MSR.

© 2025

MSR.

2025

CUSTOMER SEGMENTATION & BEHAVIORAL ANALYTICS ON

BANK DATASET

AI with R Programming Language

2025

Made by

MUHAMMAD SYAFIQQUR R.

MSR.

2025

- This campaign is aimed at credit card customers.
- In this campaign, a 5% Maxipuan will be given for grocery and fuel expenses up to 200 TL per month with Maximum credit cards.
- **The target audience** in the campaign is **customers who do not spend enough on groceries and fuel with our Bank cards but make other types of expenses.**
- SMS will be used in the communication of the campaign.

MSR.

What we will do

We'll create a list of customers who are eligible for this campaign and who will be notified about the campaign using the file 'Customer_Information_2025_English.xlsx' given below.

There are 2 pages in this file.

1) The customer information page includes basic information about 3,000 customers, including home/car ownership status, İşCep ownership status, and whether they allow SMS and e-mail communication.

2) The spending information page includes the types and amounts of spending made by these customers with their credit cards in June, July and August.

MSR.

2025

Expected Output

As mentioned before, notifying everyone about the campaign and not notifying them at all are two extreme options. The main goal is to reach an ideal point between these two. Therefore, there is no single right answer.

Our goal is to **prepare a customer list that you think is optimal** by considering spending patterns, customer profiles and legal regulations.

A Glimpse of the Datasets

	A	B	C	D	E	F	G	H	I	J	K	L
1	IAB (Customer N	Year of Bir	Monthly Incor	Educational Status	Marital Stat	Gender	Employment Stat	Owns a Hous	Owns a Ca	Is İşcep Installe	Has SMS Communication Permissio	Has E-mail Communication Per
2	100000	1990	28500	Below High School	Single	Male	Freelance	0	1	0	0	1
3	100001	1953	29500	Undergraduate	Single	Female	Civil Servant	0	0	1	1	0
4	100002	1999	31000	High School	Single	Female	Worker	1	0	1	1	1
5	100003	1959	185000	Postgraduate	Married	Female	Civil Servant	0	0	0	0	1
6	100004	1962	32000	High School	Single	Female	Civil Servant	1	1	1	0	0
7	100005	1941	29500	Below High School	Single	Male	Retired	1	1	0	0	0
8	100006	1960	215000	Undergraduate	Single	Female	Worker	0	0	0	1	0
9	100007	1991	28500	Below High School	Single	Male	Freelance	0	0	0	0	0
10	100008	1940	32000	Postgraduate	Married	Female	Retired	1	0	1	0	0
11	100009	1968	55000	High School	Single	Female	Worker	0	1	0	1	1
12	100010	1976	25000	High School	Married	Male	Civil Servant	1	0	0	1	0
13	100011	1940	25000	Below High School	Single	Male	Retired	0	1	1	0	0
14	100012	2002	29500	Below High School	Married	Female	Worker	0	1	0	1	0
15	100013	1998	215000	Postgraduate	Married	Female	Worker	0	1	0	0	1
16	100014	1959	245000	Undergraduate	Married	Female	Worker	0	1	0	1	1
17	100015	1971	115000	Postgraduate	Married	Female	Freelance	0	0	1	0	1
18	100016	1996	125000	Undergraduate	Married	Male	Freelance	1	0	0	1	0
19	100017	1960	65000	High School	Single	Female	Civil Servant	1	0	1	1	1
20	100018	1987	25000	Undergraduate	Married	Female	Worker	1	1	1	1	0
21	100019	1997	185000	Undergraduate	Single	Female	Civil Servant	0	0	0	0	0
22	100020	1980	28500	High School	Married	Female	Civil Servant	1	1	0	0	1
23	100021	1998	29500	High School	Married	Female	Civil Servant	1	1	1	0	1
24	100022	1953	55000	High School	Single	Male	Civil Servant	0	1	0	1	1
25	100023	2000	65000	Postgraduate	Married	Male	Civil Servant	0	0	0	1	0

Customer_Information_2025_English.xlsx

A Glimpse of the Datasets

	A	B	C	D	E
1	IAB (Customer No)	KKR No.	Statement Period	Expenditure Type	Expenditure Amount
2	100000	4205 3037 6098 5126	June	Electronics	0
3	100000	4205 3037 6098 5126	June	Market	0
4	100000	4205 3037 6098 5126	June	Other	0
5	100000	4205 3037 6098 5126	June	Clothing	14171
6	100000	4205 3037 6098 5126	June	Tourism	2917
7	100000	4205 3037 6098 5126	June	Automotive	6391
8	100000	4205 3037 6098 5126	June	Cosmetics	0
9	100000	4205 3037 6098 5126	June	Cash Withdrawal	4000
10	100000	4205 3037 6098 5126	June	Education	0
11	100000	4205 3037 6098 5126	June	Furniture	0
12	100000	4205 3037 6098 5126	June	Travel	0
13	100000	4205 3037 6098 5126	June	White Goods	13066
14	100000	4205 3037 6098 5126	June	Fuel	4132
15	100000	4205 3037 6098 5126	June	Tax Payments	0
16	100000	4205 3037 6098 5126	July	Fuel	2669
17	100000	4205 3037 6098 5126	July	Education	0
18	100000	4205 3037 6098 5126	July	Travel	0
19	100000	4205 3037 6098 5126	July	White Goods	4485
20	100000	4205 3037 6098 5126	July	Clothing	8516

Expenditure_Information_2025_English.xlsx

```
# Load required libraries
library(tidyverse)  # includes dplyr, ggplot2, and other useful packages
library(readxl)     # for reading Excel files
library(scales)     # for formatting plots
library(gridExtra)  # for arranging multiple plots
library(writexl)
library(dplyr)
```

```r
# Step 1: Data Loading and Cleaning
# Read Excel files
df_customer <- read_excel("Customer_Information_2025_English.xlsx")
df_expenditure <- read_excel("Expenditure_Information_2025_English.xlsx")

# Clean up the expense period (capitalize + strip spaces)
df_expenditure$`Statement Period` <- str_to_title(trimws(df_expenditure$`Statement Period`))

# Standardize the expense type (lowercase + strip spaces)
df_expenditure$`Expenditure Type` <- tolower(trimws(df_expenditure$`Expenditure Type`))

# Convert expense amounts to numeric (safe against corrupted data)
df_expenditure$`Expenditure Amount` <- as.numeric(df_expenditure$`Expenditure Amount`)
df_expenditure$`Expenditure Amount`[is.na(df_expenditure$`Expenditure Amount`)] <- 0

# Normalize all column names to remove spaces (as a precaution)
names(df_customer) <- trimws(names(df_customer))
names(df_expenditure) <- trimws(names(df_expenditure))
```

```r
# Display first few rows of customer data
print(head(df_customer))

# Display first few rows of expenditure data
print(head(df_expenditure))
```

```
# A tibble: 6 × 12
  `IAB (Customer No)` `Year of Birth` `Monthly Income` `Educational Status`
                <dbl>           <dbl>            <dbl> <chr>
1              100000            1990            28500 Below High School
2              100001            1953            29500 Undergraduate
3              100002            1999            31000 High School
4              100003            1959           185000 Postgraduate
5              100004            1962            32000 High School
6              100005            1941            29500 Below High School
# ℹ 8 more variables: `Marital Status` <chr>, Gender <chr>,
#   `Employment Status` <chr>, `Owns a House?` <dbl>, `Owns a Car?` <dbl>,
#   `Is İşcep Installed?` <dbl>, `Has SMS Communication Permission?` <dbl>,
#   `Has E-mail Communication Permission?` <dbl>
# A tibble: 6 × 5
  `IAB (Customer No)` `KKR No.`            `Statement Period` `Expenditure Type`
                <dbl> <chr>                <chr>              <chr>
1              100000 4205 3037 6098 5126 June               electronics
2              100000 4205 3037 6098 5126 June               market
3              100000 4205 3037 6098 5126 June               other
4              100000 4205 3037 6098 5126 June               clothing
5              100000 4205 3037 6098 5126 June               tourism
6              100000 4205 3037 6098 5126 June               automotive
# ℹ 1 more variable: `Expenditure Amount` <dbl>
```

Basic data summary

summary(df_customer)

summary(df_expenditure)

```
 IAB (Customer No) Year of Birth  Monthly Income   Educational Status
 Min.   :100000    Min.   :1939   Min.   : 24500   Length:3000
 1st Qu.:100750    1st Qu.:1955   1st Qu.: 29500   Class :character
 Median :101500    Median :1971   Median : 65000   Mode  :character
 Mean   :101500    Mean   :1971   Mean   :106150
 3rd Qu.:102249    3rd Qu.:1987   3rd Qu.:175000
 Max.   :102999    Max.   :2004   Max.   :345000
 Marital Status        Gender          Employment Status  Owns a House?
 Length:3000        Length:3000        Length:3000        Min.   :0.000
 Class :character   Class :character   Class :character   1st Qu.:0.000
 Mode  :character   Mode  :character   Mode  :character   Median :0.000
                                                          Mean   :0.493
                                                          3rd Qu.:1.000
                                                          Max.   :1.000
  Owns a Car?    Is İşcep Installed? Has SMS Communication Permission?
 Min.   :0.000   Min.   :0.0000      Min.   :0.0000
 1st Qu.:0.000   1st Qu.:0.0000      1st Qu.:0.0000
 Median :1.000   Median :0.0000      Median :0.0000
 Mean   :0.505   Mean   :0.4903      Mean   :0.4883
 3rd Qu.:1.000   3rd Qu.:1.0000      3rd Qu.:1.0000
 Max.   :1.000   Max.   :1.0000      Max.   :1.0000
 Has E-mail Communication Permission?
 Min.   :0.0000
 1st Qu.:0.0000
 Median :1.0000
 Mean   :0.5063
 3rd Qu.:1.0000
 Max.   :1.0000
```

df_customer

```
IAB (Customer No)   KKR No.          Statement Period   Expenditure Type
Min.   :100000    Length:126000      Length:126000      Length:126000
1st Qu.:100750    Class :character   Class :character   Class :character
Median :101500    Mode  :character   Mode  :character   Mode  :character
Mean   :101500
3rd Qu.:102249
Max.   :102999
Expenditure Amount
Min.   :    0
1st Qu.:    0
Median :    0
Mean   : 2870
3rd Qu.: 4696
Max.   :15000
```

df_expenditure

Step 2: Identifying Categorical Variable Values

```
# Education Status
print("Educational Status:")
df_customer %>%
  distinct(`Educational Status`) %>%
  pull()

# Employment Status
print("Employment Status:")
df_customer %>%
  distinct(`Employment Status`) %>%
  pull()

# Statement Period
print("Statement Period:")
df_expenditure %>%
  distinct(`Statement Period`) %>%
  pull()

# Expenditure Type
print("Expenditure Type:")
df_expenditure %>%
  distinct(`Expenditure Type`) %>%
  pull()
```

```
 [1] "Educational Status:"
'Below High School' · 'Undergraduate' · 'High School' · 'Postgraduate'
 [1] "Employment Status:"
'Freelance' · 'Civil Servant' · 'Worker' · 'Retired'
 [1] "Statement Period:"
'June' · 'July' · 'August'
 [1] "Expenditure Type:"
'electronics' · 'market' · 'other' · 'clothing' · 'tourism' · 'automotive' · 'cosmetics' · 'cash withdrawal' · 'education' · 'furniture' · 'travel' · 'white goods' · 'fuel' · 'tax payments'
```

MSR.

Step 3: Create and Summarize Spend Groups Based on Customers

```
# Create a function to group expenditure types
expenditure_group <- function(tur) {
  tur <- tolower(trimws(as.character(tur)))
  if (grepl("market", tur)) {
    return("market")
  } else if (grepl("fuel", tur)) {
    return("fuel")
  } else {
    return("other")
  }
}

# Apply the function to create new column
df_expenditure$`Expenditure Group` <- sapply(df_expenditure$`Expenditure Type`, expenditure_group)

# Calculate the number of months spent for each customer
month_numbers <- df_expenditure %>%
  group_by(`IAB (Customer No)`) %>%
  summarise(Month_Number = n_distinct(`Statement Period`))
```

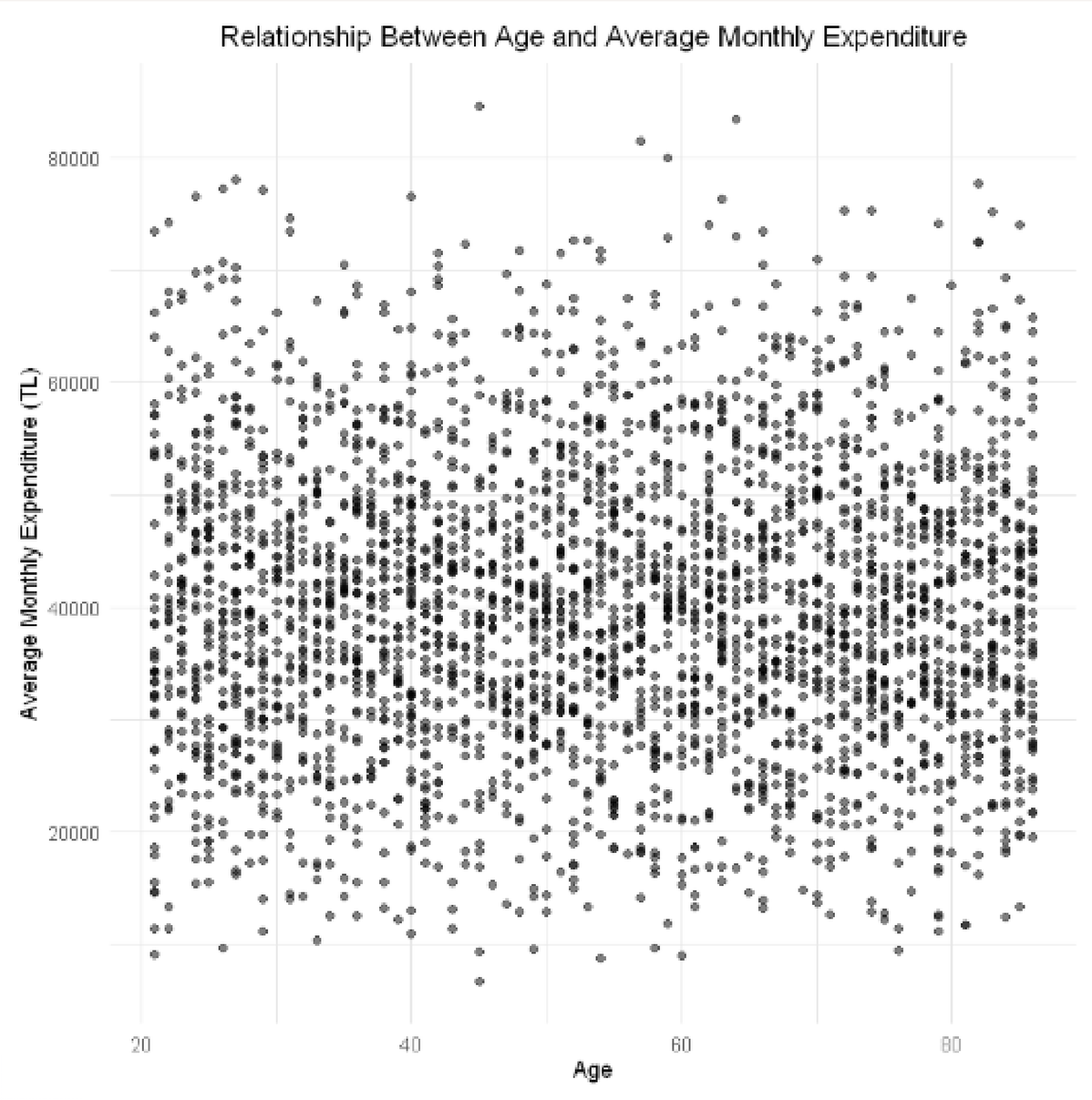
to be continued

```r
# Calculate the total amount of each spending group based on the customer
df_expenditure_summary <- df_expenditure %>%
  group_by(`IAB (Customer No)`, `Expenditure Group`) %>%
  summarise(Total_Amount = sum(`Expenditure Amount`), .groups = "drop") %>%
  pivot_wider(
    names_from = `Expenditure Group`,
    values_from = Total_Amount,
    values_fill = 0
  )

# Add possible missing groups (as columns)
required_groups <- c("market", "fuel", "other")
missing_groups <- setdiff(required_groups, names(df_expenditure_summary))
for (group in missing_groups) {
  df_expenditure_summary[[group]] <- 0
}

# Merge with the number of months
df_expenditure_summary <- left_join(df_expenditure_summary, month_numbers, by = "IAB (Customer No)")

# Calculate averages
df_expenditure_summary <- df_expenditure_summary %>%
  mutate(
    market = market / Month_Number,
    fuel = fuel / Month_Number,
    other = other / Month_Number
  )

# Remove the Month_Number column
df_expenditure_summary <- df_expenditure_summary %>%
  select(-Month_Number)

# Export to Excel
writexl::write_xlsx(df_expenditure_summary, "df_expenditure_summary_monthly.xlsx")
```

	A	B	C	D
1	IAB (Customer No)	fuel	market	other
2	100000	4110.333	0	39338.67
3	100001	5713.333	0	34261
4	100002	0	7419.333	51088.33
5	100003	0	5545.667	10869.67
6	100004	2333.333	0	34049
7	100005	3490.333	0	34293
8	100006	3682.667	0	22700.33
9	100007	0	0	24616.67
10	100008	4176.667	6855.333	33638
11	100009	0	0	21873.33
12	100010	3053.333	0	50746.33
13	100011	0	7604.667	28251.67
14	100012	0	0	24713.33
15	100013	2850.667	9759.333	20386.67
16	100014	3492.333	0	30803
17	100015	0	0	43761.67
18	100016	4447.667	5161.333	29358
19	100017	0	0	51138.67
20	100018	4132.333	0	50103.33
21	100019	4576.333	9857.333	29424.67
22	100020	0	8060.667	30723.67
23	100021	0	7093.333	16652.33
24	100022	0	0	34170

df_expenditure_summary_monthly.xlsx

```r
# Step 4: Start Data Visualization and Storytelling

# Step 4.1: Merge Customer and Expenditure Summary Data
df_customer <- read_excel("Customer_Information_2025_English.xlsx")
df_expenditure_summary <- read_excel("df_expenditure_summary_monthly.xlsx")

# Merge the dataframes
df <- left_join(df_customer, df_expenditure_summary, by = "IAB (Customer No)") %>%
  replace_na(list(market = 0, fuel = 0, other = 0))

# Step 4.2: Calculate Age and Total Expenditure Variables
df <- df %>%
  mutate(
    Age = 2025 - `Year of Birth`,
    `Total Expenditure` = market + fuel + other
  )

# Step 4.3: Age vs. Total Expenditure – Scatter Diagram
ggplot(df, aes(x = Age, y = `Total Expenditure`)) +
  geom_point(alpha = 0.5) +
  theme_minimal() +
  labs(
    title = "Relationship Between Age and Average Monthly Expenditure",
    x = "Age",
    y = "Average Monthly Expenditure (TL)"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5),
    panel.grid = element_line(color = "gray90")
  )
```



- The scatter plot suggests that there is **no strong correlation between age and average monthly expenditure**, as spending varies widely across different age groups.

- **Some individuals exhibit exceptionally high expenditures**, potentially influenced by factors such as income level, lifestyle preferences, or financial responsibilities.

- While younger individuals generally show a concentration of lower expenditures, **older individuals display a broader range**, indicating **diverse financial habits** and priorities.

```
# Step 4.4: Monthly Income vs. Monthly Average Other Expenditures – Scatter Diagram
ggplot(df, aes(x = `Monthly Income`, y = other)) +
  geom_point(alpha = 0.5) +
  theme_minimal() +
  labs(
    title = "Relationship Between Monthly Income and Monthly Average of Other Expenditures",
    x = "Monthly Income (TL)",
    y = "Monthly Average of Other Expenditures (TL)"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5),
    panel.grid = element_line(color = "gray90")
  )
```

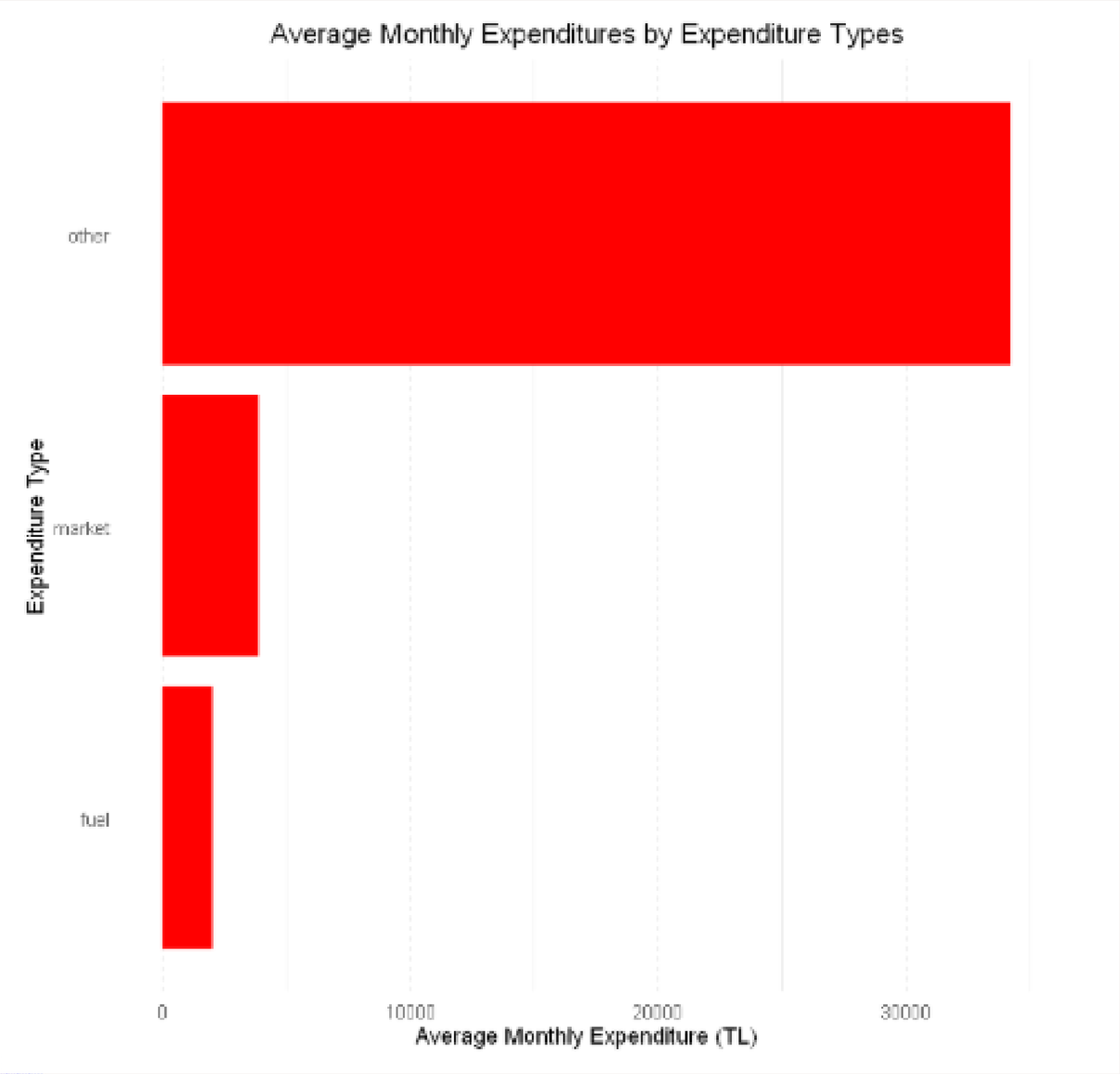
- The dense concentration of points at lower income levels suggests that a **majority of individuals earn less than 50,000 TL per month**, leading to relatively lower expenditures on other categories.

- While expenditures generally increase with income, **the scattered distribution at higher income levels indicates varied spending habits**, potentially influenced by lifestyle choices or financial priorities.

- **A few data points show significantly higher expenditures relative to income**, which might reflect financial obligations such as **loans, investments, or unusually high spending behaviors.**

Step 4.5: Monthly Average Expenditure by Expenditure Types – Horizontal Bar Graph

```
# Calculate means and create a data frame
expenditure_means <- df %>%
  summarise(
    market = mean(market),
    fuel = mean(fuel),
    other = mean(other)
  ) %>%
  pivot_longer(
    cols = everything(),
    names_to = "Expenditure_Type",
    values_to = "Mean_Expenditure"
  ) %>%
  arrange(Mean_Expenditure)

# Create horizontal bar plot
ggplot(expenditure_means, aes(x = reorder(Expenditure_Type, Mean_Expenditure), y = Mean_Expenditure)) +
  geom_bar(stat = "identity", fill = "red") +
  coord_flip() +  # This makes the bar plot horizontal
  theme_minimal() +
  labs(
    title = "Average Monthly Expenditures by Expenditure Types",
    x = "Expenditure Type",
    y = "Average Monthly Expenditure (TL)"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5),
    panel.grid.major.x = element_line(color = "gray90", linetype = "dashed"),
    panel.grid.major.y = element_blank()
  )
```

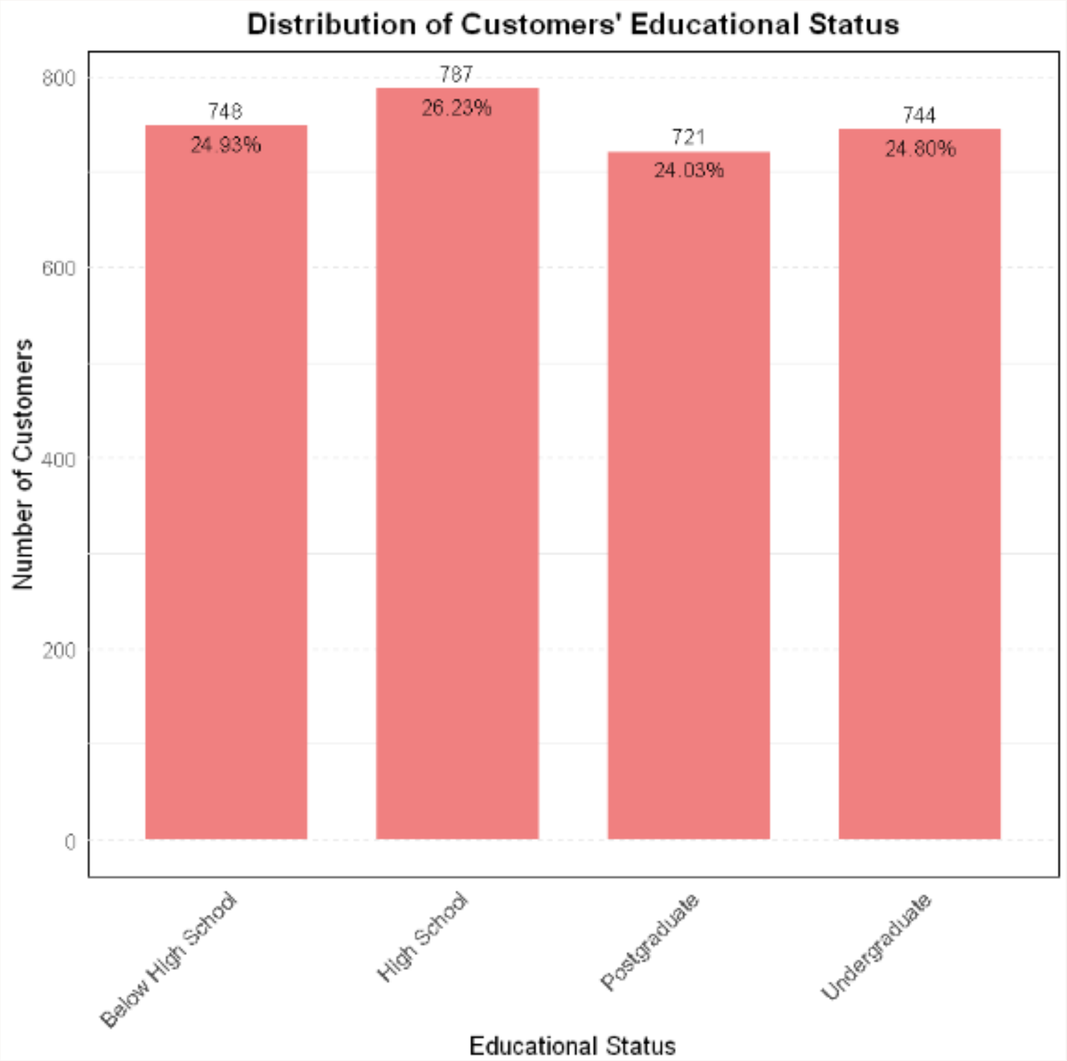


- **The "other" expenditure category has the highest monthly spending,** indicating that non-essential or **miscellaneous expenses make up a significant portion** of total costs.

- **Market-related expenses are moderately high**, suggesting that grocery and household necessities remain a consistent financial commitment.

- **Fuel expenditures are the lowest among the three categories**, which could imply efficient transportation habits or a reliance on alternative commuting methods

```r
# Step 5 – Customer Distributions According to Categorical Variables
# 5.1 – Distribution of Educational Status (Bar Chart)
ggplot(df_customer, aes(x = `Educational Status`)) +
  geom_bar(fill = "lightcoral", width = 0.7) +
  geom_text(
    aes(label = ..count..),  # Add count labels
    stat = "count",
    vjust = -0.5,  # Position labels above bars
    size = 3.5
  ) +
  theme_minimal() +
  labs(
    title = "Distribution of Customers' Educational Status",
    x = "Educational Status",
    y = "Number of Customers"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),
    panel.grid.major.y = element_line(color = "gray90", linetype = "dashed"),
    panel.grid.major.x = element_blank(),
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white")
  ) +

  # Add percentage labels
  geom_text(
    aes(label = scales::percent(..count../sum(..count..))),
    stat = "count",
    vjust = 1.5,
    size = 3.5
  )
```

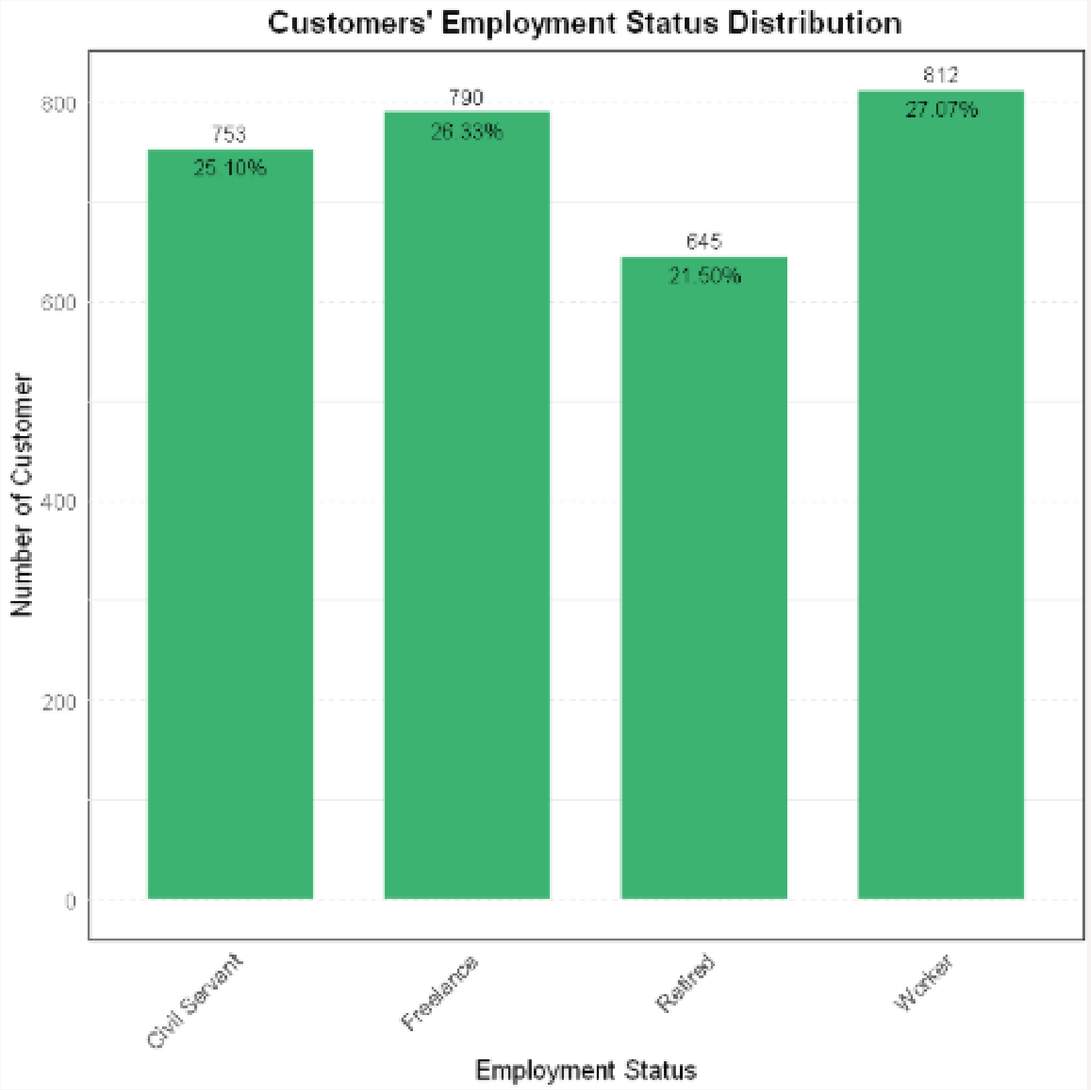


- The bar chart shows four categories: **Below High School, High School, Postgraduate, and Undergraduate.**

- **High School graduates make up the largest** group at **787 (26.23%), while Postgraduates are the smallest at 721 (24.03%),** with the **other two groups showing similar numbers.**

- This distribution helps in **understanding customer demographics, guiding marketing strategies** and service customization.

```r
# 5.2 – Employment Status Distribution (Bar Chart)
ggplot(df_customer, aes(x = `Employment Status`)) +
  geom_bar(fill = "mediumseagreen", width = 0.7) +
  geom_text(
    aes(label = ..count..),  # Add count labels
    stat = "count",
    vjust = -0.5,  # Position labels above bars
    size = 3.5
  ) +
  theme_minimal() +
  labs(
    title = "Customers' Employment Status Distribution",
    x = "Employment Status",
    y = "Number of Customer"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),
    panel.grid.major.y = element_line(color = "gray90", linetype = "dashed"),
    panel.grid.major.x = element_blank(),
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white")
  ) +
  # Add percentage labels
  geom_text(
    aes(label = scales::percent(..count../sum(..count..))),
    stat = "count",
    vjust = 1.5,
    size = 3.5
  )
```



- The bar chart shows four categories: **Below High School, High School, Postgraduate, and Undergraduate.**

- **High School graduates make up the largest** group at **787 (26.23%), while Postgraduates are the smallest at 721 (24.03%),** with the **other two groups showing similar numbers.**

- This distribution helps in **understanding customer demographics, guiding marketing strategies** and service customization.

5.3 – Home and Car Ownership Rates (Pie Chart)

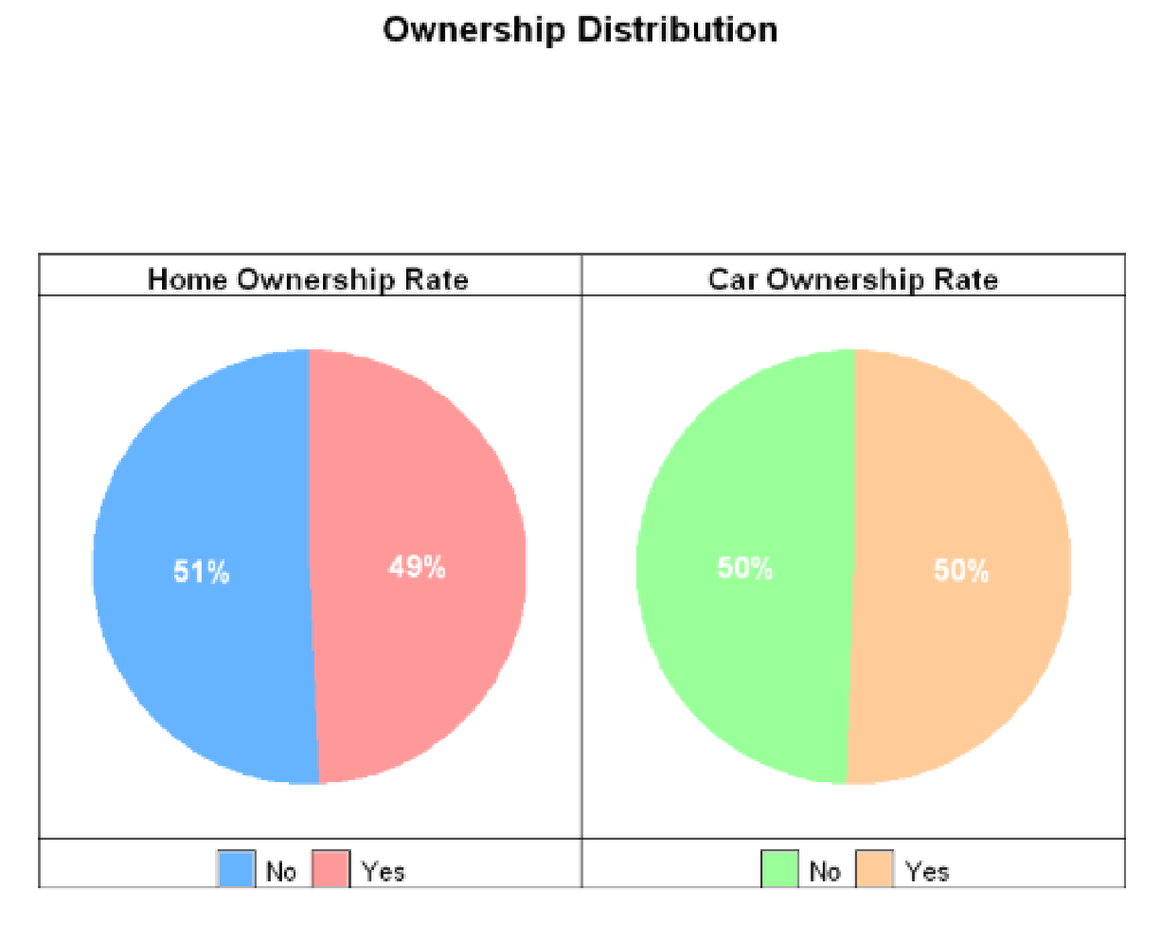
```
# Load required packages
library(ggplot2)
library(dplyr)
library(gridExtra)
library(grid)  # Add this for textGrob

# Create a function to make pie charts with more customization
create_pie_chart_custom <- function(data, column, title, colors) {
  # Calculate percentages
  pie_data <- data %>%
    count(!!sym(column)) %>%
    mutate(
      percentage = n / sum(n) * 100,
      label = paste0(round(percentage), "%"),
      category = ifelse(!!sym(column) == 1, "Yes", "No")
    )
```

to be continued

```
# Create pie chart
ggplot(pie_data, aes(x = "", y = n, fill = category)) +
geom_bar(stat = "identity", width = 1) +
coord_polar("y", start = 0) +
theme_void() +
scale_fill_manual(values = colors) +
labs(title = title) +
theme(
plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
legend.title = element_blank(),
legend.position = "bottom",
legend.text = element_text(size = 12),
plot.background = element_rect(fill = "white"),
panel.background = element_rect(fill = "white")
) +
geom_text(
aes(label = label),
position = position_stack(vjust = 0.5),
color = "white",
size = 5,
fontface = "bold"
)
}
```

to be continued

```
# Create the two pie charts
p1 <- create_pie_chart_custom(
 df_customer,
 "Owns a House?",
 "Home Ownership Rate",
 c("#66b3ff", "#ff9999")
)

p2 <- create_pie_chart_custom(
 df_customer,
 "Owns a Car?",
 "Car Ownership Rate",
 c("#99ff99", "#ffcc99")
)

# Arrange the plots side by side with a title
grid.arrange(
 p1, p2,
 ncol = 2,
 top = textGrob("Ownership Distribution", gp = gpar(fontsize =
16, fontface = "bold"))
)
```



- **Home ownership is nearly balanced,** with 49% of individuals owning a home and 51% not owning one, indicating a relatively equal split between homeowners and renters.

- **Car ownership follows a perfect 50-50 split**, suggesting that factors such as accessibility to public transportation, financial affordability, or lifestyle preferences play a crucial role in decision-making.

- **Overall ownership trends show minimal differences**, implying that both home and car ownership decisions **might be influenced by similar economic or social factors.**

Step 6 – Scatter Diagram by Customer Based on Expenditure Types

```
# Read the data
df_campaign <- read_excel("df_expenditure_summary_monthly.xlsx")

# Calculate Market + Fuel Expenditures Total
df_campaign <- df_campaign %>%
 mutate(market_fuel_total = market + fuel)

# Create scatter plot
ggplot(df_campaign, aes(x = market_fuel_total, y = other)) +
 geom_point(
  alpha = 0.6,
  color = "black",
  size = 2,
  stroke = 0.3
 ) +
 theme_minimal() +
 labs(
  title = "Distribution of Customers' Market + Fuel and Other Expenses",
  x = "Total Monthly Expenditure on Market + Fuel (TL)",
  y = "Other Monthly Average Expenditure (TL)"
 ) +
 theme(
  plot.title = element_text(size = 14, hjust = 0.5),
  axis.title = element_text(size = 12),
  panel.grid = element_line(color = "gray90", linetype = "dashed")
 )
```

- **Most customers spend relatively low amounts on market and fuel**, as shown by the dense clustering of points in the lower expenditure range.

- **Higher expenditure values exhibit a wider spread in other expenses**, suggesting that as spending on necessities increases, discretionary spending also varies significantly.

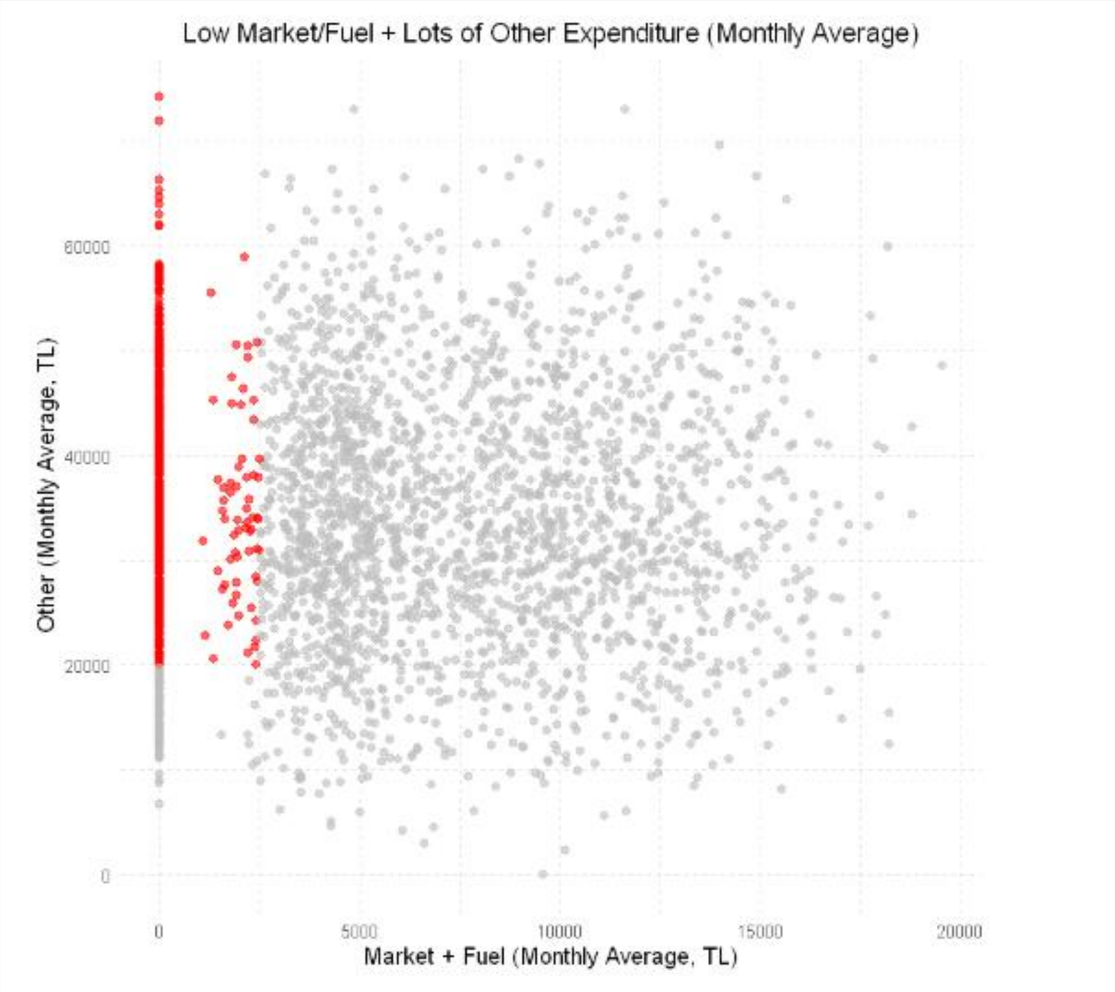
- **There is no clear linear correlation between market + fuel expenses and other expenditures**, indicating that external factors such as income levels, lifestyle choices, or financial obligations might influence spending patterns

MSR.

```
# Step 7 – Scatter Diagram Highlighting Low Market/Fuel + High Other Spenders

# Read the data if not already loaded
df_campaign <- read_excel("df_expenditure_summary_monthly.xlsx")

# Calculate monthly averages and create target customer flag
df_campaign <- df_campaign %>%
 mutate(
  market_fuel_monthly = market + fuel,
  other_monthly = other,
  target_customer = ifelse(market_fuel_monthly < 2500 & other_monthly > 20000, 1, 0)
 )
```

to be continued

```r
# Create scatter plot
ggplot(df_campaign, aes(x = market_fuel_monthly, y = other_monthly, color = factor(target_customer))) +
 geom_point(
 alpha = 0.6,
 size = 2,
 stroke = 0.3
 ) +
 scale_color_manual(
 values = c("0" = "gray", "1" = "red"), # Changed 'crimson' to 'red'
 labels = c("Other Customers", "TARGET customers")
 ) +
 theme_minimal() +
 labs(
 title = "Low Market/Fuel + Lots of Other Expenditure (Monthly Average)",
 x = "Market + Fuel (Monthly Average, TL)",
 y = "Other (Monthly Average, TL)",
 color = "" # Remove legend title
 ) +
 theme(
 plot.title = element_text(size = 14, hjust = 0.5),
 axis.title = element_text(size = 12),
 panel.grid = element_line(color = "gray90", linetype = "dashed"),
 legend.position = "top right"
 )
```



- **There is a clear pattern where individuals spend little on Market + Fuel but allocate significantly more toward Other expenses,** suggesting different spending priorities or financial flexibility.

- **The cluster of red points along the y-axis highlights a subset of individuals who consistently spend minimally on essentials but considerably on discretionary categories,** possibly indicating luxury or unpredictable spending habits.

- **The dispersed grey points suggest a broader range of expenditures,** with some individuals balancing Market + Fuel expenses more evenly, potentially reflecting different lifestyle or economic conditions.

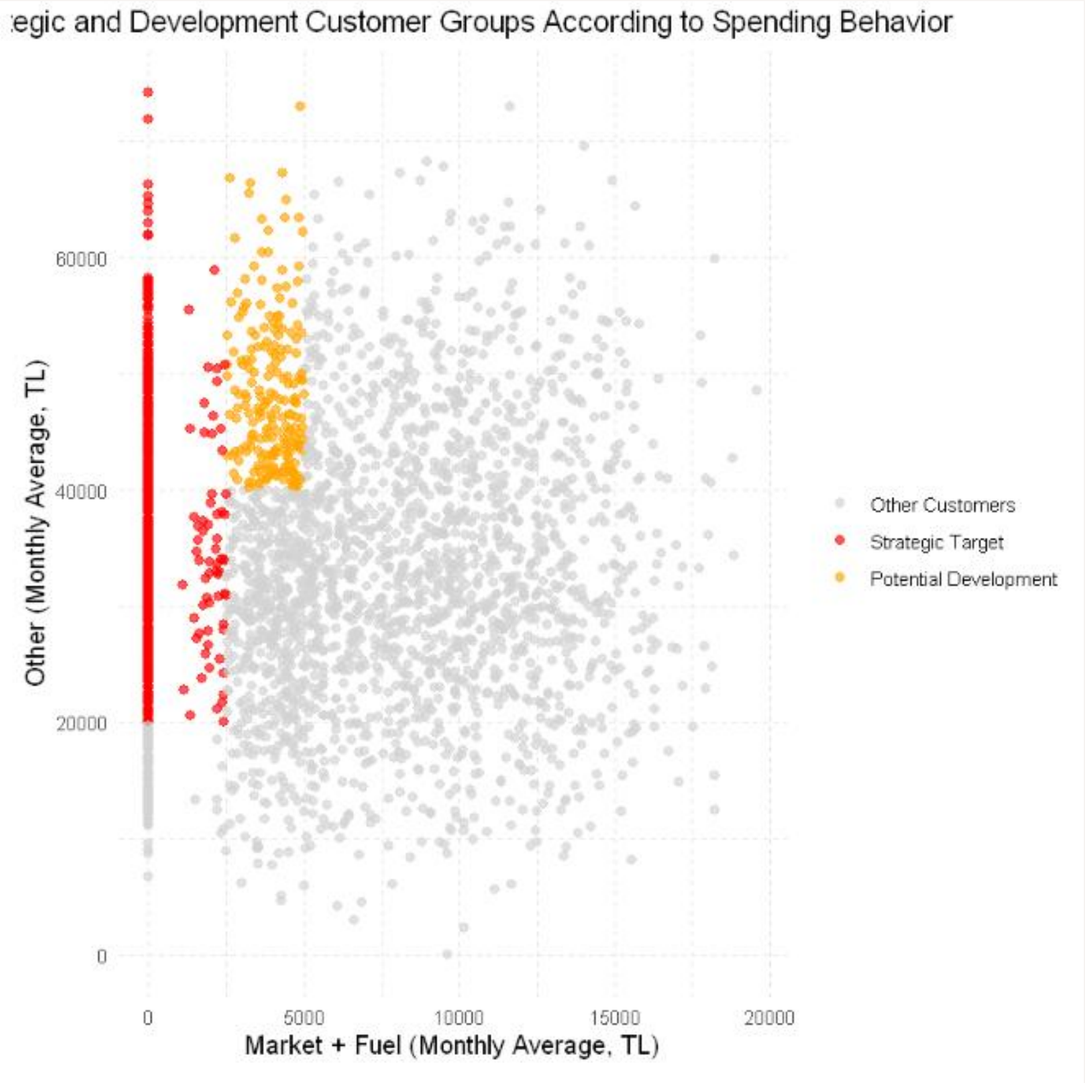
MSR.

Step 8 – Two-Level Target Customer Emphasis: "Strategic Target" and "Potential Development"

```
# Read the data if not already loaded
df_campaign <- read_excel("df_expenditure_summary_monthly.xlsx")

# Calculate monthly totals
df_campaign <- df_campaign %>%
  mutate(
    market_fuel_monthly = market + fuel,
    other_monthly = other,
    # Create segment categories
    segment_category = case_when(
      market_fuel_monthly < 2500 & other_monthly > 20000 ~ 1,  # Strategic Target
      market_fuel_monthly < 5000 & other_monthly > 40000 ~ 2,  # Potential Development
      TRUE ~ 0  # Other
    )
  )
```

to be continued

```r
# Create the plot
ggplot(df_campaign, aes(x = market_fuel_monthly, y = other_monthly, color = factor(segment_category))) +
 geom_point(
  alpha = 0.65,
  size = 2,
  stroke = 0.3
 ) +
 scale_color_manual(
  values = c("0" = "lightgray", "1" = "red", "2" = "orange"),
  labels = c("Other Customers", "Strategic Target", "Potential Development")
 ) +
 theme_minimal() +
 labs(
  title = "Strategic and Development Customer Groups According to Spending Behavior",
  x = "Market + Fuel (Monthly Average, TL)",
  y = "Other (Monthly Average, TL)",
  color = ""
 ) +
 theme(
  plot.title = element_text(size = 14, hjust = 0.5),
  axis.title = element_text(size = 12),
  panel.grid = element_line(color = "gray90", linetype = "dashed"),
  legend.position = "right"
 )
```



- **The red-highlighted points indicate a clear pattern of individuals who spend minimally on Market + Fuel but allocate significantly more toward Other expenses,** This will be our **Strategic Target** to be advertised via SMS.

- While the yellow ones will be our **Potential Development Group,** having high expenses in other but spending more on market + fuel than the Strategic Target Group.

- **The grey points are other customers that are not our target.**

MSR.

Step 9 – Scatter Plot Highlighting Strategic and Potential Groups with SMS Permission

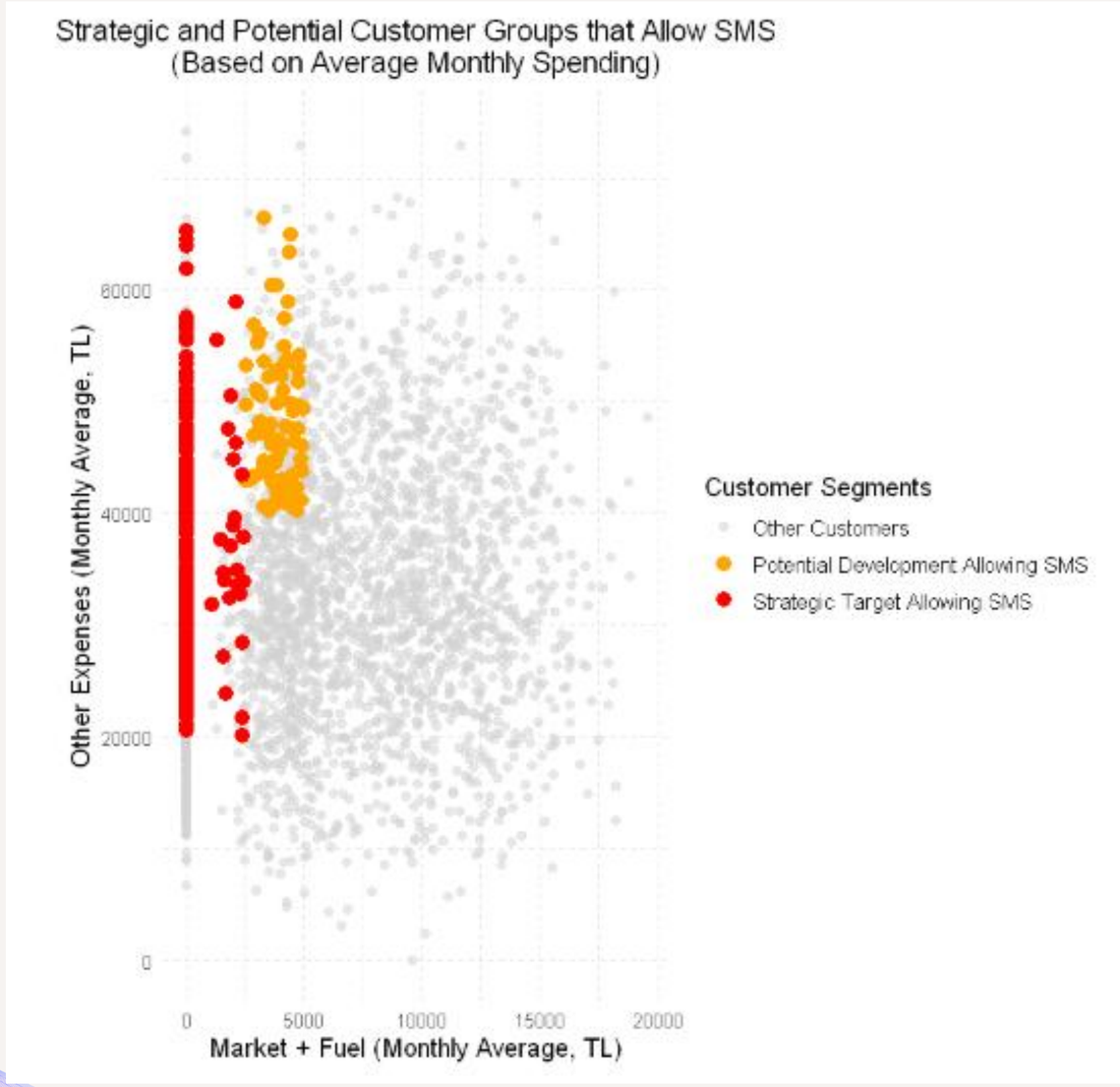
```
# 1. Read Excel files
df_customer <- read_excel("Customer_Information_2025_English.xlsx")
df_expenditure_monthly <- read_excel("df_expenditure_summary_monthly.xlsx")

# 2. Merge → campaign dataset (simplified join)
df_campaign <- left_join(
  df_customer,
  df_expenditure_monthly,
  by = "IAB (Customer No)"  # Only join on customer ID
) %>%
  replace_na(list(market = 0, fuel = 0, other = 0))

# 3. New variables: Market + Fuel and Other expenses total (monthly average)
df_campaign <- df_campaign %>%
  mutate(
    market_fuel_monthly = market + fuel,
    other_monthly = other,
    # 4. Segmentation
    segment_category = case_when(
      market_fuel_monthly < 2500 & other_monthly > 20000 ~ 1,  # Strategic Target
      market_fuel_monthly < 5000 & other_monthly > 40000 ~ 2,  # Potential Development
      TRUE ~ 0  # Other
    )
  )
```

to be continued

```r
# 5. Create subsets for different groups
strategic_sms <- df_campaign %>%
  filter(segment_category == 1 & `Has SMS Communication Permission?` == 1)

potential_sms <- df_campaign %>%
  filter(segment_category == 2 & `Has SMS Communication Permission?` == 1)

other <- df_campaign %>%
  anti_join(strategic_sms, by = "IAB (Customer No)") %>%
  anti_join(potential_sms, by = "IAB (Customer No)")

# 6. Create scatter plot with legend
ggplot() +
  # Other customers
  geom_point(
    data = other,
    aes(x = market_fuel_monthly, y = other_monthly, color = "Other Customers"),
    alpha = 0.5,
    size = 2,
    stroke = 0.2
  ) +
  # Potential Development with SMS
  geom_point(
    data = potential_sms,
    aes(x = market_fuel_monthly, y = other_monthly, color = "Potential Development Allowing SMS"),
    size = 3,
    stroke = 0.5
  ) +
```

to be continued

```r
# Strategic Target with SMS
geom_point(
  data = strategic_sms,
  aes(x = market_fuel_monthly, y = other_monthly, color = "Strategic Target Allowing SMS"),
  size = 3,
  stroke = 0.5
) +
theme_minimal() +
labs(
  title = "Strategic and Potential Customer Groups that Allow SMS\n(Based on Average Monthly Spending)",
  x = "Market + Fuel (Monthly Average, TL)",
  y = "Other Expenses (Monthly Average, TL)",
  color = "Customer Segments"  # Legend title
) +
theme(
  plot.title = element_text(size = 14, hjust = 0.5),
  axis.title = element_text(size = 12),
  panel.grid = element_line(color = "gray90", linetype = "dashed"),
  legend.position = "right",
  legend.title = element_text(size = 12),
  legend.text = element_text(size = 10)
) +
# Add legend with custom colors
scale_color_manual(
  values = c(
    "Other Customers" = "lightgray",
    "Potential Development Allowing SMS" = "orange",
    "Strategic Target Allowing SMS" = "red"
  )
)
```



- **SMS permissions are likely more prevalent among high-spending individuals on "Other Expenses",** indicating a strong correlation between discretionary spending and openness to digital engagement.

- A data-driven **campaign should prioritize the red group (strategic targets) and monitor the yellow group (potential developers).**

- Consider segmenting offers or communications based on the specific spending behaviors of each group.

Step 10: Segmentation of Strategic and Potential Customer Groups that Allow SMS and Transfer to Excel

```
# 1. Load Data
df_customer <- read_excel("Customer_Information_2025_English.xlsx")
df_expenditure_monthly <- read_excel("df_expenditure_summary_monthly.xlsx")

# Merge datasets
df_campaign <- left_join(
  df_customer,
  df_expenditure_monthly,
  by = "IAB (Customer No)"
) %>%
  replace_na(list(market = 0, fuel = 0, other = 0))

# 2. Calculate expenditure totals
df_campaign <- df_campaign %>%
  mutate(
    market_fuel_monthly = market + fuel,
    other_monthly = other
  )
```

to be continued

```r
# 3. Filter only those who permit SMS
df_sms <- df_campaign %>%
 filter(`Has SMS Communication Permission?` == 1)

# 4. Segment assignment
df_sms <- df_sms %>%
 mutate(
 Segment_Type = case_when(
 market_fuel_monthly < 2500 & other_monthly > 20000 ~ "Strategic Target",
 market_fuel_monthly < 5000 & other_monthly > 40000 ~ "Potential Development",
 TRUE ~ NA_character_ # Out of segment
 )
 )

# 5. Filter target groups (remove NA values)
df_export <- df_sms %>%
 filter(!is.na(Segment_Type))

# 6. Save to Excel
write_xlsx(df_export, "target_customer_list.xlsx")
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	(Customer	Year of Birth	Monthly Income	Educational Status	arital Statu	Gender	loyment St	wns a Hous	Owns a Car	cep Install	municatio	nmunicatic	fuel	market	other	et_fuel_mo	her_month	Segment_Type
2	100009	1968	55000	High School	Single	Female	Worker	0	1	0	1	1	0	0	21873.33	0	21873.33	Strategic Target
3	100010	1976	25000	High School	Married	Male	Civil Serva	1	0	0	1	0	3053.333	0	50746.33	3053.333	50746.33	Potential Development
4	100012	2002	29500	Below High School	Married	Female	Worker	0	1	0	1	0	0	0	24713.33	0	24713.33	Strategic Target
5	100017	1960	65000	High School	Single	Female	Civil Serva	1	0	1	1	1	0	0	51138.67	0	51138.67	Strategic Target
6	100018	1987	25000	Undergraduate	Married	Female	Worker	1	1	1	1	0	4132.333	0	50103.33	4132.333	50103.33	Potential Development
7	100022	1953	55000	High School	Single	Male	Civil Serva	0	1	0	1	1	0	0	34170	0	34170	Strategic Target
8	100024	2000	125000	High School	Married	Male	Worker	0	1	1	1	1	0	0	38539.33	0	38539.33	Strategic Target
9	100032	1945	65000	Postgraduate	Married	Male	Retired	0	1	1	1	0	0	0	28249	0	28249	Strategic Target
10	100033	1959	225000	Below High School	Married	Male	Freelance	1	1	0	1	0	0	0	41790.67	0	41790.67	Strategic Target
11	100036	1942	245000	Below High School	Single	Female	Retired	1	0	1	1	0	0	0	36757.33	0	36757.33	Strategic Target
12	100045	1985	31000	Postgraduate	Single	Female	Freelance	0	1	1	1	1	0	0	31078.67	0	31078.67	Strategic Target
13	100046	1973	25000	High School	Single	Female	Freelance	0	1	1	1	1	0	0	26615.33	0	26615.33	Strategic Target
14	100049	1942	28500	High School	Married	Female	Retired	0	1	1	1	0	0	0	38813.33	0	38813.33	Strategic Target
15	100050	1940	29500	Below High School	Married	Female	Retired	0	1	1	1	1	3514.667	0	43400.67	3514.667	43400.67	Potential Development
16	100063	2000	215000	Undergraduate	Married	Female	Civil Serva	0	0	0	1	0	0	0	42462.33	0	42462.33	Strategic Target
17	100075	1939	31000	Below High School	Single	Female	Retired	0	1	0	1	0	0	0	27517.33	0	27517.33	Strategic Target
18	100093	1975	29500	High School	Married	Female	Civil Serva	0	0	1	1	1	2451.333	0	33899.67	2451.333	33899.67	Strategic Target
19	100099	1939	35000	High School	Married	Female	Retired	0	0	1	1	0	0	0	24819.33	0	24819.33	Strategic Target
20	100103	1947	25000	Postgraduate	Married	Female	Retired	1	0	1	1	0	0	0	41310.67	0	41310.67	Strategic Target
21	100104	1953	245000	Postgraduate	Married	Male	Worker	1	0	0	1	1	0	0	20568	0	20568	Strategic Target
22	100105	1980	185000	Undergraduate	Single	Female	Worker	1	0	1	1	0	0	0	41423	0	41423	Strategic Target

target_customer_list.xlsx

01. RESULT

•**"Strategic Target"**: Customers with very **low grocery/fuel expenditures** and **very high other expenditures** are those with the highest conversion potential.

•**"Potential Development"**: Users with **low grocery/fuel expenditures, high other expenditures**, and users close to the strategic level can also be mobilized with the right timing and communication.

02. SUGGESTION

Providing campaign information to these customers **via SMS** has a **high potential for success in terms of targeted marketing** and increases the bank-wide awareness. It offers the opportunity to maximize the effectiveness of the campaign by increasing credit card usage volume

Results & Suggestions

MSR.

2025

OTHER ANALYSIS

Target Customer Profiling

2025

Now that we know our target customers, we can do a **customer profiling** on them, such as what's their gender, educational status, etc.

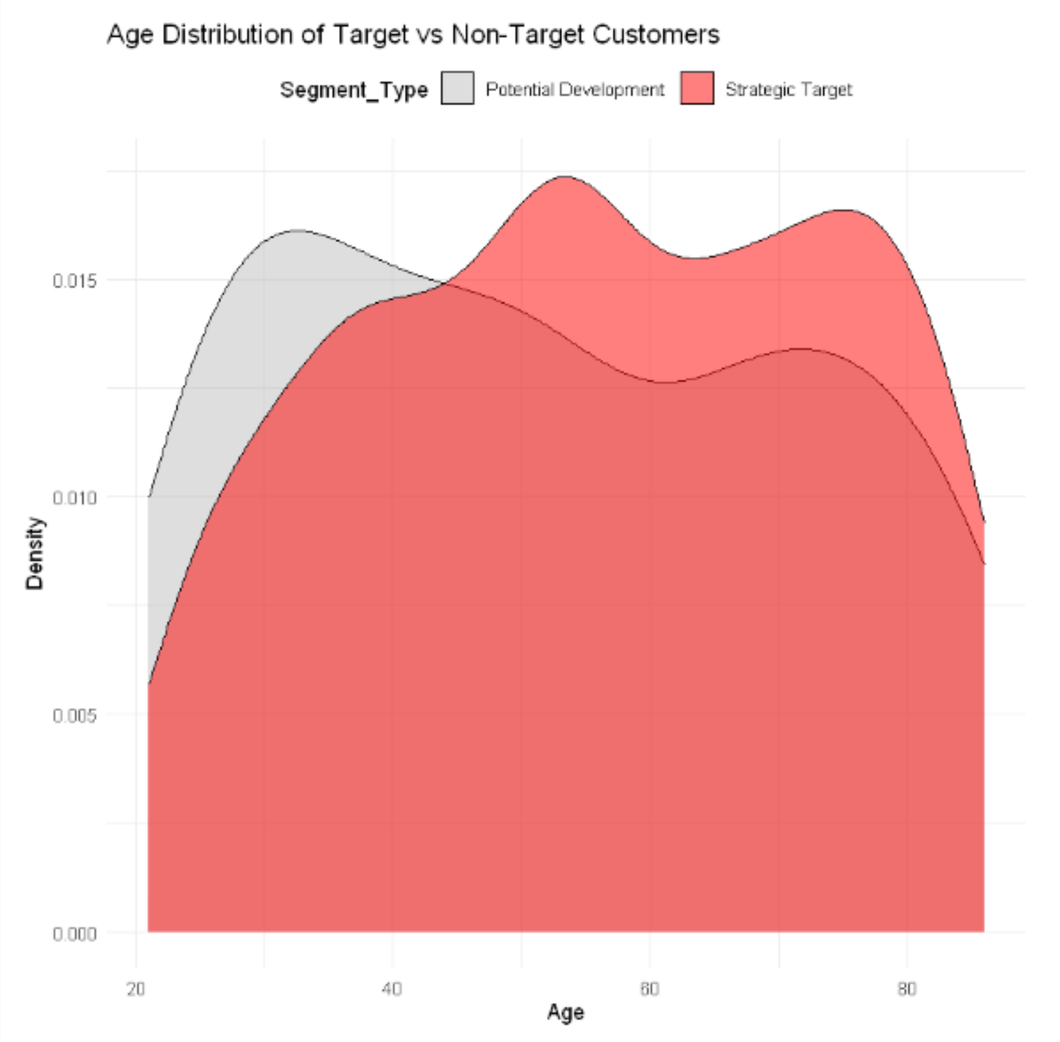
```
# Read the target data file
target_data <- read_excel("target_customer_list.xlsx")

# Calculate Age from Year of Birth
target_data$Age <- 2025 - target_data$`Year of Birth`

print(head(target_data))
```

```
# A tibble: 6 × 19
  `IAB (Customer No)` `Year of Birth` `Monthly Income` `Educational Status`
                <dbl>           <dbl>            <dbl> <chr>
1              100009            1968            55000 High School
2              100010            1976            25000 High School
3              100012            2002            29500 Below High School
4              100017            1960            65000 High School
5              100018            1987            25000 Undergraduate
6              100022            1953            55000 High School
# ℹ 15 more variables: `Marital Status` <chr>, Gender <chr>,
#   `Employment Status` <chr>, `Owns a House?` <dbl>, `Owns a Car?` <dbl>,
#   `Is İşcep Installed?` <dbl>, `Has SMS Communication Permission?` <dbl>,
#   `Has E-mail Communication Permission?` <dbl>, fuel <dbl>, market <dbl>,
#   other <dbl>, market_fuel_monthly <dbl>, other_monthly <dbl>,
#   Segment_Type <chr>, Age <dbl>
```

MSR.

1. Age Distribution Analysis

```
age_profile <- ggplot(target_data, aes(x = Age, fill = Segment_Type)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("Strategic Target" = "red", "Potential Development" = "gray")) +
  labs(title = "Age Distribution of Target vs Non-Target Customers",
       x = "Age",
       y = "Density") +
  theme_minimal() +
  theme(legend.position = "top")

# Display the plot
print(age_profile)
```



- **Strategic Target Customers (red),** who are more valuable and likely already **engage via SMS, tend to be older, with density peaks around 55 and 75 years old.**

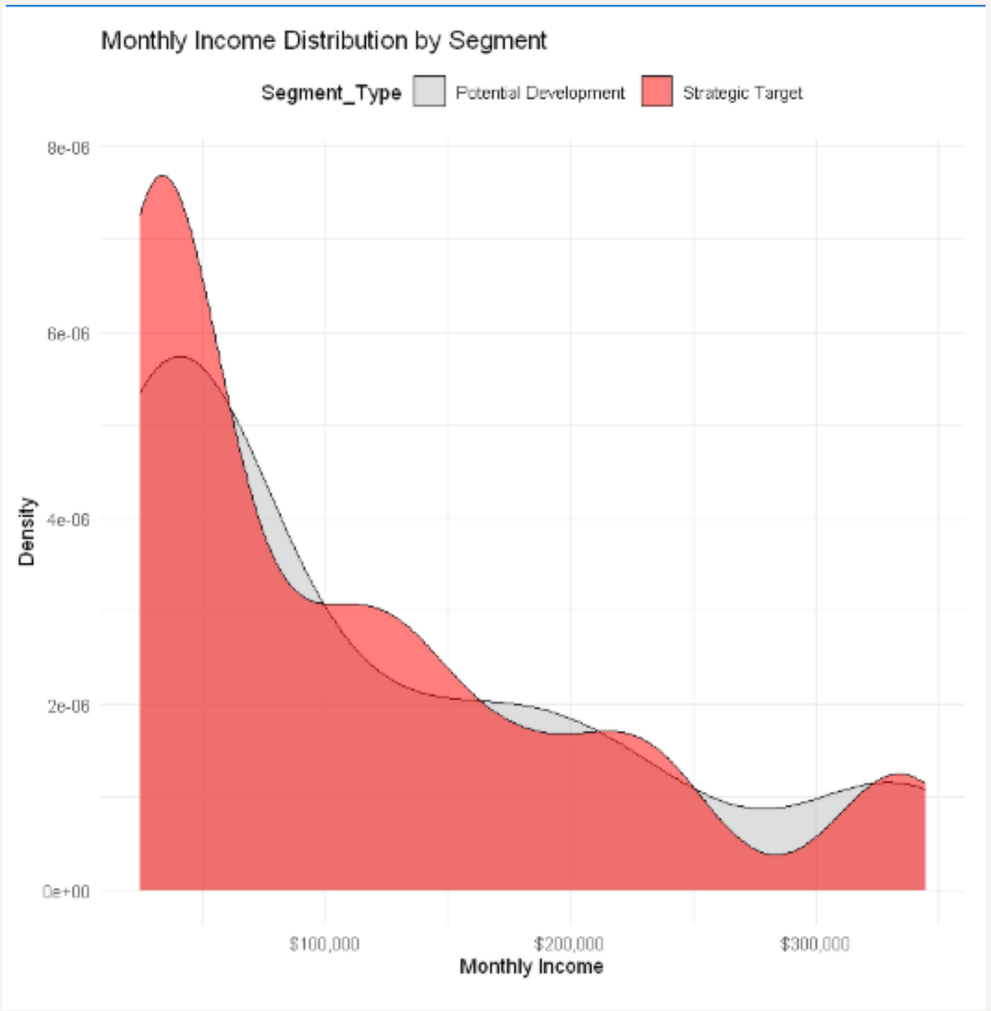
- **In contrast, Potential Development Customers (gray),** though not yet prime targets, are younger on average, with a visible peak in the 30–40 age range.

- This suggests that age plays a significant role in determining **SMS receptiveness and value potential.**

MSR.

2. Monthly Income Analysis

```
income_plot <- ggplot(target_data, aes(x = `Monthly Income`, fill = Segment_Type)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("Strategic Target" = "red", "Potential Development" = "gray")) +
  scale_x_continuous(labels = scales::dollar_format()) +
  labs(title = "Monthly Income Distribution by Segment",
       x = "Monthly Income",
       y = "Density") +
  theme_minimal() +
  theme(legend.position = "top")
print(income_plot)
```

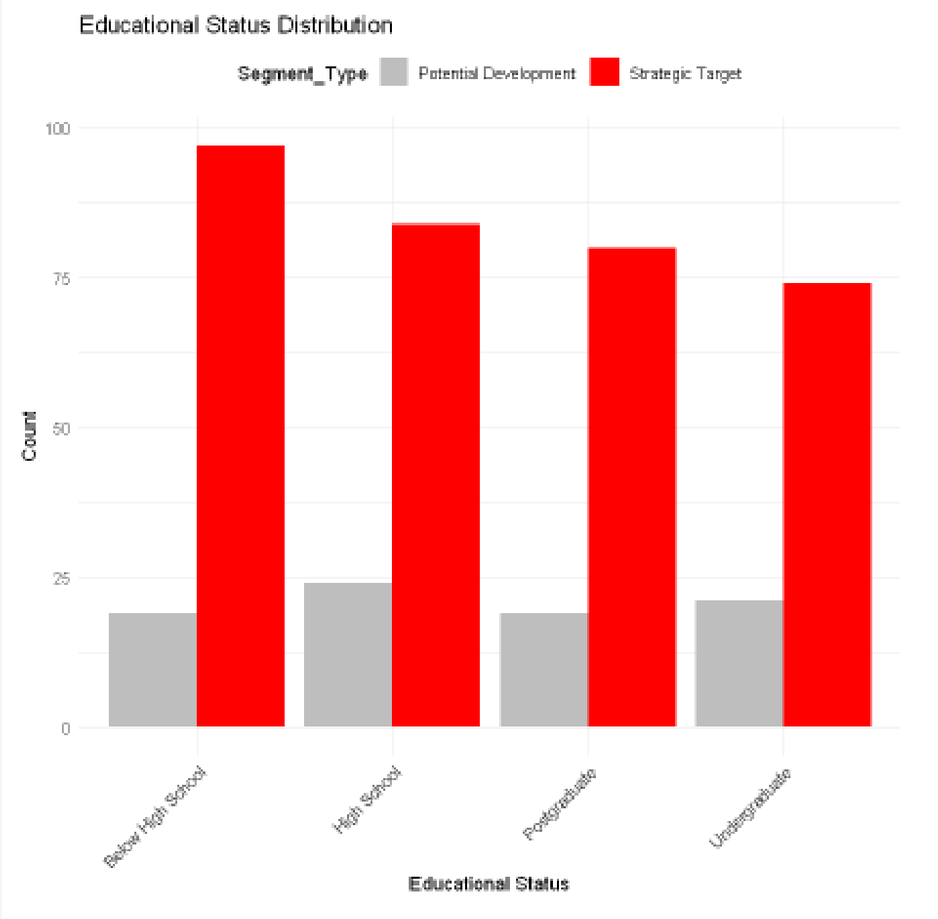


- **The majority** of both segments are clustered at the lower end of the income spectrum **(under ~₺50,000), but:**
  - **Strategic Target customers** have a broader spread across mid-to-high income ranges.
  - **The density tail for Strategic Targets stretches** further into higher income brackets **(₺100,000–₺300,000),** indicating they are disproportionately represented among higher earners.
  - **The Potential Development group tends to thin out more quickly as income increases,** signaling a concentration in the lower-middle income tiers.

MSR.

3. Educational Status

```
education_plot <- target_data %>%
  group_by(`Educational Status`, Segment_Type) %>%
  summarise(count = n(), .groups = 'drop') %>%
  ggplot(aes(x = `Educational Status`, y = count, fill = Segment_Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("Strategic Target" = "red", "Potential Development" = "gray")) +
  labs(title = "Educational Status Distribution",
       x = "Educational Status",
       y = "Count") +
  theme_minimal() +
  theme(legend.position = "top",
        axis.text.x = element_text(angle = 45, hjust = 1))
print(education_plot)
```



- The **Below High School** category has the highest count overall, meaning lower education levels might be a significant characteristic of the **Strategic Target group.**

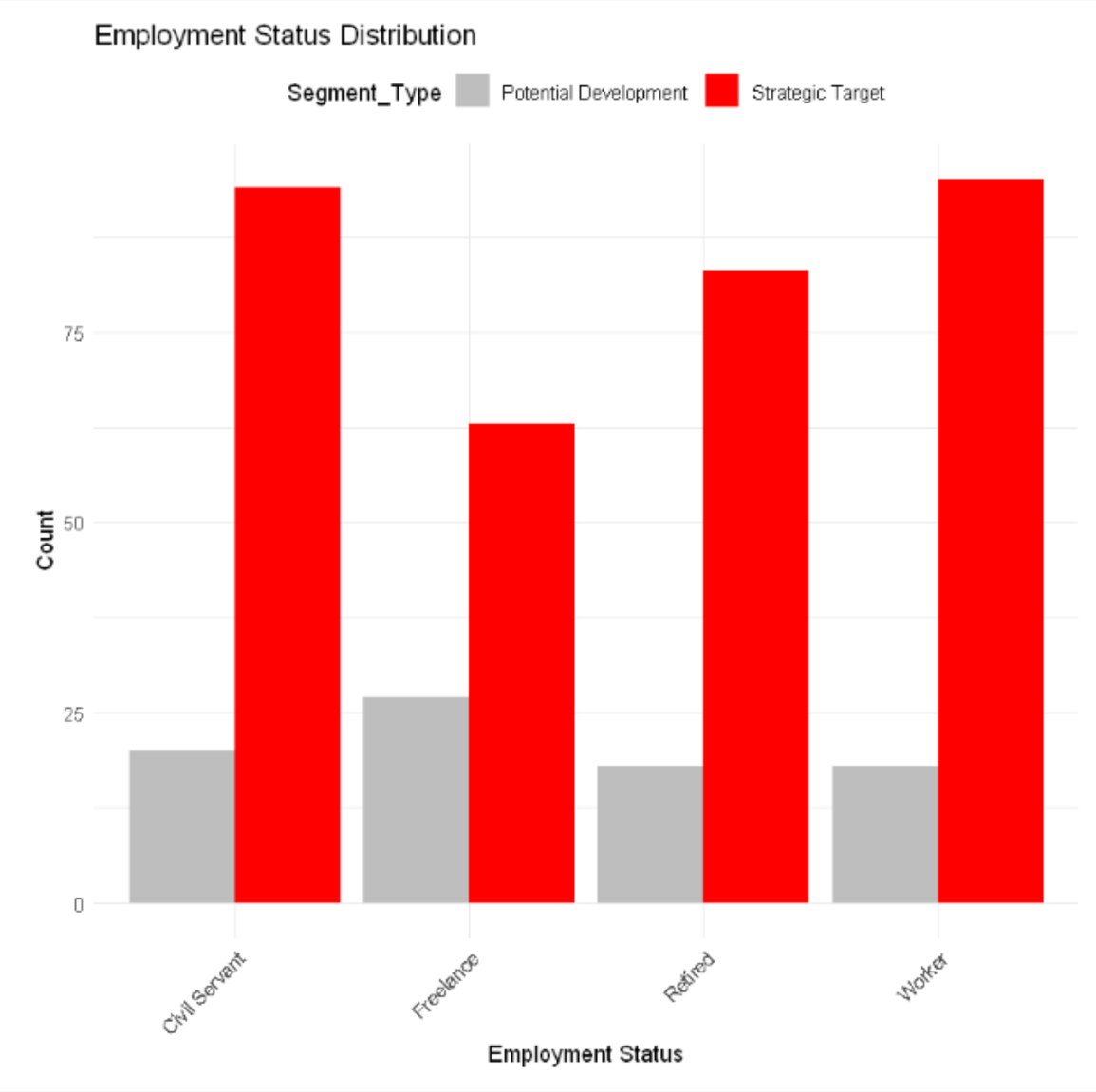
- **The Potential Development group (gray bars)** remains relatively small across all educational levels, implying that it may not be the immediate priority for action.

- There's an almost **uniform distribution across High School, Undergraduate, and Postgraduate levels** for the Strategic Target group, which could indicate a broad appeal or a diversified target audience.

MSR.

4. Employment Status Analysis

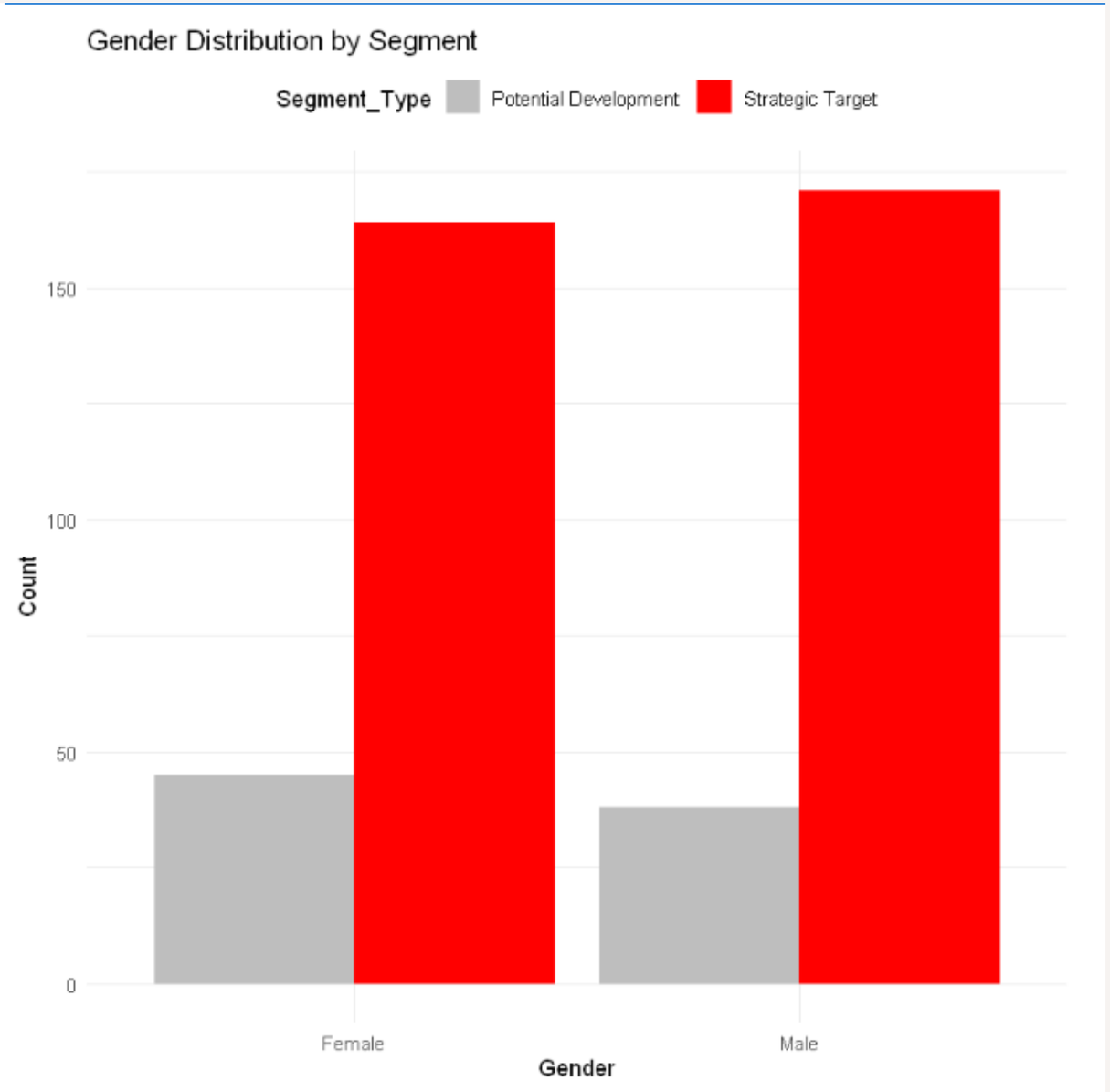
```
employment_plot <- target_data %>%
  group_by(`Employment Status`, Segment_Type) %>%
  summarise(count = n(), .groups = 'drop') %>%
  ggplot(aes(x = `Employment Status`, y = count, fill = Segment_Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("Strategic Target" = "red", "Potential Development" = "gray")) +
  labs(title = "Employment Status Distribution",
       x = "Employment Status",
       y = "Count") +
  theme_minimal() +
  theme(legend.position = "top",
        axis.text.x = element_text(angle = 45, hjust = 1))
print(employment_plot)
```


Employment Status Distribution
Segment_Type
Potential Development
Strategic Target
Count
0
25
50
75
Civil Servant
Freelance
Retired
Worker
Employment Status

MSR.

 # 5. Gender Distribution

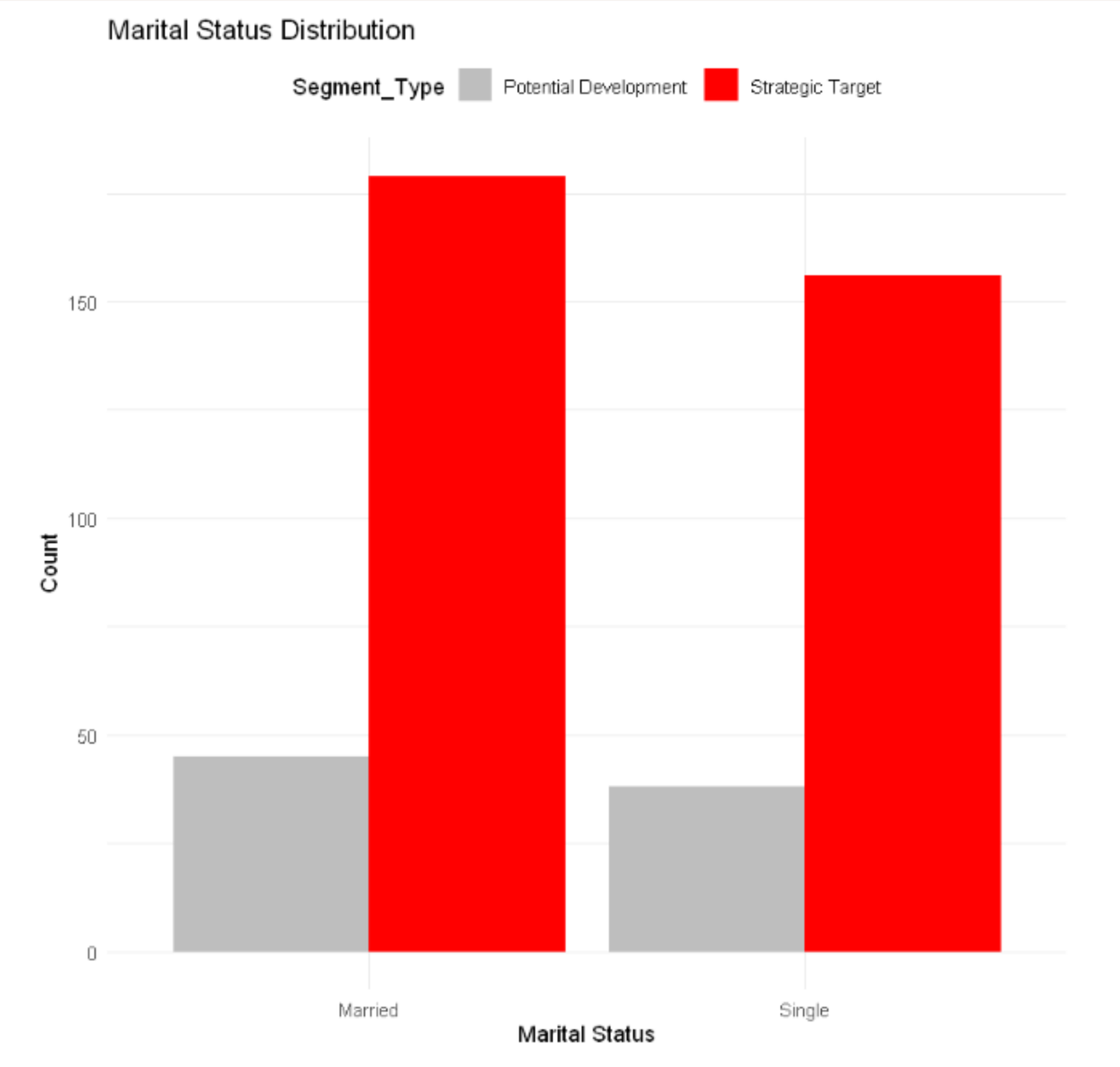
```
gender_plot <- target_data %>%
  group_by(Gender, Segment_Type) %>%
  summarise(count = n(), .groups = 'drop') %>%
  ggplot(aes(x = Gender, y = count, fill = Segment_Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("Strategic Target" = "red", "Potential Development" = "gray")) +
  labs(title = "Gender Distribution by Segment",
       x = "Gender",
       y = "Count") +
  theme_minimal() +
  theme(legend.position = "top")
print(gender_plot)
```


Gender Distribution by Segment
Segment_Type
Potential Development
Strategic Target
150
100
50
0
Count
Female
Male
Gender

MSR.

 # 6. Marital Status Plot

```
marital_plot <- target_data %>%
  group_by(`Marital Status`, Segment_Type) %>%
  summarise(count = n(), .groups = 'drop') %>%
  ggplot(aes(x = `Marital Status`, y = count, fill = Segment_Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("Strategic Target" = "red", "Potential Development" = "gray")) +
  labs(title = "Marital Status Distribution",
       x = "Marital Status",
       y = "Count") +
  theme_minimal() +
  theme(legend.position = "top")
print(marital_plot)
```


Marital Status Distribution
Segment_Type
Potential Development
Strategic Target
150
100
50
0
Count
Married
Single
Marital Status

LAST BUT NOT LEAST,

Thank You!