



GROUP ASSIGNMENT

Customer Analysis

PREPARED BY:

Navy Him

Sreylin Loun

Vichnea Phal

Chanmoni Sopheak Yeuon

Chanbormey Rous

Cherya Krorn

PREPARED FOR:

Mr. Sokmeng Soeun

DATE OF SUBMISSION:

02 July 2025

Table of Contents

Abstract.....	3
I. Company Background and Problem Statement.....	3
II. Data Description.....	3
III. Description Analysis.....	4
IV. Methodology and Analysis.....	7
1. Data Cleaning.....	7
2. Chi-square Test.....	7
❖ Internet Service vs. Churn:.....	7
❖ Phone Service vs. Churn:.....	8
❖ Contract Type vs. Churn:.....	8
3. Logistic Regression.....	9
VI. Findings and Insights.....	11
VII. Recommendations.....	11
Reference and Appendix.....	13

Abstract

This analysis examines customer churn and finds that longer contracts reduce churn, while fiber optic internet users have the highest churn rate. Phone service does not affect churn. Higher monthly charges increase the chance of customers leaving. To reduce churn, the company should improve fiber optic service, promote longer contracts, and review pricing strategies. Due to the customers being no longer interested in our services we decided to survey them and use that information to analyse what and why they use our services in a short period of time. Furthermore, there are three methods that we use in this process of analysing including data description (pie chart, bar chart), methodology and analysis, finding and insight, and solution.

I. Company Background and Problem Statement

Our company is a service provider that focuses on customer retention programs. The purpose of doing this analysis is to achieve the customers satisfaction of using our services. The point is to

II. Data Description

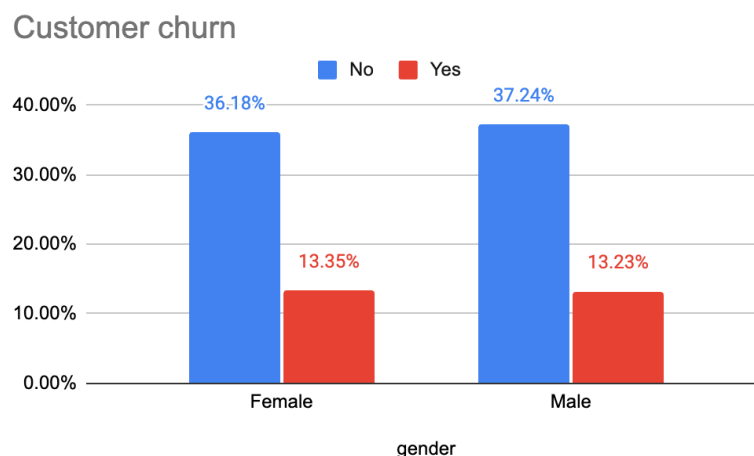
Dataset Source: Kaggle (Telco Customer Churn dataset)

- ❖ Number of Observations: 7043 rows (customers)
- ❖ Key Variables:
 - Customers who left within the last month – the column is called Churn
 - Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
 - Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
 - Demographic info about customers – gender, age range, and if they have partners and dependents
 - CustomerID: A unique ID that identifies each customer.
 - Number of Referrals: Indicates the number of referrals to date that the customer has made.
 - Tenure in Months: Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above.
 - Phone Service: Indicates if the customer subscribes to home phone service with the company: Yes, No
 - Multiple Lines: Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No
 - Internet Service: Indicates if the customer subscribes to Internet service with the company: No, DSL, Fiber Optic, Cable.

- Online Security: Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No
- Online Backup: Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No
- Device Protection Plan: Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No
- Premium Tech Support: Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No
- Streaming TV: Indicates if the customer uses their Internet service to stream television programming from a third party provider: Yes, No. The company does not charge an additional fee for this service.
- Streaming Movies: Indicates if the customer uses their Internet service to stream movies from a third party provider: Yes, No. The company does not charge an additional fee for this service.
- Contract: Indicates the customer's current contract type: Month-to-Month, One Year, Two Year.
- Paperless Billing: Indicates if the customer has chosen paperless billing: Yes, No
- Payment Method: Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check
- Monthly Charge: Indicates the customer's current total monthly charge for all their services from the company.
- Total Charges: Indicates the customer's total charges, calculated to the end of the quarter specified above.

III. Description Analysis

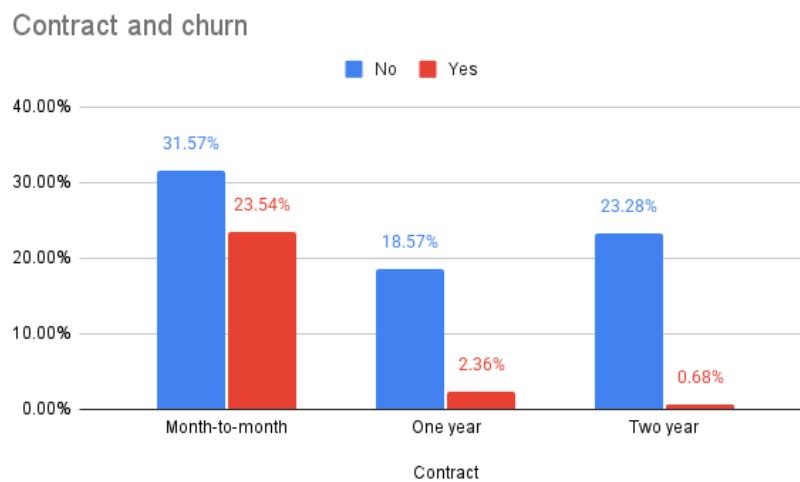
Question1: How many percent of customers continue using and stop using our services (including male and female)?



According to the chart, we observe that:

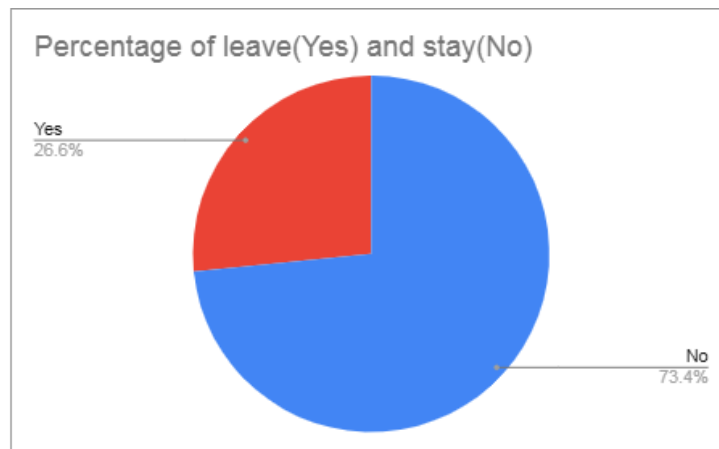
- The number of male customers is higher than female customers. So the grand total of male is (50.47%) and females (49.53%).
- Female customers who are still using our services are (36.18%) lower than male customers (37.24%) among grand total.
- Female customers who stop using our services are (13.35%) greater than male customers (13.23%) among grand total.
- Customers who are still using our services including male and female total is (73.42%).
- Customers who stop using our services including male and female is (26.58%).

Question2: Does contract type (month-to-month, one-year, two-year) affect churn rates?



The churn data presented in the chart provides valuable insights into customer behavior based on the type of internet service contract. Among customers with a **month-to-month contract**, 31.57% continue using the internet service, while a significant 23.54% have discontinued, indicating a relatively high churn rate. This suggests that customers on flexible, short-term plans are more likely to leave. In contrast, customers with a **one-year contract** show much greater retention, with 18.57% still using the service and only 2.36% having left. This clearly highlights that customers who commit to a longer-term contract are less likely to churn. The trend becomes even more evident with **two-year contracts**, where 23.28% remain active users and only a very small portion—just 0.68%—have stopped using the service. This declining trend in churn rates, from month-to-month to two-year contracts, demonstrates a strong correlation between contract length and customer loyalty. The longer customers are locked into a contract, the less likely they are to cancel the service. Therefore, this data supports the conclusion that encouraging customers to choose longer-term contracts can significantly reduce churn. It reflects positively on the company's ability to retain long-term customers, implying satisfaction with service over time and reinforcing the value of long-term commitment strategies for business growth.

Question3: Does having phone service, internet service, or both, make customers more or less likely to quit, and which specific one makes them leave the most?



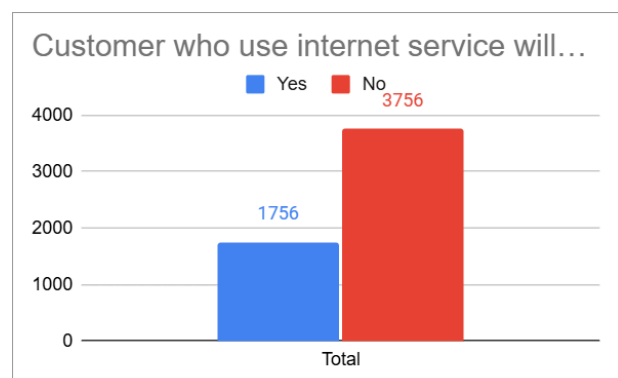
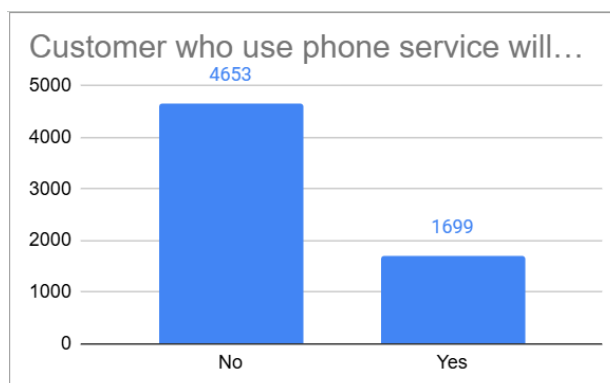
Observation:

The pie chart visually represents the overall proportion of customers who have "left" (churned) versus those who have "stayed" (not churned) within the entire customer base.

- The **blue section**, labeled "No" and representing **73.4%**, shows the percentage of customers who *did not churn* (i.e., they stayed with the company).
- The **red section**, labeled "Yes" and representing **26.6%**, shows the percentage of customers who *did churn* (i.e., they left the company).

This chart *only* provides the **overall churn rate** for the entire company's customer base. It **does not** show any specific relationships between the *types of services* (Phone Service or Internet Service) and customer churn.

Question4: Which type of service (Phone Service , DSL Internet, or Fiber Optic Internet) makes customers most likely to leave, and which one is the riskiest for companies?



Observation:

Phone service doesn't seem to affect whether customers leave—people with phone service churn at pretty much the same rate as those without it. However, internet service is a big deal when it comes to customers leaving. The biggest problem is with **Fiber Optic internet**. A

huge number of Fiber Optic customers (almost 42%) quit, which is much higher than those with regular **DSL internet** (only 19% churn) or no internet at all (just 7.4% churn). So, while the phone service is stable, there's a serious issue with the **Fiber Optic service** that's causing many customers to leave. Our company should really focus on figuring out what's wrong with its Fiber Optic offering. Maybe it's the cost, how reliable it is, whether customers feel they're getting good value, or how customer support handles it. Fixing the Fiber Optic problems should be our company's top priority to stop customers from leaving.

IV. Methodology and Analysis

1. Data Cleaning

Using R programing for cleaning and data prepared for analysis

- Converted relevant columns to factor types in R.
- No missing and duplicate values were reported in each variable

2. Chi-square Test

To determine a significant association between service features and churn we chi-square test .

State Hypothesis:

- **H₀ (Null):** Service features has no impact on churn rate.
- **H₁ (Alternative):** Service feature does impact churn rate.

❖ Internet Service vs. Churn:

By using R program, the result is given as below:

```
> # View the table
> print(internet_table)

      No  Yes
DSL    1957 459
Fiber optic 1799 1297
No      1407 113
>
> # chi-square test for Internet Service
>> chisq.test(internet_table)

Pearson's Chi-squared test with Yates' continuity correction

data:  internet_table
X-squared = 728.7, df = 2, p-value < 2.2e-16
>
```

- A chi-square test of independence shows a highly significant relationship ($p < 2.2e-16$) that means reject null hypothesis, suggesting that the type of internet service affects churn behavior.
- Customers with fiber optic service have the highest churn rate, indicating dissatisfaction or external factors such as pricing or competition.

❖ Phone Service vs. Churn:

```
# Chi-square test for Phone Service
>
> phone_table <- table(data$PhoneService, data$Churn)
> print(phone_table)

      No  Yes
No    510 170
Yes  4653 1699
> chisq.test(phone_table)

Pearson's Chi-squared test with Yates' continuity correction

data:  phone_table
X-squared = 0.87373, df = 1, p-value = 0.3499
>
```

- The p-value (0.3499) indicates no statistically significant relationship between having phone service and customer churn.
- This implies that phone service alone does not influence a customer's decision to leave.

❖ Contract Type vs. Churn:

```
> #Chi-square test for contract
> contract_table <- table(data$Contract, data$Churn)
> print(contract_table)

      No  Yes
Month-to-month 2220 1655
One year      1306 166
Two year      1637 48
> chisq.test(contract_table)

Pearson's Chi-squared test with Yates' continuity correction

data:  contract_table
X-squared = 1179.5, df = 2, p-value < 2.2e-16
>
```


- The chi-square test shows a highly significant relationship ($p < 2.2e-16$) that means reject null hypothesis.
- There is a strong relationship between contract type and churn.
- Customers on month-to-month contracts show a substantially higher churn rate than those on one-year or two-year contracts.

3. Logistic Regression

To understand how pricing affects customer churn, we applied logistic regression with MonthlyCharges as the independent variable and Churn as the dependent variable. The primary objective was to investigate whether the amount customers pay monthly is associated with their likelihood of leaving the company.

Logistic regression was chosen instead of linear regression because the target variable, *Churn*, is *binary* (Yes/No), and logistic models are well-suited for estimating probabilities of binary outcomes.

The model output in R is summarized below:

```
> summary(model_monthly)
Call:
glm(formula = Churn ~ MonthlyCharges, family = binomial, data = data)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.0619850   0.0742186  -27.78   <2e-16 ***
MonthlyCharges  0.0153680   0.0009672   15.89   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8143.4  on 7031  degrees of freedom
Residual deviance: 7873.3  on 7030  degrees of freedom
AIC: 7877.3

Number of Fisher Scoring iterations: 4
```

The logistic regression model estimates the log-odds of churn using the following equation:

$$\text{logit}(p) = \log(p/(1-p)) = -2.061985 + 0.015368 \times \text{MonthlyCharges}$$

0.074(*) 0.001(***)**

Where: p is the probability of churn.

Interpretation:

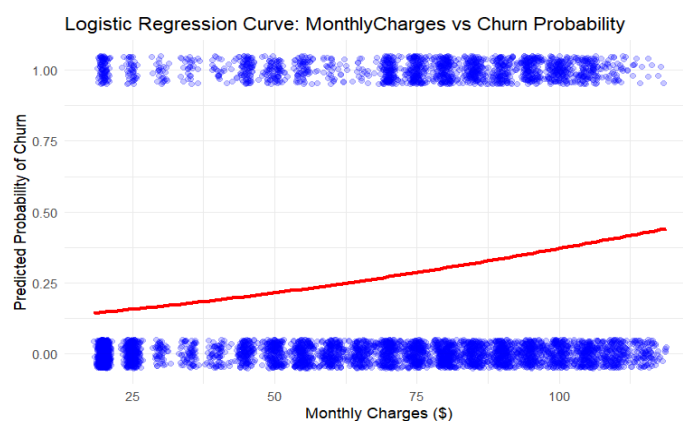
- The positive coefficient for *MonthlyCharges* (+0.01537) indicates that as the monthly charge increases, the probability of churn also increases.
- The result is statistically significant ($p < 2e-16$), confirming that pricing plays an important role in customer retention.

Example: We can use the regression coefficients in R to compute churn probabilities for specific monthly charges:

```
> # Coefficients from model
> intercept <- -2.061985
> slope <- 0.015368
>
> # Function to calculate churn probability
> determine_churn_prob <- function(monthly_charge) {
+   logit <- intercept + slope * monthly_charge
+   prob <- 1 / (1 + exp(-logit))
+   return(prob)
+ }
>
> # Example probabilities
> determine_churn_prob(50)
[1] 0.2152466
> determine_churn_prob(90)
[1] 0.3365147
```

This demonstrates that higher monthly charges are associated with a greater likelihood of customer churn. Specifically, a customer paying \$50 per month has an estimated 21.5% chance of churning, while a customer paying \$90 per month has a much higher estimated probability of 35.4%. This rising trend in churn probability with increased billing highlights the importance of evaluating perceived value relative to cost in service pricing strategies.

Visualization: A probability curve is plotted to show how churn probability increases with higher monthly charges:



V. Findings and Insights

From our analysis, we found that customers who use fiber optic internet are the most likely to leave the company. This might be because they feel the service is not good enough, or the price is too high for what they get. We also found that customers with month-to-month contracts are more likely to leave than those with one-year or two-year contracts. This shows that short contracts make it easier for people to switch to other companies. Phone service, however, doesn't seem to affect whether customers stay or leave. Lastly, we saw that customers who pay more money each month are more likely to cancel their service. This means that customers who pay more might not feel the service is worth the price. These findings can help the company understand which groups of customers are at higher risk of leaving and what might be causing them to leave.

VI. Recommendations

Based on the analysis, several strategies are recommended to reduce churn and improve customer satisfaction. First, retention efforts should prioritize fiber optic users, who show the highest likelihood of leaving. The company should consider offering targeted support such as discount upgrade offers, improved technical assistance, or loyalty rewards to retain these high-risk customers. Second, encouraging customers to switch from month-to-month plans to longer-term contracts can increase customer stability. Financial incentives—such as reduced fees, bundled services, or special discounts—may motivate customers to commit for a longer period. Third, the company should evaluate the quality of its internet services, especially for fiber optic users. Regular performance checks and proactive improvements could address customer dissatisfaction before it leads to churn. Finally, the company should review its pricing strategy, particularly for customers with higher monthly charges. These customers are more likely to churn, which may indicate they do not perceive sufficient value for the cost. To address this, the company should conduct a competitive pricing analysis to see how its rates compare to those of other providers. Offering flexible, tiered, or customizable pricing plans could help align services with customer expectations while remaining competitive in the market.

Reference and Appendix

- Dataset: <https://www.kaggle.com/blastchar/telco-customer-churn>
- Idea of Logistic Regression:
<https://www.geeksforgeeks.org/machine-learning/understanding-logistic-regression/>
- R packages used: *readxl*, *tidyverse*, *ggplot2*,
- R code for cleaning data:

```
library(tidyverse)
data <-
read_csv("D:/Paragon_Class/Semester4/BUS212/WA_Fn-UseC_-Telco-Customer-Churn.csv")

data <- na.omit(data)
data <- data %>%
  distinct()

write_csv(data,
"D:/Paragon_Class/Semester4/BUS212/WA_Fn-UseC_-Telco-Customer-Churn-clean.csv")
```

- R code for doing Chi-square test:

```
library(readxl)
library(tidyverse)

data <-
read_excel("D:/Paragon_Class/Semester4/BUS212/WA_Fn-UseC_-Telco-Customer-Churn.xlsx")

# Convert Churn and service columns to factors
data$Churn <- factor(data$Churn)
data$InternetService <- factor(data$InternetService)
data$PhoneService <- factor(data$PhoneService)
data$Contract <- factor(data$Contract)

#For Internet Service

# Create contingency table
internet_table <- table(data$InternetService, data$Churn)

# View the table
print(internet_table)

# Perform Chi-square test
chisq.test(internet_table)

# Chi-square test for Phone Service

phone_table <- table(data$PhoneService, data$Churn)
print(phone_table)
```

```
chisq.test(phone_table)

#Chi-square test for contract
contract_table <- table(data$Contract, data$Churn)
print(contract_table)
chisq.test(contract_table)
```

- R Code for doing Logistic regression:

```
library(readxl)
library(ggplot2)

# Step 1: Load data
data <-
read_excel("D:/Paragon_Class/Semester4/BUS212/WA_Fn-UseC_-Telco-Customer-Churn.xlsx")

data$Churn <- factor(data$Churn, levels = c("No", "Yes"))

model_monthly <- glm(Churn ~ MonthlyCharges, data = data, family = binomial)

summary(model_monthly)

# Create a sequence of values for MonthlyCharges
monthly_seq <- data.frame(MonthlyCharges = seq(min(data$MonthlyCharges),
                                                max(data$MonthlyCharges),
                                                length.out = 100))

# Predict probabilities using the model
monthly_seq$predicted_prob <- predict(model_monthly, newdata = monthly_seq, type =
"response")

# Plot original data (jittered for better visibility) and the curve
ggplot(data, aes(x = MonthlyCharges, y = as.numeric(Churn == "Yes"))) +
  geom_jitter(height = 0.05, alpha = 0.2, color = "blue") +
  geom_line(data = monthly_seq, aes(x = MonthlyCharges, y = predicted_prob), color
= "red", size = 1.2) +
  labs(title = "Logistic Regression Curve: MonthlyCharges vs Churn Probability",
       x = "Monthly Charges ($)",
       y = "Predicted Probability of Churn") +
  theme_minimal()

> # Coefficients from model
> intercept <- -2.061985
> slope <- 0.015368
>
> # Function to calculate churn probability
```

```
> determine_churn_prob <- function(monthly_charge) {  
+   logit <- intercept + slope * monthly_charge  
+   prob <- 1 / (1 + exp(-logit))  
+   return(prob)  
+ }  
>  
> # Example probabilities  
> determine_churn_prob(50)  
[1] 0.2152466  
> determine_churn_prob(90)  
[1] 0.3365147
```