Theran Meadows
Project 3: Churn Prediction
Milestone 1
July 28, 2024

**Topic**

My project will be to use a data set provided by IBM for a fictional company,Telco, to determine who of their customers are more likely to churn. From the same data set, I would also like to investigate what characteristics or reasons of the long term customers prevented them from leaving.

**Business Problem**

When a customer churns from a company, it means that that customer is no longer a customer of that company's services. When customers churn, it can be for many different reasons. Maybe the product stopped being reliable. Maybe the product got too expensive and was no longer worth it in the eyes of the customer or, even worse, competition from another company has attracted them to try their similar product.

In my own life, I know I have churned a couple different times for my car insurance, who I had streaming platform subscriptions with, or even where I ate my fast food at. No worries, there will always be a new customer to take a churned customer's place right? Yes and no. Yes because there are probably so many different people and types of audiences that a product could go to to attract new customers. No, because it is actually more expensive to sell to a new customer than it is to sell to an existing customer. If you take on a lot of customers but also have a high churn rate, you won't have enough money to sustain a constant flow of customers coming in and going out.

There needs to be a balance of trying to achieve low churn and maintain existing customer satisfaction. My business questions to answer will be:

***What customers are more likely to churn and why?***

**Datasets**

Kaggle.com is providing the data set I will be using. It is a dataset given by IBM for a fictional company called Telco. It has over 7000 rows of customer data including tenure, the products they have or don't have, how much they spend, and some personal information about them. This is a fictional set, any personal identifiable information is a farce.

**Methods**

I will be doing exploratory data analysis as well as 2 predictive models. I would like to try a logistic regression model and a Naive Bayes model. Logistic regression model is easy to interpret and can classify who churned and who didn't.

A Naive Bayes model would also be good because it does well with multiple classifications and can classify customers as 'low', 'medium', or 'high' risk.

**Ethical Considerations**

No ethical considerations to consider since the data I am using is open to the public and is fictional.

**Challenges/Issues**

This is my first time analyzing any data set to investigate churning rates.

**References**

https://www.kaggle.com/datasets/blastchar/telco-customer-churn