Maduabughichi Achilefu

Okeke Onyedikachukwu,

Babatunde Odumuyiwa,

Eeshaan Ali Syed

Report on Churn Data Analysis

Executive Summary

This report presents a comprehensive analysis of customer churn data from a telecommunications company. Key insights reveal the relationship between various customer metrics and churn. The findings are crucial for strategic decision-making to enhance customer retention.

Introduction

The churn_data dataset includes metrics such as churn status, account duration, contract renewal, data plan subscription, and monthly charges. The objective is to identify factors influencing customer churn.

Summary of Initial Exploratory Data Analysis (EDA)

Our initial EDA revealed a churn rate of approximately 14.5%, with the average account duration of about 101 weeks. Notably, 90.4% of customers have renewed their contracts, and around 27.6% have data plans. The average monthly charge was found to be $56.27.

Brief Overview:

The dataset comprises 6514 entries with 12 columns. Key variables include Churn status, Account Weeks, Contract Renewal, Data Plan subscription, Data Usage, Customer Service Calls, Daily Usage metrics, Monthly Charges, Overage Fees, and Roaming Minutes. Key findings:

- ✓ Churn Rate: Approximately 14.5% of customers have churned.
- ✓ Account Duration: Average account duration is about 101 weeks.
- ✓ Contract Renewal: Majority of customers (90.4%) have renewed their contracts.
- ✓ Data Plan Usage: Around 27.6% of customers have data plans.
- ✓ Customer Service Calls: Customers make an average of 1.56 service calls.
- ✓ Monthly Charges: Average monthly charge is $56.27, with a wide range from $14 to $111.30.
- ✓ Data Usage: Average data usage is 0.82 GB, indicating a significant portion of customers use data services.
- ✓ Key Metrics Examined:
- ✓ Churn Status
- ✓ Account Duration (AccountWeeks)
- ✓ Contract Renewal
- ✓ Data Plan Subscription
- ✓ Data Usage
- ✓ Customer Service Calls

✓ Monthly Charges
✓ Overage Fees
✓ Roaming Minutes

Formulated Questions and Hypotheses

We want to find out why 14.5% of customers are churning out.

Question 1: Is there a correlation between monthly charges and customer service calls?

H0: A significant correlation exists.

H1: No correlation exists.

Question 2: Does the data usage affect the monthly charge?

H_0: Data usage affects the monthly charge.

H_1: Data usage does not affect the monthly charge.

Methodology

Statistical methods included correlation tests and regression. Independent variable was monthly charge while dependent variables were customer service calls and data usage, respectively.

Results and Analysis

Monthly Charges vs. Customer Service Calls

[1] -0.02689513

```
        Pearson's product-moment correlation

data:  churn_data$CustServCalls and churn_data$MonthlyCharge
t = -2.1711, df = 6512, p-value = 0.02996
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.051146766 -0.002611796
sample estimates:
        cor
-0.02689513
```

Weak negative correlation found (r = -0.0269).

P-value marginally significant (p = 0.03).

Data Usage vs. Monthly Charge

[1] 0.78177

```
        Pearson's product-moment correlation

data:  churn_data$DataUsage and churn_data$MonthlyCharge
t = 101.17, df = 6512, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7721444 0.7910369
sample estimates:
    cor
0.78177
```

Significant correlation (r = 0.78177).
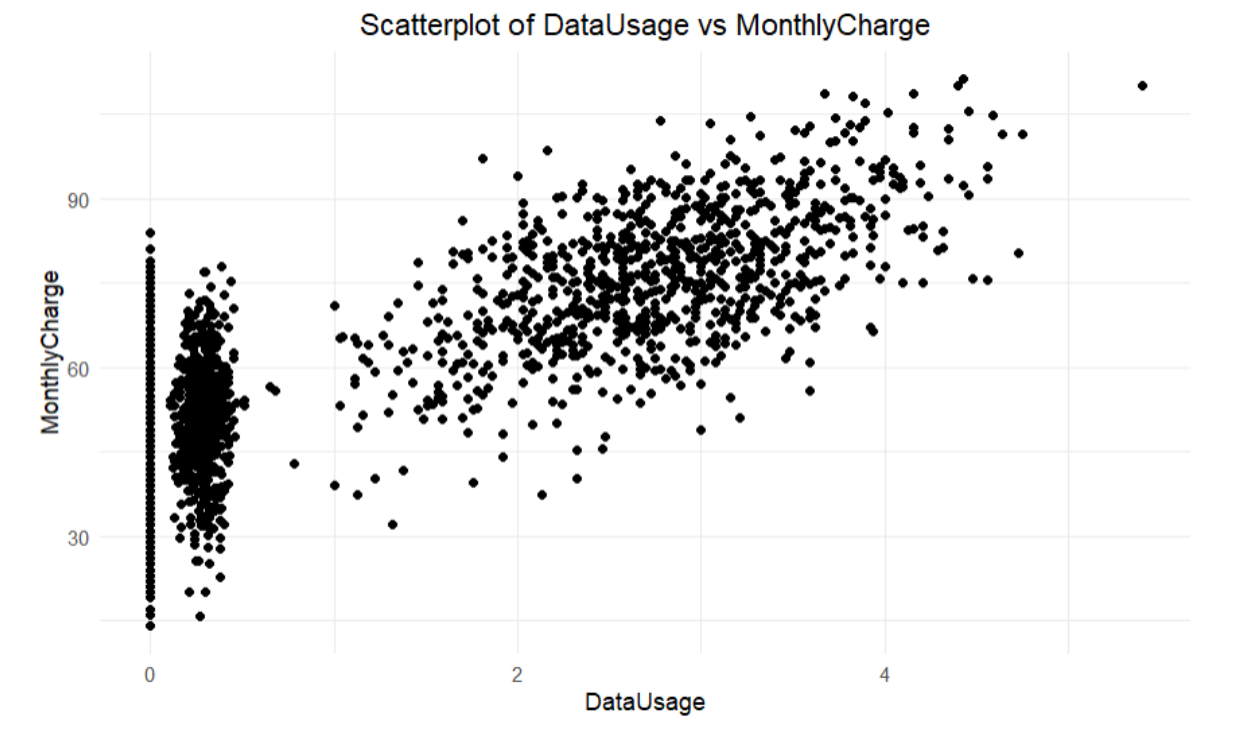P-value significant (p = 2.2e-16).

Conclusion

The analysis suggests that monthly charges have a minimal impact on customer service calls but have a significant impact on data usage.

Appendix

Scatterplots of analyzed relationships.

Scatterplot of Monthly Charges vs Customer Service Calls

Scatterplot of DataUsage vs MonthlyCharge
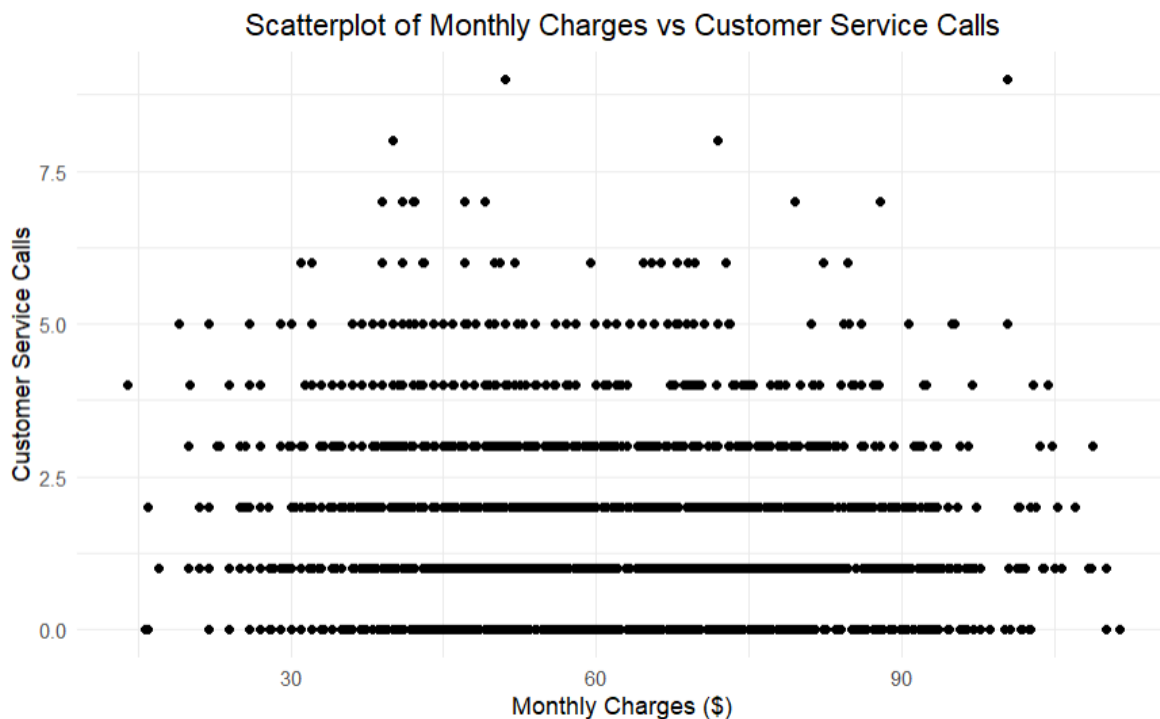
Summary statistics of the dataset.

R code used for analysis.

Data Summary

```
     Churn            AccountWeeks  ContractRenewal      DataPlan           DataUsage        CustServCalls        DayMins
 Min.   :0.0000   Min.   :  1    Min.   :0.0000    Min.   :0.000    Min.   :0.000    Min.   :0.000    Min.   :  0.0
 1st Qu.:0.0000   1st Qu.: 74    1st Qu.:1.0000    1st Qu.:0.000    1st Qu.:0.000    1st Qu.:1.000    1st Qu.:143.7
 Median :0.0000   Median :101    Median :1.0000    Median :0.000    Median :0.000    Median :1.000    Median :179.3
 Mean   :0.1448   Mean   :101    Mean   :0.9044    Mean   :0.276    Mean   :0.816    Mean   :1.562    Mean   :179.7
 3rd Qu.:0.0000   3rd Qu.:127    3rd Qu.:1.0000    3rd Qu.:1.000    3rd Qu.:1.780    3rd Qu.:2.000    3rd Qu.:216.2
 Max.   :1.0000   Max.   :243    Max.   :1.0000    Max.   :1.000    Max.   :5.400    Max.   :9.000    Max.   :350.8
    DayCalls        MonthlyCharge      OverageFee        RoamMins       DataUsagecat
 Min.   :  0.0    Min.   : 14.00    Min.   : 0.00    Min.   : 0.00    Length:6514
 1st Qu.: 87.0    1st Qu.: 45.00    1st Qu.: 8.33    1st Qu.: 8.50    Class :character
 Median :101.0    Median : 53.40    Median :10.05    Median :10.30    Mode  :character
 Mean   :100.4    Mean   : 56.27    Mean   :10.05    Mean   :10.24
 3rd Qu.:114.0    3rd Qu.: 66.00    3rd Qu.:11.77    3rd Qu.:12.10
 Max.   :165.0    Max.   :111.30    Max.   :18.19    Max.   :20.00
```

Scatterplots

```{r}
# Load necessary libraries
library(ggplot2)
library(stats)

# Read the dataset (assuming it is in the current working directory)
churn_data <- read.csv("churn_data.csv")

# Create a scatterplot for Monthly Charges vs Customer Service Calls
ggplot(churn_data, aes(x = MonthlyCharge, y = CustServCalls)) +
  geom_point() +  # Add points
  labs(title = "Scatterplot of Monthly Charges vs Customer Service Calls",
       x = "Monthly Charges ($)",
       y = "Customer Service Calls") +
  theme_minimal() +  # Use a minimal theme
  theme(plot.title = element_text(hjust = 0.5))  # Center the title

# Conduct a correlation test
correlation_test <- cor.test(churn_data$MonthlyCharge, churn_data$CustServCalls)

# Print the results
print(correlation_test)
```

```{r}
# Read the dataset (assuming it is in the current working directory)
churn_data <- read.csv("churn_data.csv")

# Create a scatterplot for Monthly Charges vs Customer Service Calls
ggplot(churn_data, aes(x = DataUsage, y = MonthlyCharge)) +
  geom_point() +  # Add points
  labs(title = "Scatterplot of DataUsage vs MonthlyCharge",
       x = "DataUsage",
       y = "MonthlyCharge") +
  theme_minimal() +  # Use a minimal theme
  theme(plot.title = element_text(hjust = 0.5))  # Center the title

# Conduct a correlation test
correlation_test <- cor.test(churn_data$DataUsage, churn_data$MonthlyCharge)

# Print the results
print(correlation_test)
```

Regression

```{r}
# Calculate the correlation coefficient between 'CustServCalls' and 'MonthlyCharge' in the 'churn_data' dataset
cor(churn_data$CustServCalls, churn_data$MonthlyCharge)

# Perform a correlation test between 'CustServCalls' and 'MonthlyCharge' in the 'churn_data' dataset.
# This test provides not only the correlation coefficient but also its statistical significance.
cor.test(churn_data$CustServCalls, churn_data$MonthlyCharge)
```

```{r}
# Calculate the correlation coefficient between 'DataUsage' and 'MonthlyCharge' in the 'churn_data' dataset
cor(churn_data$DataUsage, churn_data$MonthlyCharge)

# Perform a correlation test between 'DataUsage' and 'MonthlyCharge' in the 'churn_data' dataset.
# This test provides the correlation coefficient, along with its statistical significance, including p-value and confidence
interval.
cor.test(churn_data$DataUsage, churn_data$MonthlyCharge)

```

References

1. R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

2. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.).
3. Customer Churn. (n.d.). Www.kaggle.com.
https://www.kaggle.com/datasets/barun2104/telecom-churn/data