Data Overview

Telco Customer Churn data is used to predict behavior to retain customers.

The data includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

Initial EDA - Insights

- 26.5% of customers have churn status.
- Column TotalCharges have extreame values that can be seen from the mean value greater than the median value.

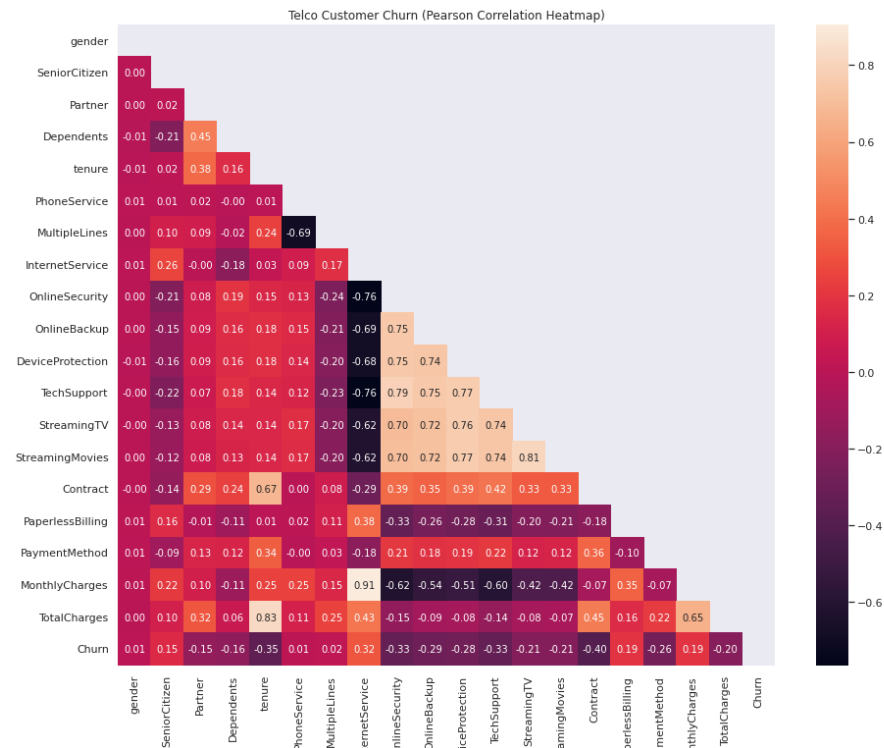|	TotalCharges	Churn
count	7043.000000	7043.000000
mean	2279.734304	0.265370
std	2266.794470	0.441561
min	0.000000	0.000000
25%	398.550000	0.000000
50%	1394.550000	0.000000
75%	3786.600000	1.000000
max	8684.800000	1.000000

Target Distribution - Insights

• There are Churn distribution of 26.5% and 73.5% for yes an no, respectively.
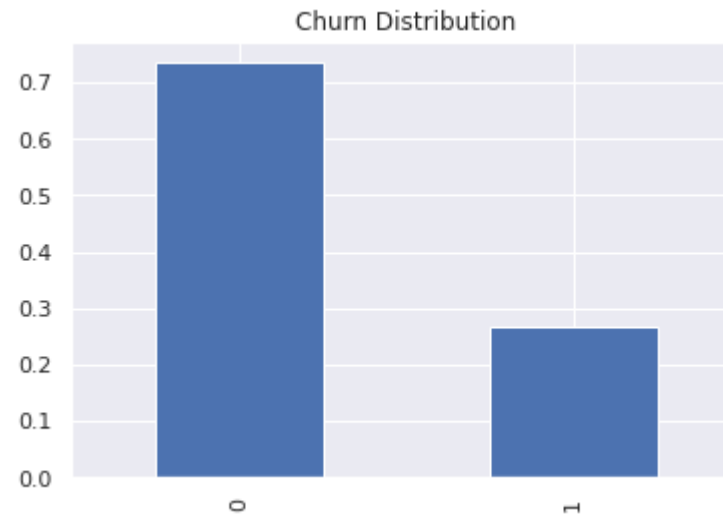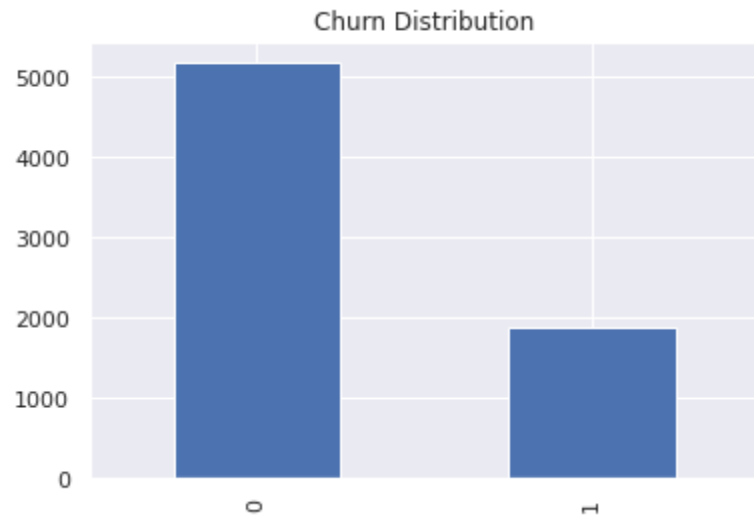
Correlation Analysis - Insights

- InternetService and MonthlyCharges have a very high correlation, which one of them will be chosen as predictor.

- Tenure and TotalCharges have a high positive correlation which make sense since the longer tenure will generate higher charges.

- InternetService has a negative impact to OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV & StreamingMovies because if customers have no internet service they will not be able to have those services.

Telco Customer Churn (Pearson Correlation Heatmap)

Outlier Analysis - Insights

- Based on outlier analysis of TotalCharges, there is no outlier found in TotalCharges.

TotalCharges Boxplot
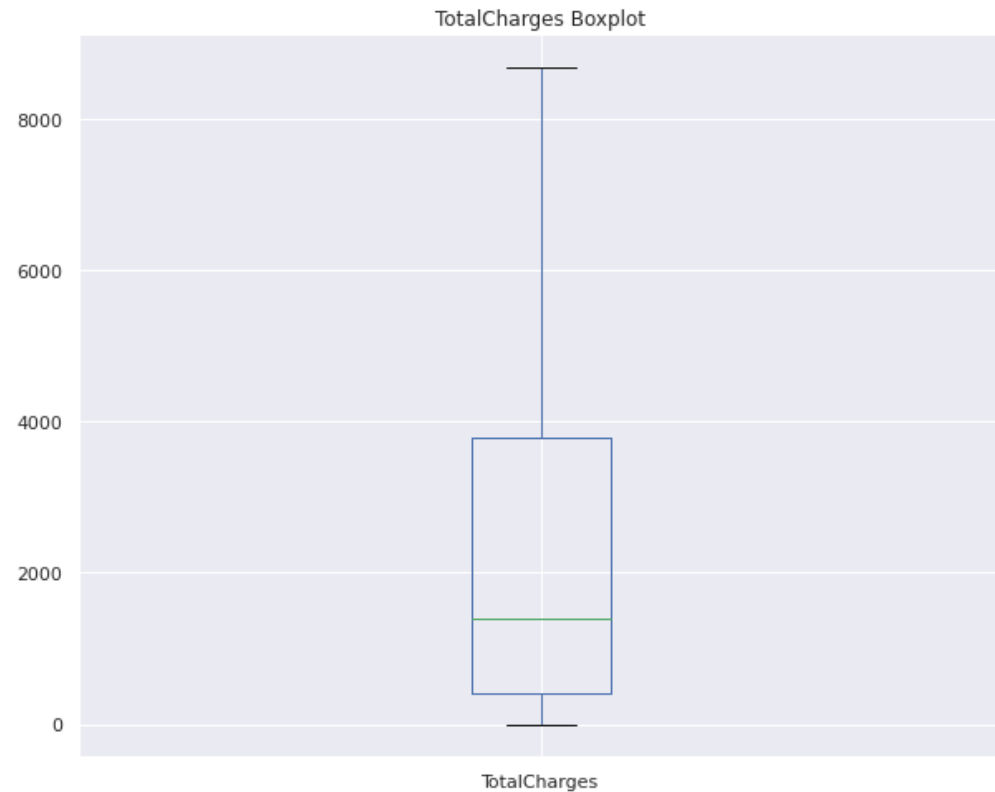
Services Analysis - Insights

- Based on InternetService, customer that used fiber optic have a higher probability of churn.

InternetService	Churn mean	count
0	0.074050	1526
1	0.189591	2421
2	0.418928	3096

- From services that each customer has signed up for, StreamingTV have the highest probability of churn.

services	Churn mean	count
DP	0.384840	686
OB	0.310054	1303
OS	0.146112	2019
SM	0.247504	2404
ST	0.606684	389
TS	0.309917	242

Demographic Analysis - Insights

- Based on demographic analysis, senior citizen have a higher probability of churn. Especially, senior citizen without partner and dependents.

SeniorCitizen	Partner	Churn mean	count
0	0	0.300130	3072
	1	0.166490	2829
1	0	0.488576	569
	1	0.345550	573

SeniorCitizen	Dependents	Churn mean	count
0	0	0.280526	3882
	1	0.150570	2019
1	0	0.431970	1051
	1	0.241758	91

Customer Account Analysis - Insights

- Based on Contract, customer with Month-to-Month contract have a higher probability of churn.

Contract	Churn mean	count
0	0.427097	3875
1	0.112695	1473
2	0.028319	1695

- In general, customers that choose paperless billing have a higher probability of churn. Especially with electronic check payment method.

PaperlessBilling	PaymentMethod	Churn mean	count
0	0	0.327448	623
	1	0.134937	956
	2	0.110260	653
	3	0.100000	640
1	0	0.497704	1742
	1	0.272866	656
	2	0.208754	891
	3	0.190476	882

Reference

https://colab.research.google.com/drive/1q0vfTejYxKMwYyy-H3oZN7Dya-7Jr_93?usp=sharing