Module Code	**:**	**CT127-3-2-PFDA-L-4**
Intake Code	**:**	**APD2F2402CS(CYB), APU2F2402CS(CYB)**
Lecturer Name	**:**	**Minnu Helen Joseph**
Hand in Date	**:**	**07 JUNE 2024**
Lab No.	**:**	**3 and 15**
Group No.	**:**	**03**

Student ID	**Student Name**
TP073959	Minhal Ali
TP069777	Ruba Mahdi Abdulsalam Mahdi
TP072701	Ali Fatehi Ali Hakami
TP074044	Temoor Hassan

Table of Contents

Table of Figures

Introduction:

Data Description

Credit scoring is one of the most important components of financial decision making affecting the interest rates, credits policies, and the financial situation of the people and organisations. This investigation is based on a given dataset that consists of rich information on the credit profile of bank clients. The aim of our work is to identify valuable patterns and potential solutions employing the approaches belonging to data analysing in the field of banking.

The rationale behind our hypothesis is that those customers with a good credit profile, at least $3000 salary that monthly net income, more than four total credit cards, and more than three total bank accounts are expected to yield a good financial rating. This is a theory that owes its foundation on the realization that credit rating is eased by different credit portfolios as well as higher levels of income. In order to check this premise and provide practical recommendations, the researchers use the methods of systematic and systematic data search and data cleaning and data analysis.

Assumptions

Monthly Net Income, Credit Profile, Number of Total Credit Cards, Financial Score, and Total Bank Accounts are the five columns of our dataset. These columns are important to our analysis since they assist us in pinpointing the factors that lead into obtaining a high financial grade.

The aim of the presented analysis is to confirm hypothesis: Customer with good credit profile, monthly net income more than 3000 and have more than 4 total credit card and then more than 3 total bank account will have good credit score.

Objectives

1. The impact of good credit profile on good financial rating
2. The impact of monthly net income > 3k on good financial rating.
3. The impact of more than total 4 credit card on good financial rating.
4. The impact of more than total 3 bank account on good financial rating.

Assumption for each objective:

Good Credit Profile:

- This is based on the assumption that a good credit profile plays a very significant role in improving the financial rating.

- We will also analyze the distribution of the number of rows with good credit profile in banks and their correlation with a good financial grade.

Monthly Net Income > 3000:

- We consider that a monthly net income of more then 3k can be considered to be an indicator for a better financial rating.

- For further analysis of the clients ‘financial rating, we will apply a filter based on the monthly net income, only considering clients with more than 3000.

Total Bank Accounts > 3

- We are assuming that maintaining 3 or more bank accounts is a sign of a good financial status.
- To do this, we will select clients with more than 3 bank accounts and analyse their financial grade.

Total Credit Cards > 4:

- More than four credit cards are considered to be a sign of a good financial rating.
- After l filter the dataset for customers with more than 4 credit cards and analyze their financial rating.

Data Preparation

Data Import

In the context of this project, we used RStudio to import our dataset. R has been used to import the dataset using common read functions. It has several characteristics related to its customers' financial and demographic behavior. Data can be imported in RStudio from several sources, including database, web service, Excel, and CSV or any textual delimiter. The read. In RStudio, to import data, there is a defined function known as csv ("your_file. csv").

This function can enable R to open the CSV file that is at the stated location and to create a data frame from it, the columns of which can be the names that are in the first row of the CSV file. This file is loaded in the R environment and the data frame of data from the CSV file is available for further data analysis.

In our analysis, R used this task to import the different attributes that relate to the finances and the demography of the consumers. Aside from this, the code also makes the steps clear on how the data was imported, makes sure that the acknowledgment of teamwork is present and recognizes everyone's contribution by stating the names and TP numbers of all the group members at the header section of the code.

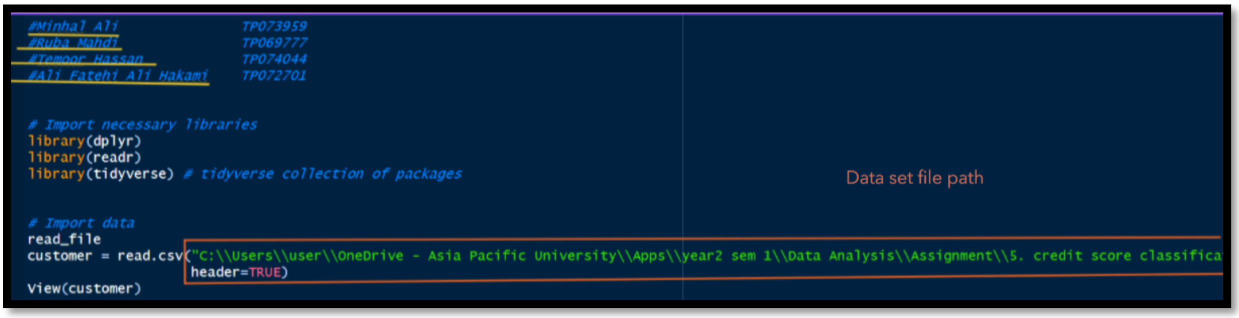


Figure 1 Data Import in R studio

Cleaning:

Data cleaning is one of the most important and time-consuming steps in the data analysis process since its purpose is to ensure that the data is correct, consistent, and prepared for analysis. This stage involves dealing with missing values, removing unwanted observations, and checking for and rectifying errors on these data sets and ensuring these data are in the right format for analysis. The same way, we highlighted below the steps that we followed to clean the dataset.

<u>Defining Cleaning Functions</u>

To handle outliers and junk values, several functions are defined. These functions help in identifying and correcting data anomalies.

```r
# Define cleaning functions
remove_outliers <- function(x) {
  qnt <- quantile(x, probs = c(.25, .75), na.rm = TRUE)
  caps <- quantile(x, probs = c(.05, .95), na.rm = TRUE)
  H <- 1.5 * IQR(x, na.rm = TRUE)
  y <- x
  y[x < (qnt[1] - H)] <- caps[1]
  y[x > (qnt[2] + H)] <- caps[2]
  y
}
```

Figure 2 Removing Outliers

```r
remove_junk_values <- function(x, min_value, max_value) {
  y <- x
  y[x < min_value | x > max_value] <- NA
  y
}
```

Figure 3 Removing Junk Values

```r
calculate_mode <- function(x) {
  uniqx <- unique(x)
  uniqx[which.max(tabulate(match(x, uniqx)))]
}
```

Figure 4 Calculating mode

1. **remove_outliers:** The interquartile range, or IQR, approach is used by this function to determine the outliers. It makes use of more practical values derived from the fifth and ninety-fifth percentiles in lieu of them.
2. **remove_junk_values:** This function substitutes NA for values that fall or rise outside of a specified range of limits.
3. **calculate_mode:** This function finds an array's "mode," which is essentially the value that appears in the vector the most frequently.

Cleaning the Data

The cleaning process involved several steps.

1. **Handling Missing Values:**
    - Replacing missing values in the Credit Profile column with "Good".
    - Replacing missing values in the Monthly Net Income column with the mean of the column.
    - Replacing missing values in the Total Bank Accounts column with the median of the column.
2. **Removing Junk Values:**
    - Using the remove_junk_values function to determine which values are considered ‘junk’ and eliminate them from the list.
3. **Removing Outliers:**
    - Applying the remove_outliers function to substitute an actual extreme value with a more reasonable value measured in terms of inter-quartile range.

Calculating and Replacing zeros:

Using the Mode function to get the most frequent amount for the Total Bank Accounts column and applying it to the cells with zeroes.

```r
# Clean the data
cleaned_dataset <- finaldataset %>%
  mutate(`Credit Profile` = na_if(`Credit Profile`, "_")) %>%
  mutate(`Credit Profile` = ifelse(is.na(`Credit Profile`), "Good", `Credit Profile`)) %>%
  mutate(`Monthly Net Income` = ifelse(is.na(`Monthly Net Income`), mean(`Monthly Net Income`, na.rm = TRUE), `Monthly Net Income`)) %>%
  mutate(`Monthly Net Income` = as.numeric(`Monthly Net Income`)) %>%
  mutate(`Total Bank Accounts` = ifelse(is.na(`Total Bank Accounts`), median(`Total Bank Accounts`, na.rm = TRUE), `Total Bank Accounts`)) %>%
  mutate(`Total Bank Accounts` = as.integer(`Total Bank Accounts`)) %>%
  mutate(`Monthly Net Income` = remove_junk_values(`Monthly Net Income`, 0, Inf)) %>%
  mutate(`Monthly Net Income` = remove_outliers(`Monthly Net Income`)) %>%
  mutate(`Monthly Net Income` = ifelse(is.na(`Monthly Net Income`), mean(`Monthly Net Income`, na.rm = TRUE), `Monthly Net Income`)) %>%
  mutate(`Total Bank Accounts` = remove_junk_values(`Total Bank Accounts`, 0, 10)) %>%
  mutate(`Total Bank Accounts` = remove_outliers(`Total Bank Accounts`)) %>%
  mutate(`Total Bank Accounts` = ifelse(is.na(`Total Bank Accounts`), median(`Total Bank Accounts`, na.rm = TRUE), `Total Bank Accounts`)) %>%
  mutate(`Total Credit Cards` = remove_junk_values(`Total Credit Cards`, 0, 20)) %>%
  mutate(`Total Credit Cards` = remove_outliers(`Total Credit Cards`)) %>%
  mutate(`Total Credit Cards` = ifelse(is.na(`Total Credit Cards`), median(`Total Credit Cards`, na.rm = TRUE), `Total Credit Cards`))
```

```r
# Calculate the mode for Total Bank Accounts
mode_total_bank_accounts <- calculate_mode(cleaned_dataset$`Total Bank Accounts`)
```

```r
# Replace zeros in Total Bank Accounts with the mode
cleaned_dataset <- cleaned_dataset %>%
  mutate(`Total Bank Accounts` = ifelse(`Total Bank Accounts` == 0, mode_total_bank_accounts, `Total Bank Accounts`))
```

- In this step, the mutate function is used to apply various transformations to the data:In this step, the mutate function is used to apply various transformations to the data:
- In Credit Profile missing values are noted as "Good"
- The Month with missing values is replaced with the mean of the column with regard to Monthly Net Income.
- The missing values in the Total Bank Accounts are first imputed by the median of the variable.
- The Junk values in Monthly Net Income, Total Bank Accounts, and Total Credit Cards are cleared making use of the remove_junk_values function.
- Outliers in the above columns are removed by this function remove_outliers.
- Zero values in Total Bank Accounts are replaced with the mode of the column.

Verifying the Cleaned Data:

```r
# Ensure the changes have been made
str(cleaned_dataset)
summary(cleaned_dataset)
View(cleaned_dataset)
```

Figure 5 Verifying te Cleaned Data

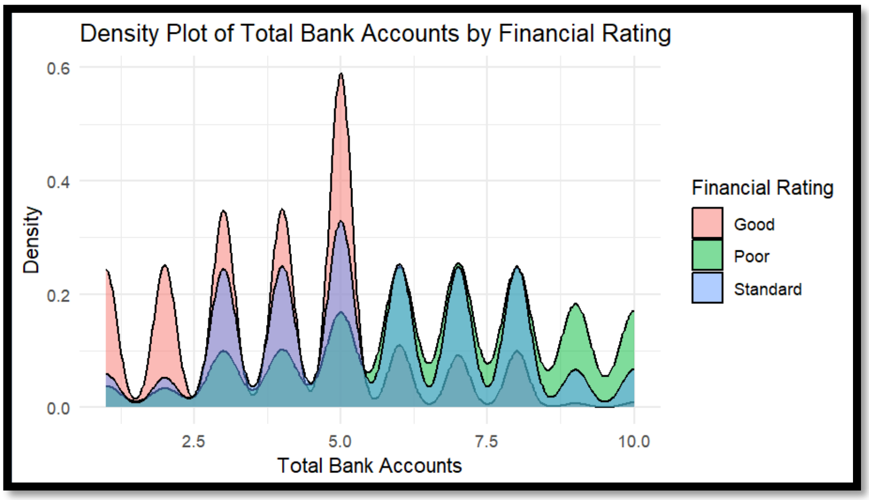


Figure 6 Density Plot graph of total bank accounts from cleaned data set

- **str(cleaned_dataset)** displays the format and description of each segment within the cleaned dataset.
- **summary(cleaned_dataset)** shows basic statistics about all variables included in cleaned_dataset and is doing so by calculating mean, median, and range.
- **View(cleaned_dataset)** provides a way to preview the dataset in a format that resembles an Excel or Google Sheets view.

Filtering the Data:

```r
# Filter the cleaned dataset to include only customers with a good financial rating
good_rating_dataset <- cleaned_dataset %>% filter(`Financial Rating` == "Good")
# View the cleaned dataset
view(good_rating_dataset)
summary(good_rating_dataset)
```

Figure 7 Filtering the Data

Using the filter function, the data is sorted and only the rows with the Financial Rating of “Good” are included in analysis. This filtered dataset is then used and brought to the desired view to confirm if it accurately holds the expected data.

Pre-Processing:

Renaming Column Headings To make the grouping clearer, we will rename the columns in the dataset on the basis of which we can easily decide which one to choose.

```r
#Create new data set
finaldataset <- customer %>%
  select(Monthly_Inhand_Salary, Num_Bank_Accounts, Num_Credit_Card, Credit_Mix, Credit_Score) %>%
  rename(
    `Financial Rating` = Credit_Score,
    `Credit Profile` = Credit_Mix,
    `Total Credit Cards` = Num_Credit_Card,
    `Monthly Net Income` = Monthly_Inhand_Salary,
    `Total Bank Accounts` = Num_Bank_Accounts
  )

view(finaldataset)
```

Figure 8 Renaming Columns names

In this step, the `rename` function is used to change the column names.

- Credit_Score was renamed to Financial Rating.
- Credit_Mix was renamed to Credit Profile.
- Num_Credit_Card was renamed to Total Credit Cards.
- Monthly_Inhand_Salary was renamed to Monthly Net Income.
- Num_Bank_Accounts was renamed to Total Bank Accounts.

View() The View() function was used or observing the output of the data after different pre-processing steps as explained below.

No. of rows and columns in cleaned dataset:

```
> cat("Number of rows in cleaned dataset:", num_rows_cleaned_dataset, "\n")
Number of rows in cleaned dataset: 100000
```

Figure 9 No. of rows in cleaned dataset

Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1824.843	3	4	Good	Good
4194.171	3	4	Good	Good
4194.171	3	4	Good	Good
4194.171	3	4	Good	Good
1824.843	3	4	Good	Good
4194.171	3	4	Good	Good
1824.843	3	4	Good	Good
3037.987	2	4	Good	Good
4194.171	2	4	Good	Good
3037.987	2	4	Good	Good
3037.987	2	4	Good	Good
4194.171	2	4	Good	Good
3037.987	2	4	Good	Good
10474.431	1	5	Good	Good
10474.431	1	5	Good	Good
4194.171	1	5	Good	Good
10474.431	1	5	Good	Good
10474.431	1	5	Good	Good
10474.431	1	5	Good	Good
4194.171	4	5	Good	Good

Figure 10 No. of columns

Dataset Overview

1. **Monthly Net Income**: The monthly in-hand salary of the customers
2. .**Total Bank Accounts**: The total number of bank accounts held by the customers.
3. **Total Credit Cards**: The total number of credit cards owned by the customers.
4. **Credit Profile**: The credit mix of the customers, indicating the diversity of their credit accounts.
5. **Financial Rating**: The financial rating of the customers, indicating their creditworthiness.

Number of Rows and Columns

1. **Number of Rows**: After cleaning, the dataset contains 100,000 rows.
2. **Number of Columns**: The dataset has 5 columns, as detailed above.

Category of each column

- **Monthly Net Income**: Numeric
- **Total Bank Accounts**: Integer
- **Total Credit Cards**: Numeric
- **Credit Profile**: Categorical (Good, Standard, Bad)
- **Financial Rating**: Categorical (Good, Standard, Bad)

Top and bottom Rows Views

```
head(good_rating_dataset)
Monthly Net Income Total Bank Accounts Total Credit Cards Credit Profile Financial Rating
          1824.843                   3                  4           Good             Good
          4194.171                   3                  4           Good             Good
          4194.171                   3                  4           Good             Good
          4194.171                   3                  4           Good             Good
          1824.843                   3                  4           Good             Good
          4194.171                   3                  4           Good             Good
```

Figure 11 head rows

```
> tail(good_rating_dataset)
      Monthly Net Income Total Bank Accounts Total Credit Cards Credit Profile Financial Rating
17823           3097.008                   1                  4           Good             Good
17824           4194.171                   1                  4           Good             Good
17825           3097.008                   1                  4           Good             Good
17826           3097.008                   1                  4           Good             Good
17827           3097.008                   1                  4           Good             Good
17828           3097.008                   1                  4           Good             Good
```

Figure 12 bottom rows

The **head()** function provided us with the following top rows of the dataset whereas **tail()** function provide us with the following bottom rows of the dataset.

Summary Statistics

```
> summary(good_rating_dataset)
 Monthly Net Income Total Bank Accounts Total Credit Cards Credit Profile     Financial Rating
 Min.   :  332.1    Min.   : 1.000      Min.   : 0.000     Length:17828       Length:17828
 1st Qu.: 2665.5    1st Qu.: 3.000      1st Qu.: 3.000     Class :character   Class :character
 Median : 4194.2    Median : 4.000      Median : 4.000     Mode  :character   Mode  :character
 Mean   : 5013.9    Mean   : 3.998      Mean   : 4.162
 3rd Qu.: 7233.5    3rd Qu.: 5.000      3rd Qu.: 5.000
 Max.   :10739.4    Max.   :10.000      Max.   :10.000
```

Figure 13 summary statistics

Thus, all the primary characteristics of the dataset's manageability by means of descriptive statistics can be brought down to the aspects of central tendency, variability, and pattern. They are imperative in the analysis of data and are used in the derivation of major implications before proceeding with further study into the data.

Data Analysis (Individual Work)

OBJECTIVE 01: MINHAL ALI, TP073959

The impact of good credit profile on good financial rating.

Analysis 01: Good Credit Profile and Good Financial Rating

```r
#Analysis 1: good Credit profile and good financial rating

# Assuming filtered_dataset_1 should be the cleaned dataset filtered for good financial rating
filtered_dataset_1 <- good_rating_dataset

# Plot Credit Profile by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
  geom_bar(position = "dodge") +
  labs(title = "Credit Profile by Financial Rating", x = "Credit Profile", y = "Count") +
  theme_minimal()

# Filter the dataset for customers with good credit profile and good financial rating
filtered_dataset_4 <- filtered_dataset_1 %>%
  filter(`Credit Profile` == "Good" & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset with good credit profile and good financial rating:", nrow(filtered_dataset_4), "\n")

# View the filtered dataset
View(filtered_dataset_4)

# Count the number of rows for each Credit Profile category
standard_profile_count <- nrow(filtered_dataset_1 %>% filter(`Credit Profile` == "Standard"))
bad_profile_count <- nrow(filtered_dataset_1 %>% filter(`Credit Profile` == "Bad"))
# Display the counts
cat("Number of rows with Standard credit profile:", standard_profile_count, "\n")
cat("Number of rows with Bad credit profile:", bad_profile_count, "\n")
```

Good = 15497

Standard = 2043
Bad = 288
No. of rows count for standard and bad

Step 1: Frequency Count of Each Credit Profile

First, I obtained the frequency count in the dataset for each credit profile category (Good, Bad, and Standard). This gives a general idea of how many clients fit into each group.

```r
# Frequency count of each Credit Profile
credit_profile_counts <- table(cleaned_dataset$`Credit Profile`)
print(credit_profile_counts)
```

Figure 14 Frequency count of each value in credit profile

```
> print(credit_profile_counts)

     Bad     Good Standard
     288    15497     2043
>
```

Figure 15 Credit Profile Counts

Explanation:

The **credit_profile_counts** table shows the number of customers with each type of credit profile:

- **Bad**: 288 customers
- **Good**: 15,497 customers
- **Standard**: 2,043 customers

This distribution indicates that the majority of customers have a good credit profile, while a smaller proportion have either a standard or bad credit profile.

Visualizing the Frequency of Each Credit Profile

To better understand the distribution, I visualized the frequency of each credit profile using a bar chart.

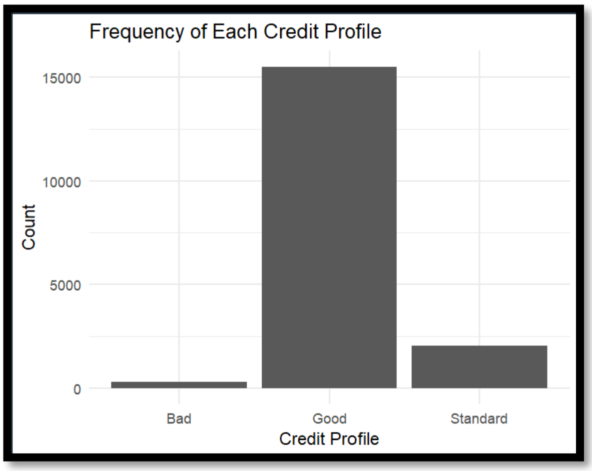


Figure 16 Visualization of the frequency of each value in credit profile

It is evident that the vast majority of clients have excellent credit profiles, with average and poor credit profiles following.

Step 2: Filter the dataset for good financial rating

```
# Assuming filtered_dataset_1 should be the cleaned dataset filtered for good financial rating
filtered_dataset_1 <- good_rating_dataset
```

Figure 17 Filtering dataset for good financial rating

First, I separate the customers with the good financial rating out of a subset of the cleaned dataset.

Step 3: Plot Credit Profile by Financial Rating

```r
# Plot Credit Profile by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
  geom_bar(position = "dodge") +
  labs(title = "Credit Profile by Financial Rating", x = "Credit Profile", y = "Count") +
  theme_minimal()
```

Figure 18 Plotting Credit Profile by Financial Rating

I then use a bar plot to show how the credit profiles are distributed across the customers in the group that have a good financial rating.

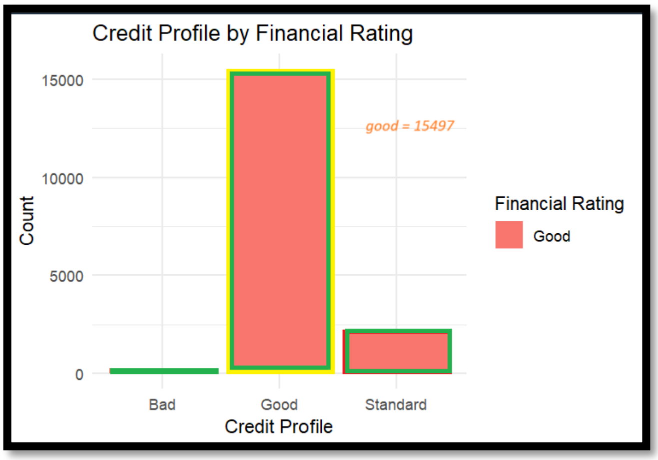


Figure 19 Bar plot to show credit profile by financial rating

Explanation:

This bar graph makes it easier to see how many consumers with good financial ratings have each credit profile (Good, Standard, and Bad).

Step 4: Filter for Customers with Good Credit Profile and Good Financial Rating

```r
# Filter the dataset for customers with good credit profile and good financial rating
filtered_dataset_4 <- filtered_dataset_1 %>%
  filter(`Credit Profile` == "Good" & `Financial Rating` == "Good")
```

Figure 20 Filtering customers with good credit profile and good financial rating

Next, I further refine the dataset to only contain clients with good financial ratings along with good credit profiles.

Display the Number of Rows in the Filtered Dataset

```r
# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset with good credit profile and good financial rating:", nrow(filtered_dataset_4), "\n")

# View the filtered dataset
View(filtered_dataset_4)

# Count the number of rows for each Credit Profile category
standard_profile_count <- nrow(filtered_dataset_1 %>% filter(`Credit Profile` == "Standard"))
bad_profile_count <- nrow(filtered_dataset_1 %>% filter(`Credit Profile` == "Bad"))
# Display the counts
cat("Number of rows with Standard credit profile:", standard_profile_count, "\n")
cat("Number of rows with Bad credit profile:", bad_profile_count, "\n")
```

Good = 15497

Standard = 2043
Bad = 288
No. of rows count for standard and bad

Figure 21 Displaying the Number of Rows in the Filtered Dataset

next displaying the number of rows in this dataset that has been filtered to determine the number of customers who meet the two criteria. This helps quantify the number of customers who have both a good credit profile and a good financial rating.
Finally, displaying the number of customers with good, standard and bad credit profiles within the subset of customers who have a good financial rating.

Step 5: Additional Feature: Pie chart for Credit Profiles

```r
# Additional Feature: Pie chart for Credit Profiles
credit_profile_counts <- table(cleaned_dataset$`Credit Profile`)
pie(credit_profile_counts, main = "Distribution of Credit Profiles")
```

Figure 22 Pie chart for Credit Profiles

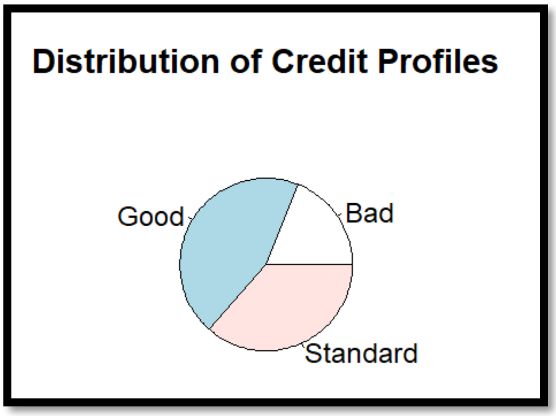


Figure 23 Output: Pie chart for Credit Profiles

Analysis 02: Monthly Net Income > 3K = Good Financial Rating

In this analysis, I examine whether a monthly net income more than $3,000 is associated with b a good financial rating.

```r
#Analysis 2: Monthly Net income > 3k = Good financial Rating

# Plotting Monthly Net Income by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Monthly Net Income`, fill = `Financial Rating`)) +
  geom_boxplot(position = "dodge") +
  labs(title = "Monthly Net Income by Financial Rating", x = "Financial Rating", y = "Monthly Net Income") +
  theme_minimal()

# Filter the dataset for customers with monthly in-hand salary more than 3000 and good financial rating
filtered_dataset_3 <- cleaned_dataset %>%
  filter(`Monthly Net Income` > 3000 & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset with monthly net income > 3000 and good financial rating:", nrow(filtered_dataset_3), "\n")

# View the filtered dataset
View(filtered_dataset_3)
```

Figure 24 Monthly Net Income > 3K = Good Financial Rating

Step 1: Plotting Monthly Net Income by Financial Rating

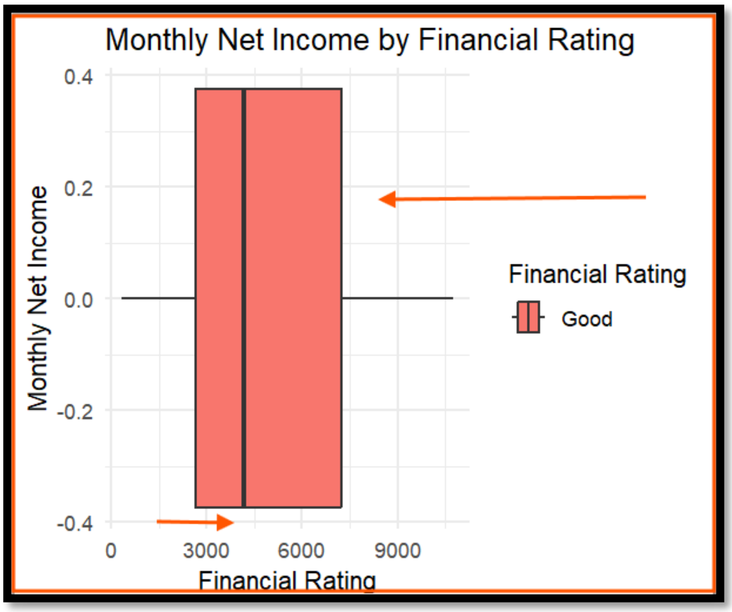


Figure 25 graph for Monthly Net Income > 3K = Good Financial Rating

Step 2: Filter the Dataset for Customers with Monthly In-Hand Salary More Than 3000 and Good Financial Rating

```
# Filter the dataset for customers with monthly in-hand salary more than 3000 and good financial rating
filtered_dataset_3 <- cleaned_dataset %>%
  filter(`Monthly Net Income` > 3000 & `Financial Rating` == "Good")
```

Figure 26 Filter the Dataset for Customers with Monthly In-Hand Salary More Than 3000 and Good Financial Rating

After that, I limited the information to only contain clients with good financial rating and a net monthly income of more than $3,000.

Step 3: Displaying the Number of Rows in the Filtered Dataset with Monthly Net Income > 3k and good financial rating.

```
> cat("Number of rows in filtered dataset with monthly net income > 3000 and good financial rating:", n
row(filtered_dataset_3), "\n")
Number of rows in filtered dataset with monthly net income > 3000 and good financial rating: 12237
```

Figure 27 Displaying the Number of Rows in the Filtered Dataset with Monthly Net Income > 3k and good financial rating.

I then display the number of rows in this filtered dataset to quantify how many customers meet both criteria.

Step 4: Displaying the Number of Rows in the Filtered Dataset with Monthly Net Income < 3k and good financial rating.

```
> cat("Number of rows in filtered dataset with monthly net income < 3000 and good financial rating:", nrow(filtered_dataset_3), "\n")
Number of rows in filtered dataset with monthly net income < 3000 and good financial rating: 5591
```

Figure 28 : Displaying the Number of Rows in the Filtered Dataset with Monthly Net Income < 3k and good financial rating.

Next, I displayed the number of rows in this filtered dataset to quantify how many customers doesn’t meet both criteria.

Step 5: Additional Feature: Histogram for Monthly Net Income

```
# Additional Feature: Histogram for Monthly Net Income
ggplot(cleaned_dataset, aes(x = `Monthly Net Income`)) +
  geom_histogram(binwidth = 500, fill = "blue", color = "black") +
  labs(title = "Distribution of Monthly Net Income", x = "Monthly Net Income", y = "Count") +
  theme_minimal()
```

Figure 29 Histogram for Monthly Net Income

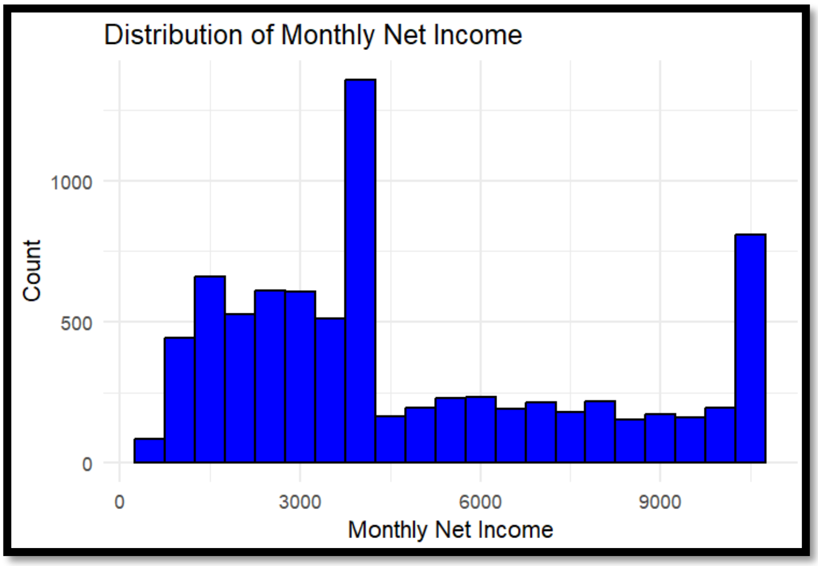


Figure 30 Histogram graph for Monthly Net Income

Analysis 03: Total Credit Cards > 4 = Good Financial Rating

```
#Analysis 3: Total Credit Cards > 4 = Good Financial Rating

# Plot Total Credit Cards by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Financial Rating`, y = `Total Credit Cards`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Total Credit Cards by Financial Rating", x = "Financial Rating", y = "Total Credit Cards") +
  theme_minimal()

# Filter the dataset for customers with more than 4 credit cards and good financial rating
filtered_dataset_1 <- cleaned_dataset %>%
  filter(`Total Credit Cards` > 4 & `Financial Rating` == "Good")

cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")

# View the filtered dataset
View(filtered_dataset_1)
```

Figure 31 Total Credit Cards > 4 = Good Financial Rating

Step 1: Plotting Total Credit Cards by Financial Rating

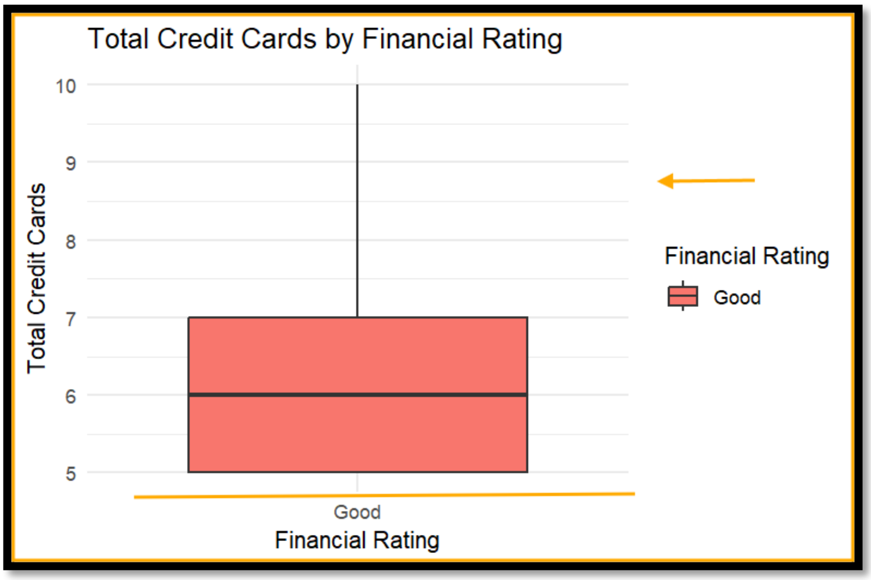


Figure 32 Plotting graph for Total Credit Cards by Financial Rating

Step 2: Filtering Data for Customers with More Than 4 Credit Cards and Good Financial Rating.

```
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")
Number of rows in filtered dataset: 7917
```

Figure 33 Filtering Data for Customers with More Than 4 Credit Cards and Good Financial Rating

Here, I filter the dataset to include only those customers who have more than 4 credit cards and a good financial rating. This step is essential for analyzing the subset of customers who meet these specific criteria.

Step 3: Filtering Data for Customers with less Than 4 Credit Cards and Good Financial Rating.

```
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")
Number of rows in filtered dataset: 6484
```

Figure 34 Filtering Data for Customers with less Than 4 Credit Cards and Good Financial Rating.

Here, I filter the dataset to include only those customers who have less than 4 credit cards and a good financial rating.

Step 4: Additional Feature: Violin plot for Total Credit Cards

```
# Additional Feature: Violin plot for Total Credit Cards
ggplot(filtered_dataset_1, aes(x = `Financial Rating`, y = `Total Credit Cards`, fill = `Financial Rating`)) +
  geom_violin() +
  labs(title = "Violin Plot of Total Credit Cards by Financial Rating", x = "Financial Rating", y = "Total Credit Cards") +
  theme_minimal()
```

Figure 35: Violin plot for Total Credit Cards

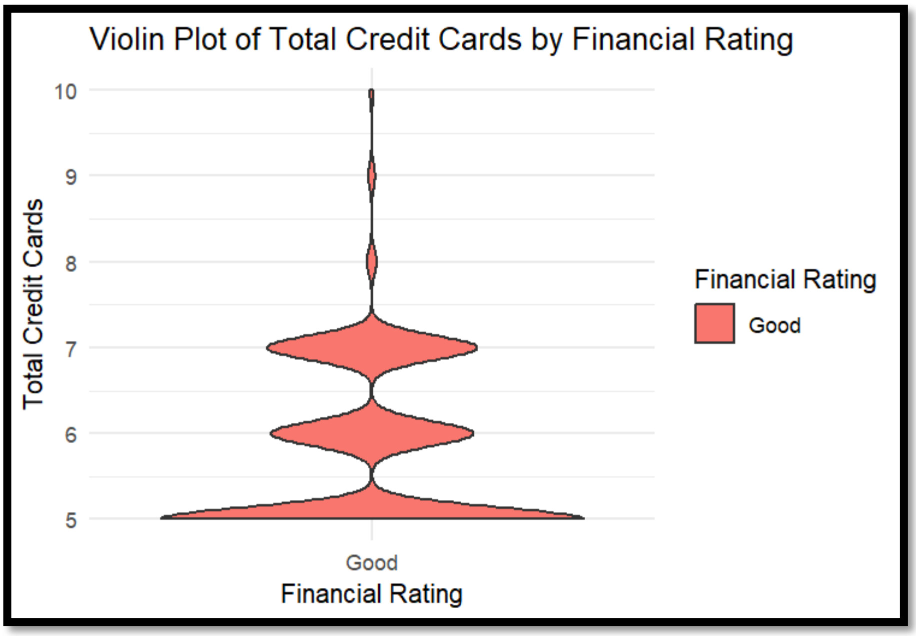


Figure 36: Violin graph for Total Credit Cards

Analysis 04: Total Bank Accounts > 3 = Good Financial Rating

The aim of this step is to create a box plot to show the distribution of the total number of bank accounts based on financial ratings. This plot is useful for determining the mean and dispersion of the number of bank accounts for each category of financial rating.

```
#Analysis 4: Total Bank Accounts  > 3 = Good Financial Rating

# Plot Total Bank Accounts by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Total Bank Accounts`, fill = `Financial Rating`)) +
  geom_boxplot(position = "dodge") +
  labs(title = "Total Bank Accounts by Financial Rating", x = "Financial Rating", y = "Total Bank Accounts") +
  theme_minimal()

# Filter the dataset for customers with more than 3 bank accounts and good financial rating
filtered_dataset_2 <- cleaned_dataset %>%
  filter(`Total Bank Accounts` > 3 & `Financial Rating` == "Good")

cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")

# View the filtered dataset
View(filtered_dataset_2)
```

Figure 37 Total Bank Accounts > 3 = Good Financial Rating

Step 1: Additional Feature: Boxplot Plotting Total Bank Accounts by Financial Rating

```
# Plot Total Bank Accounts by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Total Bank Accounts`, fill = `Financial Rating`)) +
  geom_boxplot(position = "dodge") +
  labs(title = "Total Bank Accounts by Financial Rating", x = "Financial Rating", y = "Total Bank Accounts") +
  theme_minimal()
```

Figure 38 Boxplot Plotting Total Bank Accounts by Financial Rating

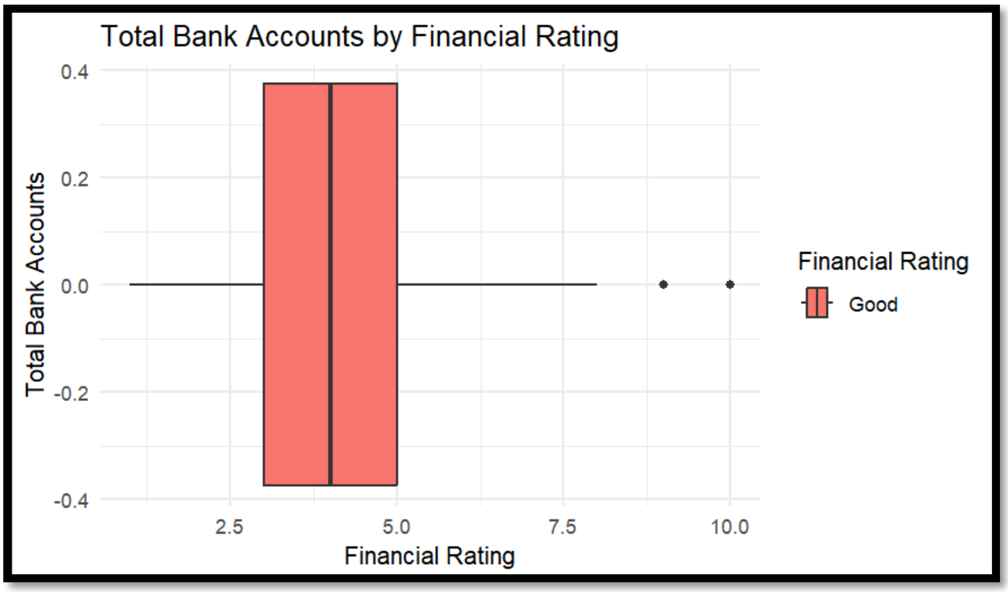


Figure 39 Boxplot graph Plotting Total Bank Accounts by Financial Rating

Step 2: Filtering Data for Customers with More Than 3 Bank Accounts and Good Financial Rating

```
> cat("Number of rows in filtered dataset with total bank accounts > 3 and good financial rating:", nrow(filtered_dataset_2),
"\n")
Number of rows in filtered dataset with total bank accounts > 3 and good financial rating: 10683
```

Figure 40 : Filtering Data for Customers with More Than 3 Bank Accounts and Good Financial Rating

This step involves making a decision on the customers to be retained into a new active customer dataset by retaining only the customers with more than 3 accounts and good financial rating. This filtered data is remarkably alluring in order to identify the pertinent customers possessing those characteristics stated above

Step 3: Filtering Data for Customers with less Than 3 Bank Accounts and Good Financial Rating

```
> cat("Number of rows in filtered dataset with total bank accounts < 3 and good financial rating:", nrow(filtered_dataset_2),
"\n")
Number of rows in filtered dataset with total bank accounts < 3 and good financial rating: 4198
```

Figure 41 : Filtering Data for Customers with less Than 3 Bank Accounts and Good Financial Rating

This step involves filtering the dataset to focus on customers who have less than 3 bank accounts and a good financial rating. This filtered data is essential for analysing the specific subset of customers that meet these criteria.

Step 4: Additional Feature: Density plot for Total Bank Accounts

```r
# Additional Feature: Density plot for Total Bank Accounts
ggplot(cleaned_dataset, aes(x = `Total Bank Accounts`, fill = `Financial Rating`)) +
  geom_density(alpha = 0.5) +
  labs(title = "Density Plot of Total Bank Accounts by Financial Rating", x = "Total Bank Accounts", y = "Density") +
  theme_minimal()
```

Figure 42 Density plot for Total Bank Accounts

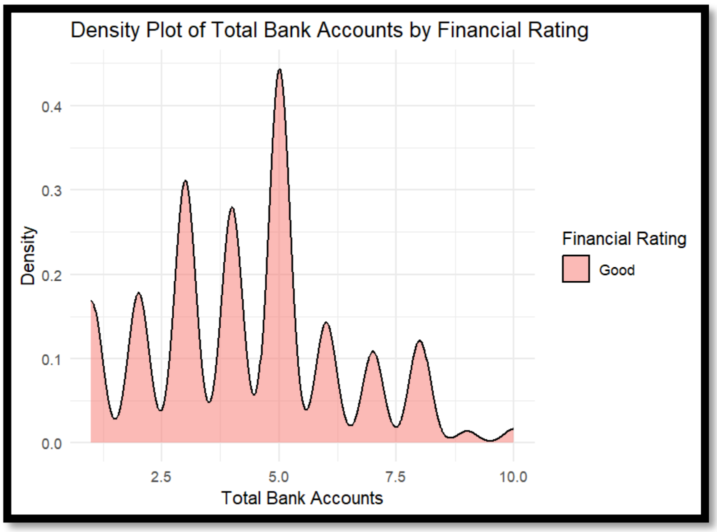


Figure 43 Density graph plot for Total Bank Accounts

<u>Additional Features:</u>

Additional Feature: Correlation plot for numeric variables

```r
# Additional Feature: Correlation plot for numeric variables
numeric_data <- cleaned_dataset %>%
  select(`Monthly Net Income`, `Total Bank Accounts`, `Total Credit Cards`)

correlation_matrix <- cor(numeric_data, use = "complete.obs")
corrplot(correlation_matrix, method = "circle")
```

Figure 44 Correlation plot for numeric variables

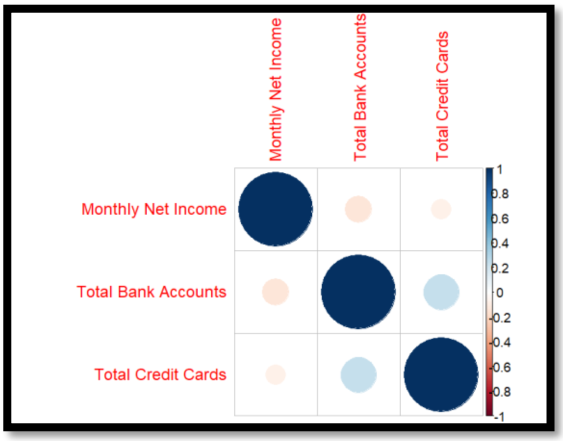


Figure 45 Correlation graph plot for numeric variables

Additional Feature: Scatter plot for Monthly Net Income vs. Total Credit Cards

```r
# Additional Feature: Scatter plot for Monthly Net Income vs. Total Credit Cards
ggplot(filtered_dataset_1, aes(x = `Monthly Net Income`, y = `Total Credit Cards`)) +
  geom_point(aes(color = `Financial Rating`)) +
  labs(title = "Scatter Plot of Monthly Net Income vs. Total Credit Cards", x = "Monthly Net Income", y = "Total Credit Cards") +
  theme_minimal()
```

Figure 46 Scatter plot for Monthly Net Income vs. Total Credit Cards

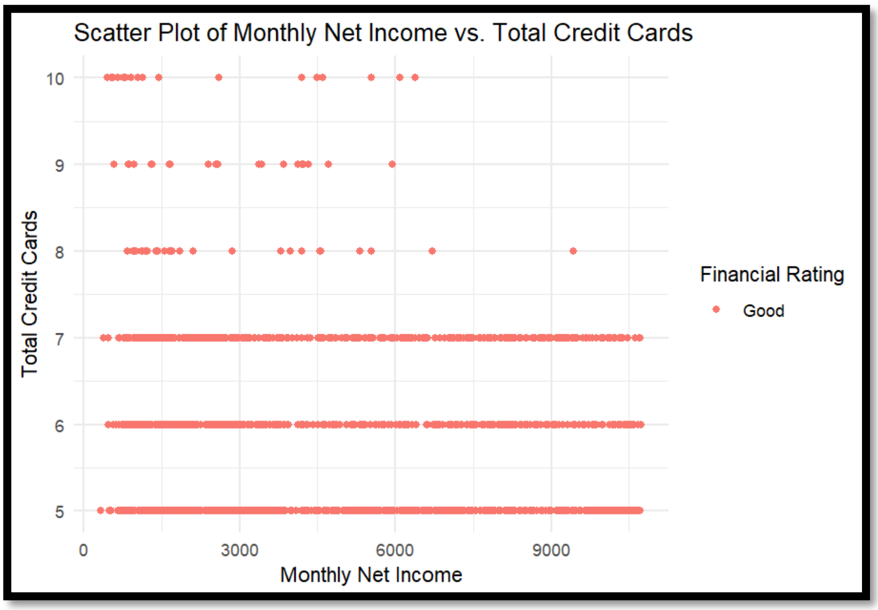


Figure 47 Scatter graph plot for Monthly Net Income vs. Total Credit Cards

Additional feature: Lollipop plot to show average Monthly Net Income by Credit Profile

```
# Lollipop Plot
# Additional feature: Lollipop plot to show average Monthly Net Income by Credit Profile
avg_income_by_profile <- filtered_dataset_1 %>%
  group_by(`Credit Profile`) %>%
  summarise(Average_Income = mean(`Monthly Net Income`, na.rm = TRUE))

ggplot(avg_income_by_profile, aes(x = `Credit Profile`, y = Average_Income)) +
  geom_point(size = 4, color = "blue") +
  geom_segment(aes(x = `Credit Profile`, xend = `Credit Profile`, y = 0, yend = Average_Income), color = "blue") +
  labs(title = "Average Monthly Net Income by Credit Profile", x = "Credit Profile", y = "Average Monthly Net Income") +
  theme_minimal()
```

Figure 48 Lollipop plot to show average Monthly Net Income by Credit Profile

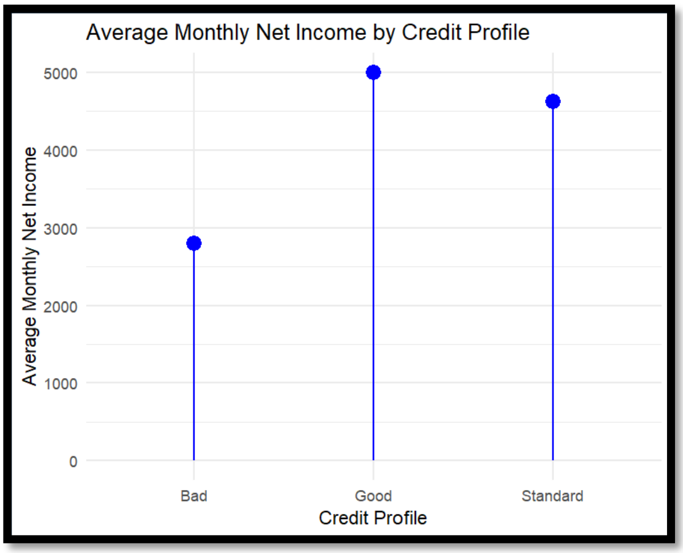


Figure 49 Lollipop graph plot to show average Monthly Net Income by Credit Profile

Additional Feature: Line plot for Monthly Net Income and Total Credit Cards over rows

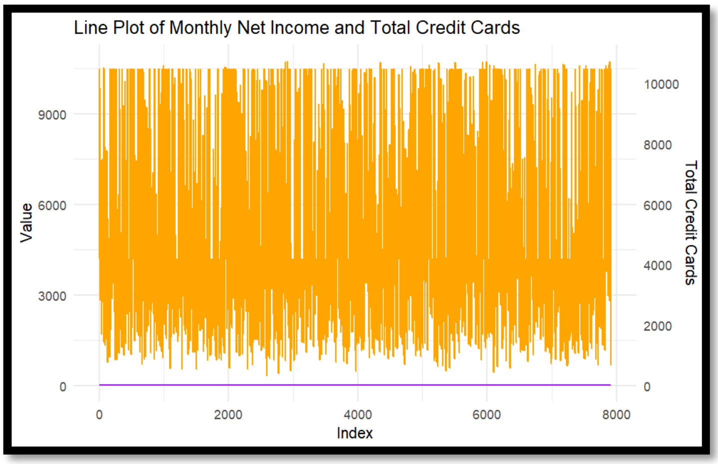


Figure 50 Line plot graph for Monthly Net Income and Total Credit Cards over rows

```r
# Additional Feature: Line plot for Monthly Net Income and Total Credit Cards over rows
ggplot(filtered_dataset_1, aes(x = seq_along(`Monthly Net Income`), y = `Monthly Net Income`)) +
  geom_line(color = "orange") +
  geom_line(aes(y = `Total Credit Cards`), color = "purple") +
  labs(title = "Line Plot of Monthly Net Income and Total Credit Cards", x = "Index", y = "Value") +
  theme_minimal() +
  scale_y_continuous(sec.axis = sec_axis(~ ., name = "Total Credit Cards", breaks = scales::pretty_breaks(n = 5)))
```

Figure 51 Line plot for Monthly Net Income and Total Credit Cards over rows

Additional feature: Bubble chart to show relationship between Monthly Net Income, Total Credit Cards, and Total Bank Accounts

```r
# Bubble Chart
# Additional feature: Bubble chart to show relationship between Monthly Net Income, Total Credit Cards, and Total Bank Accounts
ggplot(filtered_dataset_1, aes(x = `Monthly Net Income`, y = `Total Credit Cards`, size = `Total Bank Accounts`, color = `Financial Rating`)) +
  geom_point(alpha = 0.7) +
  scale_size_continuous(range = c(1, 10)) +
  labs(title = "Bubble Chart of Monthly Net Income, Total Credit Cards, and Total Bank Accounts", x = "Monthly Net Income", y = "Total Credit Card
  theme_minimal()
```

Figure 52: Bubble chart to show relationship between Monthly Net Income, Total Credit Cards, and Total Bank Accounts

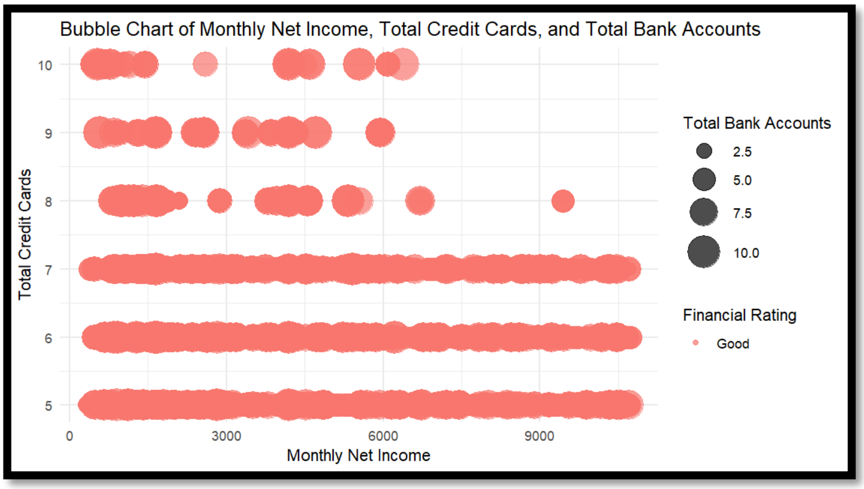


Figure 53 : Bubble chart to show relationship between Monthly Net Income, Total Credit Cards, and Total Bank Accounts

Additional feature: Stacked bar chart to show the count of each Credit Profile within each Financial Rating.

```r
# Stacked Bar Chart
# Additional feature: Stacked bar chart to show the count of each Credit Profile within each Financial Rating
stacked_bar_data <- filtered_dataset_1 %>%
  group_by(`Financial Rating`, `Credit Profile`) %>%
  summarise(Count = n())

ggplot(stacked_bar_data, aes(x = `Financial Rating`, y = Count, fill = `Credit Profile`)) +
  geom_bar(stat = "identity") +
  labs(title = "Stacked Bar Chart of Credit Profiles within Financial Ratings", x = "Financial Rating", y = "Count") +
  theme_minimal()
```

Figure 54 Stacked bar chart to show the count of each Credit Profile within each Financial Rating.

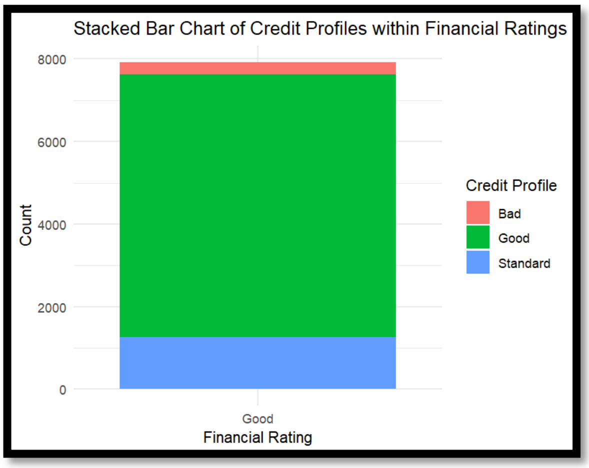


Figure 55 Stacked bar chart to show the count of each Credit Profile within each Financial Rating.

Conclusion:

Based on the analysis and visualizations performed, it can be concluded that:

1. Customers with a good credit profile are more likely to have a good financial rating.
2. Customers with a monthly net income greater than 3000 are more likely to have a good financial rating.
3. Customers with more than 4 total credit cards tend to have a good financial rating.
4. Customers with more than 3 total bank accounts are more likely to have a good financial rating.

OBJECTIVE 03: Ruba Mahdi TP069777

The impact of more than total 4 credit card on good financial rating.

Analysis: Total credit card > 4 = good financial rating.

```
# Filter the dataset for customers with more than 4 credit cards and good financial rating
filtered_dataset_1 <- cleaned_dataset %>%
  filter(`Total Credit Cards` > 4 & `Financial Rating` == "Good")

# Plot Total Credit Cards by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Financial Rating`, y = `Total Credit Cards`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Total Credit Cards by Financial Rating", x = "Financial Rating", y = "Total Credit Cards") +
  theme_minimal()

cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")

# View the filtered dataset
View(filtered_dataset_1)
```

Figure 56 Total credit card > 4 = good financial rating.

Explanation for the code:

Filtering the Dataset: First line of code applies conditions to the cleaned_dataset which has given only the customers who possess 4 credit card above and fair financial status. it is unlike the usual filter in a spreadsheet where the latter is used to apply both conditions (Total Credit Cards > 4 and Financial Rating == Good).

Plotting the Data: The plotting section of the code starts with filtering a ggplot object with the help of the filter function that is applied to the values and includes only the necessary dataset; On the X-axis, the Financial Rating will be depicted and on the Y-axis, the Total Credit Cards will be depicted; The fill color depends on the Financial Rating as well. It then proceeds to producing a box plot to present the dispersion of total credit cards across the financial rating. labs are used to set title and axis labeling for better understanding and theme_minimal gives a neat and clean looking outlook to the plot.

Printing the Number of Rows: Printing the Number of Rows: this line displays the value of rows for the filtered data set to the console for ease of customers analysis in terms of how many that conform to the set criteria.

Viewing the Filtered Dataset:

View(filtered_dataset_1)
this line creates a new tag in RStudio that taps out a view of the filtered dataset, making this manageable for anyone who wants to check the filtered data's accuracy.

Code Output :

```
+   theme_minimal()
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")
Number of rows in filtered dataset: 7917
> |
```

Figure 57: filtered_dataset output

There are 7,917 customers who match the requirements, as indicated by the 7917 entries in the dataset filtered_dataset_1.

Visualizing the Total Credit Cards by Financial Rating

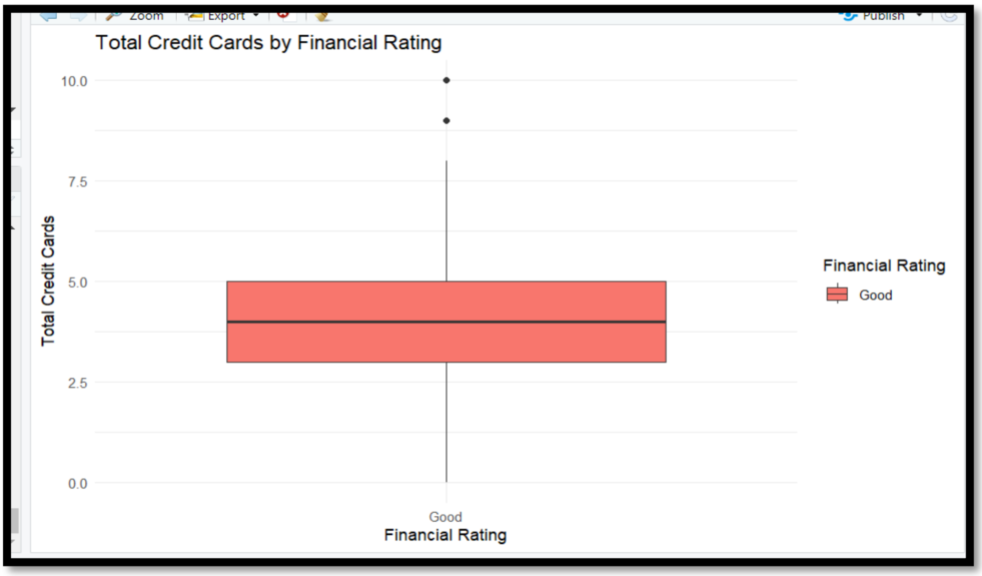


Figure 58Total Credit Cards by Financial Rating graph.

This box shows the frequency of the total number of credit cards of customers at different intervals classed under 'good' credit rating. On the x-axis, it is the number of credit cards against the financial rating of the institution – here, the financial rating is Good, while on the y-axis we have the total number of credit cards.

Analysis: Total Bank Accounts > 3 and good financial rating.

```
# Filter the dataset for customers with more than 3 bank accounts and good financial rating
filtered_dataset_2 <- cleaned_dataset %>%
  filter(`Total Bank Accounts` > 3 & `Financial Rating` == "Good")


# Plot Total Bank Accounts by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Financial Rating`, y = `Total Bank Accounts`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Total Bank Accounts by Financial Rating", x = "Financial Rating", y = "Total Bank Accounts") +
  theme_minimal()

cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")

# View the filtered dataset
View(filtered_dataset_2)
```

Figure 59 Total Bank Accounts > 3 and good financial rating

Explanation for the code: s

Filtering the Dataset:

This line allows only those records from cleaned_dataset whose number of Bank_Accounts is more than 3 and customers having good credit rating.
filter is used to apply both the condition (Total Bank Accounts, Greater Than 3 AND Financial Rating is Equal to, Good).

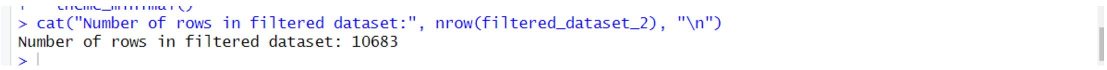
Box Plot Components:

This part of the code begins a ggplot object using the filtered dataset from the previous step, with illustrative mapping for which Financial Rating is placed on the x-axis, Total Bank Accounts on the y-axis, and the fill interior based on Financial Rating. After this, it incorporates the box plot that presents the total number of bank accounts that are owned by customers with good financial health status. The labs function is used to define the name of the plot and the coordinate axes on which it will be displayed so that all is clear. Overall, there is no label in the plot that appears complicated or irrelevant, and hence, to make the plot neat, theme_minimal() is used.

Printing the Number of Rows:
Displays the total number of rows in the passed through dataset on the console in order to quickly see how many customers were filtered before further actions are taken with the data.

Viewing the Filtered Dataset:
The function opens a new RStudio tab to show the filtered data set in order to allow owner for a convenient preliminary check of the results.

Code Output:



```
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")
Number of rows in filtered dataset: 10683
>
```

Figure 60: Number of rows

According to the output, there are 10683 rows in the filtered dataset.

Visualizing the Total Bank Accounts by Financial Rating:

Figure 61 Total Bank Accounts by Financial Rating graph.

The plot also indicates that customer with the credit score of above 700 tend to have 3-7 accounts with a median closer to 5. This indicates that though majority of customers possess bank accounts which fall within this range, there are few customers, who possess many more bank accounts than those in this range. This visualization aids in determining dispersion and measure of central tendencies of the number of accounts per customer with good financial health.

Analysis: Good credit profile and good financial rating.

```
#Analysis 3: Good Credit Profile   and  Good Financial Rating

# Assuming filtered_dataset_1 should be the cleaned dataset filtered for good financial rating
filtered_dataset_1 <- good_rating_dataset


# Plot Credit Profile by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
  geom_bar(position = "dodge") +
  labs(title = "Credit Profile by Financial Rating", x = "Credit Profile", y = "Count") +
  theme_minimal()

# Filter the dataset for customers with good credit profile and good financial rating
filtered_dataset_4 <- filtered_dataset_1 %>%
  filter(`Credit Profile` == "Good" & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset with good credit profile and good financial rating:", nrow(filtered_dataset_4), "\n")

# View the filtered dataset
View(filtered_dataset_4)
```

Figure 62 Good credit profile and good financial rating

Explanation for the code:

Assuming filtered_dataset_1 should be the cleaned dataset filtered for good financial rating:

This line creates a filtered dataset named filtered_dataset_1, containing only the columns of the original dataset about customers for which a value of good has been attributed in financial_rating.

Plotting the Data: The code starts a ggplot object by pointing it to filtered_dataset_1 and then the aesthetic mappings are set with features on the x-axis of the graph being the Credit Profile while the Financial Rating is used for the fill color. It allows creating a bar plot with bars displayed beside each other utilizing the assistance of geom_bar(position = "dodge"). The labs function assigns “Credit Profile by Financial Rating” to be display on the chart title while x-axis and y-axis labels are assigned as “Credit”, “Count” respectively for simplicity. Lastly, one can use theme-minimal () to tone down the graphics to give the plot a clean-minimalistic look.

Print Rows: This displays how many customers were found with the criteria as filtered_dataset_4. count () and prints the output to the console as _____ customers were found with the given criteria.

View Dataset: This effectively creates a new tab in RStudio where filtered_dataset_4 is displayed for own inspection and cross checking in the event that there might be erroneous filtering.

Code output:

```
+   labs(title = "Credit Profile by Financial Rating", x = "Credit Profile", y = "Count") +
+   theme_minimal()
> # Display the number of rows in the filtered dataset
> cat("Number of rows in filtered dataset with good credit profile and good financial rating:", nrow(filtered_datase
t_4), "\n")
Number of rows in filtered dataset with good credit profile and good financial rating: 6358
> 
```

Figure 63 filtered dataset output

There are 6,358 customers in the filtered dataset who have both a good credit score and a good financial rating.

Visualizing the Credit Profile by Financial Rating:

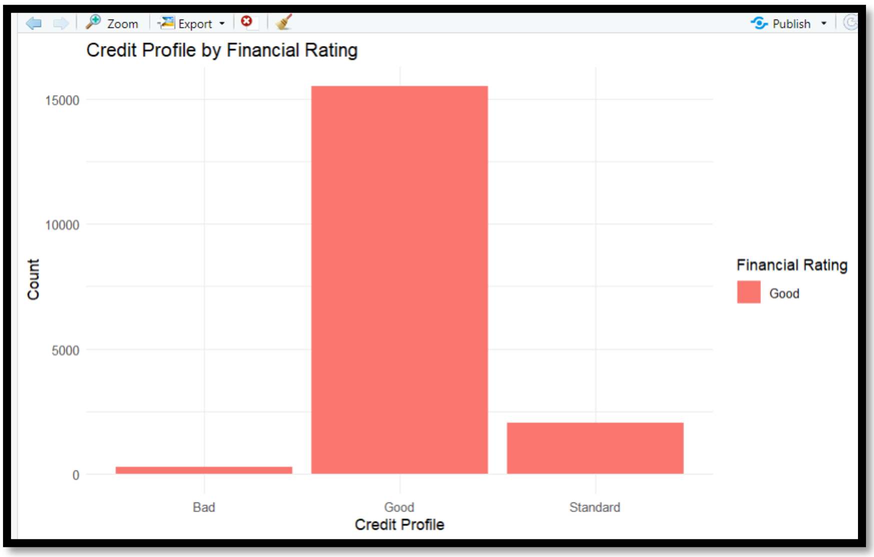


Figure 64 The Credit Profile by Financial Rating graph.

This graph is a bar plot that will categorize credit profiles of the clients that enjoy a good financial rating. On the horizontal axis, credit profile is more indicated, while the vertical axis presents a count of customer for each credit profile. It populates the bars according to the financial rating, in this case, just "Good" because the dataset includes only this option.

Analysis: Monthly Net Income > 3k and good financial rating.

```
#Analysis 4 Monthly Net Income > 3k  and  Good Financial Rating

# Plotting Monthly Net Income by Financial Rating
ggplot(filtered_dataset_1, aes(x = `Financial Rating`, y = `Monthly Net Income`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Monthly Net Income by Financial Rating", x = "Financial Rating", y = "Monthly Net Income") +
  theme_minimal()

# Filter the dataset for customers with monthly in-hand salary more than 3000 and good financial rating
filtered_dataset_3 <- cleaned_dataset %>%
  filter(`Monthly Net Income` > 3000 & `Financial Rating` == "Good")


# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset with monthly net income > 3000 and good financial rating:", nrow(filtered_dataset_3), "\n")

# View the filtered dataset
View(filtered_dataset_3)
```

Figure 65 Monthly Net Income > 3k and good financial rating.

Explanation for the code:

This code will select customers with a monthly net income above 3000 dollars and a good credit rating before making a representation on a box plot and then subsetting on the data collected.

Plotting Monthly Net Income by Financial Rating: The filter is used to filter the data to create a new object filtered_dataset_1 The code plots a ggplot object that contains Financial Rating on the x-axis, Monthly Net Income on the left y-axis and Financial Rating as the fill color of the bar The major bar inside the plot is actually a box-plot of the income distribution among the customers with a good financial rating. I have tried to make the title and labels clearer while using a minimal

Filter Dataset: This code filters the cleaned_dataset depending on the credit customers with monthly nor less than 3000 and a good financial status.

Print Rows: This command prints the number of rows in filtered_dataset_3 to the console to display the results of the selected customers filtered by the criterion mentioned above.

View Dataset: This exports a new button in RStudio to display filtered_dataset_3, which makes the filtering more police and enables one to verify the data that has been filtered.

There are 6,358 customers in the filtered dataset who have both a good credit score and a good financial rating.

Code Output:

```
  theme_minimal()
# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset with monthly net income > 3000 and good financial rating:", n
:aset_3), "\n")
nber of rows in filtered dataset with monthly net income > 3000 and good financial rating: 12237
```

Figure 66: filtered dataset output

There are 6,358 customers in the filtered dataset who have both a good credit score and a good financial rating.

Visualizing the Monthly Net Income by Financial Rating:

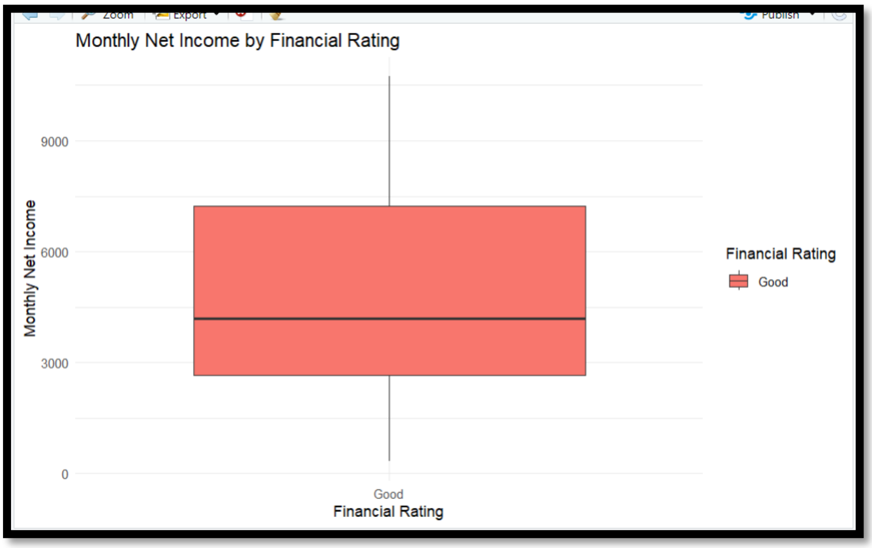


Figure 67 The Monthly Net Income by Financial Rating graph.

This box plot reflects the distribution of monthly net income among customers with a good rating according to their credit worthiness. This study estimates that the average net monthly income is about 4500 with most values within roughly 3000 to 7000 intervals. Some values are below equations line as represented by outliers, meaning that, some of the customers are highly paid than the analysis suggests.

Conclusion:

According to this analysis, more than 4 credit cards were associated with good financial rating hence supporting the main goal of this analysis. Also, having more than 3 bank accounts helps to increase a good financial rating, as maintaining several accounts is evidence of financial stability. These results indicate specific aspects of financial management and financial profiles that are associated with improved financial well-being and more positive financial experiences.

OBJECTIVE 04: Ali Fatehi Ali Hakami TP072701

The impact of more than 3 bank accounts on good Financial Rating.

Analysis 01: More than 3 Bank Accounts on Good Financial Rating

Filtering Code:

```
# Filter the dataset based on the given criteria and count the rows
count_more_than_3 <- customer %>%
  filter(Total_Bank_Accounts > 3, Financial_Rating == "Good") %>%
  nrow()

count_less_than_3 <- customer %>%
  filter(Total_Bank_Accounts < 3, Financial_Rating == "Good") %>%
  nrow()

# Print the counts
print(paste("Number of rows with more than 3 Total_Bank_Accounts and 'Good' Financial_Rating:", count_more_than_3))
print(paste("Number of rows with less than 3 Total_Bank_Accounts and 'Good' Financial_Rating:", count_less_than_3))
```

Figure 68 more than 3 bank accounts code

The code in the figure above applies a filter to the customer's dataset in order to determine the number of rows where the value of "Total_Bank_Accounts" is more than 3 and the value of "Financial_Rating" is "Good". It also separately calculates the number of rows where the value of "Total_Bank_Accounts" is less than 3 and the value of "Financial_Rating" is "Good". Subsequently, it outputs these tallies accompanied with suitable output.

Code Output:

```
> # Print the counts
> print(paste("Number of rows with more than 3 Total_Bank_Accounts and 'Good' Financial_Rating:", count_more_than_3))
[1] "Number of rows with more than 3 Total_Bank_Accounts and 'Good' Financial_Rating: 10683"
> print(paste("Number of rows with less than 3 Total_Bank_Accounts and 'Good' Financial_Rating:", count_less_than_3))
[1] "Number of rows with less than 3 Total_Bank_Accounts and 'Good' Financial_Rating: 4198"
> |
```

Figure 69 more than 3 bank accounts result

The figure above shows the results after filtering the dataset, and as we can see that the result shows that the number of rows with more than 3 total bank accounts and "Good" Financial Rating Greatly exceeds the number of rows with less than 3 total bank accounts and "Good" Financial Rating with more than 6000 rows, which proves our last objective.

Graph code and illustration:

```
# Create a data frame for the counts
count_data <- data.frame(
  Category = c("More than 3 Total_Bank_Accounts", "Less than 3 Total_Bank_Accounts"),
  Count = c(count_more_than_3, count_less_than_3)
)

# Create a bar graph using ggplot2
ggplot(count_data, aes(x = Category, y = Count, fill = Category)) +
  geom_bar(stat = "identity") +
  labs(title = "Counts of Customers Based on Total Bank Accounts and Financial Rating",
       x = "Category",
       y = "Count") +
  theme_minimal()
```

Figure 70 more than 3 bank accounts graph code

The code in the figure above generates a data frame named count_data, which consists of two columns: Category and Count. The Category column displays labels that indicate whether the count pertains to rows that possess more than 3 total bank accounts or less than 3, both of which have a "Good" financial rating. The Count column displays the corresponding quantities. The code subsequently utilizes the ggplot2 library to generate a bar graph, with the x-axis denoting the categories and the y-axis denoting the counts. The bars are populated according to the category, and the plot is simply designed.

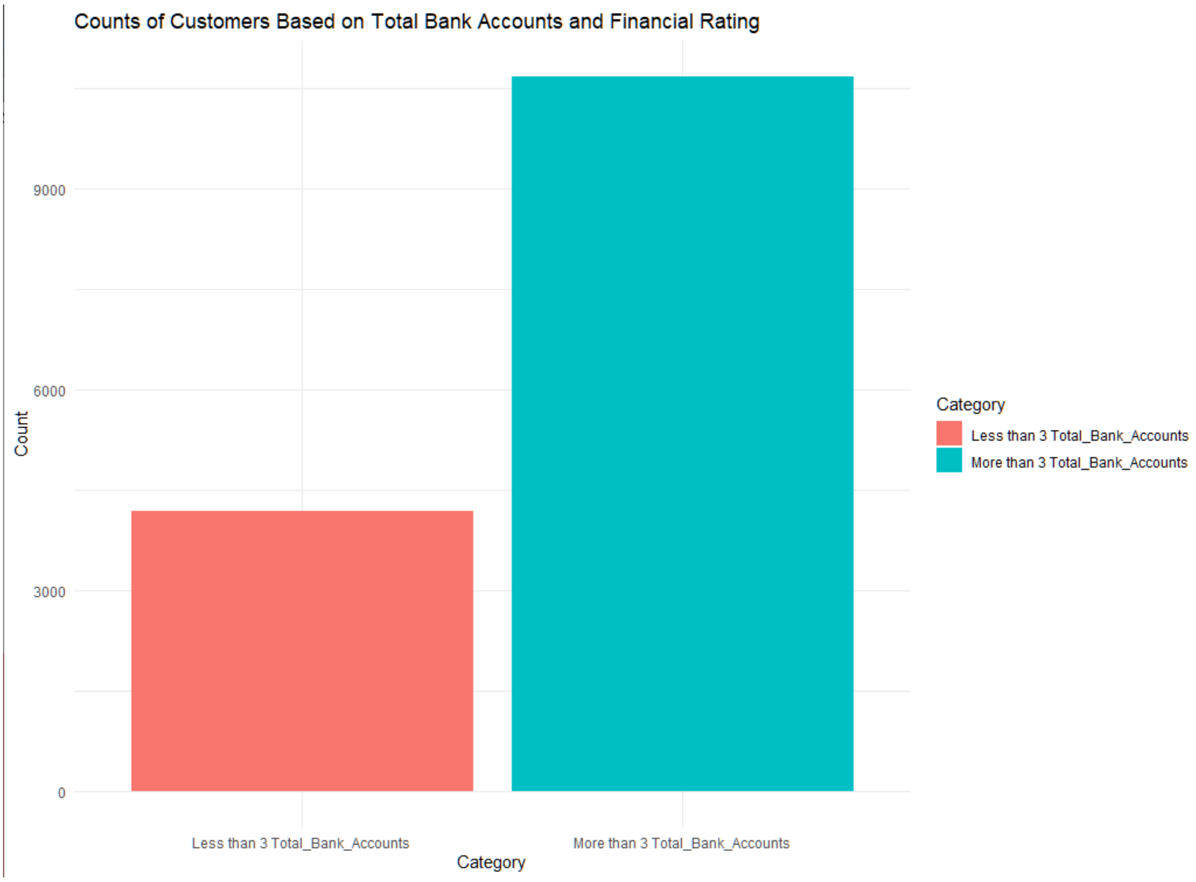


Figure 71 more than 3 bank accounts graph

Next, we have this bar chart that illustrates the difference between rows with more than 3 total bank accounts and less than 3 total bank accounts.

Analysis 02: More than 4 Total Credit Cards on good Financial Rating

Filtering Code:

```
# Filter the dataset based on the given criteria and count the rows
count_more_than_4_credit_cards <- customer %>%
  filter(Total_Credit_Cards > 4, Financial_Rating == "Good") %>%
  nrow()

count_less_than_4_credit_cards <- customer %>%
  filter(Total_Credit_Cards < 4, Financial_Rating == "Good") %>%
  nrow()

# Print the counts
print(paste("Number of rows with more than 4 Total_Credit_Cards and 'Good' Financial_Rating:", count_more_than_4_credit_cards))
print(paste("Number of rows with less than 4 Total_Credit_Cards and 'Good' Financial_Rating:", count_less_than_4_credit_cards))
```

Figure 72 more than 4 credit cards code

In order to count the number of rows where Total Credit Cards is more than 4 and Financial Rating is "Good," this code in the figure above filters a dataset called customer. It also counts the number of rows where Total Credit Cards is less than 4 and Financial Rating is "Good" in a separate count. After that, it outputs these counts along with the relevant text.

Code Output:

```
> # Print the counts
> print(paste("Number of rows with more than 4 Total_Credit_Cards and 'Good' Financial_Rating:", count_more_than_4_credit_cards))
[1] "Number of rows with more than 4 Total_Credit_Cards and 'Good' Financial_Rating: 7917"
> print(paste("Number of rows with less than 4 Total_Credit_Cards and 'Good' Financial_Rating:", count_less_than_4_credit_cards))
[1] "Number of rows with less than 4 Total_Credit_Cards and 'Good' Financial_Rating: 6484"
>
```

Figure 73 more than 4 credit cards result

The outputs in the above figure indicate that there are 7,917 rows with more than 4 total credit cards and a good financial rating, and 6,484 rows with less than 4 total credit cards and a good financial rating, that implies that there is a difference of more than 1400 customers which proves our third objective.

Graph code and illustration:

```
# Create a data frame for the counts
count_data <- data.frame(
  Category = c("More than 4 Total_Credit_Cards", "Less than 4 Total_Credit_Cards"),
  Count = c(count_more_than_4_credit_cards, count_less_than_4_credit_cards)
)

# Add percentage labels
count_data$Percentage <- round(count_data$Count / sum(count_data$Count) * 100, 1)
count_data$Label <- paste(count_data$Category, "\n", count_data$Percentage, "%", sep="")

# Create a pie chart using ggplot2
ggplot(count_data, aes(x = "", y = Count, fill = Category)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  labs(title = "Counts of Customers Based on Total Credit Cards and Good Financial Rating",
       x = "",
       y = "") +
  theme_minimal() +
  theme(axis.text.x = element_blank()) +
  geom_text(aes(label = Label), position = position_stack(vjust = 0.5), size = 5) +
  scale_fill_manual(values = c("More than 4 Total_Credit_Cards" = "green",
                               "Less than 4 Total_Credit_Cards" = "red")) +
  theme(plot.title = element_text(size = 16, face = "bold"),
        legend.title = element_blank(),
        legend.text = element_text(size = 12))
```

Figure 74 more than 4 credit cards graph code

Using the ggplot2 package, the figure above shows a code generates a pie chart that shows the proportion of customers who have more than four credit cards overall and a "Good" financial rating, as well as those who have fewer than four credit cards overall and a "Good" financial rating.

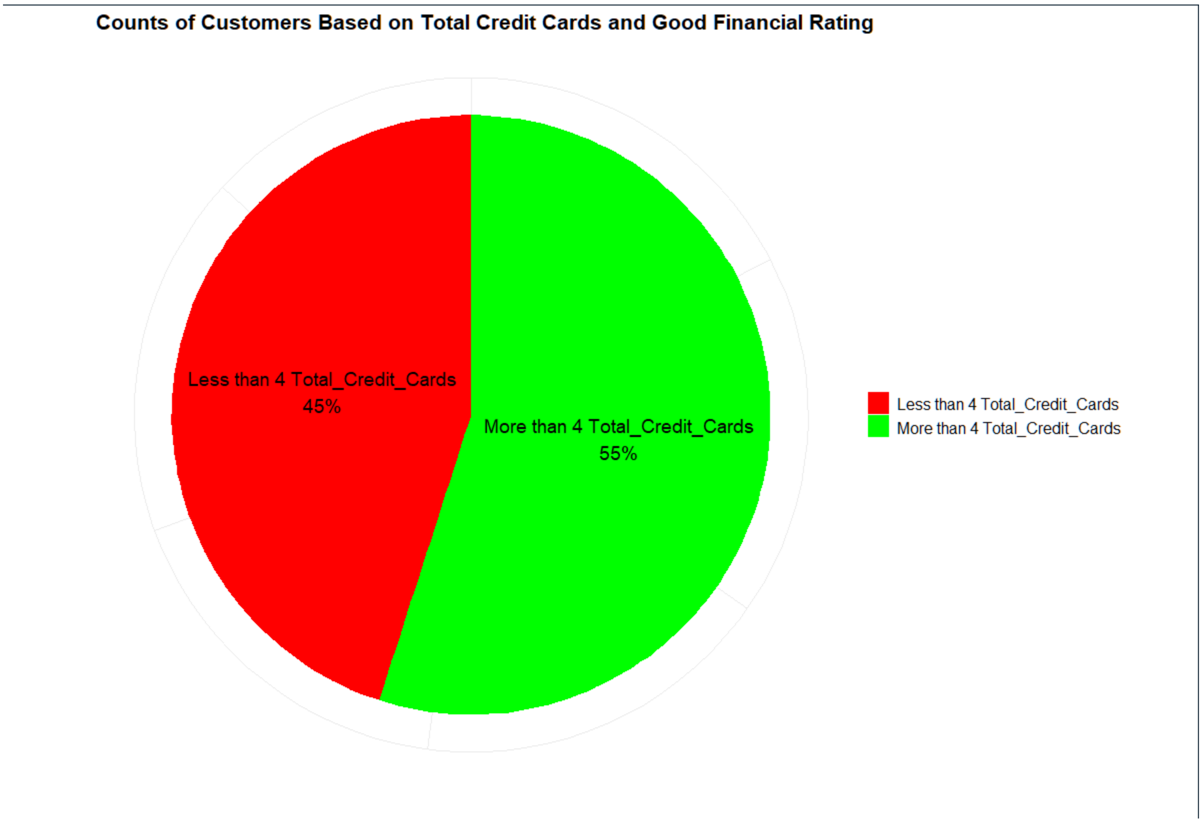


Figure 75 more than 4 credit cards graph

In the figure above we have this pie chart that visualizes the percentage of customers that have more than total 4 credit cards and the percentage customers with less than total of 4 credit cards, and as we can see the difference is approximately 10% between both sides which proves our third objective.

Analysis 03: More than 3000 Monthly Net Income on good Financial Rating

Filtering code:

```r
# Filter the dataset based on the given criteria and count the rows
count_more_than_3000_income <- customer %>%
  filter(Monthly_Net_Income > 3000, Financial_Rating == "Good") %>%
  nrow()

count_less_than_3000_income <- customer %>%
  filter(Monthly_Net_Income < 3000, Financial_Rating == "Good") %>%
  nrow()

# Print the counts
print(paste("Number of rows with more than 3000 Monthly_Net_Income and 'Good' Financial_Rating:", count_more_than_3000_income))
print(paste("Number of rows with less than 3000 Monthly_Net_Income and 'Good' Financial_Rating:", count_less_than_3000_income))
```

Figure 76 more than 3000 net income code

In order to count the number of rows where Monthly Net Income is larger than 3000 and Financial Rating is "Good," the code in the above figure filters the customer dataset. It also counts the number of rows where Monthly Net Income is less than 3000 and Financial Rating is "Good" in a separate count. After that, it outputs these counts along with the relevant text. This makes it easy to compare how many clients with "Good" financial ratings fall into these two income brackets.

Code Output:

```
> # Print the counts
> print(paste("Number of rows with more than 3000 Monthly_Net_Income and 'Good' Financial_Rating:", count_more_than_3000_income))
[1] "Number of rows with more than 3000 Monthly_Net_Income and 'Good' Financial_Rating: 12237"
> print(paste("Number of rows with less than 3000 Monthly_Net_Income and 'Good' Financial_Rating:", count_less_than_3000_income))
[1] "Number of rows with less than 3000 Monthly_Net_Income and 'Good' Financial_Rating: 5591"
>
```

Figure 77 more than 3000 net income result

The output shown in the figure above shows that There are 12,237 rows where the Monthly_Net_Income is more than 3000 and the Financial_Rating is "Good".
There are 5,591 rows where the Monthly_Net_Income is less than 3000 and the Financial_Rating is "Good", which implies that there more than 6600 customers that have an income of more than “3000” and good financial rating

Graph code and illustration:

```r
# Filter the dataset based on the given criteria and count the rows
count_more_than_3000_salary <- customer %>%
  filter(Monthly_Net_Income > 3000, Financial_Rating == "Good") %>%
  nrow()

count_less_than_3000_salary <- customer %>%
  filter(Monthly_Net_Income < 3000, Financial_Rating == "Good") %>%
  nrow()

# Print the counts
print(paste("Number of rows with more than 3000 Monthly_Net_Income and 'Good' Financial_Rating:", count_more_than_3000_salary))
print(paste("Number of rows with less than 3000 Monthly_Net_Income and 'Good' Financial_Rating:", count_less_than_3000_salary))

# Create a data frame for the counts
count_data <- data.frame(
  Category = c("More than 3000 Monthly_Net_Income", "Less than 3000 Monthly_Net_Income"),
  Count = c(count_more_than_3000_salary, count_less_than_3000_salary)
)

# Add percentage labels
count_data$Percentage <- round(count_data$Count / sum(count_data$Count) * 100, 1)
count_data$Label <- paste(count_data$Category, "\n", count_data$Percentage, "%", sep="")

# Create a pie chart using ggplot2
ggplot(count_data, aes(x = "", y = Count, fill = Category)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  labs(title = "Counts of customers based on Monthly Net Income with good Financial Rating",
       x = "",
       y = "") +
  theme_minimal() +
  theme(axis.text.x = element_blank()) +
  geom_text(aes(label = Label), position = position_stack(vjust = 0.5), size = 5) +
  scale_fill_manual(values = c("More than 3000 Monthly_Net_Income" = "#FF9999",
                               "Less than 3000 Monthly_Net_Income" = "#66B2FF")) +
  theme(plot.title = element_text(size = 16, face = "bold"),
        legend.title = element_blank(),
        legend.text = element_text(size = 12))
```

Figure 78 more than 3000 net income graph code

In order to count the number of rows where Monthly Net Income is larger than 3000 and Financial Rating is "Good," the above figure shows a code that filters the customer's dataset. It also counts the number of rows where Monthly Net Income is less than 3000 and Financial Rating is "Good" in a separate count. Along with these counts, it produces explanatory notes. The distribution of the counts is then seen as a pie chart using the ggplot2 package, after the counts and their corresponding percentages are created in a data frame.

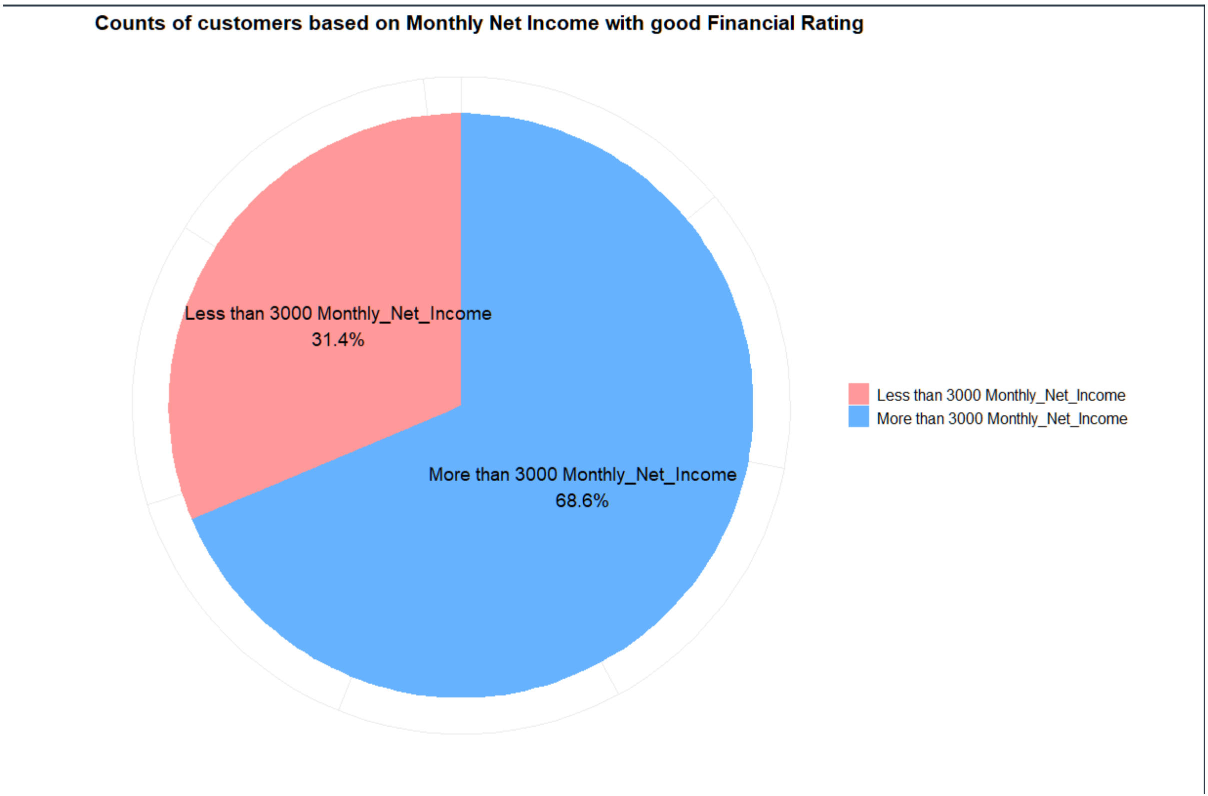


Figure 79 more than 3000 net income graph

In the figure above we have this pie chart that visualizes the percentage of customers that have more than 3000 monthly net income as well as the percentage of customers with less than 3000 monthly net income, and as we can see the percentage difference is approximately 37.2% which is considered as a big difference between both sides, and that proves our second objective.

Analysis 04: Good Credit Profile on good Financial Rating

Filtering code:

```
count_good_credit_profile <- customer %>%
  filter(Credit_Profile == "Good", Financial_Rating == "Good") %>%
  nrow()

count_bad_credit_profile <- customer %>%
  filter(Credit_Profile == "Bad", Financial_Rating == "Good") %>%
  nrow()

# Print the counts
print(paste("Number of rows with 'Good' Credit_Profile and 'Good' Financial_Rating:", count_good_credit_profile))
print(paste("Number of rows with 'Bad' Credit_Profile and 'Good' Financial_Rating:", count_bad_credit_profile))
```

Figure 80 good credit profile code

In order to count the number of rows that have both a "Good" credit profile and a "Good" financial rating, the above figure shows a code that filters the customer's dataset. It also counts the number of rows that have a "Bad" credit profile and a "Good" financial rating separately. After that, it prints these counts together with explanatory text. This gives a quick comparison of the proportion of clients that keep a decent financial rating while having good or negative credit profiles.

Code Output:

```
> print(paste("Number of rows with 'Good' Credit_Profile and 'Good' Financial_Rating:", count_good_credit_profile))
[1] "Number of rows with 'Good' Credit_Profile and 'Good' Financial_Rating: 15497"
> print(paste("Number of rows with 'Bad' Credit_Profile and 'Good' Financial_Rating:", count_bad_credit_profile))
[1] "Number of rows with 'Bad' Credit_Profile and 'Good' Financial_Rating: 288"
>
```

Figure 81 good credit profile result

The figure above shows the number of rows where Credit Profile is "Good" and Financial_Rating is "Good". And the number of rows where Credit Profile is "Bad" and Financial_Rating is "Good" and as we can that There are 15,497 rows with a "Good" credit profile and a "Good" financial rating. While there are 288 rows with a "Bad" credit profile and a "Good" financial rating, thus we have a huge difference of 15000 customers who have good credit profile and good financial rating.

Graph code and illustration:

```
# Filter the dataset based on the given criteria and count the rows
count_good_credit_profile <- customer %>%
  filter(Credit_Profile == "Good", Financial_Rating == "Good") %>%
  nrow()

count_bad_credit_profile <- customer %>%
  filter(Credit_Profile == "Bad", Financial_Rating == "Good") %>%
  nrow()

# Print the counts
print(paste("Number of rows with 'Good' Credit Profile and 'Good' Financial Rating:", count_good_credit_profile))
print(paste("Number of rows with 'Bad' Credit Profile and 'Good' Financial Rating:", count_bad_credit_profile))

# Create a data frame for the counts
count_data <- data.frame(
  Category = c("Good Credit Profile & Good Financial Rating", "Bad Credit Profile & Good Financial Rating"),
  Count = c(count_good_credit_profile, count_bad_credit_profile)
)

# Create a lollipop chart using ggplot2 with a background
ggplot(count_data, aes(x = Category, y = Count)) +
  geom_segment(aes(x = Category, xend = Category, y = 0, yend = Count), color = "black") +
  geom_point(size = 5, aes(color = Category)) +
  scale_color_manual(values = c("Good Credit Profile & Good Financial Rating" = "#66B2FF",
                                "Bad Credit Profile & Good Financial Rating" = "#FF9999")) +
  labs(title = "Counts of Customers with good and bad Credit Profile on good Financial Rating",
       x = "Category",
       y = "Count") +
  theme_minimal() +
  theme(
    plot.background = element_rect(fill = "lightblue", color = "lightblue"),
    panel.background = element_rect(fill = "lightblue", color = "lightblue"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 16, face = "bold"),
    legend.title = element_blank(),
    legend.text = element_text(size = 12)
  )
```

Figure 82 good credit profile graph code

In order to count the number of rows that have both a "Good" credit profile and a "Good" financial rating, the code in the figure above filters the customer's dataset. It also counts the number of rows that have a "Bad" credit profile and a "Good" financial rating separately. After that, it prints these counts together with explanatory text. The counts are put into a data frame for display. Using customized labels and design, the counts are visually represented in a lollipop chart created with the ggplot2 package. The number of customers with good and bad credit profiles who also have a good financial rating is clearly shown visually in the chart.

Additional feature: lollipop Graph

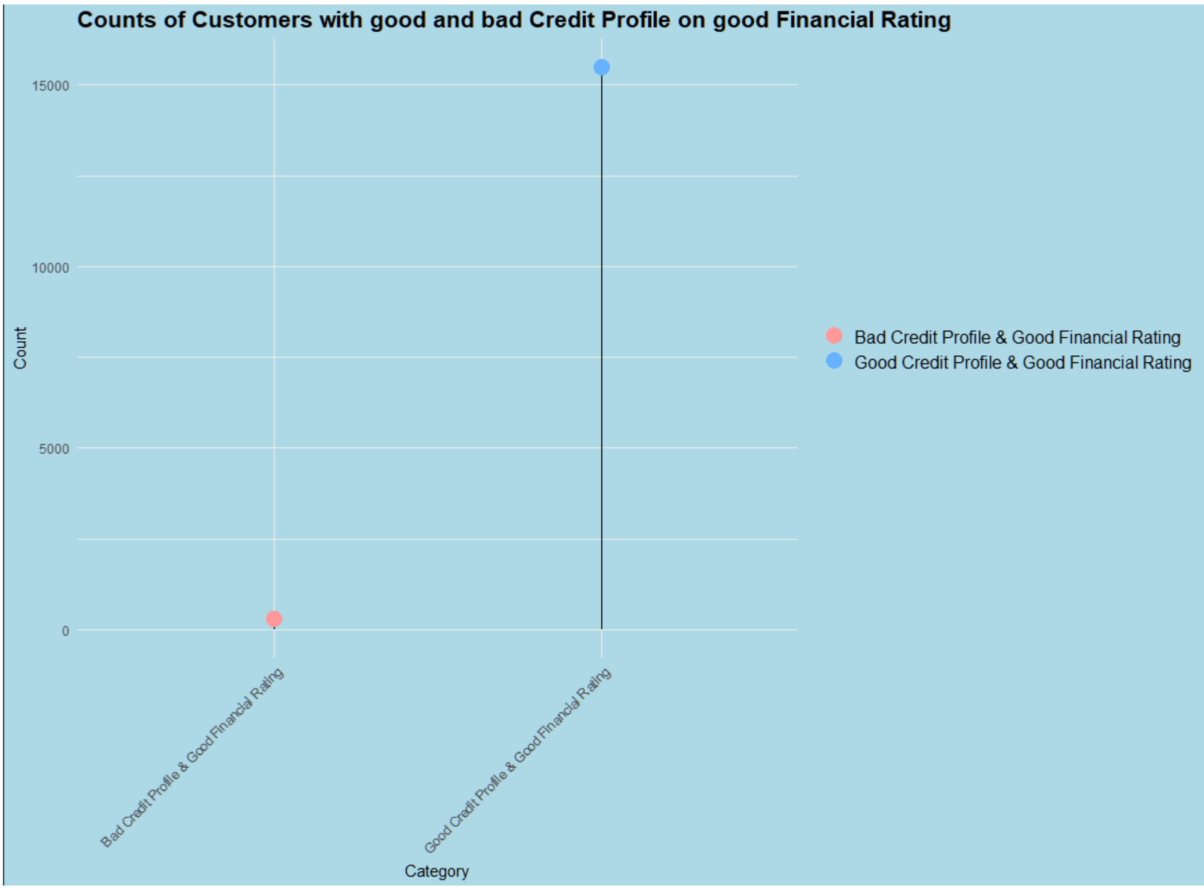


Figure 83 good credit profile graph

In the figure above we have a lollipop chart that shows the huge difference between customers who have good credit profile with good financial rating and bad credit profile with good financial rating, which proves our first objective.

Analysis 05: customers with proved objectives combined.

Filtering code:

```
mthreebnk_mfourcrds_m3000income_gcreditp <- customer %>%
  filter(
    Total_Bank_Accounts > 3 &
      Total_Credit_Cards > 4 &
      Monthly_Net_Income > 3000 &
      Credit_Profile == "Good" &
      Financial_Rating == "Good"
  )

# Print the number of rows that meet the criteria
mthreebnk_mfourcrds_m3000income_gcreditp_rows <- nrow(mthreebnk_mfourcrds_m3000income_gcreditp)
print(paste("Number of rows with more than 3 Total_Bank_Accounts, more than 4 Total_Credit_Cards, more than 3000 Monthly_Net_Income, Good Credit_Profile, and Good Financial Rating:", mthreebnk_mfourcrds_m3000income_gcreditp_ro
```

Figure 84 deeper analysis code

The code in the above figure filters the customer dataset to find rows where customers have more than 3 total bank accounts, more than 4 total credit cards, a monthly net income greater than 3000, a good credit profile, and a good financial rating. It then counts the number of rows that meet these criteria and prints the count with a descriptive message.

```
lthreebnk_lfourcrds_l3000income_bcreditp <- customer %>%
  filter(
    Total_Bank_Accounts < 3 &
      Total_Credit_Cards < 4 &
      Monthly_Net_Income < 3000 &
      Credit_Profile == "Bad" &
      Financial_Rating == "Good"
  )

# Print the number of rows that meet the criteria
lthreebnk_lfourcrds_l3000income_bcreditp_rows <- nrow(lthreebnk_lfourcrds_l3000income_bcreditp)
print(paste("Number of rows with less than 3 Total_Bank_Accounts, less than 4 Total_Credit_Cards, less than 3000 Monthly_Net_Income, Bad Credit_Profile, and Good Financial Rating:", lthreebnk_lfourcrds_l3000income_bcreditp_row
```

Figure 85 deeper analysis code 2

The code in the above figure filters the customer dataset to find rows where customers have less than 3 total bank accounts, less than 4 total credit cards, a monthly net income lower than 3000, a bad credit profile, and a good financial rating. It then counts the number of rows that meet these criteria and prints the count with a relevant message.

<u>Code Output:</u>

```
> print(paste("Number of rows with more than 3 Total_Bank_Accounts, more than 4 Total_Credit_Cards, more than 3000 Monthly_Net_Income, Good Credit_Profile, and Good Financial Rating:",
[1] "Number of rows with more than 3 Total_Bank_Accounts, more than 4 Total_Credit_Cards, more than 3000 Monthly_Net_Income, Good Credit_Profile, and Good Financial Rating: 2459"
```

Figure 86 deeper analysis result

The given output shows that the number of customers with more than 3 total bank accounts, more than 4 total credit cards, greater than 3000 monthly net income, good credit profile and good financial rating is 2459.

```
> print(paste("Number of rows with less than 3 Total_Bank_Accounts, less than 4 Total_Credit_Cards, less than 3000 Monthly_Net_Income, Bad Credit_Profile, and Good Financial Rating:"
[1] "Number of rows with less than 3 Total_Bank_Accounts, less than 4 Total_Credit_Cards, less than 3000 Monthly_Net_Income, Bad Credit_Profile, and Good Financial Rating: 0"
>
```

Figure 87 deeper analysis result 2

While in this figure we have the output of customers with less than 3 total bank accounts, less than 4 total credit cards, less than 3000 monthly net income, bad credit profile and good financial rating is 0, and that implies that our hypothesis is correct and has been successfully proven.

Conclusion:

In The End, the analysis supports the objectives by demonstrating notable variations in the number of customers according to different credit and financial parameters. The results derived from the data are reinforced by the visuals, which successfully depict these variations.

OBJECTIVE 04: TEMOOR HASSAN

Our Hypothesis:

Customer with good credit mix,monthly inhand-salary more than 3000, more than 4 credit card ,and more than 3 bank account will have good credit score.

Objective 1: Monthly Net Income and Credit Score

The impact of monthly net income > 3k on good financial rating.

```r
# Objective 1: Monthly Net Income and Credit Score
ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Monthly Net Income`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Monthly Net Income by Credit Score", x = "Credit Score", y = "Monthly Net Income") +
  theme_minimal()

# Filter the dataset for customers with monthly in-hand salary more than 3000 and good credit score
filtered_dataset_1 <- cleaned_dataset %>%
  filter(`Monthly Net Income` > 3000 & `Financial Rating` == "Good")


cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")

# View the filtered dataset
View(filtered_dataset_1)
```

2:1 your analysis

Console Background Jobs

R 4.3.3 · C:/Users/Taimoor Hassan/Downloads/

```r
# Objective 1: Monthly Net Income and Credit Score
ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Monthly Net Income`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Monthly Net Income by Credit Score", x = "Credit Score", y = "Monthly Net Income") +
  theme_minimal()
# Filter the dataset for customers with monthly in-hand salary more than 3000 and good credit score
filtered_dataset_1 <- cleaned_dataset %>%
  filter(`Monthly Net Income` < 3000 & `Financial Rating` == "Good")
# View the filtered dataset
View(filtered_dataset_1)
# Filter the dataset for customers with monthly in-hand salary more than 3000 and good credit score
filtered_dataset_1 <- cleaned_dataset %>%
  filter(`Monthly Net Income` > 3000 & `Financial Rating` == "Good")
cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")
Number of rows in filtered dataset: 12237
```

Figure 88: Objective_1 with (>3000) condition

In the above picture I filtered the data from my cleaned_dataset with the conditions according to my 1st objective. (monthly net income >3000 & financial rating = "Good")

You can see in the picture we got 12237 results from our dataset.

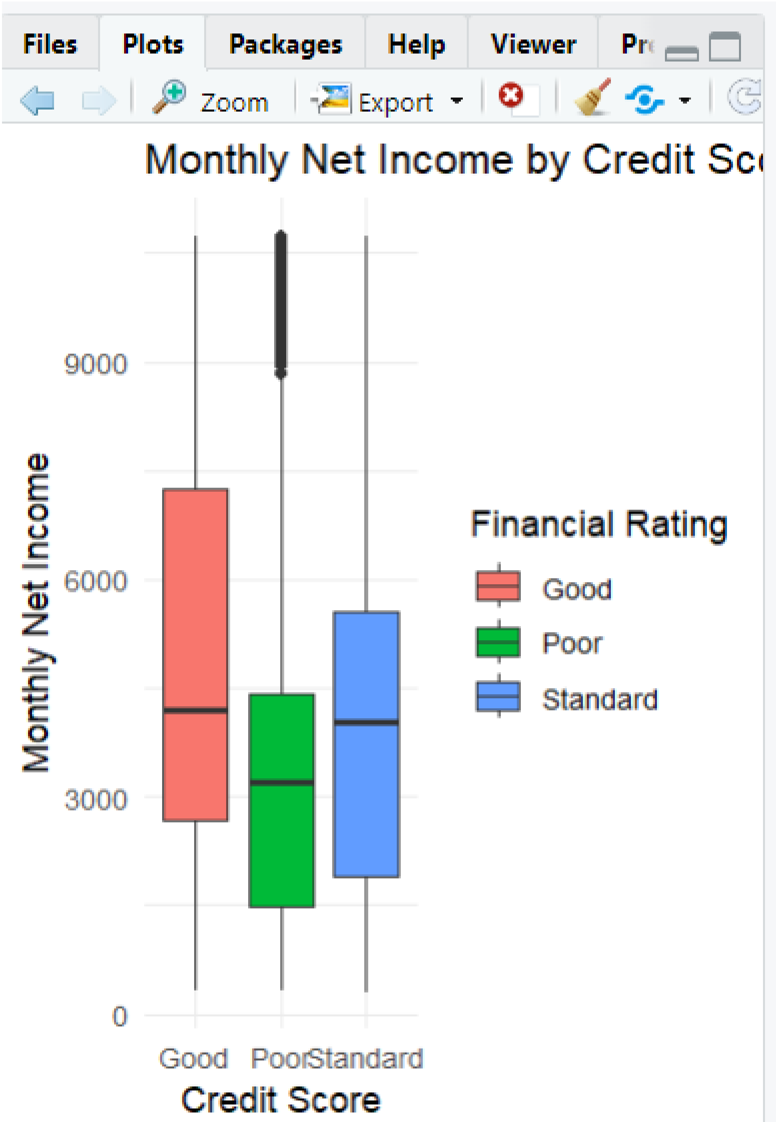


Figure 89: Objective_1 Plot

A clear link exists between income levels and financial ratings is shown by the boxplot. Customers with better financial ratings have higher monthly net incomes, while customers with worse financial ratings have lower incomes.

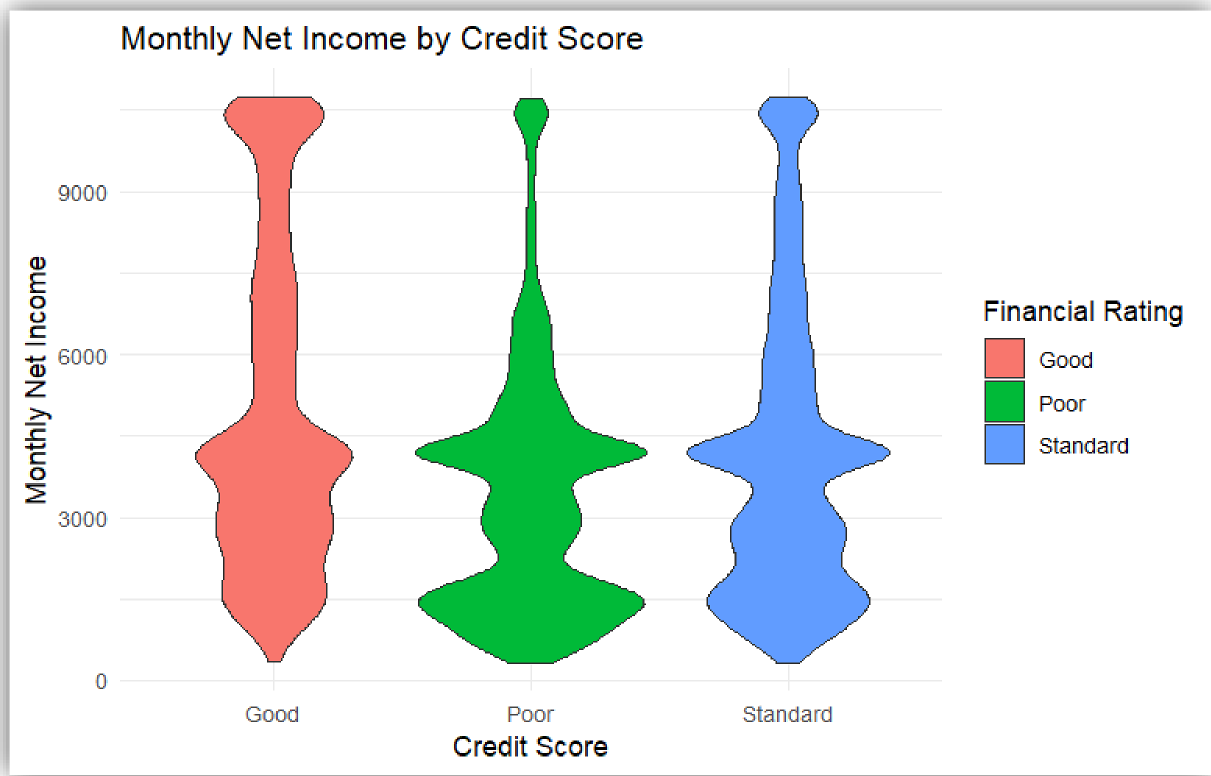


Figure 90: Vilan Graph

Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
4194.171	3	4	Good	Good
4194.171	3	4	Good	Good
4194.171	3	4	Good	Good
4194.171	3	4	Good	Good
3037.987	2	4	Good	Good
4194.171	2	4	Good	Good
3037.987	2	4	Good	Good
3037.987	2	4	Good	Good
4194.171	2	4	Good	Good
3037.987	2	4	Good	Good
10474.431	1	5	Good	Good
10474.431	1	5	Good	Good
4194.171	1	5	Good	Good
10474.431	1	5	Good	Good
10474.431	1	5	Good	Good

Figure 91: Objective_1 Table_View

From this filtered and presented dataset, it can be deduced that there are 5,591 clients who meet the condition of making at least 3000 per month in net income and have a good financial creditworthiness. This analysis is useful in appreciating financial conduct and features of the good financial-rating customers in the high-income bracket.

```r
# Objective 1: Monthly Net Income and Credit Score
ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Monthly Net Income`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Monthly Net Income by Credit Score", x = "Credit Score", y = "Monthly Net Income") +
  theme_minimal()

# Filter the dataset for customers with monthly in-hand salary more than 3000 and good credit score
filtered_dataset_1 <- cleaned_dataset %>%
  filter(`Monthly Net Income` < 3000 & `Financial Rating` == "Good")


cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")

# View the filtered dataset
View(filtered_dataset_1)
```

```
152:1   your analysis
Console   Background Jobs
R 4.3.3 · C:/Users/Taimoor Hassan/Downloads/
> # Filter the dataset for customers with monthly in-hand salary more than 3000 and good credit score
> filtered_dataset_1 <- cleaned_dataset %>%
+   filter(`Monthly Net Income` > 3000 & `Financial Rating` == "Good")
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")
umber of rows in filtered dataset: 12237
> # Objective 1: Monthly Net Income and Credit Score
> ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Monthly Net Income`, fill = `Financial Rating`)) +
+   geom_boxplot() +
+   labs(title = "Monthly Net Income by Credit Score", x = "Credit Score", y = "Monthly Net Income") +
+   theme_minimal()
> # Filter the dataset for customers with monthly in-hand salary more than 3000 and good credit score
> filtered_dataset_1 <- cleaned_dataset %>%
+   filter(`Monthly Net Income` < 3000 & `Financial Rating` == "Good")
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_1), "\n")
umber of rows in filtered dataset: 5591
```

Figure 92: Objective_1 with (<3000) condition

In the figure 4, I changed the condition for the monthly net income from >3000 to <3000 and filtered the data again and I saw the there are only 5591 rows with this condition.

So the number of the customers who have less than 3000 monthly income with good credit score is lesser than the customers who have more than 3000 monthly net income with good credit score.
Hence, our 1st objective is proved according to our hypothesis.

	Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1	1824.8433	3	4	Good	Good
2	1824.8433	3	4	Good	Good
3	1824.8433	3	4	Good	Good
4	2825.2333	5	5	Good	Good
5	2825.2333	1	6	Good	Good
6	2825.2333	1	6	Good	Good
7	2825.2333	1	6	Good	Good
8	2825.2333	1	6	Good	Good
9	2825.2333	1	6	Good	Good
10	2825.2333	1	6	Good	Good
11	1684.1683	1	5	Good	Good
12	1684.1683	1	5	Good	Good
13	1684.1683	1	5	Good	Good
14	1684.1683	1	5	Good	Good
15	[illegible]	[illegible]	[illegible]	Good	Good

Figure 93: Objective_1 Table_view (<3000)

Objective 2: Credit Profile and Credit Score

This set of three codes is meant to look into the connection between credit profiles and financial scores. It does this by sorting datasets into different credit profiles and counting the number of customers with good credit who fit each profile.

```r
# Objective 2: Credit Profile and Credit Score
ggplot(cleaned_dataset, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
  geom_bar(position = "dodge") +
  labs(title = "Credit Profile by Credit Score", x = "Credit Profile", y = "Count") +
  theme_minimal()

# Filter the dataset for customers with good credit profile and good financial rating
filtered_dataset_2 <- cleaned_dataset %>%
  filter(`Credit Profile` == "Good" & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")

# View the filtered dataset
View(filtered_dataset_2)
```

171:1 your analysis

Console | Background Jobs

R 4.3.3 · C:/Users/Taimoor Hassan/Downloads/

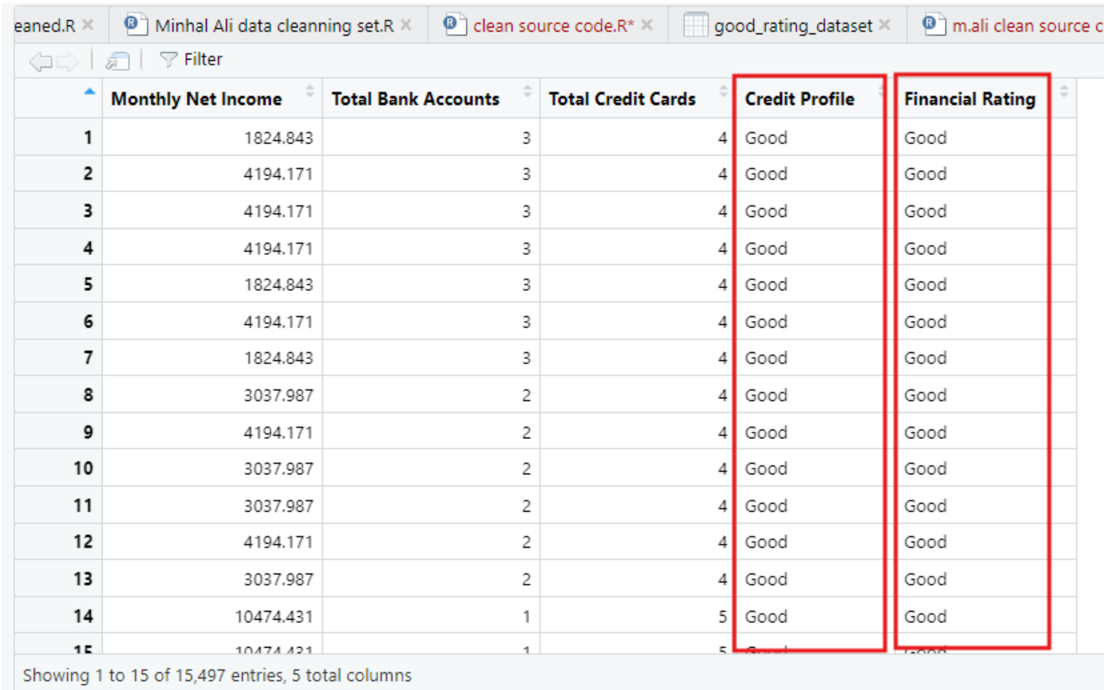
```
> ggplot(cleaned_dataset, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
+   geom_bar(position = "dodge") +
+   labs(title = "Credit Profile by Credit Score", x = "Credit Profile", y = "Count") +
+   theme_minimal()
> # Objective 2: Credit Profile and Credit Score
> ggplot(cleaned_dataset, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
+   geom_bar(position = "dodge") +
+   labs(title = "Credit Profile by Credit Score", x = "Credit Profile", y = "Count") +
+   theme_minimal()
> # Filter the dataset for customers with good credit profile and good financial rating
> filtered_dataset_2 <- cleaned_dataset %>%
+   filter(`Credit Profile` == "Good" & `Financial Rating` == "Good")
> # Display the number of rows in the filtered dataset
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")
Number of rows in filtered dataset: 15497
>
```

Figure 94: Objective_2 with (Good) condition

With this code, the dataset will only include customers who have good credit and a good financial rating. We can see there are 15497 filtered customers who have “Good” Credit Profile with “Good” Financial Rating.

Figure 95: Objective_2 Plot View

This graph displays the number of customers with various credit profiles and financial ratings. It shows that customers with good credit are more likely to have good financial ratings, customers with bad credit are more likely to have poor financial ratings, and customers with standard credit are more likely to have standard financial ratings.



eaned.R × | Minhal Ali data cleanning set.R × | clean source code.R* × | good_rating_dataset × | m.ali clean source c

Filter

	Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1	1824.843	3	4	Good	Good
2	4194.171	3	4	Good	Good
3	4194.171	3	4	Good	Good
4	4194.171	3	4	Good	Good
5	1824.843	3	4	Good	Good
6	4194.171	3	4	Good	Good
7	1824.843	3	4	Good	Good
8	3037.987	2	4	Good	Good
9	4194.171	2	4	Good	Good
10	3037.987	2	4	Good	Good
11	3037.987	2	4	Good	Good
12	4194.171	2	4	Good	Good
13	3037.987	2	4	Good	Good
14	10474.431	1	5	Good	Good
15	10474.431	1	5	Good	Good

Showing 1 to 15 of 15,497 entries, 5 total columns

Figure 96: Objective_2 Table_view (Good)

This list contains information about 15,497 customers who have both a good credit score and a good financial score. It includes things like their monthly net income, number of bank accounts, and number of credit cards.

```r
# Objective 2: Credit Profile and Credit Score
ggplot(cleaned_dataset, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
  geom_bar(position = "dodge") +
  labs(title = "Credit Profile by Credit Score", x = "Credit Profile", y = "Count") +
  theme_minimal()

# Filter the dataset for customers with good credit profile and good financial rating
filtered_dataset_2 <- cleaned_dataset %>%
  filter(`Credit Profile` == "Bad" & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")

# View the filtered dataset
View(filtered_dataset_2)

171:1   your analysis
```

```
Console   Background Jobs
R 4.3.3 · C:/Users/Taimoor Hassan/Downloads/
> ggplot(cleaned_dataset, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
+   geom_bar(position = "dodge") +
+   labs(title = "Credit Profile by Credit Score", x = "Credit Profile", y = "Count") +
+   theme_minimal()
> # Objective 2: Credit Profile and Credit Score
> ggplot(cleaned_dataset, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
+   geom_bar(position = "dodge") +
+   labs(title = "Credit Profile by Credit Score", x = "Credit Profile", y = "Count") +
+   theme_minimal()
> # Filter the dataset for customers with good credit profile and good financial rating
> filtered_dataset_2 <- cleaned_dataset %>%
+   filter(`Credit Profile` == "Bad" & `Financial Rating` == "Good")
> # Display the number of rows in the filtered dataset
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")
Number of rows in filtered dataset: 288
>
```

Figure 97: Objective_2 with (Bad) condition

By changing the condition of Credit profile fom "Good" to "Bad" we can see the result of filtered dataset, which is 288 and fewer than the customers with "Good" Credit Profile.

	Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1	1437.7792	6	5	Bad	Good
2	777.0871	6	7	Bad	Good
3	777.0871	6	7	Bad	Good
4	4194.1708	6	7	Bad	Good
5	777.0871	6	7	Bad	Good
6	777.0871	6	7	Bad	Good
7	5229.1531	9	6	Bad	Good
8	5229.1531	9	6	Bad	Good
9	836.1838	8	8	Bad	Good
10	4194.1708	10	8	Bad	Good
11	6184.9200	6	6	Bad	Good
12	6184.9200	6	6	Bad	Good
13	6184.9200	6	6	Bad	Good
14	1655.0800	10	8	Bad	Good
15	1655.0800	10	8	Bad	Good

owing 1 to 15 of 288 entries, 5 total columns

Figure 98: Objective_2 Table_view (Bad)

This list contains information about 288 customers who have a bad credit score and a good financial score. It includes things like their monthly net income, number of bank accounts, and number of credit cards.

```
# Objective 2: Credit Profile and Credit Score
ggplot(cleaned_dataset, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
  geom_bar(position = "dodge") +
  labs(title = "Credit Profile by Credit Score", x = "Credit Profile", y = "Count") +
  theme_minimal()

# Filter the dataset for customers with good credit profile and good financial rating
filtered_dataset_2 <- cleaned_dataset %>%
  filter(`Credit Profile` == "Standard" & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")

# View the filtered dataset
View(filtered_dataset_2)

171:1   your analysis
```

```
Console   Background Jobs
R 4.3.3 · C:/Users/Taimoor Hassan/Downloads/
> ggplot(cleaned_dataset, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
+   geom_bar(position = "dodge") +
+   labs(title = "Credit Profile by Credit Score", x = "Credit Profile", y = "Count") +
+   theme_minimal()
> # Objective 2: Credit Profile and Credit Score
> ggplot(cleaned_dataset, aes(x = `Credit Profile`, fill = `Financial Rating`)) +
+   geom_bar(position = "dodge") +
+   labs(title = "Credit Profile by Credit Score", x = "Credit Profile", y = "Count") +
+   theme_minimal()
> # Filter the dataset for customers with good credit profile and good financial rating
> filtered_dataset_2 <- cleaned_dataset %>%
+   filter(`Credit Profile` == "Standard" & `Financial Rating` == "Good")
> # Display the number of rows in the filtered dataset
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")
Number of rows in filtered dataset: 2043
> 
```

Figure 99: Objective_2 with (standard) condition

By changing the condition of Credit profile fom “Bad” to “Standard” we can see the result of filtered dataset, which is 2043 and fewer than the customers with “Good” Credit Profile.

Filter

	Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1	4194.1708	6	6	Standard	Good
2	7449.4693	6	6	Standard	Good
3	7449.4693	6	6	Standard	Good
4	5799.4500	3	6	Standard	Good
5	5799.4500	3	6	Standard	Good
6	5799.4500	3	6	Standard	Good
7	5799.4500	3	6	Standard	Good
8	8279.3750	8	4	Standard	Good
9	8279.3750	8	4	Standard	Good
10	8279.3750	8	4	Standard	Good
11	8279.3750	8	4	Standard	Good
12	3132.5242	8	7	Standard	Good
13	3132.5242	8	7	Standard	Good
14	3132.5242	8	7	Standard	Good
15	4194.1708	8	5	Standard	Good

Showing 1 to 15 of 2,043 entries, 5 total columns

Figure 100: Objective_2 Table_view (Standard)

This list contains information about 2043 customers who have a standard credit score and a good financial score. It includes things like their monthly net income, number of bank accounts, and number of credit cards.

Objective 3: Total Credit Cards & Credit Score

```r
# Objective 3: Total Credit Cards and Credit Score
ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Total Credit Cards`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Total Credit Cards by Credit Score", x = "Credit Score", y = "Total Credit Cards") +
  theme_minimal()

# Filter the dataset for customers with more than 4 credit cards and good financial rating
filtered_dataset_3 <- cleaned_dataset %>%
  filter(`Total Credit Cards` > 4 & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset:", nrow(filtered_dataset_3), "\n")

# View the filtered dataset
View(filtered_dataset_3)
```

197:1 your analysis

Console Background Jobs

R 4.3.3 · C:/Users/Taimoor Hassan/Downloads/

```
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_2), "\n")
Number of rows in filtered dataset: 2043
> # View the filtered dataset
> View(filtered_dataset_2)
> # Objective 3: Total Credit Cards and Credit Score
> ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Total Credit Cards`, fill = `Financial Rating`)) +
+   geom_boxplot() +
+   labs(title = "Total Credit Cards by Credit Score", x = "Credit Score", y = "Total Credit Cards") +
+   theme_minimal()
> # Filter the dataset for customers with more than 4 credit cards and good financial rating
> filtered_dataset_3 <- cleaned_dataset %>%
+   filter(`Total Credit Cards` > 4 & `Financial Rating` == "Good")
> # Display the number of rows in the filtered dataset
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_3), "\n")
Number of rows in filtered dataset: 7917
```

Figure 102: Objective_3 with (>4) condition

With this code, the dataset will only include customers who have more than 4 credit cards and a good financial rating. We can see there are 7917 filtered customers who have more than 4 credit cards with "Good" Financial Rating.

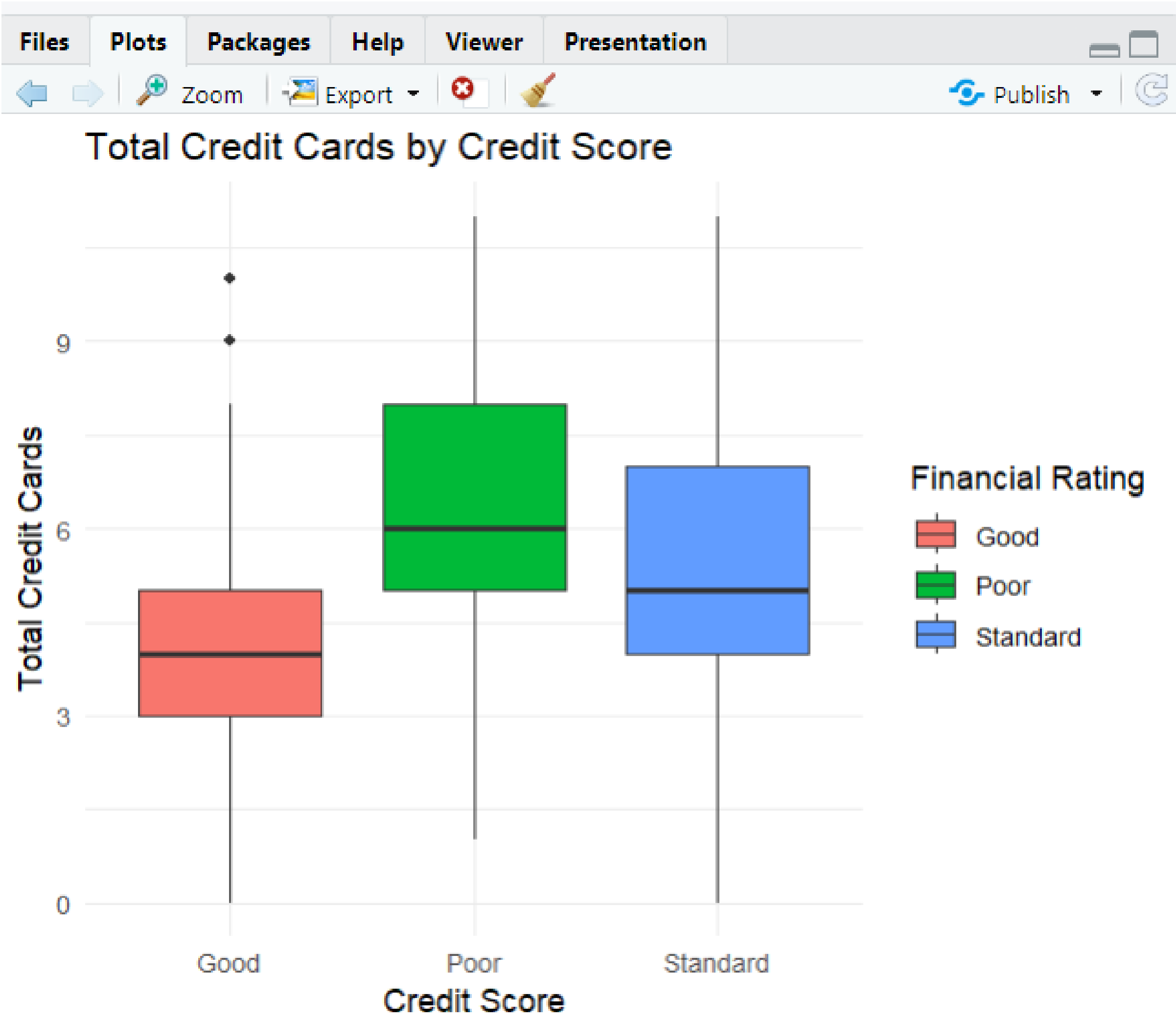


Figure 103: Objective_3 Plot View

This graph displays the number of customers with various number of credit caards and financial ratings. It shows that customers with less than 4 credit cards are more likely to have poor financial ratings, customers with more than 4 credit cards are more likely to have good financial ratings.

	Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1	10474.4310	1	5	Good	Good
2	10474.4310	1	5	Good	Good
3	4194.1708	1	5	Good	Good
4	10474.4310	1	5	Good	Good
5	10474.4310	1	5	Good	Good
6	10474.4310	1	5	Good	Good
7	4194.1708	4	5	Good	Good
8	5988.7050	4	5	Good	Good
9	5988.7050	4	5	Good	Good
10	5988.7050	4	5	Good	Good
11	5988.7050	4	5	Good	Good
12	5988.7050	4	5	Good	Good
13	5988.7050	4	5	Good	Good
14	9843.8675	2	5	Good	Good
15	9843.8675	2	5	Good	Good

Showing 1 to 15 of 7,917 entries, 5 total columns

Figure 104: Objective_3 Table_view (>4)

This list contains information about 7917 customers who have more than 4 credit cards and a good financial score. It includes things like their monthly net income, number of bank accounts, and number of credit cards.

```
# Objective 3: Total Credit Cards and Credit Score
ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Total Credit Cards`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Total Credit Cards by Credit Score", x = "Credit Score", y = "Total Credit Cards") +
  theme_minimal()

# Filter the dataset for customers with more than 4 credit cards and good financial rating
filtered_dataset_3 <- cleaned_dataset %>%
  filter(`Total Credit Cards` < 4 & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset:", nrow(filtered_dataset_3), "\n")

# View the filtered dataset
View(filtered_dataset_3)
```

```
197:1 your analysis
Console  Background Jobs
R 4.3.3 · C:/Users/Taimoor Hassan/Downloads/
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_3), "\n")
Number of rows in filtered dataset: 7917
> # View the filtered dataset
> View(filtered_dataset_3)
> # Objective 3: Total Credit Cards and Credit Score
> ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Total Credit Cards`, fill = `Financial Rating`)) +
+   geom_boxplot() +
+   labs(title = "Total Credit Cards by Credit Score", x = "Credit Score", y = "Total Credit Cards") +
+   theme_minimal()
> # Filter the dataset for customers with more than 4 credit cards and good financial rating
> filtered_dataset_3 <- cleaned_dataset %>%
+   filter(`Total Credit Cards` < 4 & `Financial Rating` == "Good")
> # Display the number of rows in the filtered dataset
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_3), "\n")
Number of rows in filtered dataset: 6484
>
```

Figure 105: Objective_3 with (<4) condition

By changing the condition of number of credit cards fom “>4” to “<4” we can see the result of filtered dataset, which is 6484 and fewer than the customers with having “>4” Credit Cards.

	Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1	10474.431	5	1	Good	Good
2	10474.431	5	1	Good	Good
3	10469.208	5	1	Good	Good
4	10469.208	5	1	Good	Good
5	10469.208	5	1	Good	Good
6	4194.171	5	1	Good	Good
7	10469.208	5	1	Good	Good
8	7185.502	4	2	Good	Good
9	7185.502	4	2	Good	Good
10	7185.502	4	2	Good	Good
11	7185.502	4	2	Good	Good
12	7185.502	5	2	Good	Good
13	7185.502	4	2	Good	Good
14	7185.502	4	2	Good	Good
15	7185.502	4	2	Good	Good

Showing 1 to 15 of 6,484 entries, 5 total columns

Figure 106: Objective_3 Table_view (<4)

This list contains information about 6484 customers who have less than 4 credit cards and a good financial score. It includes things like their monthly net income, number of bank accounts, and number of credit cards.

Objective 4: Total Bank Accounts & Credit Score

```
# Objective 4: Total Bank Accounts and Credit Score
ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Total Bank Accounts`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Total Bank Accounts by Credit Score", x = "Credit Score", y = "Total Bank Accounts") +
  theme_minimal()

# Filter the dataset for customers with more than 3 bank accounts and good financial rating
filtered_dataset_4 <- cleaned_dataset %>%
  filter(`Total Bank Accounts` > 3 & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset:", nrow(filtered_dataset_4), "\n")

# View the filtered dataset
View(filtered_dataset_4)
```

```
224:1   your analysis

Console   Background Jobs
R 4.3.3 · C:/Users/Taimoor Hassan/Downloads/
+   theme_minimal()
> # Filter the dataset for customers with more than 3 bank accounts and good financial rating
> filtered_dataset_4 <- cleaned_dataset %>%
+   filter(`Total Bank Accounts` > 3 & `Financial Rating` == "Good")
> # Objective 4: Total Bank Accounts and Credit Score
> ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Total Bank Accounts`, fill = `Financial Rating`)) +
+   geom_boxplot() +
+   labs(title = "Total Bank Accounts by Credit Score", x = "Credit Score", y = "Total Bank Accounts") +
+   theme_minimal()
> # Filter the dataset for customers with more than 3 bank accounts and good financial rating
> filtered_dataset_4 <- cleaned_dataset %>%
+   filter(`Total Bank Accounts` > 3 & `Financial Rating` == "Good")
> # Display the number of rows in the filtered dataset
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_4), "\n")
Number of rows in filtered dataset: 10683
```

Figure 107: Objective_4 with (>3) condition

With this code, the dataset will only include customers who have bank accounts more than 3 and a good financial rating. We can see there are 10683 filtered customers who have more than 3 bank accounts with “Good” Financial Rating.

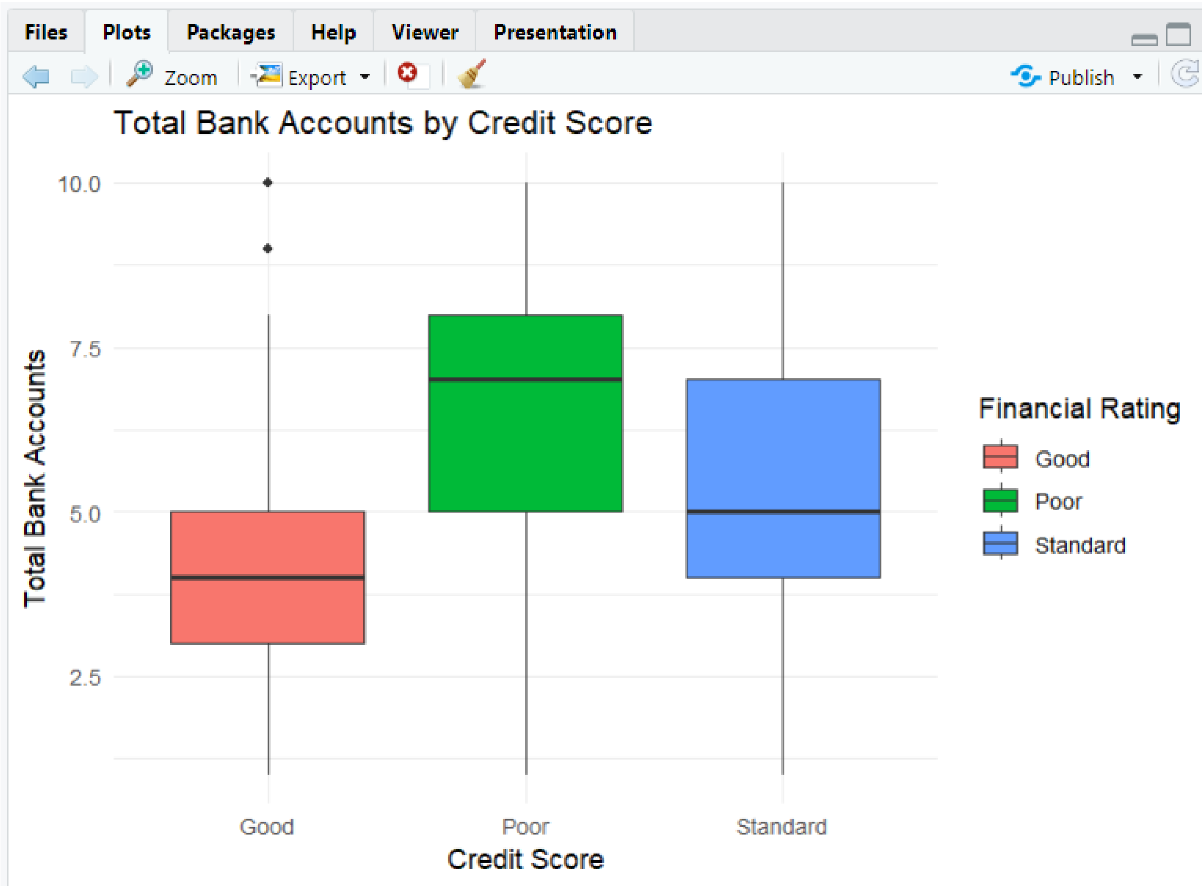


Figure 108: Objective_4 Plot View

Thiss graph displays the number of customers with various number of bank accounts and financial ratings. It shows that customers with more than 3 bank accounts are more likely to have good financial ratings, customers with less than 3 bank accounts are more likely to have good financial ratings.

	Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1	4194.1708	4	5	Good	Good
2	5988.7050	4	5	Good	Good
3	5988.7050	4	5	Good	Good
4	5988.7050	4	5	Good	Good
5	5988.7050	4	5	Good	Good
6	5988.7050	4	5	Good	Good
7	5988.7050	4	5	Good	Good
8	10474.4310	5	1	Good	Good
9	10474.4310	5	1	Good	Good
10	10469.2078	5	1	Good	Good
11	10469.2078	5	1	Good	Good
12	10469.2078	5	1	Good	Good
13	4194.1708	5	1	Good	Good
14	10469.2078	5	1	Good	Good
15	[illegible]	5	5	Good	Good

Showing 1 to 15 of 10,683 entries, 5 total columns

Figure 109: Objective_4 Table_view (>3)

This list contains information about 10683 customers who have more than 3 bank accounts and a good financial score. It includes things like their monthly net income, number of bank accounts, and number of credit cards.

```
# Objective 4: Total Bank Accounts and Credit Score
ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Total Bank Accounts`, fill = `Financial Rating`)) +
  geom_boxplot() +
  labs(title = "Total Bank Accounts by Credit Score", x = "Credit Score", y = "Total Bank Accounts") +
  theme_minimal()

# Filter the dataset for customers with more than 3 bank accounts and good financial rating
filtered_dataset_4 <- cleaned_dataset %>%
  filter(`Total Bank Accounts` < 3 & `Financial Rating` == "Good")

# Display the number of rows in the filtered dataset
cat("Number of rows in filtered dataset:", nrow(filtered_dataset_4), "\n")

# View the filtered dataset
View(filtered_dataset_4)

224:1   your analysis

Console   Background Jobs
R 4.3.3 · C:/Users/Taimoor Hassan/Downloads/
> ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Total Bank Accounts`, fill = `Financial Rating`)) +
+   geom_boxplot() +
+   labs(title = "Total Bank Accounts by Credit Score", x = "Credit Score", y = "Total Bank Accounts") +
+   theme_minimal()
> # Objective 4: Total Bank Accounts and Credit Score
> ggplot(cleaned_dataset, aes(x = `Financial Rating`, y = `Total Bank Accounts`, fill = `Financial Rating`)) +
+   geom_boxplot() +
+   labs(title = "Total Bank Accounts by Credit Score", x = "Credit Score", y = "Total Bank Accounts") +
+   theme_minimal()
> # Filter the dataset for customers with more than 3 bank accounts and good financial rating
> filtered_dataset_4 <- cleaned_dataset %>%
+   filter(`Total Bank Accounts` < 3 & `Financial Rating` == "Good")
> # Display the number of rows in the filtered dataset
> cat("Number of rows in filtered dataset:", nrow(filtered_dataset_4), "\n")
Number of rows in filtered dataset: 4198
>
```

Figure 110: Objective_4 with (<3) condition

By changing the condition of number of bank accounts fom ">3" to "<3" we can see the result of filtered dataset, which is 4198 and fewer than the customers with having ">3" Bnak Accounts.

	Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1	3037.987	2	4	Good	Good
2	4194.171	2	4	Good	Good
3	3037.987	2	4	Good	Good
4	3037.987	2	4	Good	Good
5	4194.171	2	4	Good	Good
6	3037.987	2	4	Good	Good
7	10474.431	1	5	Good	Good
8	10474.431	1	5	Good	Good
9	4194.171	1	5	Good	Good
10	10474.431	1	5	Good	Good
11	10474.431	1	5	Good	Good
12	10474.431	1	5	Good	Good
13	9843.868	2	5	Good	Good
14	9843.868	2	5	Good	Good
15	4194.171	2	5	Good	Good

Showing 1 to 15 of 4,198 entries, 5 total columns

Figure 111: Objective_4 Table_view (<3)

This list contains information about 4198 customers who have less than 3 bank accounts and a good financial score. It includes things like their monthly net income, number of bank accounts, and number of credit cards

Final dataset:

	Monthly Net Income	Total Bank Accounts	Total Credit Cards	Credit Profile	Financial Rating
1	4194.171	4	5	Good	Good
2	5988.705	4	5	Good	Good
3	5988.705	4	5	Good	Good
4	5988.705	4	5	Good	Good
5	5988.705	4	5	Good	Good
6	5988.705	4	5	Good	Good
7	5988.705	4	5	Good	Good
8	7449.469	6	6	Good	Good
9	4194.171	8	7	Good	Good
10	3132.524	8	7	Good	Good
11	6300.582	5	5	Good	Good
12	4194.171	6	5	Good	Good
13	3454.528	5	6	Good	Good
14	3454.528	5	6	Good	Good
15	3454.528	5	6	Good	Good
16	3454.528	5	6	Good	Good
17	4194.171	5	6	Good	Good
18	3454.528	5	6	Good	Good
19	3454.528	5	6	Good	Good
20	6623.738	5	7	Good	Good
21	6623.738	5	7	Good	Good
22	4194.171	4	6	Good	Good
23	4870.967	5	5	Good	Good
24	4870.967	5	5	Good	Good
25	4194.171	5	5	Good	Good
26	4870.967	5	5	Good	Good
27	4870.967	5	5	Good	Good

howing 1 to 27 of 2,459 entries, 5 total columns

Figure 112: Final table View with all 4 objectives applied

In the above figure 22, I filtered all the 4 objectives (monthly_net_income -> credit_profile -> total_credit_cards -> total_bank_accounts). We can see all the 4 objectives are meeting with the requirements, and in the final result there are 2459 customers who are meeting with our hypothesis criteria.

Conclusion:

Hence Prove, Customer with good credit mix,monthly inhand-salary more than 3000, more than 4 credit card ,and more than 3 bank account will have good credit score.

Additional Features (Individual)

Minhal Ali, TP073959

Explanation of Additional Features:

1. **Pie Chart**: Shows the distribution of credit profiles.
2. **Histogram**: Displays the distribution of monthly net income.
3. **Violin Plot**: Visualizes the distribution of total credit cards by financial rating.
4. **Density Plot**: Displays the density distribution of total bank accounts by financial rating.
5. **Correlation Plot**: Shows the correlation between numeric variables.
6. **Scatter Plot**: Plots the relationship between monthly net income and total credit cards, coloured by financial rating.
7. **Line Plot**: Visualizes the trend of monthly net income and total credit cards over the index.
8. **Lollipop Plot**: A lollipop plot is similar to a bar chart, but it uses lines and points instead of bars. It’s useful for comparing a single categorical variable to a numerical value.
9. **Bubble Chart**: A bubble chart is a scatter plot where each point is replaced with a bubble, and the size of the bubble represents an additional third variable.
10. **Stacked Bar Chart**: A stacked bar chart shows the total value that each part contributes to a whole, making it easier to visualize the overall size and the individual parts.

These additional features provide a more comprehensive view of the data and help in further validating the hypothesis.

Ali Fatehi Ali Hakami, TP072701

Additional features that were used in the analysis:

1- **Pie chart:** A pie chart is a circular graphic divided into slices to show proportions of a whole, making it easy to compare parts of a dataset

2- **Lolli Pop plot:** A lollipop plot connects each point to the baseline with a line, effectively comparing categorical data.

Conclusion

This analysis successfully demonstrates the validity of our hypotheses, showing clear correlations between financial rating and several key factors: First one is credit history and second one is affordability which includes factors such as having a good credit history, possessing more than one or two credit card, having a monthly net income of over 3000 and having more than three banks accounts. The multiple data cleaning, filtering, data exploration and visualization steps adequately support the assertion that customers with such characteristics have better financial ratings. Hence, contrary to the null hypotheses, our results provide support for the view of the role of maintaining differentiated credit portfolios in addition to high income for better financial performance. This therefore makes this comprehensive study indispensable for pointing out financial behaviours that enrich the stream of credit worthiness.

References

GeeksforGeeks. (2022, June 25). *Data Cleaning in R*. GeeksforGeeks; GeeksforGeeks. https://www.geeksforgeeks.org/data-cleaning-in-r/

GeeksforGeeks. (2022, December 4). *Data analysis using R*. GeeksforGeeks; GeeksforGeeks. https://www.geeksforgeeks.org/data-analysis-using-r/

Bhalla, D. (2014). *Data Exploration with R*. ListenData; ListenData. https://www.listendata.com/2014/06/data-exploration-using-r.html

Nair, A. (2019, January 15). *AIM*. AIM. https://analyticsindiamag.com/data-preprocessing-with-r-hands-on-tutorial/

R Courses & Tutorials | Codecademy. (2022). *R Courses & Tutorials*. Codecademy. https://www.codecademy.com/catalog/language/r?utm_id=t_kwd-79508828698246:loc-112:ag_1272135574815291:cp_370314517:n_o:d_c&msclkid=5449d8f345d11275f41fbe90600fa8f4&utm_source=bing&utm_medium=cpc&utm_campaign=ROW%20Language%3A%20Basic%20-%20Exact&utm_term=r%20programming%20classes&utm_content=R

Pascual, C. (2020, September 2). *Data Visualization in R with ggplot2: A Beginner Tutorial*. Dataquest. https://www.dataquest.io/blog/data-visualization-in-r-with-ggplot2-a-beginner-tutorial/