April 2, 2025

Shindelman, Robin

CS 406-Projects

Oregon State University

Churn analysis

IBM Telco Dataset

Table of Contents

[What is Churn? 2](#_Toc196384814)

[Problem Background 2](#_Toc196384815)

[Question Statement 2](#_Toc196384816)

[Data 2](#_Toc196384817)

[Warehousing 3](#_Toc196384818)

[Exploratory Data Analysis – EDA 3](#_Toc196384819)

[Technology 3](#_Toc196384820)

[Findings 3](#_Toc196384821)

[Inference 3](#_Toc196384822)

[Technology 3](#_Toc196384823)

[Procedure 3](#_Toc196384824)

[Discussion 3](#_Toc196384825)

[Recommendations 4](#_Toc196384826)

[References 4](#_Toc196384827)

# A group of people next to each other AI-generated content may be incorrect.What is Churn?

Image Source: https://freetools.textmagic.com/churn-rate-calculator

A common business problem is the question of customer retention. How do we, as an organization, keep customers from leaving to find another product or service? Viable answers to this issue can be difficult to come by, require access to large amounts of customer data, and an understanding of the market or domain.

Given these conditions are met, a data professional can mine this information for insights which could then be transformed into valuable drivers of business strategy. While new customers can always be found via marketing campaigns and promotions, these strategies are expensive and time consuming. Current customers cost far less to maintain and can be depended on to continue bringing revenue to your business. Therefore, a valuable use of company resources is to conduct churn analysis.

# Problem Background

## Question Statement

The Telco Customer Churn dataset represents customer data from a fictional telecommunications company. In Q3, Telco provided home and internet services to over 7000 customers in California. Information is provided on each customer’s location, demographic, and services received. Most importantly, whether or not the customer signed up, left, or continued using Telco during the quarter is included in the data [1].

Our goal with this dataset is to first identify any interesting customer trends related to churn rate. Following that, we’ll respond to this insight with recommendations for strategies to increase retention in key areas.

## Data

Detailed information from IBM on the Telco Customer Churn dataset can be found [here](https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113). This dataset is ideal for a student conducting a formalized churn analysis. A zip file can be downloaded containing five spreadsheets which act as tables. Since SQL is an important aspect of this demonstration, these spreadsheets were uploaded to Google Cloud Platform’s BigQuery as tables in a single relational database.

# Warehousing

Data warehousing is an essential component in the analysis of enterprise data. In the typical process, data will undergo a process of Extract-Transform-Load (ETL) or Extract-Load-Transform (ELT) before being usable by analysts, engineers, or leadership. In this case, the data was given a very abbreviated treatment of ELT. The data was first “extracted” from IBM’s Cognos Analytics platform as a series of disjoint Excel spreadsheets. Each sheet contained a variety of information to build a complete picture of each customer profile and their churn status.

These spreadsheets then needed to be converted over to CSV files in order to be “loaded” to the chosen data warehousing solution, Google BigQuery. While there are a vast plethora of very strong choices for housing the churn data, Google Cloud Platform offers a robust integration with a variety of services such as Looker, which will be used during the reporting phase of this analysis. BigQuery is a fully administrated relational database system, allowing users to focus on the data itself rather than engineering the SQL server and worrying about things like scaling resources during peak access times. Of the 5 spreadsheets downloaded and converted to CSVs, each was uploaded to BigQuery as a separate table, connected through foreign and primary keys.

Finally, the data was “transformed” through a series of cleansing operations done with SQL queries on BigQuery. Missing values were dropped and a couple of columns needed to have their data type altered, such as zip codes being changed from integers to strings. For the most part, the data provided by IBM was extremely clean and free from errors. During the EDA step in the next stage of this process, some statistical techniques will be used to discover any significant outliers in the data.

# Exploratory Data Analysis – EDA

## Technology

## Findings

# Inference

## Technology

## Procedure

## Discussion

# Recommendations

# References

1. S. Macko, “Telco Customer Churn,” IBM Business Analytics Blog, Jul. 11, 2019. [Online]. Available: <https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113>