End-to-End Churn Prediction Data Analysis Project

By Mehrdad Mansourdehghan

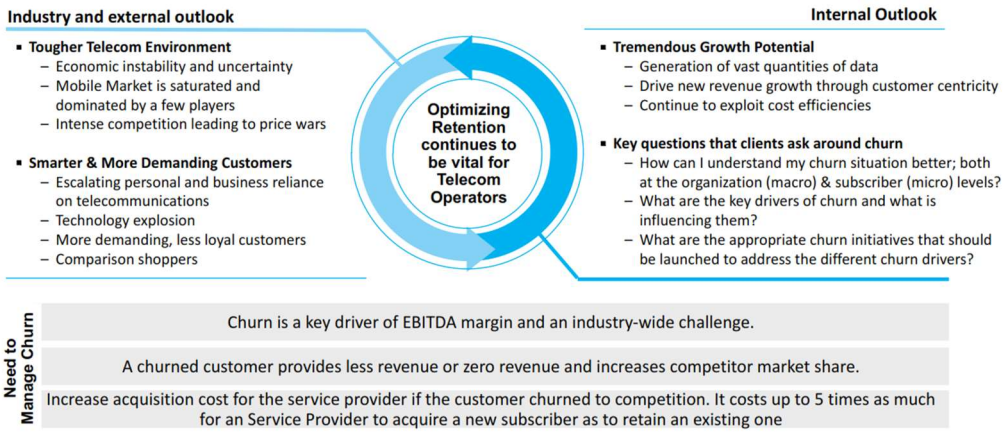
Contents

What does churn mean?

In the telecommunications industry, the term "churn" describes the frequency at which customers switch or terminate their services with a specific telecom provider and either move to a competitor or cease using the service altogether. This important metric measures customer attrition or the decline in subscribers. Additionally, if a customer changes their plan from for example $50 per month to $30 per month, they are considered a churned customer once again.

What is the problem?

Customer churn is a common occurrence across all industries. In today's competitive landscape, retaining customers and ensuring their satisfaction is crucial. Failure to do so could result in customers switching to our competitors. In the telecom industry, for example, if you are a customer of Rogers and decide to switch to another carrier like Fido, this is known as churning Rogers because Rogers has lost a customer. There are various reasons why this may happen, such as higher plan prices at Rogers compared to Fido or inadequate coverage in certain areas. Analyzing customer churn can help identify problem areas within our company or services and improve customer retention. Business concepts suggest that there are two approaches, one external and out of the operator's control, and another internal. From an external perspective, a challenging telecom environment may cause customer churn due to economic uncertainty. Additionally, intense competition among telecom carriers, with the emergence of numerous potential companies offering similar services, may further exacerbate the situation. While not all of these companies may be well-known, they could pose a threat in the future. Other factors that contribute to market share loss can be found in the image below.



Type of customer churning

There are some different customers churning that's in this section I'm going to describe all of them.

Tariff Plan Churn

Tariff plan churn refers to the situation where a customer who is currently paying for example $50 per month for a service, wishes to obtain the same service at a lower cost.

Service churn

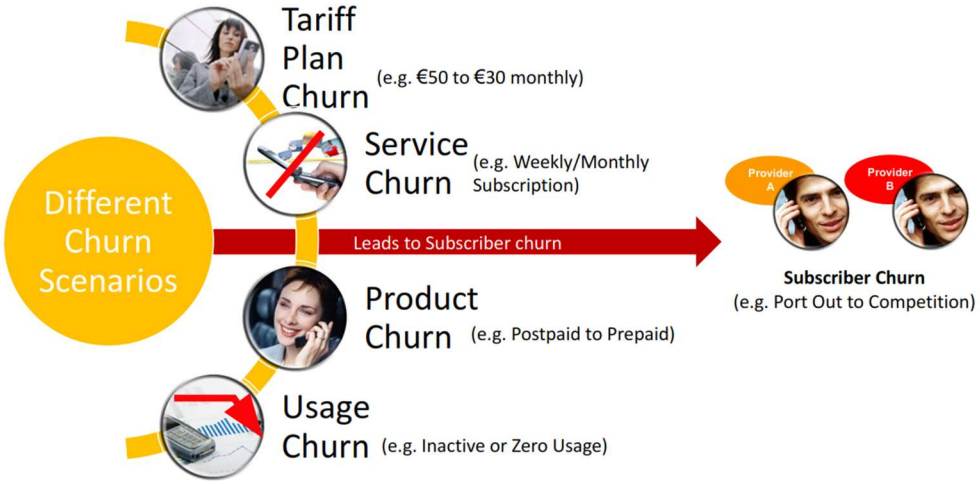
Service churn refers to a situation where a customer seeks to modify their billing cycle. This could involve switching from weekly billing to monthly billing, or vice versa, in order to better suit their needs.

Product Churn

In terms of payment options, there are two main types of customer preferences: product churn and postpaid. With product churn, customers prefer to pay their bill at the end of their billing cycle, either at the end of the month or on a specific date. On the other hand, some customers prefer a prepaid option where they can charge their account whenever they need to use the service. This option allows for more flexibility, as customers can pay more in months when they require more data or internet usage. However, having more postpaid customers can be more advantageous for the company.

Usage churn

"User churn" refers to customers who no longer wish to use the Internet or the associated service, but who still maintain a subscription with the company.



Decision Cycle for subscription

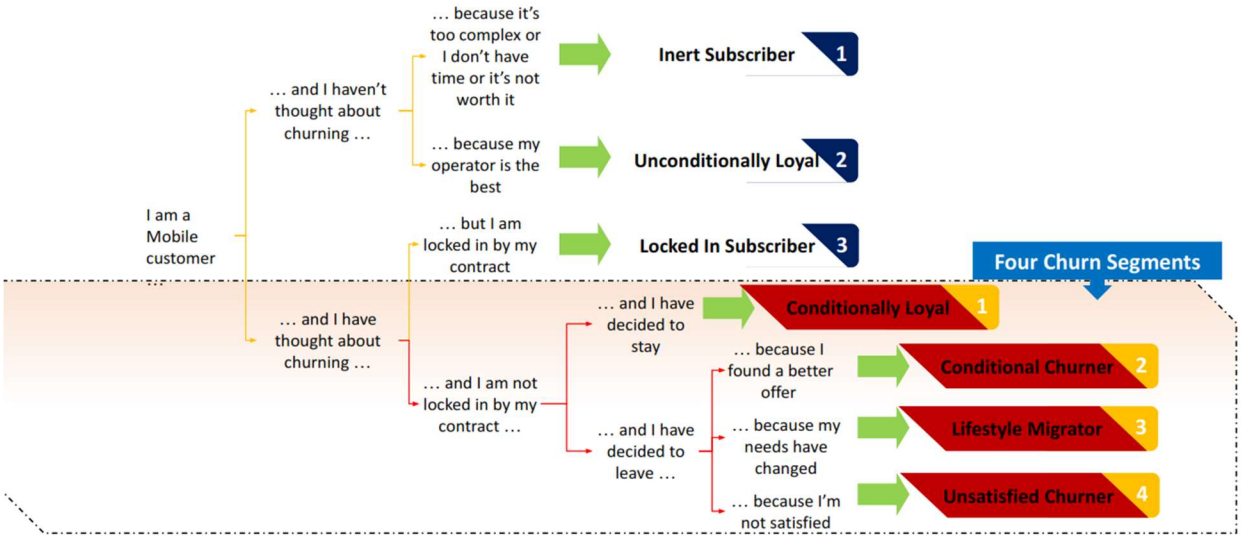
To determine if a customer has churned, refer to the diagram below. As a customer, I have two options: I may not have considered churning, or I may have considered it. If I haven't thought about churning, there are two reasons. Firstly, I may be lazy and assume that switching to another carrier involves too much documentation and portability is too complex. Therefore, I prefer to stay with my current mobile carrier, which is referred to as an "inert subscriber." Secondly, I believe that my operator is the best and I am extremely satisfied with their services. This is known as an "unconditionally loyal customer.".

Shall we consider the second scenario, which involves the possibility of churning? There are two subgroups to consider. First, if I am currently in a contract with the company, I am a locked subscriber and cannot terminate my contract even if I wanted to. However, if I am not locked in a contract, I have two options: I can choose to leave the company and immigrate elsewhere, or I can choose to stay. If I choose to stay, I am conditionally loyal to the company, meaning that I may churn in the future. But for now, I have reasons to stay, such as waiting for the next two or three months to see how things develop.

Should I choose to leave the company, there are a few reasons that would contribute to my decision. For instance, if I happen to come across a better offer, such as another carrier providing the same plan at a lower price, I will consider it. This would make me a conditional churner, as the decision to leave is based on a specific condition or circumstance.

One reason for my desire to switch carriers is due to my changing needs. My current lifestyle doesn't require as much internet as before, but my carrier doesn't offer a lower amount in their

services. Therefore, I want to migrate to another carrier as a lifestyle migrator. Another reason for churning is my dissatisfaction with my current carrier, which makes me an unsatisfied churner.



What are our solutions and goals?

This project aims to analyze a telecom company's data set and determine whether customers wish to remain with the company or leave. This is an end-to-end project that covers all aspects, from initial analysis to deploying the model. First, we will understand the problem, including why we need to analyze data for churn management. Next, we will clean the data set using Python programming language to make it more manageable. We will then use Microsoft SQL Server to run SQL queries for better insight into the data. After that, we will delve deeper into exploratory data analysis using Python programming language, visualizing and analyzing variant and invariant data. We will also run statistical analysis to determine the correlation and relationship between variables. Using a supervised machine learning model, we will construct a model for predicting customer behavior based on past pattern and data. Finally, we will create an interactive dashboard using Microsoft Power BI, allowing managers and stakeholders to gain more insight and make better decisions about customers. It is worth noting that the data complexity in this data set may pose a challenge, but this is typical of real-world projects.

Data documentation

This dataset contains some variables that you can find their description in the below:

customerID	*Identification number of each customer*	*String - Label*	*-*
gender	*The gender of each customer*	*String - Categorical*	*Female/Male*
SeniorCitizen	*Whether the customer is older than 65 years old or not?*	*Number - Binary*	*-*
Dependents	*Whether the customer has any individuals who rely on them financially, such as children or other family members.*	*String - Binary*	*Yes/No*
tenure	*The number of month that customer have been using the service non-stop.*	*Number - Numerical*	*-*
MultipleLines	*Whether the customer has more than one line or not*	*String - Binary*	*Yes/No*
InternetService	*They type of internet service that customer has*	*String - Categorical*	*DSL/Fiber/No*
OnlineSecurity	*Whether the customer has security service for online internet service or not*	*String - Binary*	*Yes/No*
OnlineBackup	*Whether the customer has online backup for online internet service or not*	*String - Binary*	*Yes/No*
StreamingMovies	*Whether the customer has streaming movie service for online internet service or not*	*String - Binary*	*Yes/No*
DeviceProtection	*Whether the customer has device protection or not*	*String - Binary*	*Yes/No*
TechSupport	*Whether the customer has technical support or not*	*String - Binary*	*Yes/No*
Contract	*They type of the contract*	*String - Categorical*	*Month/Year/BiYear*
PaperlessBilling	*Whether the customer active paper billing or not*	*String - Binary*	*Yes/No*
PaymentMethod	*The type of payment that customer pays*	*String - Categorical*	*Bank Transfer/Credit Card/E-Cheque/M-Cheque*
MonthlyCharges	*The amount of money that customer pays each month*	*Number - Numerical*	*-*

Churn	*Whether the customer churned or not during the past year*	*String - Binary*	*Yes/No*

1-Data cleaning in Python

Data cleaning is a fundamental aspect of this project. It improves the quality of the data by identifying and rectifying errors, inconsistencies, and inaccuracies present in the dataset. This involves handling missing values, dealing with outliers, and resolving issues related to data formatting or coding. By addressing these problems, data cleaning ensures that subsequent analysis is based on accurate and reliable information.

2-More insights in SQL

In this project I use Microsoft SQL Server as well. Using SQL queries is useful for gaining insights from data because it allows us to interactively explore and analyze the data in a database. With SQL, we can write queries to retrieve specific subsets of data, filter and sort rows based on conditions, aggregate data to calculate statistics, and perform various data manipulations. By leveraging SQL's capabilities, we can gain valuable insights that can inform decision-making, identify opportunities, detect problems, and drive improvements.

3-EDA in Python

In this section we want to find some characteristics between the customers who are about to churn or not churn. It is the main goal in exploratory data analysis.

Categorical variable analysis

Invariant analysis

During exploratory data analysis, we conduct univariate analysis. This involves examining each categorical variable separately to determine the number of customers who have churned and those who have not. Essentially, univariate analysis focuses on analyzing churned customers for each individual categorical variable.

Bivariant analysis

This type of analysis involves considering two or more variables simultaneously and analyzing their impact on churn. By doing so, we can observe how variables affect each other in terms of churning.

Numerical variable analysis

To determine whether numerical variables impact customer churn, we utilize correlation analysis. This allows us to examine the extent to which these variables are related to one another.

4-Predictive Modelling in Python

Our goal is to construct a predictive model that accurately predicts customer churn and categorizes them into specific churning types. This will enable the telecom company to take necessary measures to retain customers through targeted campaigns and marketing strategies. By ensuring data integrity and reliability through effective data cleaning, we can provide actionable insights to optimize customer retention efforts and minimize revenue loss associated with churn.

5-Dashboard in Power BI

An interactive dashboard can provide telecom managers with valuable insights into customer churn in a concise and user-friendly manner. By visualizing churn metrics and facilitating real-time data exploration, managers can quickly identify patterns, trends, and correlations related to customer churn. This enables them to make informed decisions and implement targeted strategies to mitigate churn effectively.