Date
July 2023

An ML Approach for Analyzing Customer Churn provided by a Teleco Company

- A Capstone Project Report Prepared by [Mahmuda Yasmin] for DATA SCIENCE CAREER TRACK Certification by Springboard.

Context:

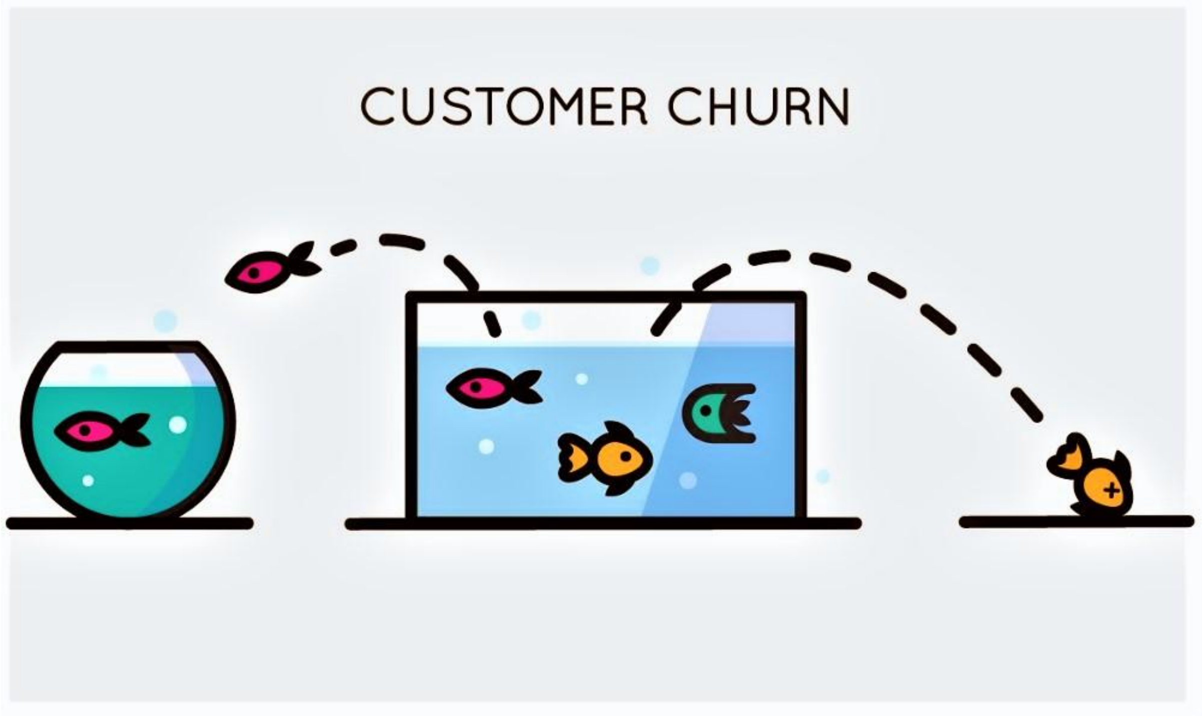
Prediction of Customer Behaviour and Customer Retention has always been a challenge for any business. It costs more to acquire new customers than it does to retain existing customers. This Business Metric helps to understand the reason behind the churn and to take effective initiatives to deal with the churn percentage.

Customer churn is the percentage of customers that stopped using the company's product or service during a certain time frame. It costs more to acquire new customers than it does to retain existing customers.

Customer churn impacts heavily on any business. There are other business metrics involved in the customer churn. A high customer churn rate indicates that a large percentage of your customers no longer want to purchase your products or services for various reasons, which can be a sign that your business is lacking in certain departments.

In this analysis, we are dealing with data provided by a Teleco Company in California Q2 2022 on Customer Churn. The purpose of this study is to analyze the provided data, try to find some meaningful interpretations and focus on finding a relevant ML model for predictive purpose.

Objective:



The goal of this study focuses on - “Why customers churn? How can you improve customer retention?”.

We will try to find out what factors may contribute to the Churn of the customers and provide supporting analysis to build an action plan to reduce the Churn.

Our Scope to find Solution for the following:

- After the subscription for the service, at which time point the Customers tend to leave most?
- Do the services provided by the company like-Monthly Bills, Internet Speed, Competitor’s Offers contribute to Churn?
- Do Customers' Demography Contribute to churn? (e.g. Location, Family Size, etc)

About the Dataset:

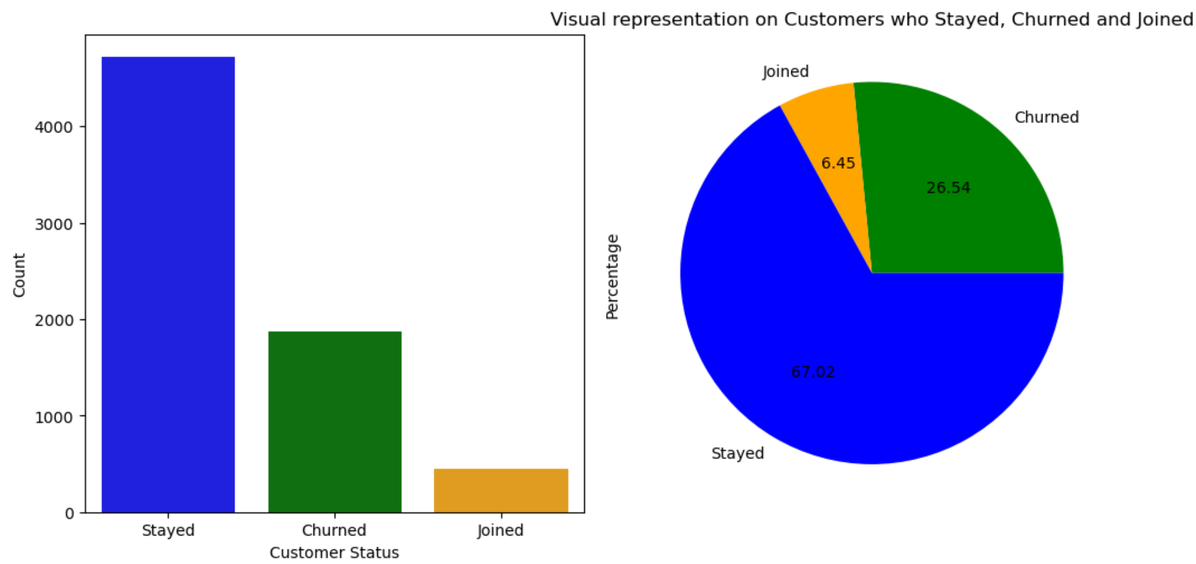
The source of the data is here. As mentioned in this Kaggle website- this dataset contains 3 tables, in CSV format:

- The Customer Churn table (7043 rows, 38 columns) contains information on all 7,043 customers from a Telecommunications company in California in Q2 2022
- Each record represents one customer, and contains details about their demographics, location, tenure, subscription services, status for the quarter (joined, stayed, or churned), and more!
- There is another file with the Data Dictionary and Zip Code. The Zip Code Population table contains complementary information on the estimated populations for the California zip codes in the Customer Churn table.

Our interest was focused on the ‘Customer Churn table’ Dataset. The “Customer Status - Joined, Stayed or Churned” Column is our event of interest.

Exploring the Reason for Churn:

The exploration of the Churn reason among churned customers revealed some interesting findings. At the end of Q2 2022, 67% of the customers stayed with the service provider and 26.54% of the Customers Churned (a customer churn rate of 25% is considered high).



While leaving, they were asked the reason. The ranking of the reasons were as below-

	Category of Reason for Leaving
5	Attitude
1	Competitor
2	Dissatisfaction
3	Other
4	Price
0	NaN

	Specific Reason for Leaving
8	Attitude of service provider
9	Attitude of support person
1	Competitor had better devices
5	Competitor made better offer
10	Competitor offered higher download speeds
11	Competitor offered more data
13	Deceased
6	Don't know
19	Extra data charges
12	Lack of affordable download/upload speed
17	Lack of self-service on Website
4	Limited range of services
7	Long distance charges
14	Moved
3	Network reliability
18	Poor expertise of online support
20	Poor expertise of phone support
16	Price too high

The majority of the Churned customers noted the Attitude of the Customer Representative as their reason to churn.

The second major cause for their churn was competitor's offers and services.

The detailed wrangling on the dataset can be found here and EDA can be found here.

Data Preprocessing:

1. Missing Values:

The Data Frame had a good chunk of missing values. It turned out that some variables were related to the response to another feature and left blank. Those missing cells had to be replaced with some corresponding values. The Data Dictionary file was very useful for this part for replacing the missing values.

2. Drop Unnecessary Columns:

For our ML based Analysis, there were some unnecessary columns (i.e. "CustomerID", "Zip Code","Latitude","Longitude","Churn Category","Churn Reason" etc.) which I had to drop.

After further investigation, several other variables seemed to be related with each other (High Multicollinearity). I dropped those corresponding variables ("Total Revenue","Total Charges", "Total Long Distance Charges") and proceeded with the independent variables in the model.

3. Converting String to Float:

While processing the data, there were some variables I had to convert from string type to float and integer.

4. Dealing with Categorical Features:

There were 20 categorical features in the dataset where some of them had multiple categories. Hot-Encoding or Dummy encoding would have increased the dimensionality of the dataframe. So I used ‘LabelBinarizer’ from Scikitlearn to encode the Categorical Variables. This helped to keep control on the high dimensionality problem.

5. Dealing with Numerical Values:

6. Designing Pipeline with Scikitlearn

Acknowledgement:

- Dataset: The details on the dataset can be found here at Kaggle and here at MavenAnalytics.
- Analysis: The Github Repo of this Analysis is here.
- Tools and Software: Python (Libraries listed below):

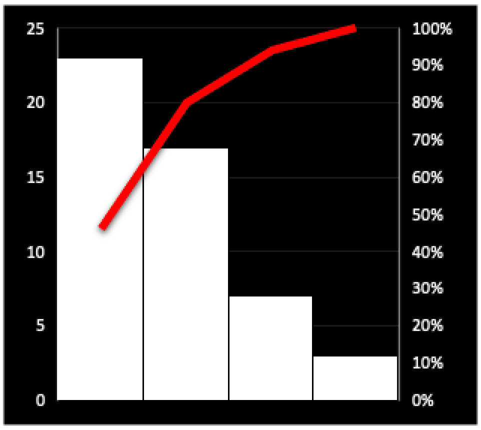
SUMMARY

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

RECOMMENDATION



LOREM IPSUM

- Lorem ipsum dolor sit amet, consectetur adipiscing elit
- Lorem ipsum dolor sit amet, consectetur adipiscing elit
- Lorem ipsum dolor sit amet, consectetur adipiscing elit
- Lorem ipsum dolor sit amet, consectetur adipiscing elit
- Lorem ipsum dolor sit amet, consectetur adipiscing elit