

assignment02

Executive Summary

Executive Summary This report investigates the drivers of credit card churn at Tifosi Bank using predictive modelling. We identified transaction frequency and account activity as the strongest predictors of churn. By applying SMOTE and XGBoost, we improved the model's recall to 82%. Based on our findings, we recommend that Tifosi Bank implement proactive outreach strategies for low-activity, high-limit customers to reduce attrition risk.

Introduction Credit card attrition represents a critical threat to Tifosi Bank's long-term profitability. With up to 25% of revenue tied to credit card operations, losing high-value customers not only reduces immediate income but undermines future cross-sell opportunities. This report investigates what drives customer churn and how analytics can inform more effective retention strategies.

Introduction

Credit card churn has a direct impact on the profitability and strategic growth of retail banks. Credit card operations can account for up to 25% of a bank's retail revenue, driven by interest earnings, interchange fees, annual fees, and cross-selling potential of other product and service. According to McKinsey's global payments insights, credit card operations can contribute significantly up to 25% to a bank's retail revenue.

For Tifosi Bank, the rise of attrition in credit card segment signals more than a loss of revenue but also reflects a shift of customer's satisfaction or loyalty. Hence, it is crucial to identify at-risk customers and understand the key drivers of churn.

In this report, we will explore the key factors contributing to credit card churn at Tifosi Bank by applying predictive modeling techniques. These insights are then translated into actionable strategies to support customer retention.

Analytical Overview

```
#package
library(tidyverse)
library(tidymodels)
library(janitor)
library(skimr)
library(here)
library(readr)
tidymodels_prefer()
```

```
raw_data<-read_csv(here("Data","bank_churners.csv")) %>%
  clean_names()
```

```
skim(raw_data)
```

Table 1: Data summary

Name	raw_data
Number of rows	10127
Number of columns	21
Column type frequency:	
character	6
numeric	15
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
attrition_flag	0	1	17	17	0	2	0
gender	0	1	1	1	0	2	0
education_level	0	1	7	13	0	7	0
marital_status	0	1	6	8	0	4	0
income_category	0	1	7	14	0	6	0
card_category	0	1	4	8	0	4	0

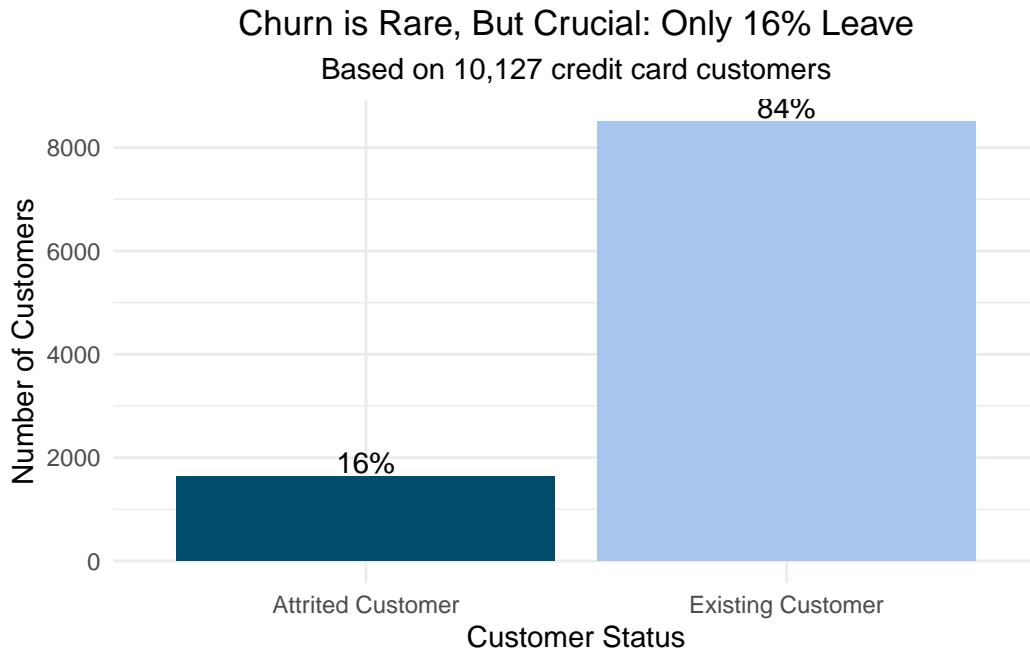
Variable type: numeric

skim_variable	n_missing	n_complete	mean	sd	p0	p25	p50	p75	p100	hist
clientnum	0	1	739177606	903783	0820833	30367701	5026358700	435382804	3083.00	
customer_age	0	1	46.33	8.02	26.0	41.00	46.00	52.00	73.00	
dependent_count	0	1	2.35	1.30	0.0	1.00	2.00	3.00	5.00	
months_on_book	0	1	35.93	7.99	13.0	31.00	36.00	40.00	56.00	
total_relationship_count	0	1	3.81	1.55	1.0	3.00	4.00	5.00	6.00	
months_inactive_12_mon	0	1	2.34	1.01	0.0	2.00	2.00	3.00	6.00	
contacts_count_12_mon	0	1	2.46	1.11	0.0	2.00	2.00	3.00	6.00	
credit_limit	0	1	8631.95	9088.78	1438.3	2555.00	4549.00	11067.50	34516.00	
total_revolving_bal	0	1	1162.81	814.99	0.0	359.00	1276.00	1784.00	2517.00	
avg_open_to_buy	0	1	7469.14	9090.69	3.0	1324.50	3474.00	9859.00	34516.00	
total_amt_chng_q4_q1	0	1	0.76	0.22	0.0	0.63	0.74	0.86	3.40	
total_trans_amt	0	1	4404.09	3397.13	510.0	2155.50	3899.00	4741.00	18484.00	
total_trans_ct	0	1	64.86	23.47	10.0	45.00	67.00	81.00	139.00	
total_ct_chng_q4_q1	0	1	0.71	0.24	0.0	0.58	0.70	0.82	3.71	
avg_utilization_ratio	0	1	0.27	0.28	0.0	0.02	0.18	0.50	1.00	

Data overview

The dataset from Tifosi Bank consists of 10127 credit card customers ,characterised by three main feature groups : demographic information (e.g., age, income, education), behavioural indicators (e.g., transaction counts, credit usage), and account relationship metrics (e.g., months on book, product count).The target variable for this report is attrition_flag which is a binary indicator that represents whether the customer remained or left

Attrition_Flag



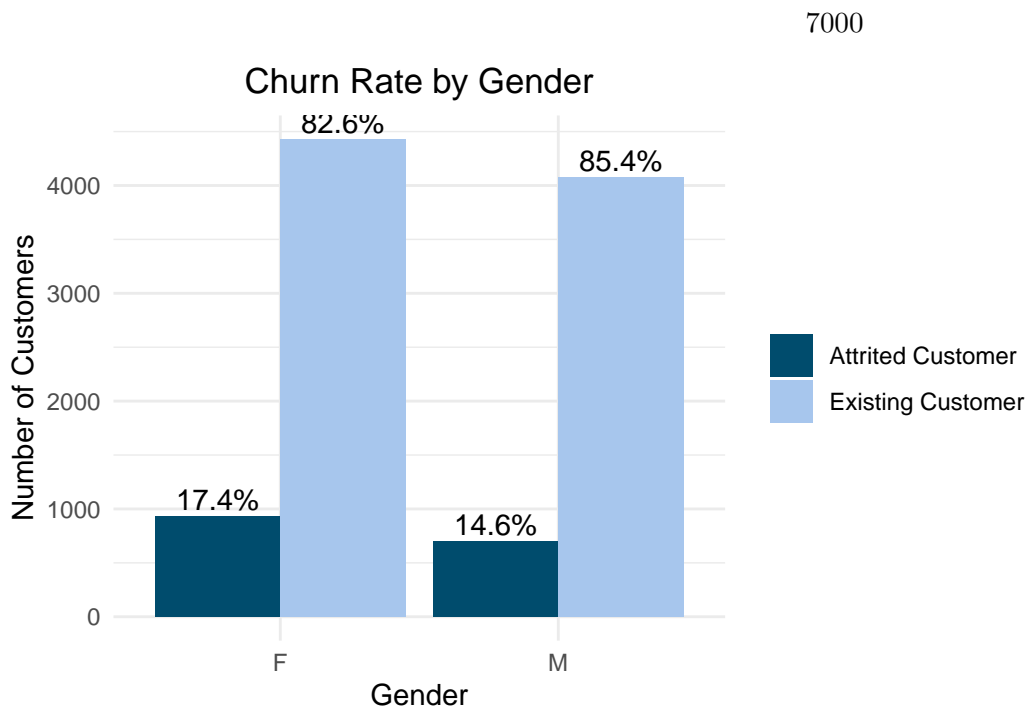
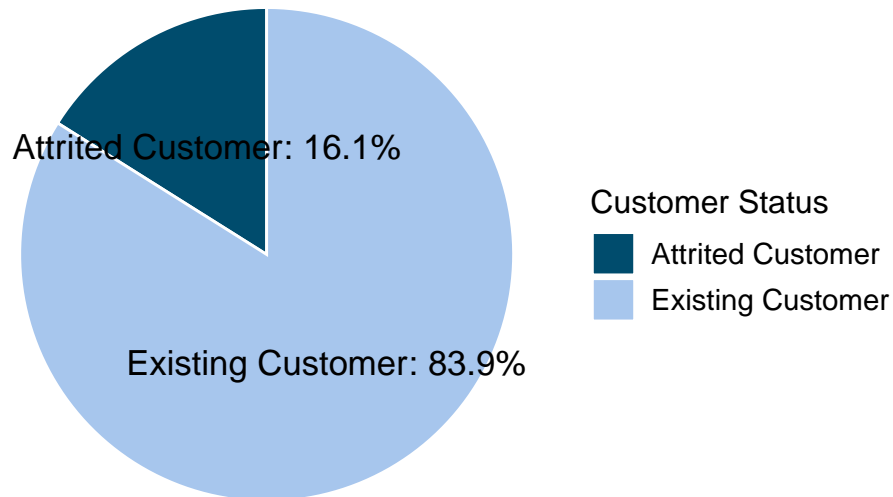
```
library(ggplot2)
library(dplyr)

#
pie_data <- raw_data %>%
  count(attrition_flag) %>%
  mutate(percentage = n / sum(n),
         label = paste0(attrition_flag, ": ", scales::percent(percentage, accuracy = 0.1)))

# pie chart
ggplot(pie_data, aes(x = "", y = percentage, fill = attrition_flag)) +
  geom_col(width = 1, color = "white") +
  coord_polar(theta = "y") +
  geom_text(aes(label = label), position = position_stack(vjust = 0.5), size = 4.5) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Customer Attrition Distribution",
    fill = "Customer Status"
  ) +
  theme_void() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    legend.title = element_text(size = 12),
```

```
legend.text = element_text(size = 11)
)
```

Customer Attrition Distribution



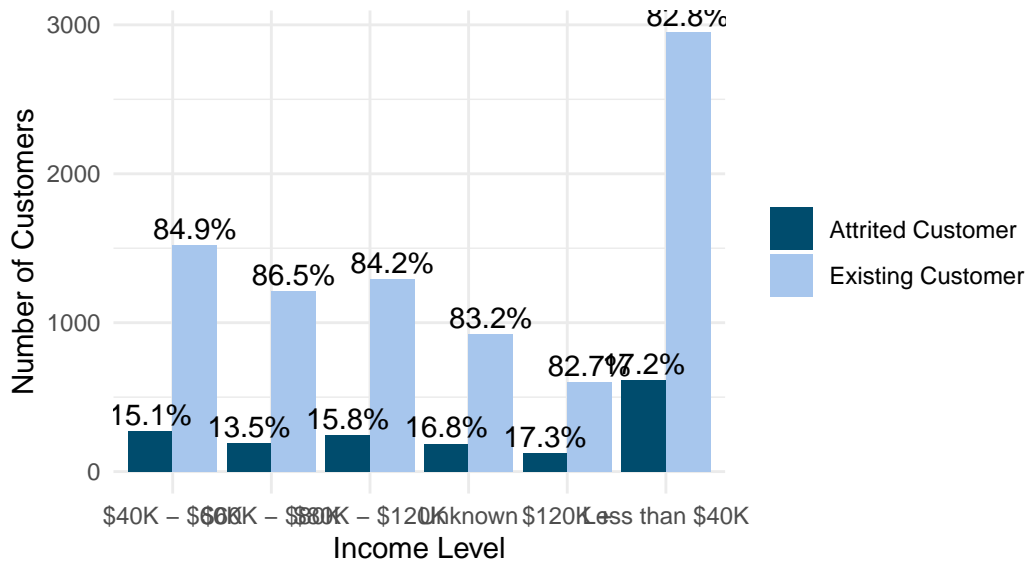
```
Warning: There was 1 warning in `mutate()`.
i In argument: `income_category = fct_relevel(...)`.
```

Caused by warning:

```
! 2 unknown levels in `f`: < $40K and > $120K
```

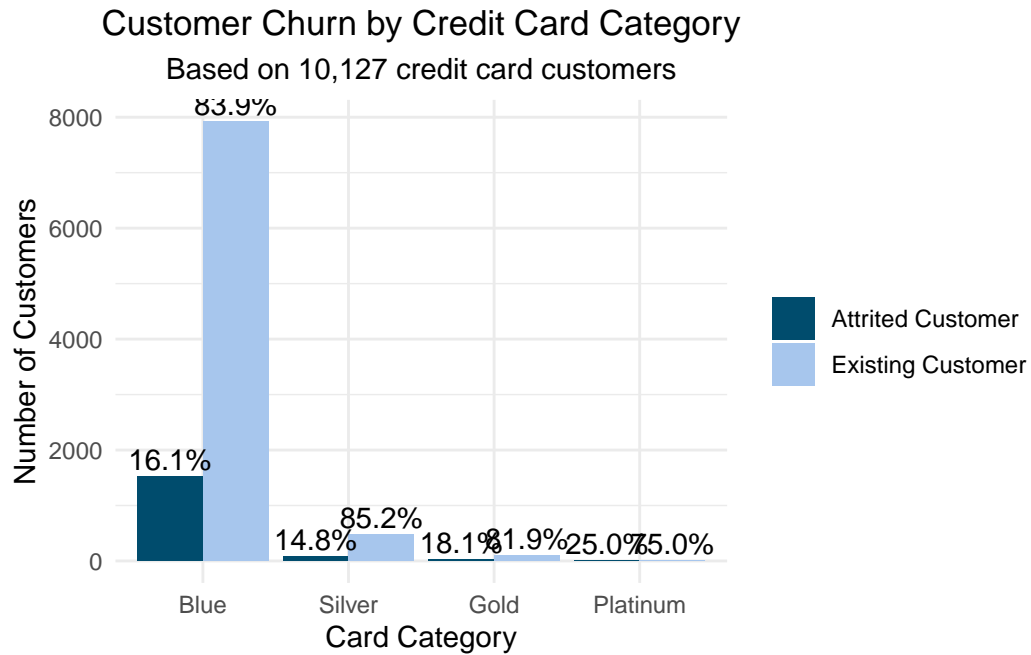
Lower-income Customers Show Higher Churn Rates

Based on 10,127 credit card customers

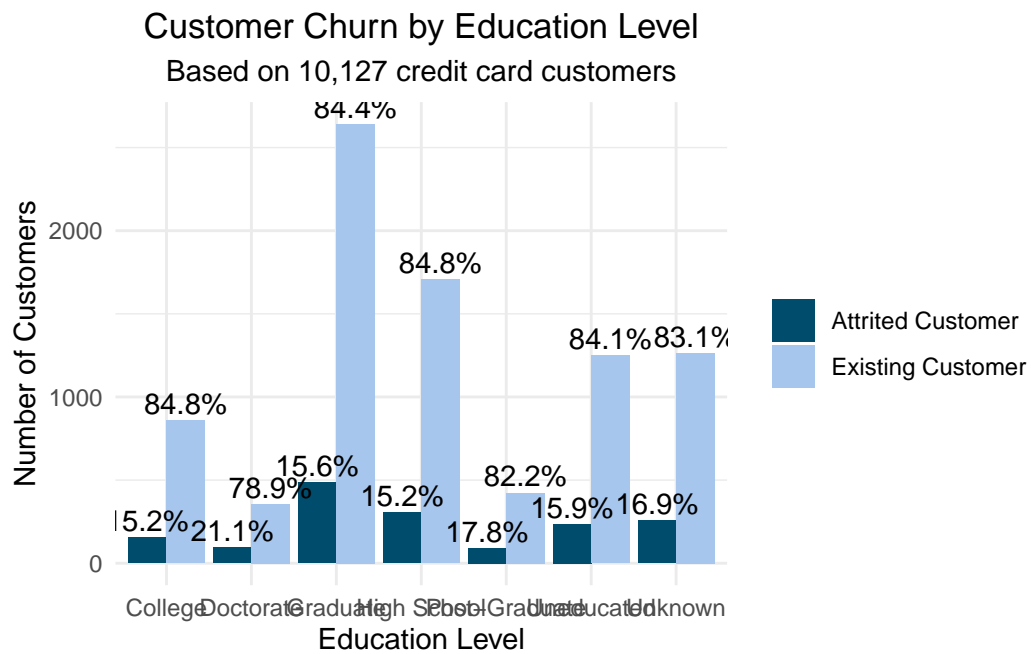


card category

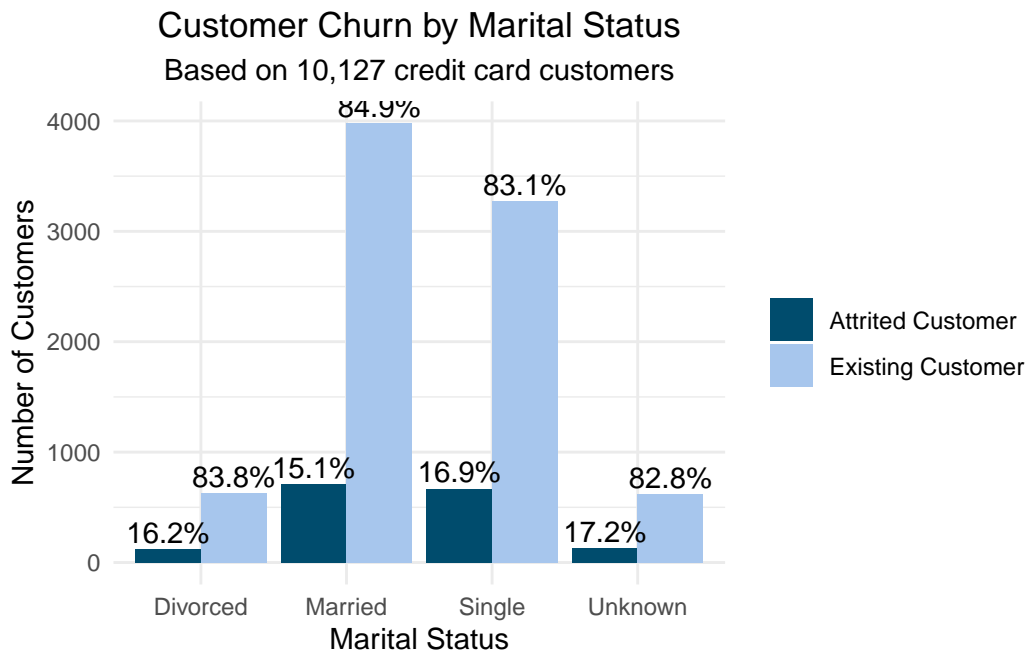
```
[1] "Blue"      "Gold"      "Silver"    "Platinum"
```



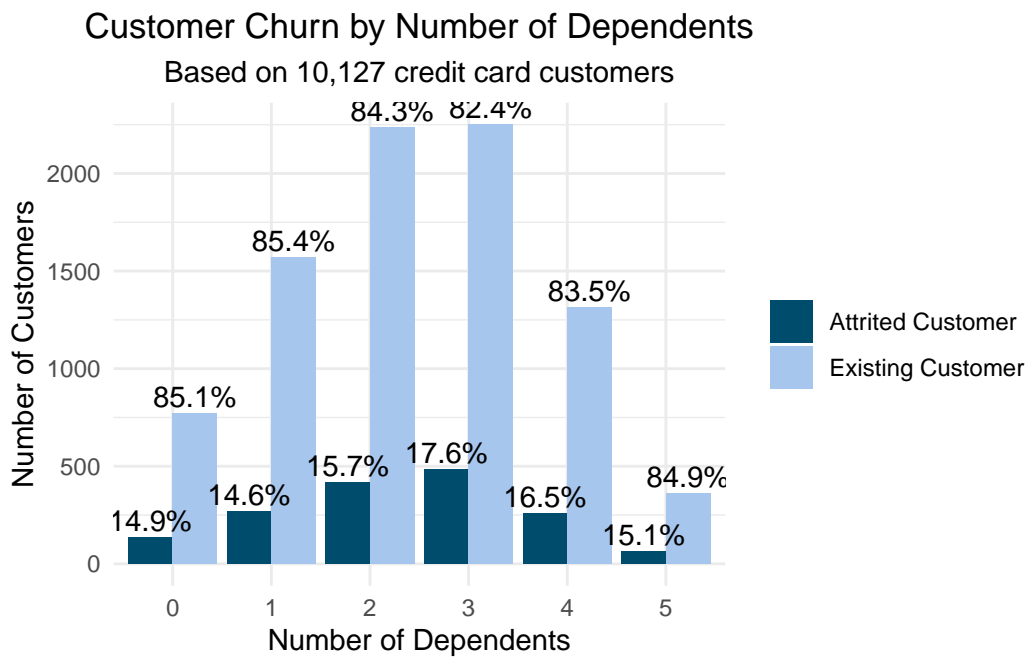
education_level



marital_status

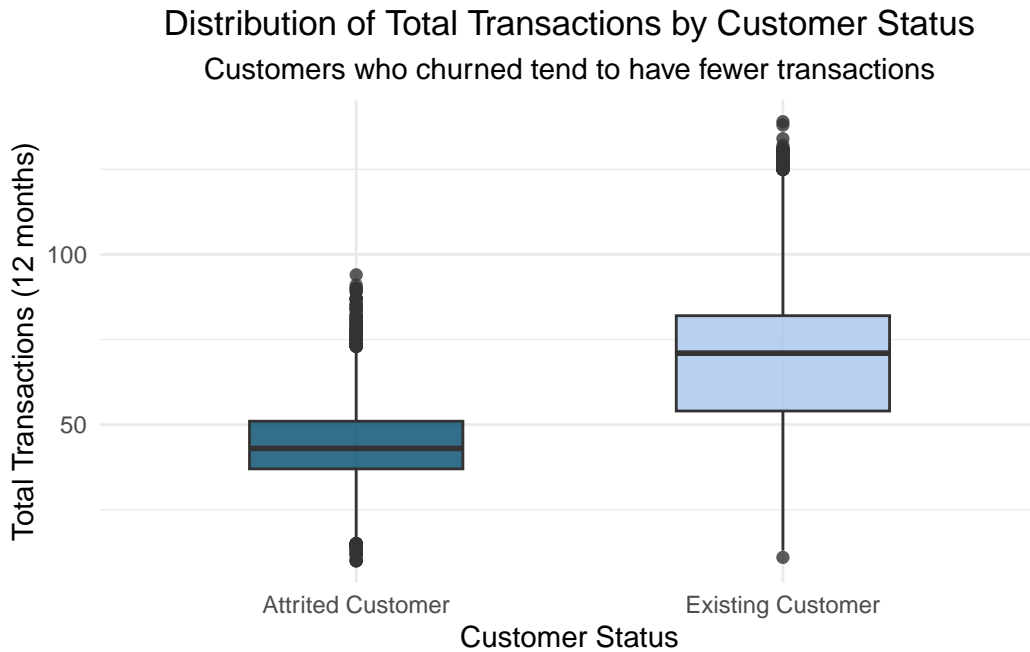


dependent



Total Transactions (12 months)

```
ggplot(raw_data, aes(x = attrition_flag, y = total_trans_ct, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Distribution of Total Transactions by Customer Status",
    subtitle = "Customers who churned tend to have fewer transactions",
    x = "Customer Status",
    y = "Total Transactions (12 months)"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5))
```

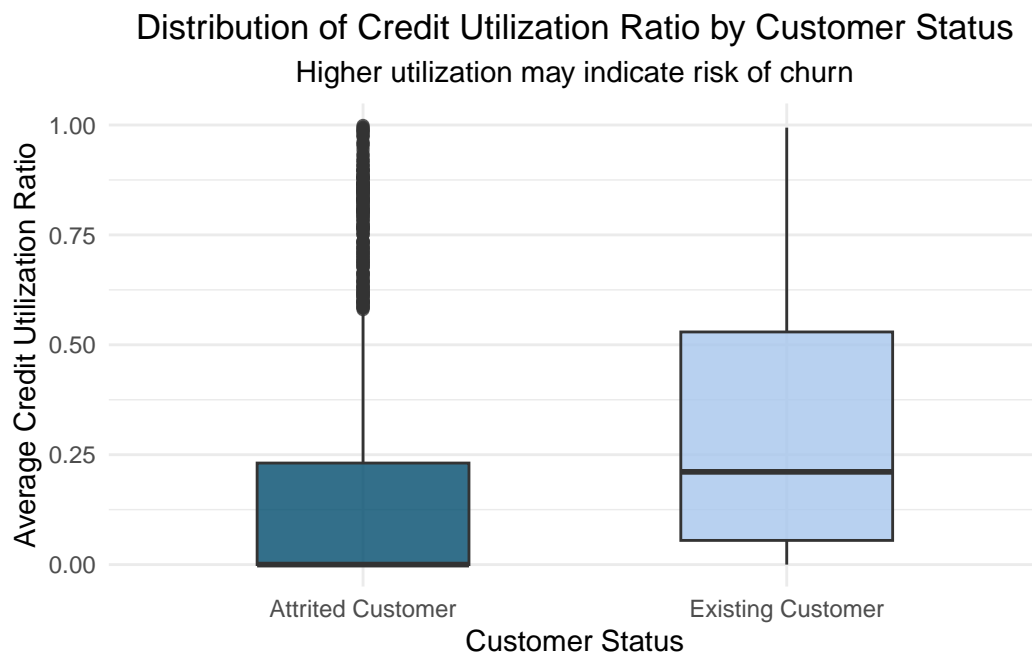
**avg_utilization_ratio**

```
ggplot(raw_data, aes(x = attrition_flag, y = avg_utilization_ratio, fill = attrition_flag))
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
```

```

title = "Distribution of Credit Utilization Ratio by Customer Status",
subtitle = "Higher utilization may indicate risk of churn",
x = "Customer Status",
y = "Average Credit Utilization Ratio"
) +
theme_minimal() +
theme(legend.position = "none",
      plot.title = element_text(hjust = 0.5),
      plot.subtitle = element_text(hjust = 0.5))

```



credit_limit

```

ggplot(raw_data, aes(x = attrition_flag, y = credit_limit, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Distribution of Credit Limit by Customer Status",
    subtitle = "Customers with lower limits may churn more",
    x = "Customer Status",
    y = "Credit Limit ($)"
  ) +

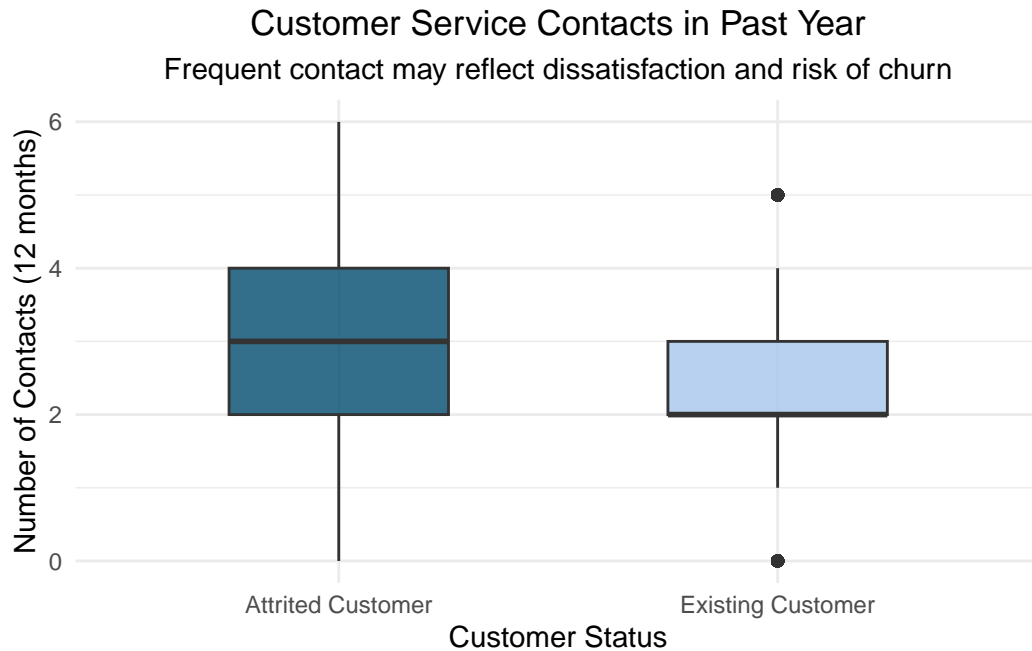
```

```
theme_minimal() +
theme(legend.position = "none",
      plot.title = element_text(hjust = 0.5),
      plot.subtitle = element_text(hjust = 0.5))
```



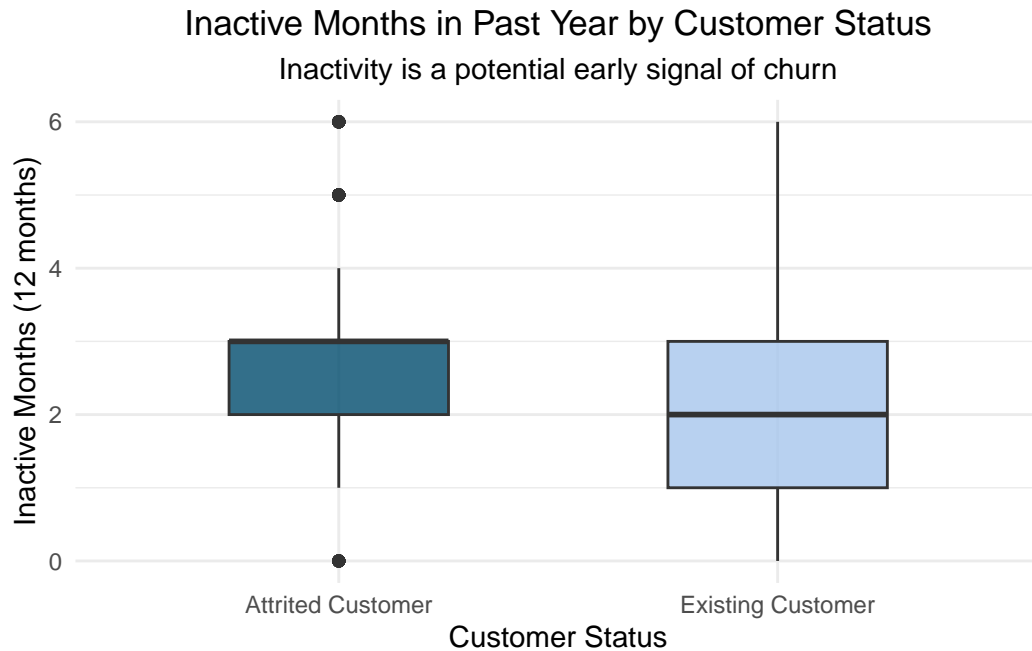
contacts_count_12_mon

```
ggplot(raw_data, aes(x = attrition_flag, y = contacts_count_12_mon, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Customer Service Contacts in Past Year",
    subtitle = "Frequent contact may reflect dissatisfaction and risk of churn",
    x = "Customer Status",
    y = "Number of Contacts (12 months)"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



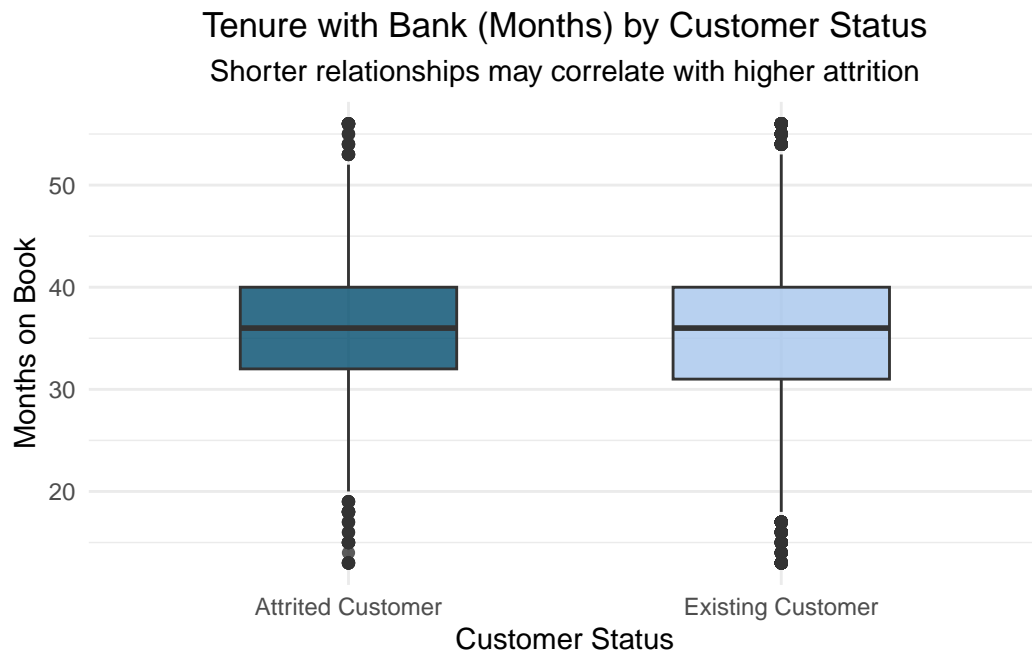
months_inactive_12_mon

```
ggplot(raw_data, aes(x = attrition_flag, y = months_inactive_12_mon, fill = attrition_flag))
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Inactive Months in Past Year by Customer Status",
    subtitle = "Inactivity is a potential early signal of churn",
    x = "Customer Status",
    y = "Inactive Months (12 months)"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



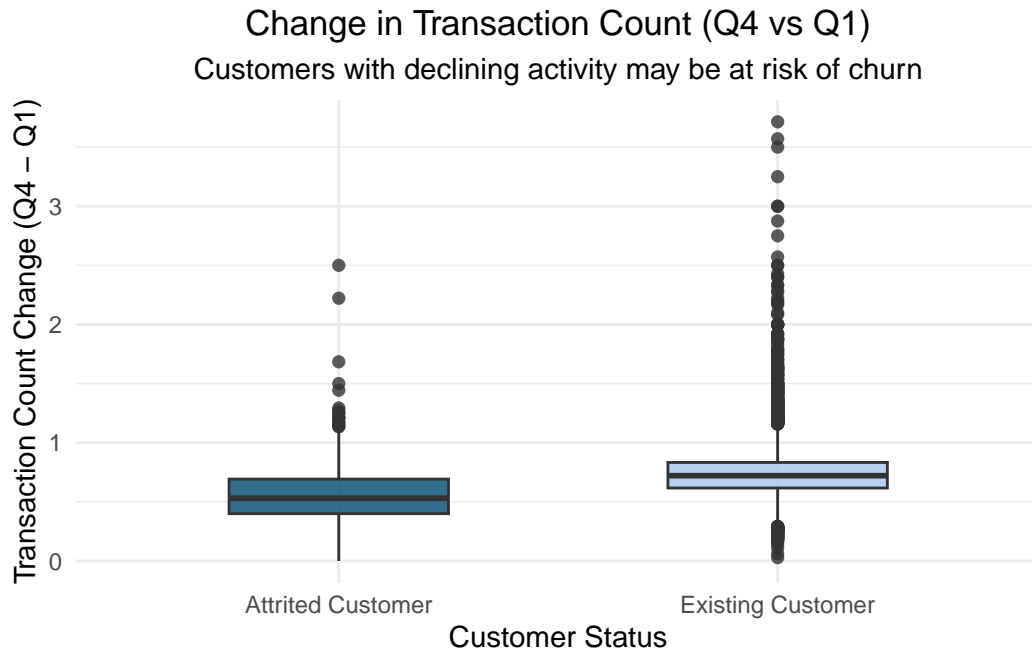
months_on_book

```
ggplot(raw_data, aes(x = attrition_flag, y = months_on_book, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Tenure with Bank (Months) by Customer Status",
    subtitle = "Shorter relationships may correlate with higher attrition",
    x = "Customer Status",
    y = "Months on Book"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



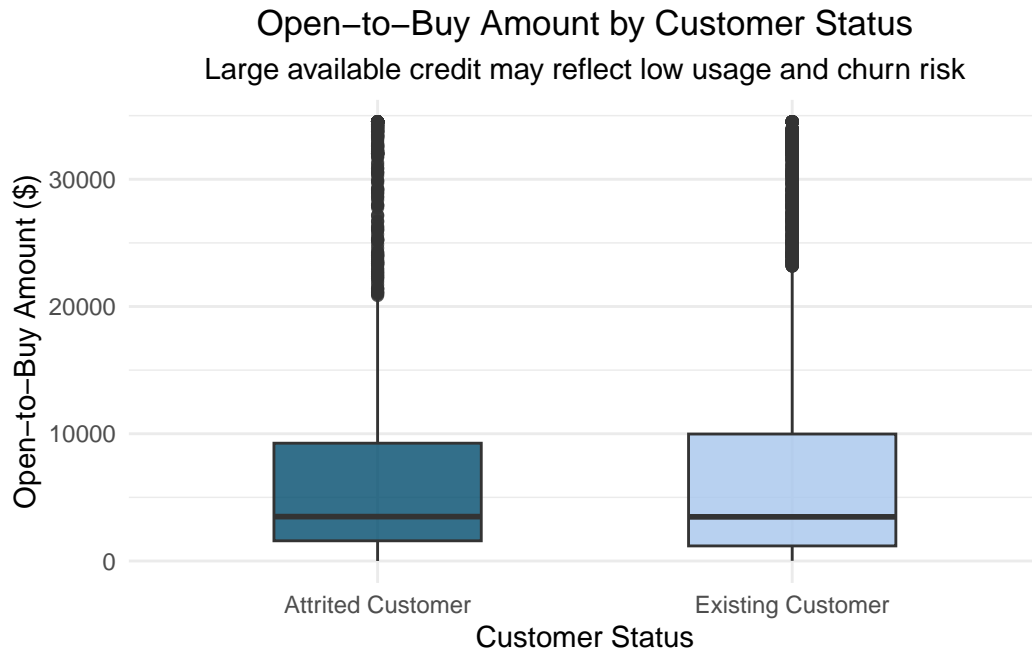
total_ct_chng_q4_q1 Q4 vs Q1

```
ggplot(raw_data, aes(x = attrition_flag, y = total_ct_chng_q4_q1, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Change in Transaction Count (Q4 vs Q1)",
    subtitle = "Customers with declining activity may be at risk of churn",
    x = "Customer Status",
    y = "Transaction Count Change (Q4 - Q1)"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



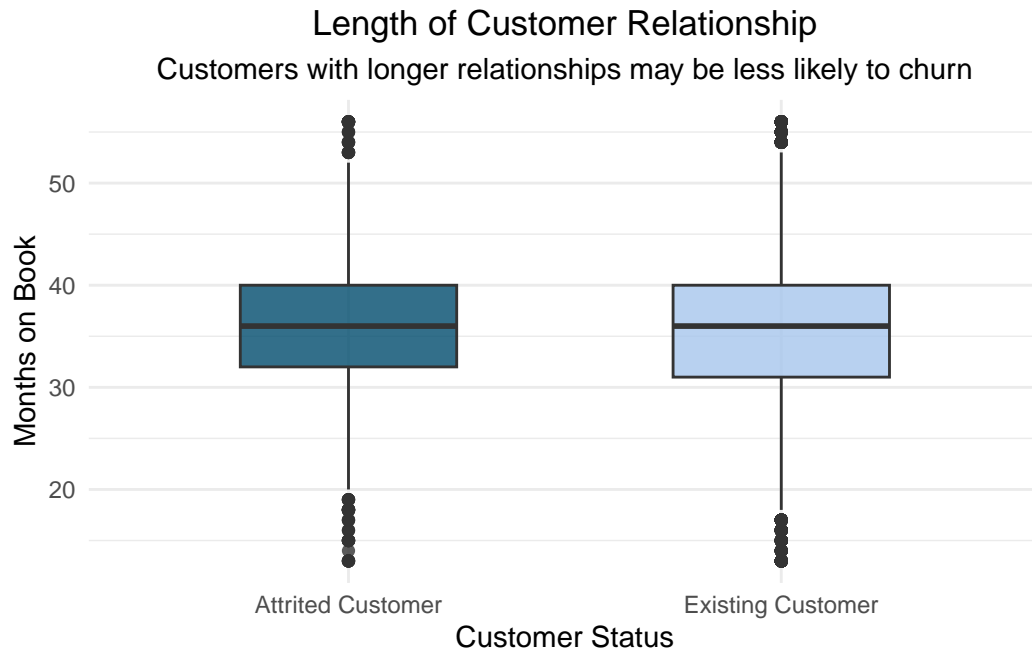
avg_open_to_buy = limit - balance

```
ggplot(raw_data, aes(x = attrition_flag, y = avg_open_to_buy, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Open-to-Buy Amount by Customer Status",
    subtitle = "Large available credit may reflect low usage and churn risk",
    x = "Customer Status",
    y = "Open-to-Buy Amount ($)"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



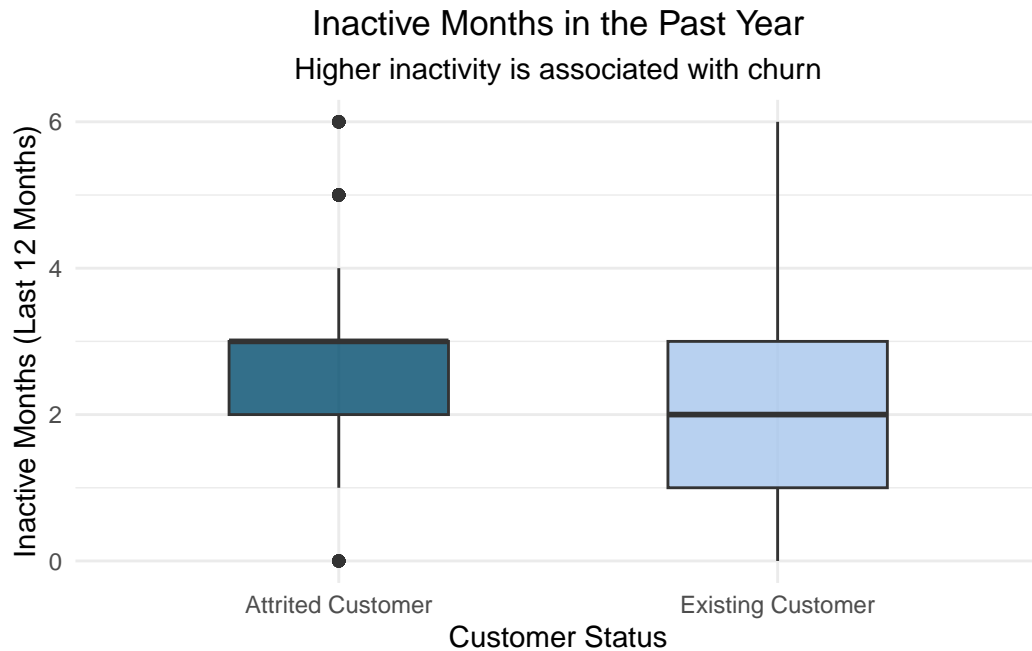
months_on_book

```
ggplot(raw_data, aes(x = attrition_flag, y = months_on_book, fill = attrition_flag)) +  
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +  
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +  
  labs(  
    title = "Length of Customer Relationship",  
    subtitle = "Customers with longer relationships may be less likely to churn",  
    x = "Customer Status",  
    y = "Months on Book"  
  ) +  
  theme_minimal() +  
  theme(legend.position = "none",  
        plot.title = element_text(hjust = 0.5),  
        plot.subtitle = element_text(hjust = 0.5))
```

months_inactive_12_mon

```
ggplot(raw_data, aes(x = attrition_flag, y = months_inactive_12_mon, fill = attrition_flag))
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Inactive Months in the Past Year",
    subtitle = "Higher inactivity is associated with churn",
    x = "Customer Status",
    y = "Inactive Months (Last 12 Months)"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



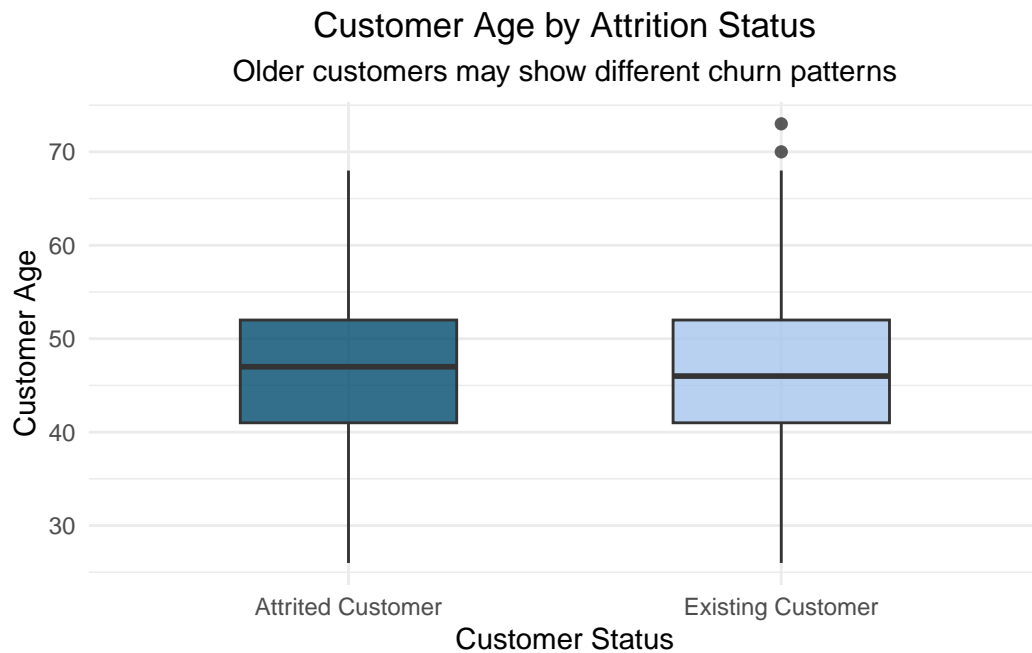
avg_open_to_buy

```
ggplot(raw_data, aes(x = attrition_flag, y = avg_open_to_buy, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Available Credit (Open to Buy)",
    subtitle = "Low usage of available credit may indicate disengagement",
    x = "Customer Status",
    y = "Average Open to Buy ($)"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



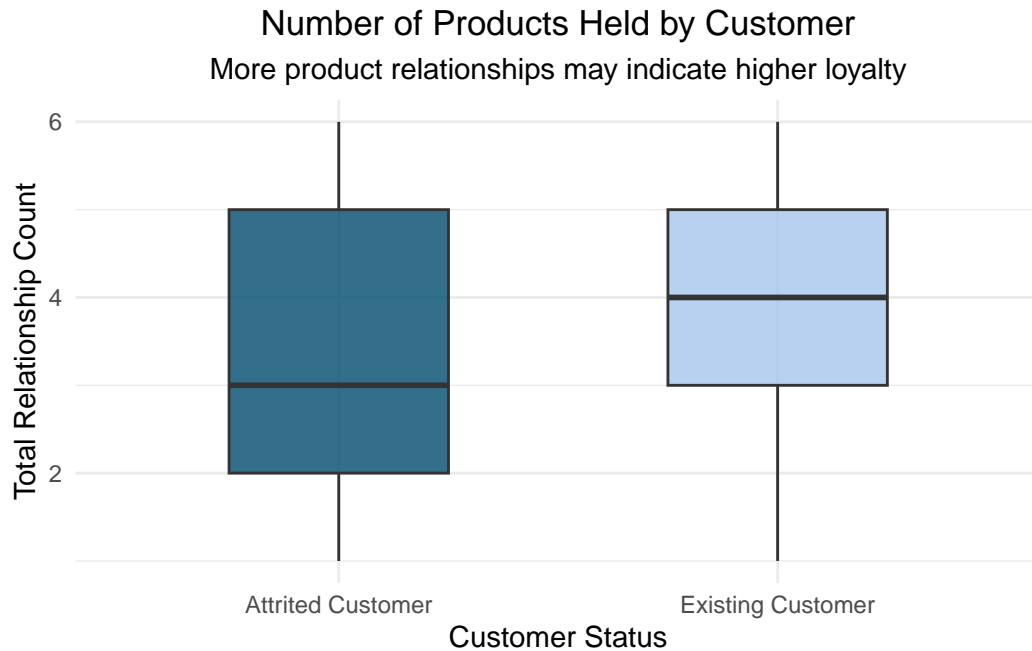
customer_age

```
ggplot(raw_data, aes(x = attrition_flag, y = customer_age, fill = attrition_flag)) +  
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +  
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +  
  labs(  
    title = "Customer Age by Attrition Status",  
    subtitle = "Older customers may show different churn patterns",  
    x = "Customer Status",  
    y = "Customer Age"  
  ) +  
  theme_minimal() +  
  theme(legend.position = "none",  
        plot.title = element_text(hjust = 0.5),  
        plot.subtitle = element_text(hjust = 0.5))
```



total_relationship_count

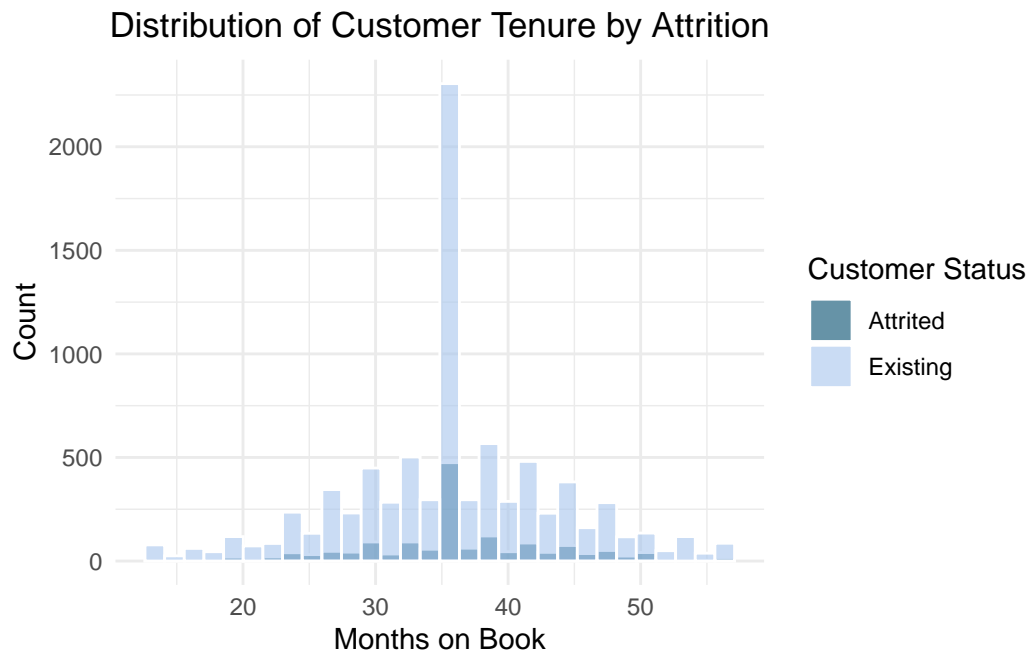
```
ggplot(raw_data, aes(x = attrition_flag, y = total_relationship_count, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Number of Products Held by Customer",
    subtitle = "More product relationships may indicate higher loyalty",
    x = "Customer Status",
    y = "Total Relationship Count"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



dis

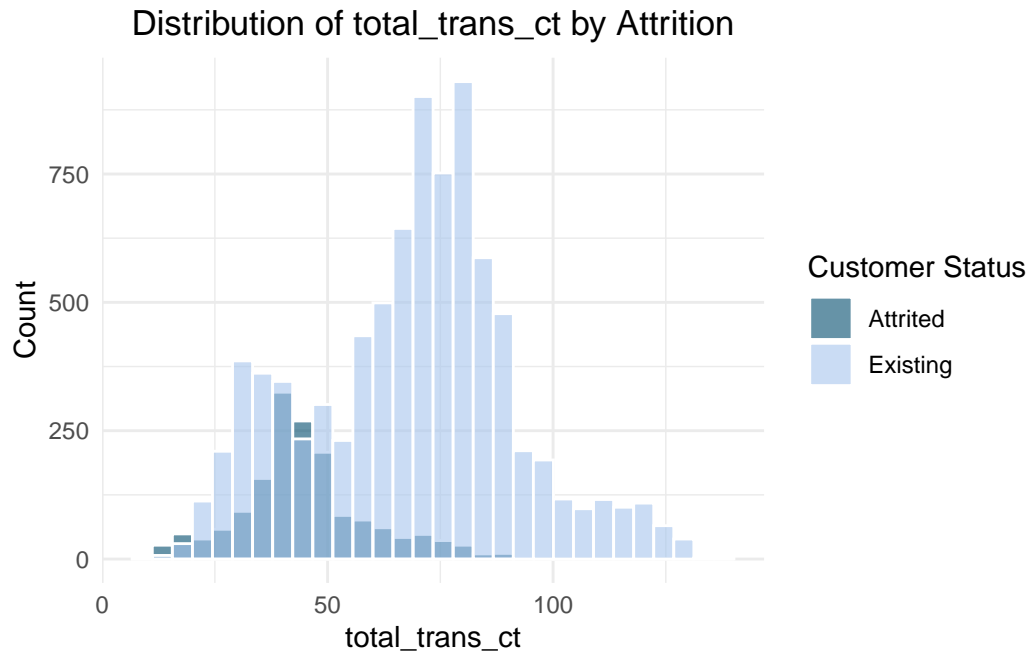
```
library(ggplot2)
library(dplyr)

ggplot(raw_data, aes(x = months_on_book, fill = attrition_flag)) +
  geom_histogram(position = "identity", alpha = 0.6, bins = 30, color = "white") +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed"), labels = c("Attrited", "Existing")) +
  labs(
    title = "Distribution of Customer Tenure by Attrition",
    x = "Months on Book",
    y = "Count",
    fill = "Customer Status"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.title = element_text(size = 11)
  )
```



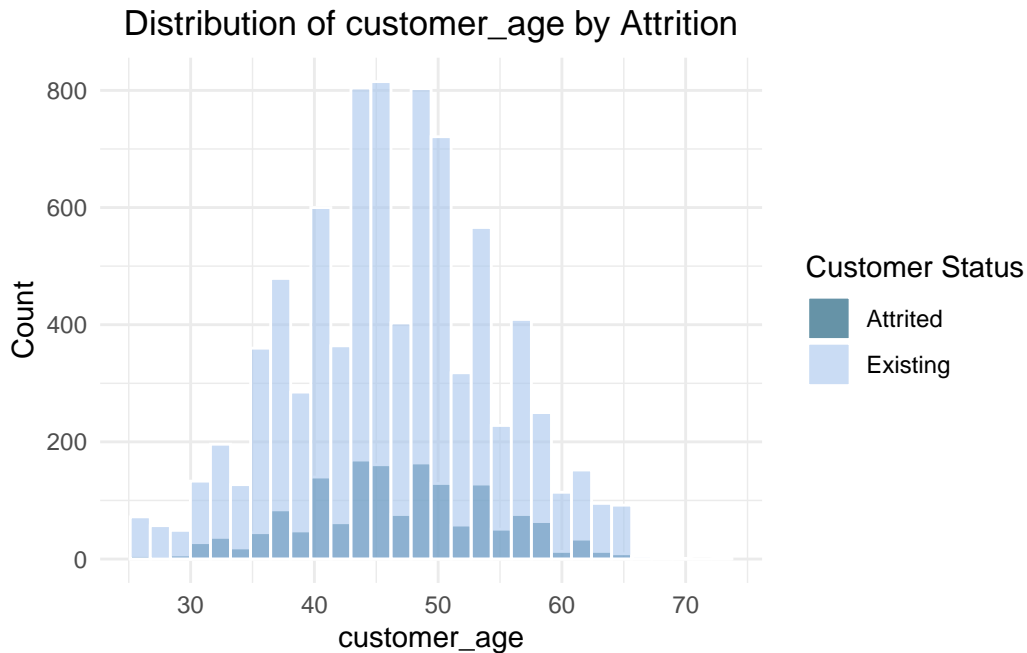
```
library(ggplot2)
library(dplyr)

ggplot(raw_data, aes(x = total_trans_ct , fill = attrition_flag)) +
  geom_histogram(position = "identity", alpha = 0.6, bins = 30, color = "white") +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed"), labels = c("Attrited", "Existing")) +
  labs(
    title = "Distribution of total_trans_ct by Attrition",
    x = "total_trans_ct",
    y = "Count",
    fill = "Customer Status"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.title = element_text(size = 11)
  )
)
```



```
library(ggplot2)
library(dplyr)

ggplot(raw_data, aes(x = customer_age , fill = attrition_flag)) +
  geom_histogram(position = "identity", alpha = 0.6, bins = 30, color = "white") +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed"), labels = c("Attrited", "Existing")) +
  labs(
    title = "Distribution of customer_age by Attrition",
    x = "customer_age",
    y = "Count",
    fill = "Customer Status"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.title = element_text(size = 11)
  )
)
```



```
library(ggplot2)
library(dplyr)

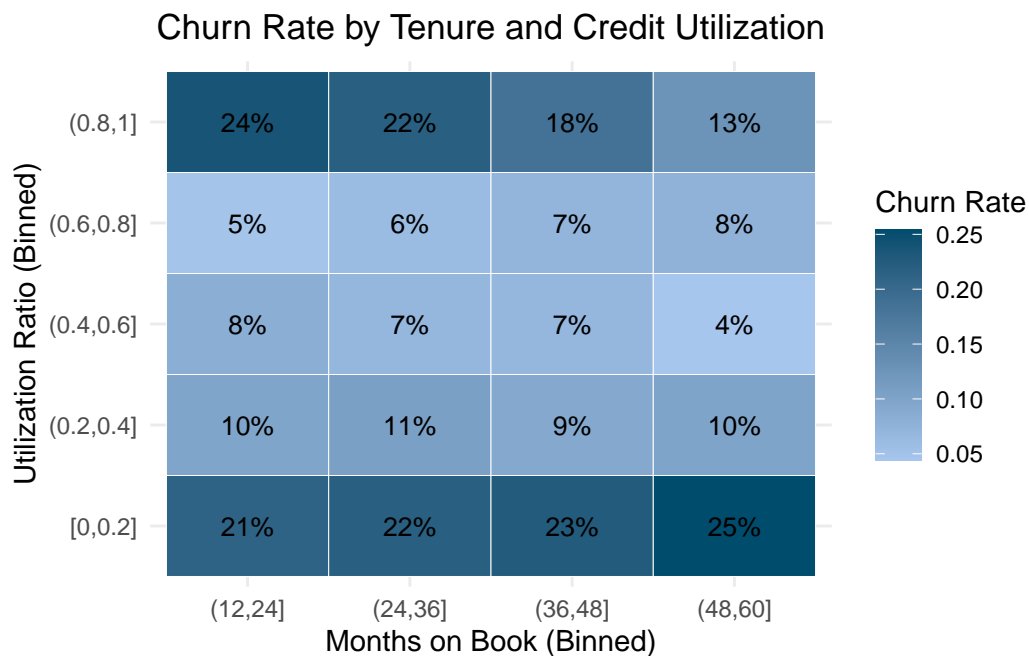
#
binned_data <- raw_data %>%
  mutate(
    tenure_bin = cut(months_on_book, breaks = c(0, 12, 24, 36, 48, 60), include.lowest = TRUE),
    util_bin = cut(avg_utilization_ratio, breaks = c(0, 0.2, 0.4, 0.6, 0.8, 1), include.lowest = TRUE)
  ) %>%
  group_by(tenure_bin, util_bin) %>%
  summarise(
    churn_rate = mean(attrition_flag == "Attrited Customer"),
    n = n()
  ) %>%
  ungroup()
```

`summarise()` has grouped output by 'tenure_bin'. You can override using the `groups` argument.

```
#
ggplot(binned_data, aes(x = tenure_bin, y = util_bin, fill = churn_rate)) +
  geom_tile(color = "white") +
```



```
geom_text(aes(label = scales::percent(churn_rate, accuracy = 1)), size = 3.5) +
scale_fill_gradient(low = "#a7c6ed", high = "#004c6d", name = "Churn Rate") +
labs(
  title = "Churn Rate by Tenure and Credit Utilization",
  x = "Months on Book (Binned)",
  y = "Utilization Ratio (Binned)"
) +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5))
```



Customer Activity Level

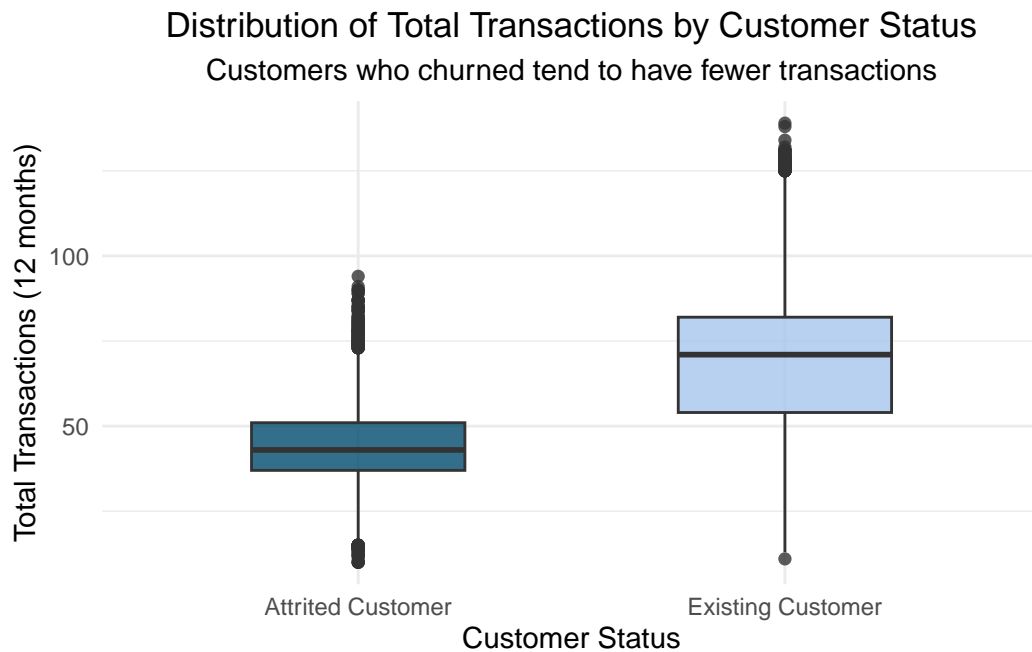
total_trans_ct (?)

```
ggplot(raw_data, aes(x = attrition_flag, y = total_trans_ct, fill = attrition_flag)) +
geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
labs(
  title = "Distribution of Total Transactions by Customer Status",
  subtitle = "Customers who churned tend to have fewer transactions",
  x = "Customer Status",
```

```

y = "Total Transactions (12 months)"
) +
theme_minimal() +
theme(legend.position = "none",
      plot.title = element_text(hjust = 0.5),
      plot.subtitle = element_text(hjust = 0.5))

```



```

library(ggplot2)
library(dplyr)

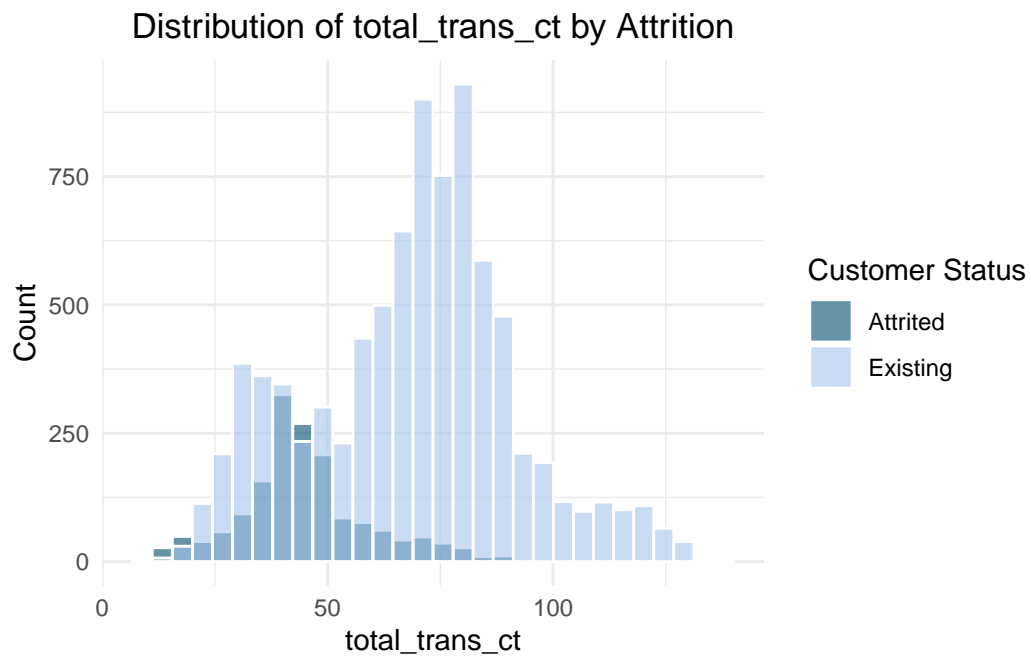
ggplot(raw_data, aes(x = total_trans_ct , fill = attrition_flag)) +
  geom_histogram(position = "identity", alpha = 0.6, bins = 30, color = "white") +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed"), labels = c("Attrited", "Existing")) +
  labs(
    title = "Distribution of total_trans_ct by Attrition",
    x = "total_trans_ct",
    y = "Count",
    fill = "Customer Status"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),

```

```

legend.title = element_text(size = 11)
)

```



40 75

30-60

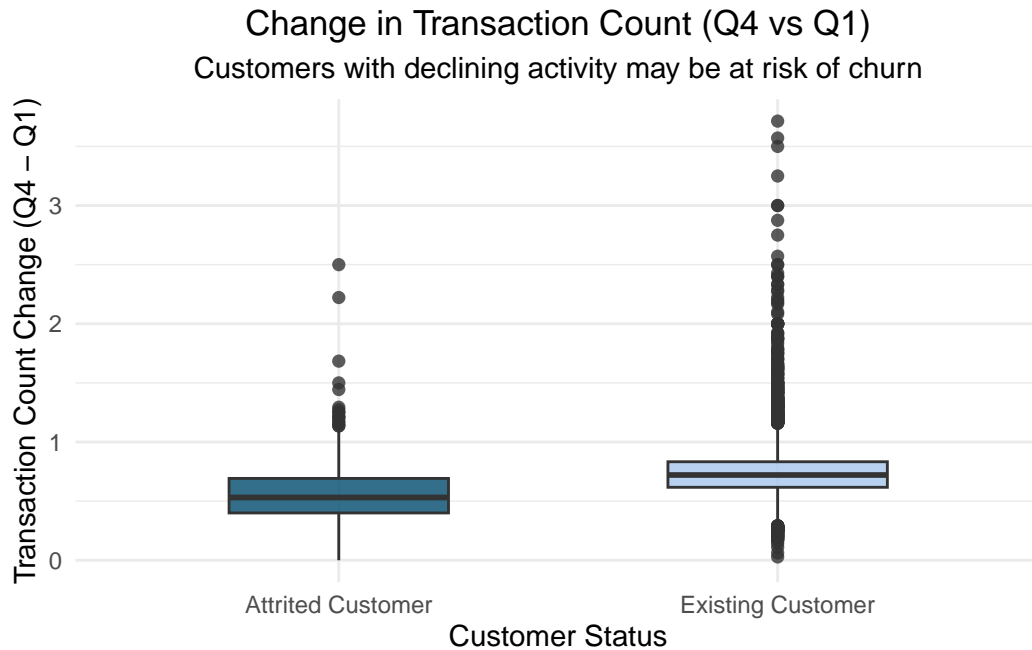
100

12 40

total_trans_ct

```
total_ct_chng_q4_q1 ( 4 1 )
```

```
ggplot(raw_data, aes(x = attrition_flag, y = total_ct_chng_q4_q1, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Change in Transaction Count (Q4 vs Q1)",
    subtitle = "Customers with declining activity may be at risk of churn",
    x = "Customer Status",
    y = "Transaction Count Change (Q4 - Q1)"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

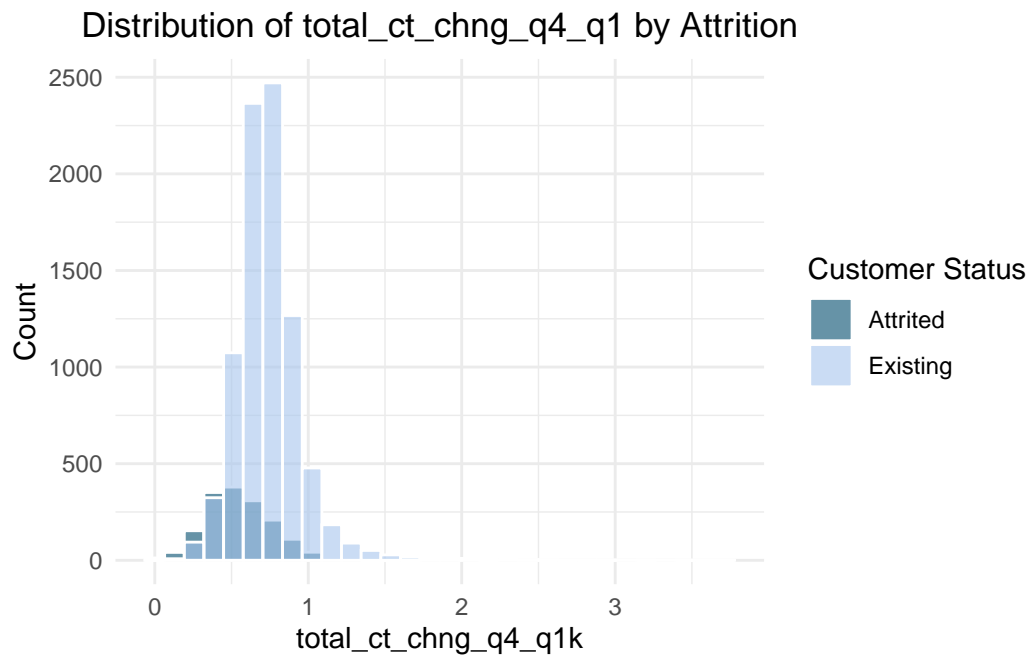


```
ggplot(raw_data, aes(x = total_ct_chng_q4_q1, fill = attrition_flag)) +
  geom_histogram(position = "identity", alpha = 0.6, bins = 30, color = "white") +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed"), labels = c("Attrited", "Existing")) +
  labs(
    title = "Distribution of total_ct_chng_q4_q1 by Attrition",
    x = "total_ct_chng_q4_q1k",
```

```

    y = "Count",
    fill = "Customer Status"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.title = element_text(size = 11)
  )

```



total_ct_chng_q4_q1 attrited existing

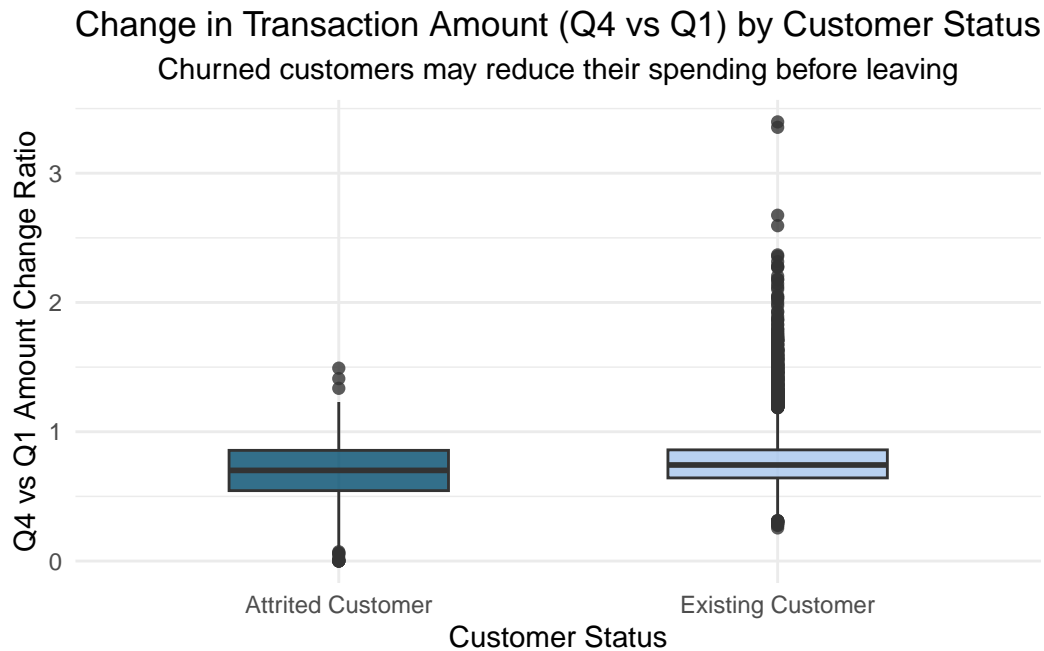
total_amt_chng_q4_q1

```

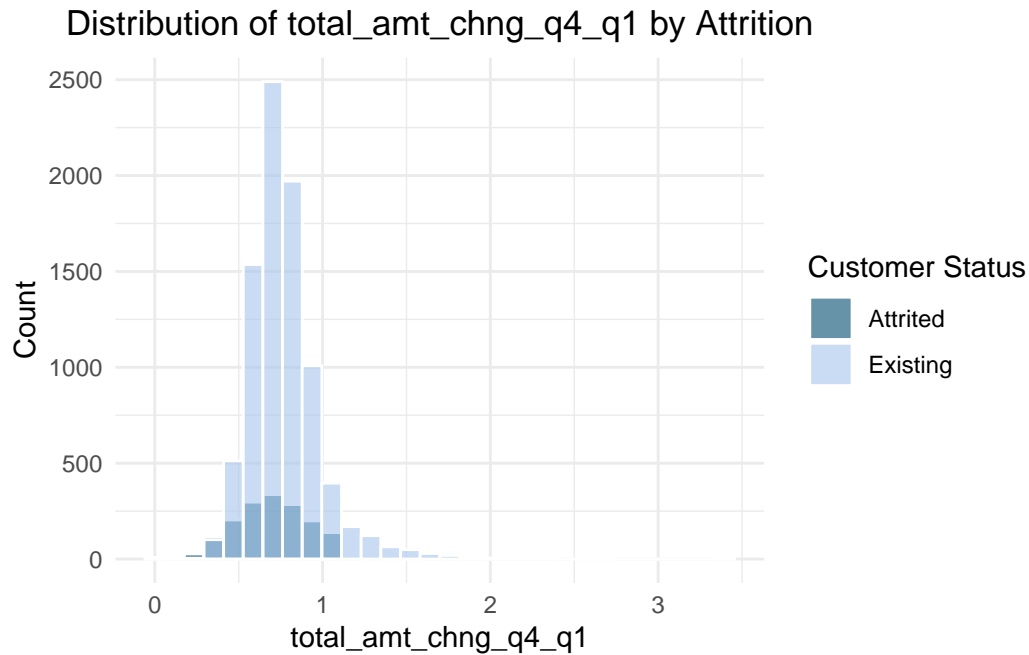
ggplot(raw_data, aes(x = attrition_flag, y = total_amt_chng_q4_q1, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "Change in Transaction Amount (Q4 vs Q1) by Customer Status",
    subtitle = "Churned customers may reduce their spending before leaving",
    x = "Customer Status",
    y = "Q4 vs Q1 Amount Change Ratio"
  ) +

```

```
theme_minimal() +
theme(legend.position = "none",
      plot.title = element_text(hjust = 0.5),
      plot.subtitle = element_text(hjust = 0.5))
```



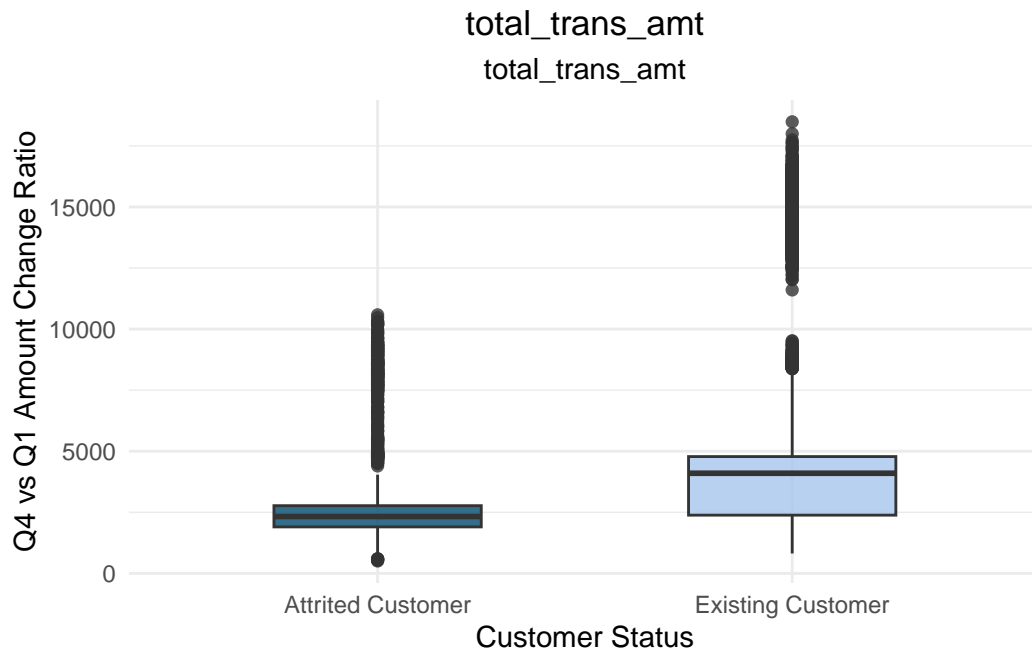
```
ggplot(raw_data, aes(x = total_amt_chng_q4_q1, fill = attrition_flag)) +
  geom_histogram(position = "identity", alpha = 0.6, bins = 30, color = "white") +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed"), labels = c("Attrited", "Existing")) +
  labs(
    title = "Distribution of total_amt_chng_q4_q1 by Attrition",
    x = "total_amt_chng_q4_q1",
    y = "Count",
    fill = "Customer Status"
  ) +
theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5),
  legend.title = element_text(size = 11)
)
```



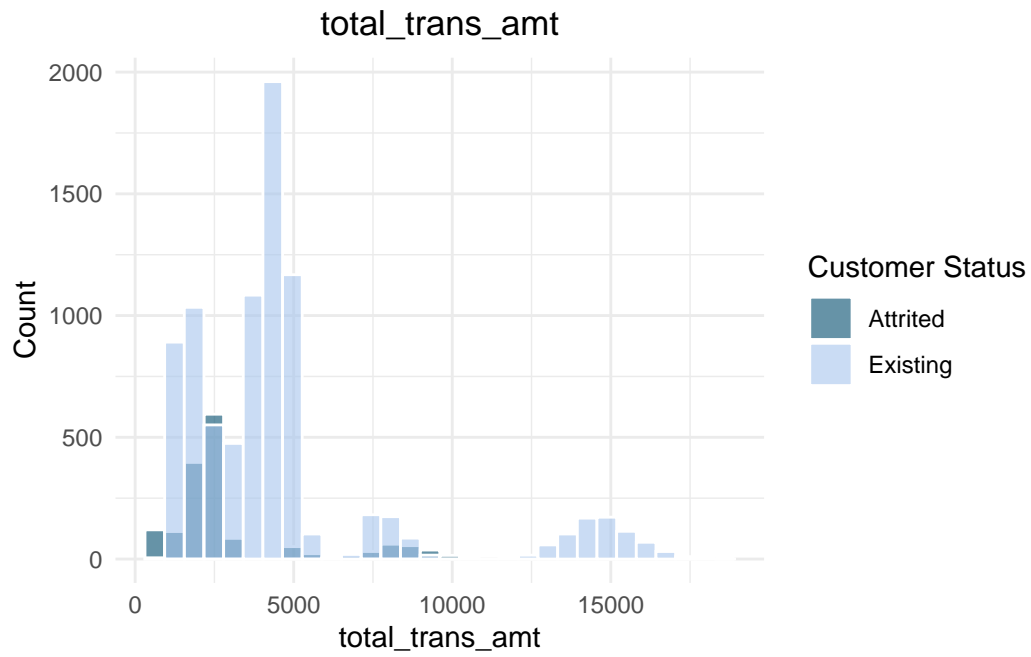
total_amt_chng_q4_q1 attrited existing

total_trans_amt

```
ggplot(raw_data, aes(x = attrition_flag, y = total_trans_amt, fill = attrition_flag)) +
  geom_boxplot(width = 0.5, alpha = 0.8, outlier.shape = 16, outlier.size = 2) +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed")) +
  labs(
    title = "total_trans_amt",
    subtitle = "total_trans_amt",
    x = "Customer Status",
    y = "Q4 vs Q1 Amount Change Ratio"
  ) +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



```
ggplot(raw_data, aes(x = total_trans_amt, fill = attrition_flag)) +
  geom_histogram(position = "identity", alpha = 0.6, bins = 30, color = "white") +
  scale_fill_manual(values = c("#004c6d", "#a7c6ed"), labels = c("Attrited", "Existing")) +
  labs(
    title = "total_trans_amt",
    x = "total_trans_amt",
    y = "Count",
    fill = "Customer Status"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.title = element_text(size = 11)
  )
```

total_trans_amt

Customer Activity Level

total_trans_amt

total_trans_ct

