##############################################################################

title: 'Telecom Customer Churn Prediction'

author: Ravindra Neralla

date: "05/29/2020"

##############################################################################

**Introduction:**

Customer churn is one of the most important concerns and major problem for large companies. Customer churn directly impacts revenues of large corporations, especially in subscription based industries like telecom filed. So, companies are working on developing technologies to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce customer churn. The telecommunications sector has become one of the main industries in developed countries. One of the main strategies in generating more revenue is increase the retention period of customers.

**Objective:**

Primary objective of the analysis is predicting customer churn. I started the analysis by investigating the Telco Customer data set to see if there are any characteristics that would help to predict the customer churn. So, these variables could be single or multiple variables going to impact customer churn.

**Data and Methodology:**

**Data:**

The dataset used in this analysis is taken from Kaggle, Telco-Customer-Churn.csv. Dataset consists of 7043 records with 21 variables, each row corresponds to one telecom customer. Description for each of the variable is listed below.

customerID - Customer ID

gender - Customer gender (female, male)

SeniorCitizen - Whether the customer is a senior citizen or not (1, 0)

Partner -Whether the customer has a partner or not (Yes, No)

Dependents - Whether the customer has dependents or not (Yes, No)

tenure - Number of months the customer has stayed with the company

PhoneService - Whether the customer has a phone service or not (Yes, No)

MultipleLines - Whether the customer has multiple lines or not (Yes, No, No phone service)

InternetService - Customer’s internet service provider (DSL, Fiber optic, No)

OnlineSecurity - Whether the customer has online security or not (Yes, No, No internet service)

OnlineBackup - Whether the customer has online backup or not (Yes, No, No internet service)

DeviceProtection - Whether the customer has device protection or not (Yes, No, No internet service)

TechSupport - Whether the customer has tech support or not (Yes, No, No internet service)

StreamingTV - Whether the customer has streaming TV or not (Yes, No, No internet service)

StreamingMovies - Whether the customer has streaming movies or not (Yes, No, No internet service)

Contract - The contract term of the customer (Month-to-month, One year, Two year)

PaperlessBilling - Whether the customer has paperless billing or not (Yes, No)

PaymentMethod - The customer’s payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))

MonthlyCharges - The amount charged to the customer monthly

TotalCharges - The total amount charged to the customer

Churn - Whether the customer churned or not (Yes or No), stop using service vs. continued service

**Methodology:**

Before starting analysis, I focused on identifying the bad data and removing it from further analysis. Correlation plots were created to see, how the variables are related to each other. Logistic regression model is developed.

**Analysis:**

In this customer churn analysis, I was interested to see the most influencing variables on customer churn and ultimately predicting customer churn. I started with understanding data and removing unwanted variables in the data for further analysis, for example customerID does not have any influence on churn analysis. Correlation analysis helped me to see the relationship between the variables, Totalcharges and Monthlycharges are correlated positively, so I removed Monthlycharges from further analysis.

# Submit a link to your repository to the assignment link during the final week of class.

All files associated with project can be found at:

<https://github.com/rneralla1/DSC530>/

<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0191-6>