**

Prepare slides for 10-15 minutes presentation. Every group members should participate during the presentation session.

Presentation date will be on 21st June 2023 (Section 01G & 02G) or 22nd June 2023 (Section 03G). The detail schedule will be given later.

Submit the slide in .pdf version through KALAM by **21st June 2023, 5:00 PM (Section 01G & 02G)** or **22nd June 2023 (Section 03G), 5:00 PM**.

9. The project report will be graded based on the given rubrics attached in the Marking Sheet.
10. Progress of project will be checked regularly to avoid last minute attempt.
11. **LATE** submission of project will not be entertained and it will not be graded.

STUDENT'S INTEGRITY DECLARATION

I declare that this is my original work and own effort. I understand that copying from others and any act of cheating are strictly prohibited. If the incidents occurred, it is considered plagiarism and I will be penalized.

GROUP PROJECT PLAN & APPROVAL

Note: Print this form and obtain your lecturer's approval before starting the project's tasks.

<table>
<tr><td>SECTION NO.</td><td colspan="2">01G</td></tr>
<tr><td>GROUP NAME</td><td colspan="2">DOLLAR $IGN</td></tr>
<tr><td rowspan="5">GROUP MEMBERS
(Leader's name is the first in the list)</td><td>Id. No.</td><td>Name</td></tr>
<tr><td>SD22061</td><td>ANG MEI YING</td></tr>
<tr><td>SD22049</td><td>THAM ZHI YUN</td></tr>
<tr><td>SD22007</td><td>MIZA SYAZWANA BINTI SAFIAN</td></tr>
<tr><td>SD22035</td><td>NURUL NAJWA BINTI NORHISHAM</td></tr>
<tr><td>PROJECT TITLE</td><td colspan="2">Credit Card Customer Segmentation</td></tr>
<tr><td>PROJECT DESCRIPTION</td><td colspan="2">We will analyse customer demographics to understand the composition of our customer base. Additionally, we will examine the distribution and factors influencing credit card limits among our customers. The analysis will also cover revolving balances to identify patterns and trends. Finally, we will explore the relationships between various attributes to gain deeper insights into customer behaviour and financial metrics.</td></tr>
<tr><td>DATA DESCRIPTION</td><td colspan="2">This dataset contains a wealth of customer information collected within a customer credit card portfolio, with the aim of analysing the customer segmentation. Include age, gender, marital status, income category, card category, number months on book and inactive periods.</td></tr>
<tr><td>APPROVED BY
(signature)</td><td colspan="2">Associate Professor Dr. Roslinazairimah Zakaria</td></tr>
<tr><td>DATE</td><td colspan="2">3 APRIL 2024</td></tr>
</table>

	DATA SCIENCE PROGRAMMING II (BSD2223)		**MARKS: 120 (30%)**
	GROUP LEADER: ANG MEI YING	**ID NO: SD22061**	
	GROUP PROJECT	**SECTION NO: 01G** **DUE DATE: 19/6/2023**	/120

Marking Sheet (For Lecturer's use only)

RUBRICS FOR CLO2/PLO2

CLO2: Analyse and summarise data using appropriate programming tools	PLO2: Cognitive Skills and Functional work skills with focus on Numeracy skills. **C4: Analysis**	/40	/10

Criteria	Achievement Level						Weightage	Score
	0	1	2	3	4	5		
	Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Ability to **obtain appropriate data** for the project	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	2	10
Ability to **summarise data numerically and graphically using R**.	Unable to do the task.	Limited ability to do the task.	Able to do the task with errors.	Able to do the task with no errors but wrong answer.	Able to do the task with no errors.	Able to do the task efficiently and correctly with no errors.	2	10
Ability to **analyse data using R and obtain data insights**.	Unable to do the task.	Limited ability to do the task.	Able to do the task with errors.	Able to do the task with no errors but wrong answer.	Able to do the task with no errors.	Able to do the task efficiently and correctly with no errors.	2	10
Ability to **provide conclusion** and recommendation from the project.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	2	10

RUBRICS FOR CLO3/PLO3

CLO3: Develop programming codes to solve problems	PLO3: Functional work skills with focus on Practical, and Digital skills. **P4: Mechanism**	/40	/10

Criteria	Achievement Level						Weightage	Score
	0	1	2	3	4	5		
	Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Ability to **construct R codes to summarise data numerically and graphically**.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	2	10
Ability to **construct R codes for data analysis**.	Unable to do the task.	Limited ability to do the task.	Able to do the task with errors.	Able to do the task with no errors but wrong answer.	Able to do the task with no errors.	Able to do the task efficiently and correctly with no errors.	2	10
Ability to **develop a dashboard (GUI) to present the data using Rshiny.**	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	4	20

RUBRICS FOR CLO4/PLO5

CLO4: Demonstrate verbal and written communication skills	PLO5:Functional work skills with focus on communication skills. **A3: Valuing**	/20	/5

	Criteria	Achievement Level						Weightage	Score
		0	1	2	3	4	5		
		Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Writter	Ability to write the report findings coherently.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5
	Ability to present the project report in the given format which include data description, data analysis, results and discussion.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5
Communicati	Ability to present the project proficiently by organizing and communicating the results in a clear, logical, and easy-tofollow manner.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5
	Ability to deliver the dashboard to summarise the project findings.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5

RUBRICS FOR CLO5/PLO8

CLO5: Relate entrepreneur skills in assigned task	PLO8:Entrepreneural skills **A4: Organising values**	/20	/5

Criteria	Achievement Level						Weightage	Score
	0	1	2	3	4	5		
	Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Ability to articulate the given/ chosen case study related to entrepreneurship.	Unable to articulate the given/ chosen case study.	Able to articulate the given/ chosen case study fairly weak.	Able to articulate the given/ chosen case study fairly well.	Able to articulate the given/ chosen case study well.	Able to articulate the given/ chosen case study reasonably well.	Able to articulate the given/ chosen case study excellently.	2	10
Ability to deliver entrepreneur ideas.	Unable to deliver any entrepreneur idea.	Delivery of idea is unclear, vague and not systematic.	Delivery of idea is less clear, vague and systematic.	Delivery of idea is moderately clear, vague and systematic.	Delivery of idea is clear and systematic.	Delivery of idea is very clear and systematic.	2	10