**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?

H0: Data usage affects the monthly charge.

H1: 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.

A graph of a number of black dots

Description automatically generated with medium confidence

A diagram of a scatter plot

Description automatically generated

Summary statistics of the dataset.

R code used for analysis.

Data Summary

A screenshot of a computer screen

Description automatically generated

Scatterplots

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

**Regression**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**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](http://Www.kaggle.com). https://www.kaggle.com/datasets/barun2104/telecom-churn/data