Telco Customer Churn

● ● ●

Mohammad Erwinsyah Hartono

Overview

This sample data module tracks a fictional telco company's customer churn based on various factors.

The data set 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 movie
- 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

The goal is to predict if customers will churn using machine learning model

We will first explore what the data has to offer

Features

Customers demographic info

- customerID — Customer ID
- Gender — Whether the customer is a male or a female
- 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)

Customer account information

- Tenure — Number of months the customer has stayed with the company
- Contract — The contract term of the customer (Month-to-month, One year, Two year)
- PaperlessBilling — Whether the customer has paperless billing (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

Features

Customer services booked

- PhoneService — Whether the customer has a phone service (Yes, No)
- MultipleLines — Whether the customer has multiple lines (Yes, No, No phone service)
- InternetService — Customer's internet service provider (DSL, Fiber optic, No)
- OnlineSecurity — Whether the customer has online security (Yes, No, No internet service)
- OnlineBackup — Whether the customer has online backup (Yes, No, No internet service)
- DeviceProtection — Whether the customer has device protection (Yes, No, No internet service)
- TechSupport — Whether the customer has tech support (Yes, No, No internet service)
- StreamingTV — Whether the customer has streaming TV (Yes, No, No internet service)
- StreamingMovies — Whether the customer has streaming movies (Yes, No, No internet service)

Customer Classification

- Churn — Whether the customer churned or not (Yes or No)

Basic Data Information

- There are 7043 observations with 21 columns
- The target variable is Churn
- Data has 3 numerical variables in:
  - Tenure,
  - Monthly Charges, and
  - Total Charges
- Data has 18 categorical variables in:
  - Customer ID with 7043 unique values
  - Target variable Churn
  - 6 binary features (Gender, Senior Citizen, Partner, Dependents, Phone Service, Paperless Billing)
  - 4 features with 3-4 categories (Multiple Lines, Internet Service, Contract, Payment Method)
  - 6 Additional features for customers with Internet Service (Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, Streaming Movies)
- Data has no missing value, although there are 11 customers who has 0 tenure and therefore 0 Total Charges, these observations should be removed when training Machine Learning model but we'll keep it for exploration for now
- Among the numerical variables there is no outlier

Initial Hypothesis

Based on out initial understanding of the dataset, we can form some hypothesis about customer churn:

- Customers with longer contract duration are less likely to churn as they are not often faced with decision to prolong or terminate the contract, and potentially pay less attention to the contract over long period of time, thus
- Tenure and contract duration are assumed to be main driving force of churn
- Customers with more expensive monthly charges can lead to churn as they might save money switching providers
- Customers with partner and dependents might churn less to keep the services for their family
- Senior citizens might churn less due to the effort of terminating existing contracts
- Customers who employ Additional Services are less likely to churn since they are more dedicated to make use of the additional services

The most important variable, Churn

We are trying to predict if the client left the company in the previous month.

Therefore we have a binary classification problem with a slightly unbalanced target:

- Churn: No - 5184, 73.5% of customers
- Churn: Yes - 1869, 26.5% of customers

We will focus in this variable when deriving insights
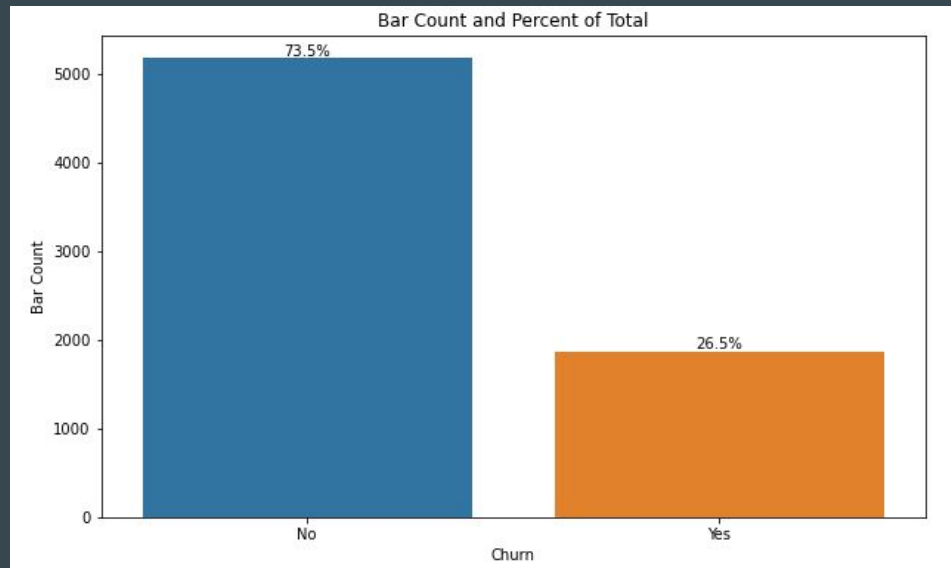
Numerical Variables

Tenure

Number of months the customer has stayed with the company

Monthly Charges

The amount charged to the customer monthly

Total Charges

The total amount charged to the customer

Tenure

- Does not follow the normal distribution
- Most common values are 1 (613 | 8.7%) and 72(362 | 5.1%)
- The distribution peaked on 1-5 months then decrease and stagnate, followed by another small peak around 65-70 months
- Why is there a lot of 72? Probably has something to do with customers who stayed with the company from the very start
- On the left side of the chart we have customers who just tested the product, which should have high churn
- The right side of the chart are loyal customers

Distribution of Customer Tenure
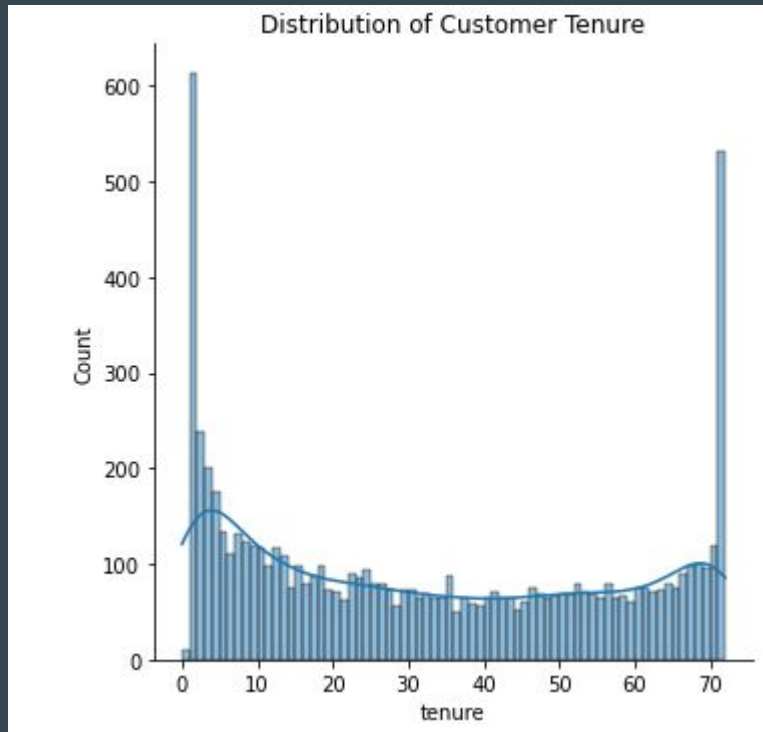
Tenure and Contract

- The tenure varies in distribution because of different contract terms available to customers, from Month-to-Month, One Year, and Two Years
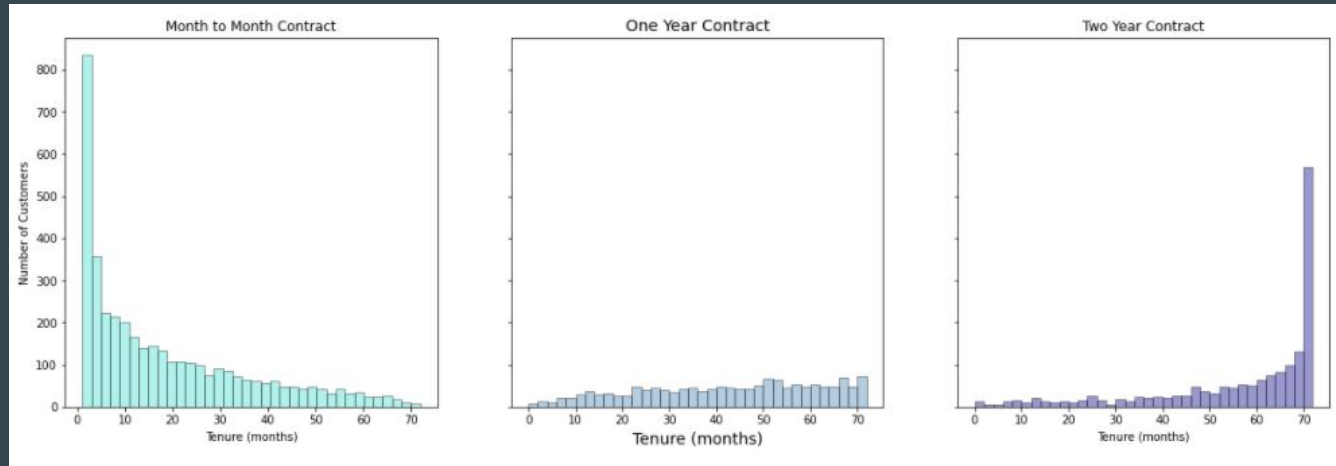
- Most of the monthly contracts last for 1-2 months, while the 2 year contracts tend to last for about 70 months. This shows that the customers taking a longer contract are more loyal to the company and tend to stay for a longer period of time.

Tenure and Churn
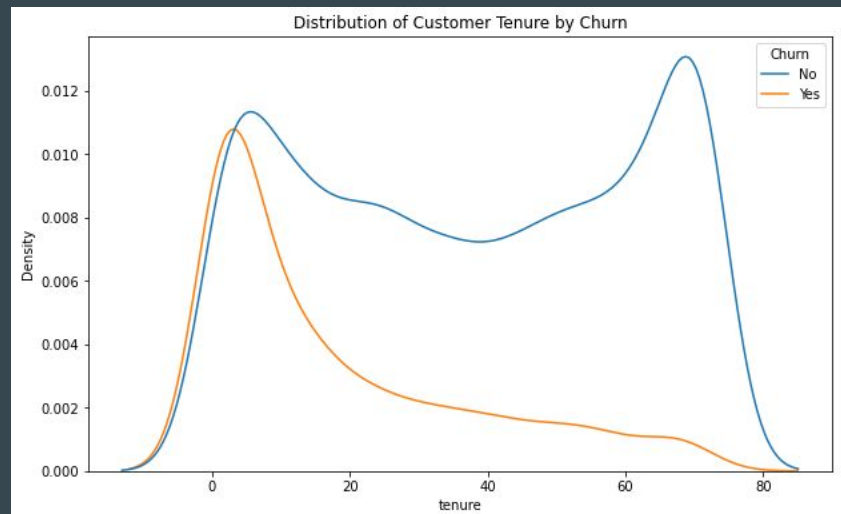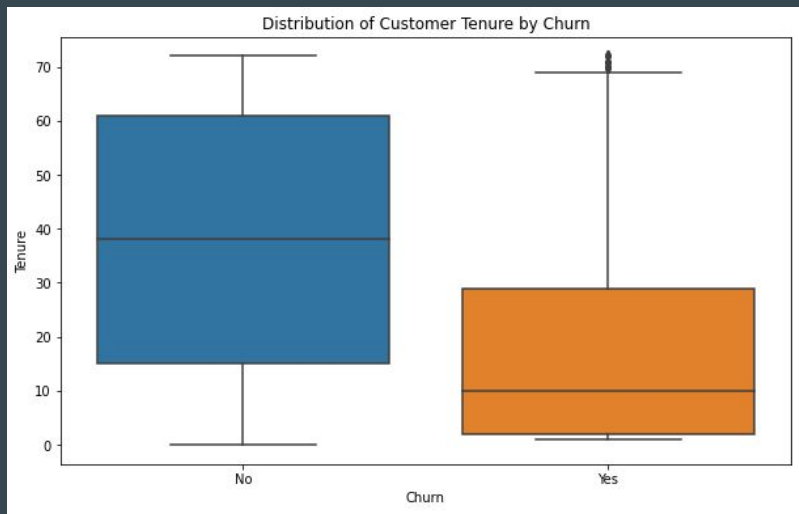
- Churn customers have lower tenure median of 10 months compared to non-churners with median of 38 months.
- Newer clients seem more likely to Churn

Monthly Charges

- Does not follow the normal distribution
- Highest peak and most common value is around 20
- The distribution peaked on 20-30$, decreases, and then steadily increases and stabilize around 40-110$ before decreasing as it reached 120$
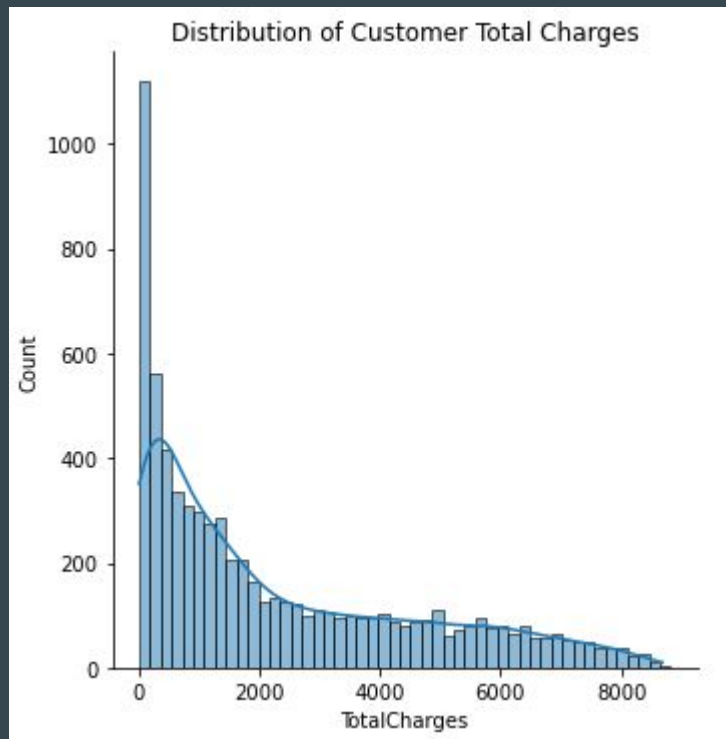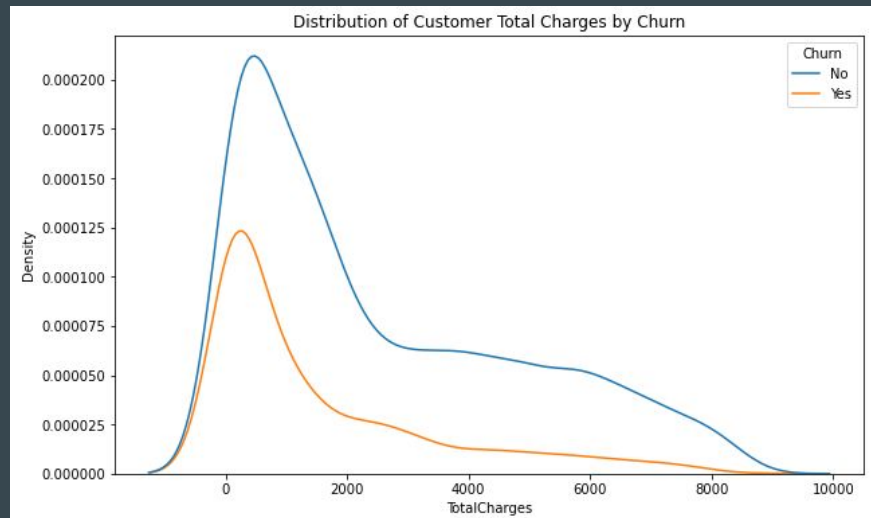- Median of Monthly Charges is 70.35$ while the mean is 64.7$

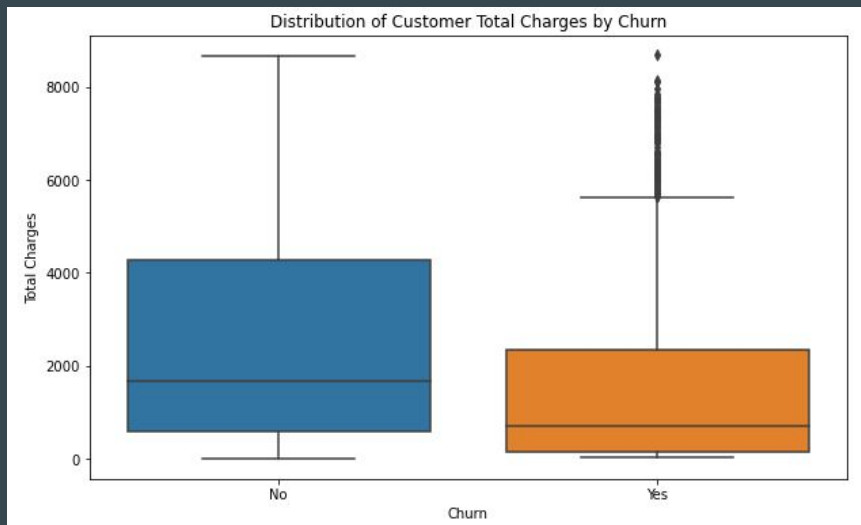Distribution of Customer Monthly Charges
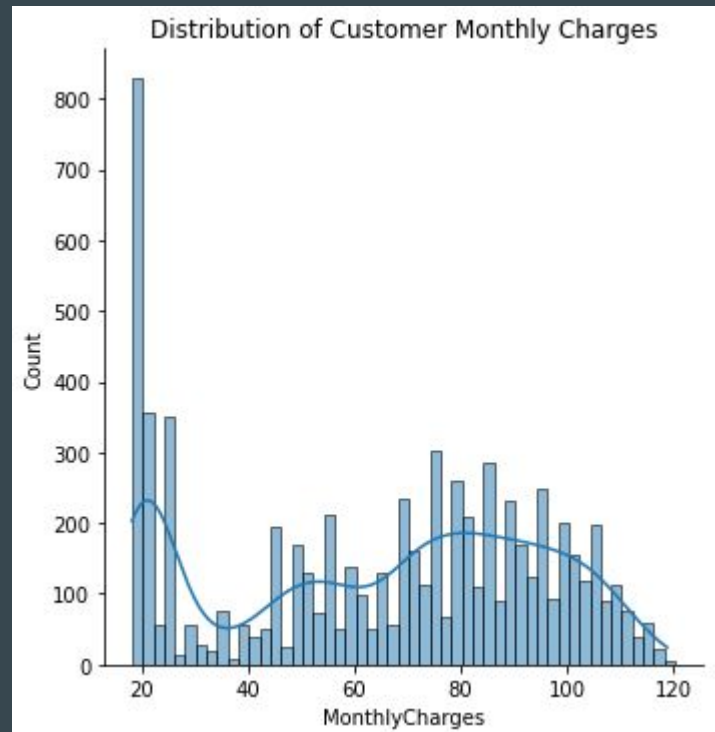
Monthly Charges and Churn
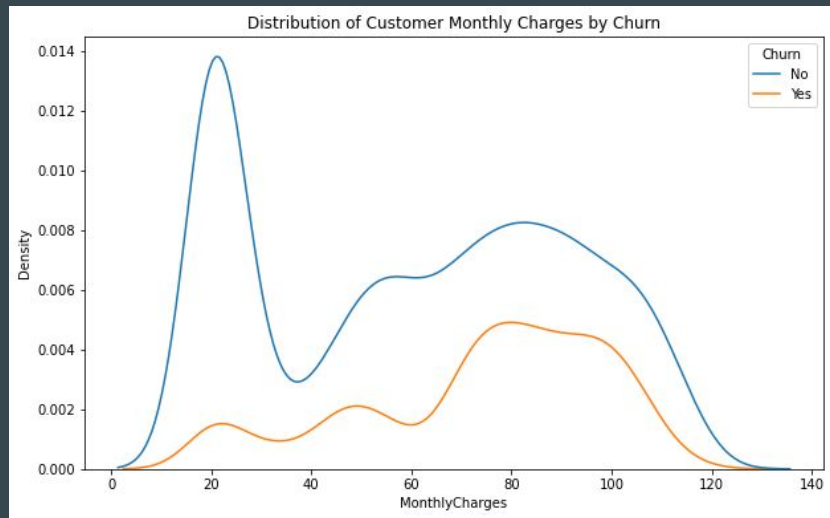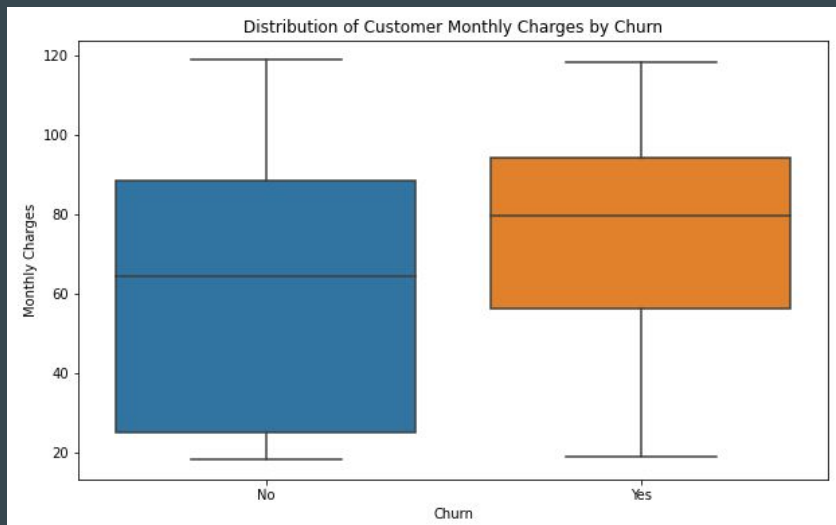
- Churn customers have higher Monthly Charges median of 80$ compared to non-churners with median of around 64$
- as MonthlyCharges go higher, customers are also more likely to churn

Total Charges

- Does not follow the normal distribution
- Highest peak and most common value is around 20
- The distribution peaked on 0-500$, decreases pretty sharply until 2000$, then steadily decrease slowly until 8000$
- Median of Total Charges is 1394.5$ while the mean is 2279.7$

Distribution of Customer Total Charges

Total Charges and Churn

Distribution of Customer Total Charges by Churn

Distribution of Customer Total Charges by Churn

- Churn customers has lower Total Charges median compared to non-churn
- It seems that there is higher churn when the total charges are lower, but
- This insight is not really intuitive, we had better patterns looking at Tenure and Monthly Charges individually compared to Total Charges as a whole but it might still be a relatively helpful variable in predicting churn

Total Charges and Monthly Charges

Total Charges and Monthly Charges Relationship
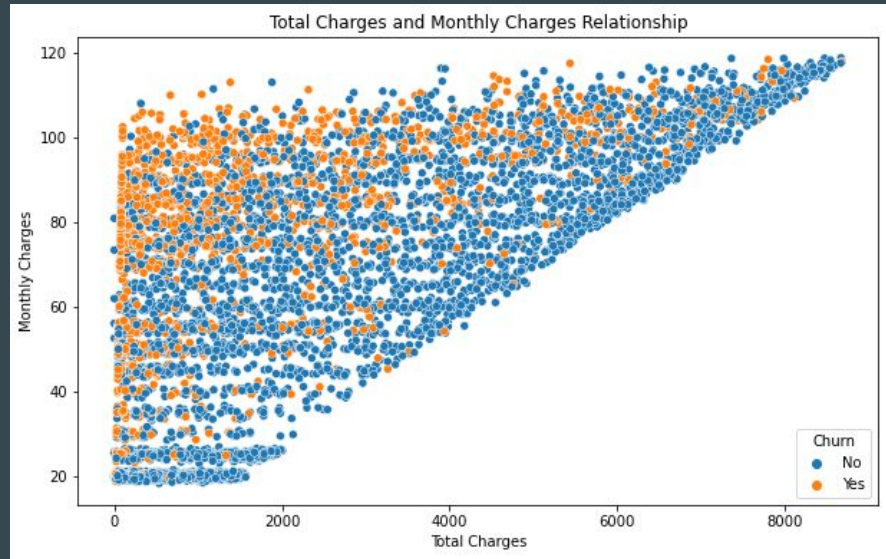
- Total Charges and Monthly Charges has a clear line boundary which is likely affected by tenure
- Churns seems more apparent on low Total Charges and High Monthly Charges which shows that low tenure customers are more likely to churn

Categorical Variables

Demographic Info

- Gender — Whether the customer is a male or a female
- 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)

Contracts and Services

- PhoneService — Whether the customer has a phone service (Yes, No)
- MultipleLines — Whether the customer has multiple lines (Yes, No, No phone service)
- InternetService — Customer's internet service provider (DSL, Fiber optic, No)
- Contract — The contract term of the customer (Month-to-month, One year, Two year)
- PaperlessBilling — Whether the customer has paperless billing (Yes, No)
- PaymentMethod — The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))

Additional Services

- OnlineSecurity — Whether the customer has online security (Yes, No, No internet service)
- OnlineBackup — Whether the customer has online backup (Yes, No, No internet service)
- DeviceProtection — Whether the customer has device protection (Yes, No, No internet service)
- TechSupport — Whether the customer has tech support (Yes, No, No internet service)
- StreamingTV — Whether the customer has streaming TV (Yes, No, No internet service)
- StreamingMovies — Whether the customer has streaming movies (Yes, No, No internet service)
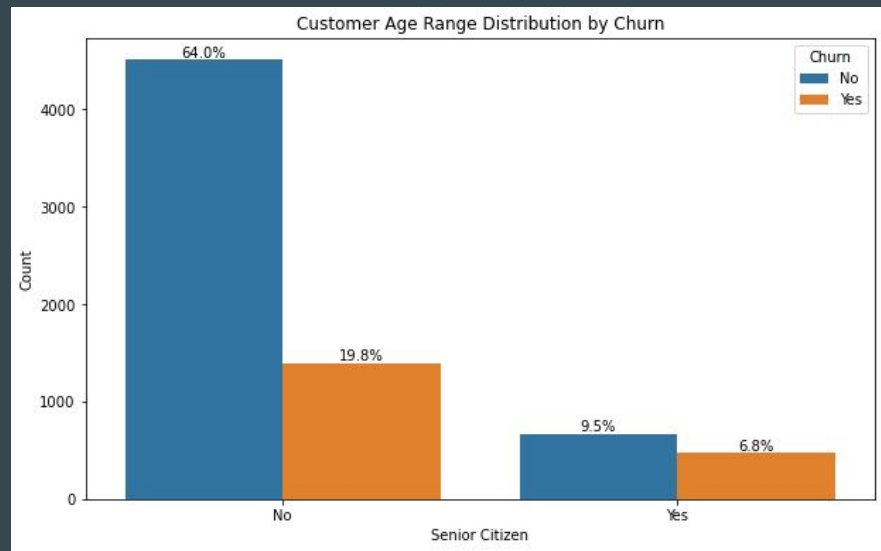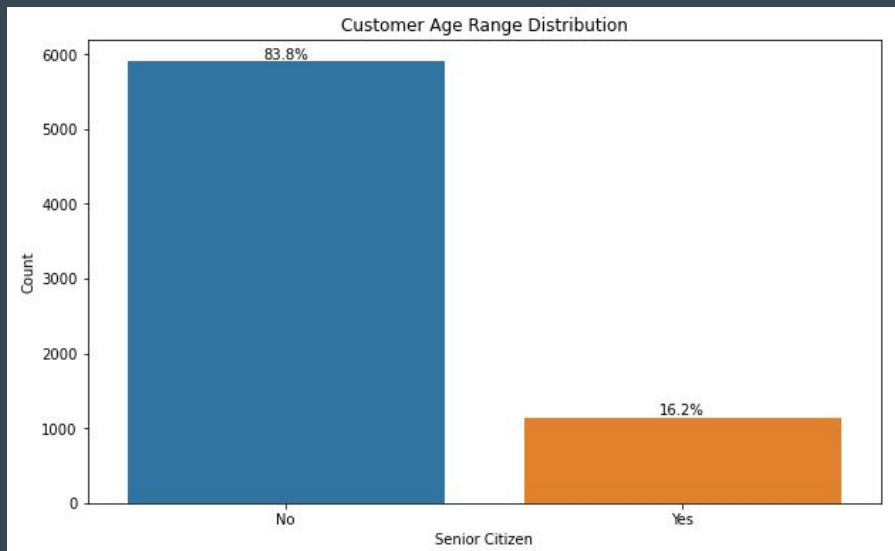
Demographic Info

gender, age range, and if they have partners and dependents

____

Gender and Churn
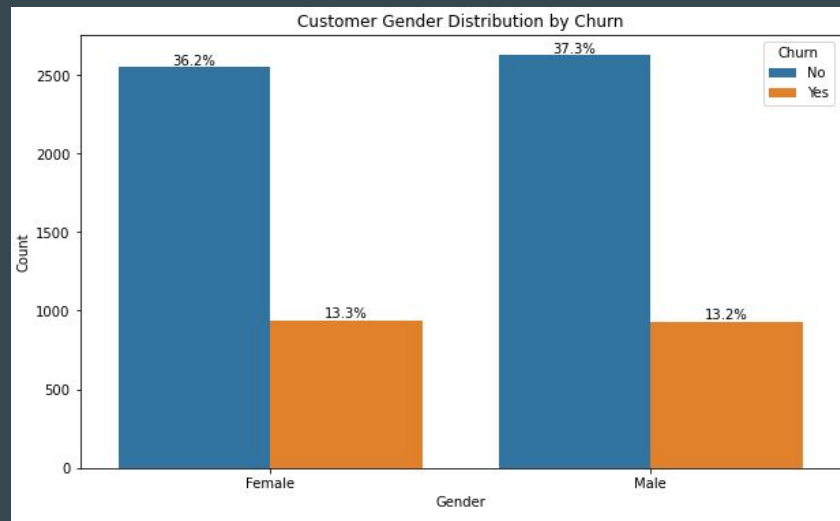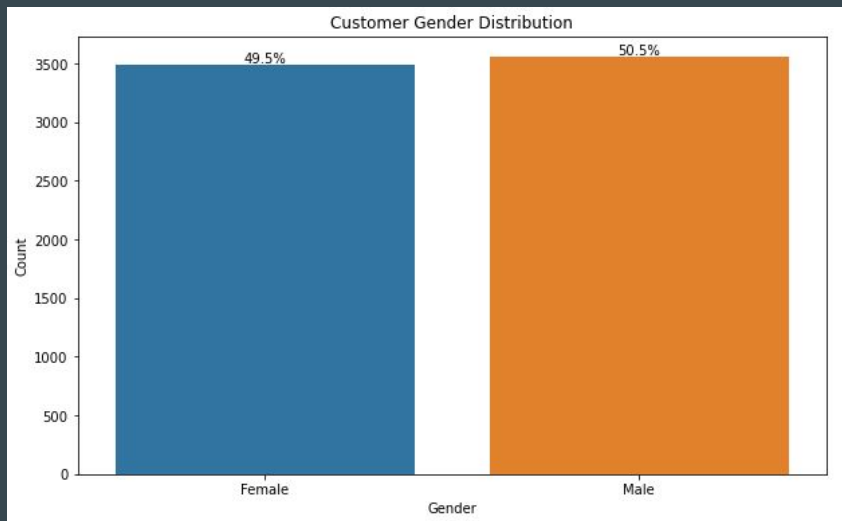
- Male and Female in this dataset is balanced, 3555(50.5%) and 3488(49.5%)
- Churn distribution among gender is also relatively balanced, thus gender is not indicative of churn

Senior Citizen and Churn

Customer Age Range Distribution



Customer Age Range Distribution by Churn
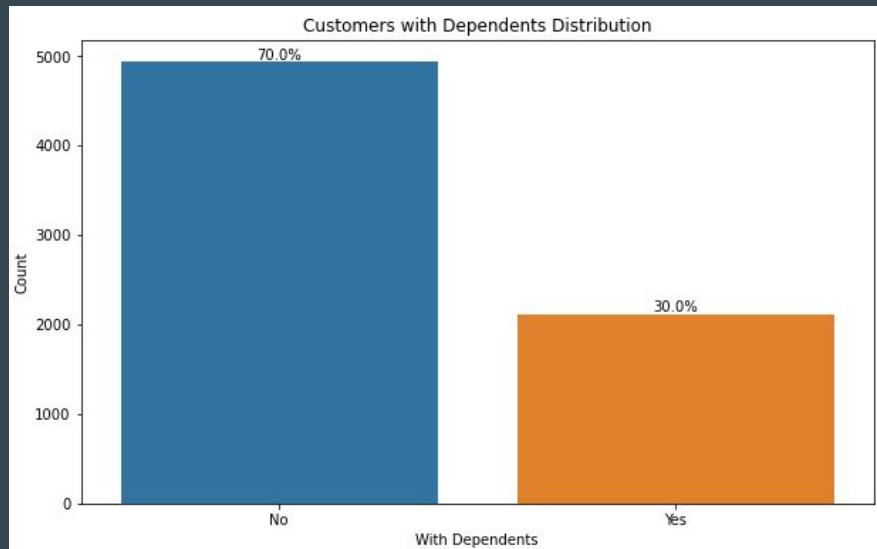
*pardon 0.1% rounding error

- Only small part, 1142 (16.2%) of customers of the customers are senior citizens, most customers are not senior citizen
- Senior Citizen are actually more likely to churn than non-senior customers, which is different from the initial hypothesis

Partner and Churn
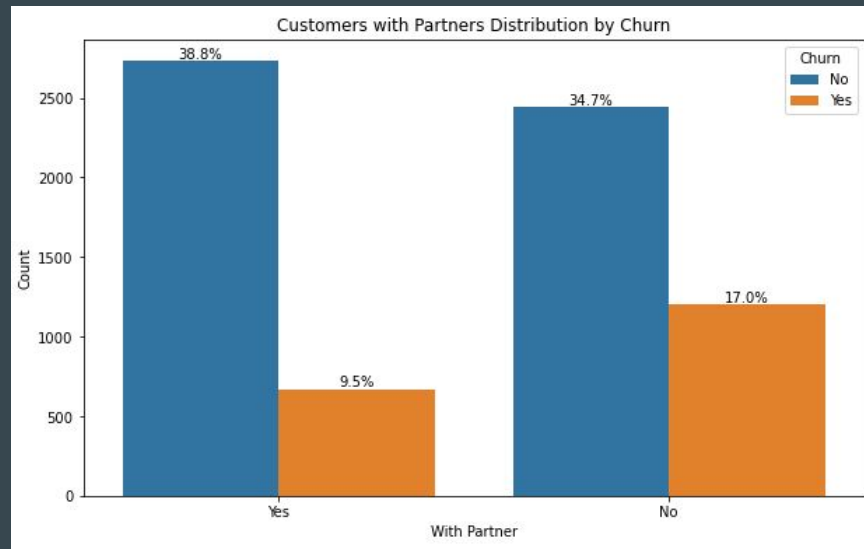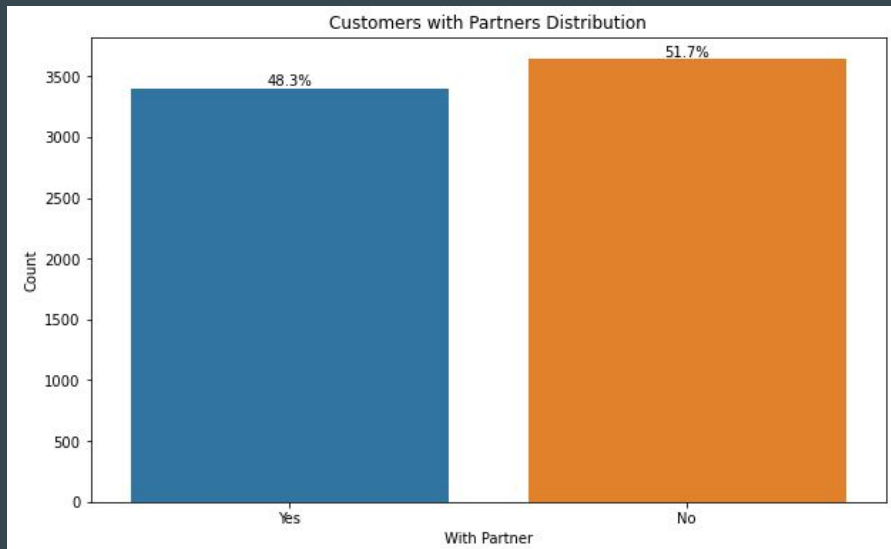
- Relatively balanced spread of customers with partner and without, 3402(48.3%) and 3641(51.7%)
- Customers with no partner are more likely to churn

Dependents and Churn

Customers with Dependents Distribution

Customers with Dependents Distribution by Churn
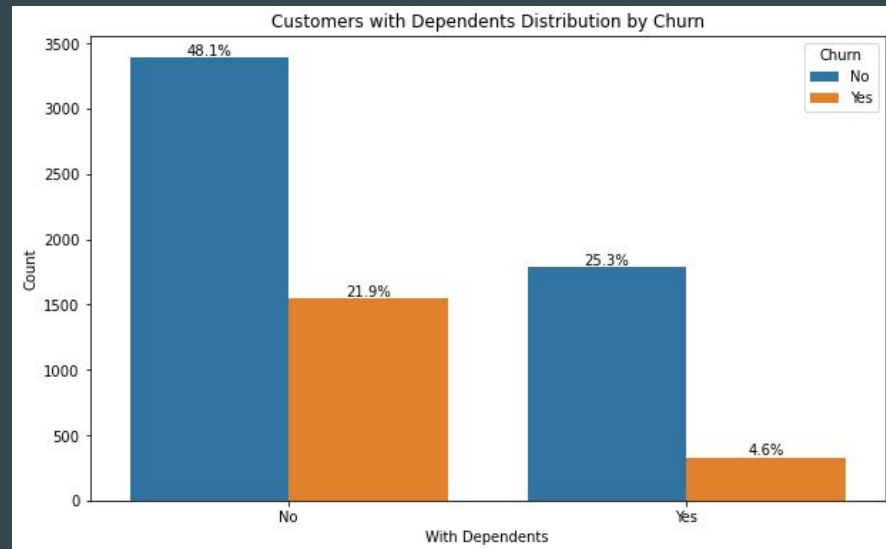
*pardon 0.1% rounding error

- 2110(30%) customers are with dependents, while 4933(70%) customers without, 70:30 slightly unbalanced distribution
- Customers with no dependents are relatively more likely to churn

Contract and Services

What main services does the customer employ, their contract, and billing information

Phone Service

PhoneService and MultipleLines variables basically has the same information with MultipleLines having more complete information, making PhoneService redundant
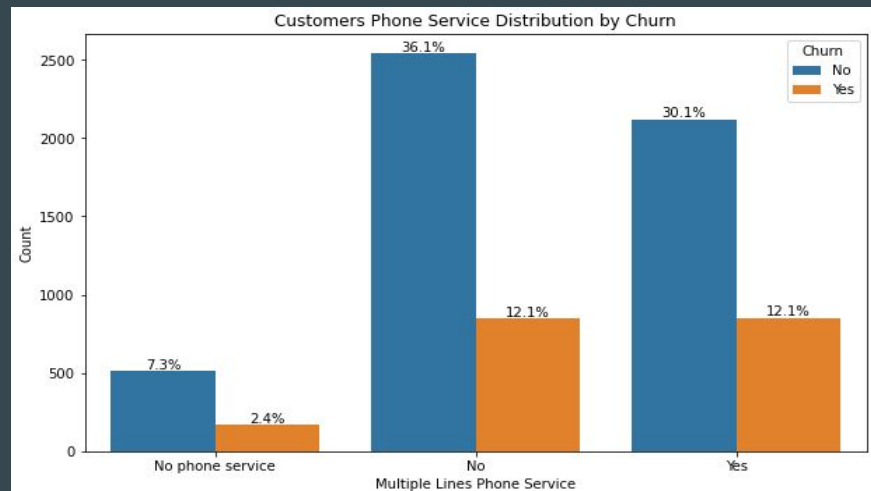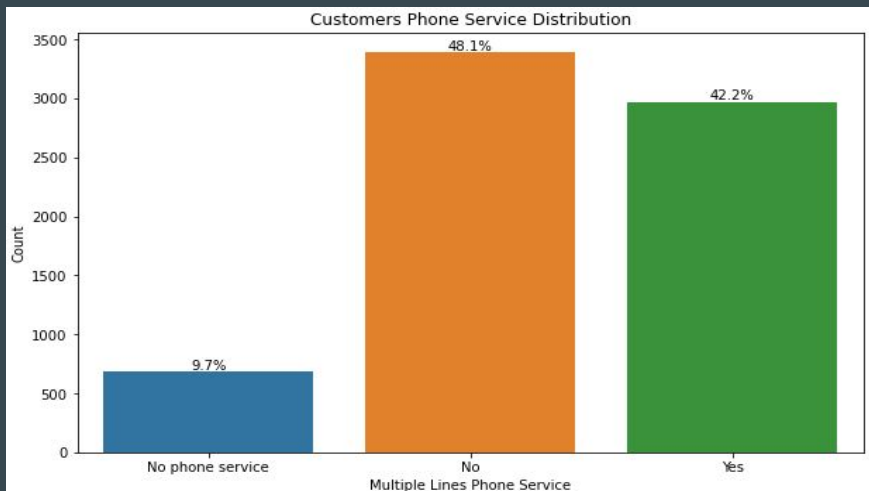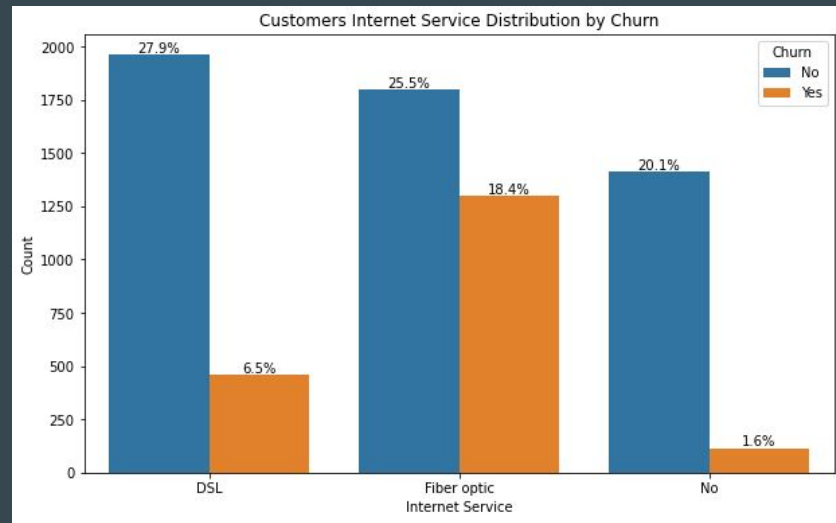

- Most customers (90.3%) use phone service, with more customer not using multiple lines (48.1%) than those who does (42.2%)
- Customers with multiple lines has slightly higher churn rate

Internet Service

Customers Internet Service Distribution

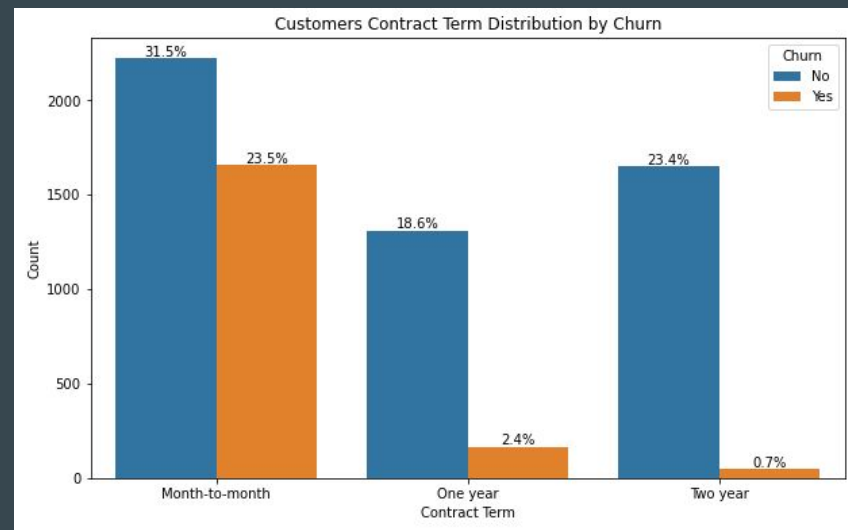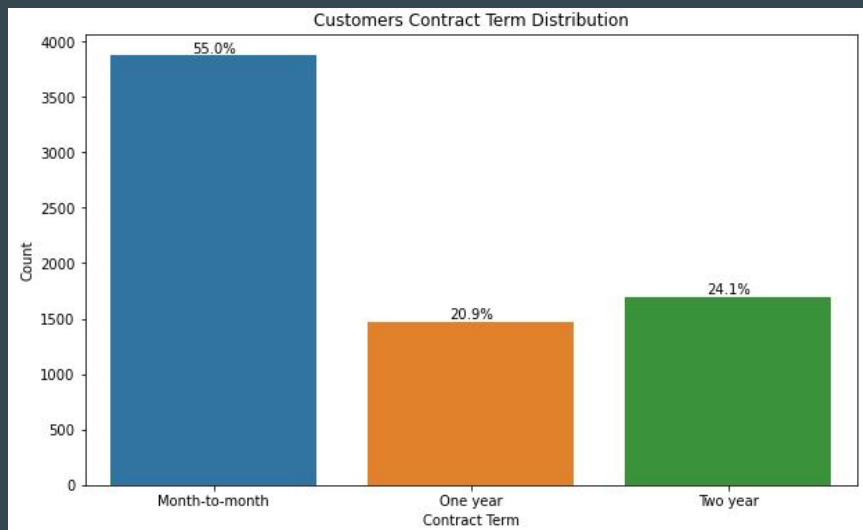Customers Internet Service Distribution by Churn
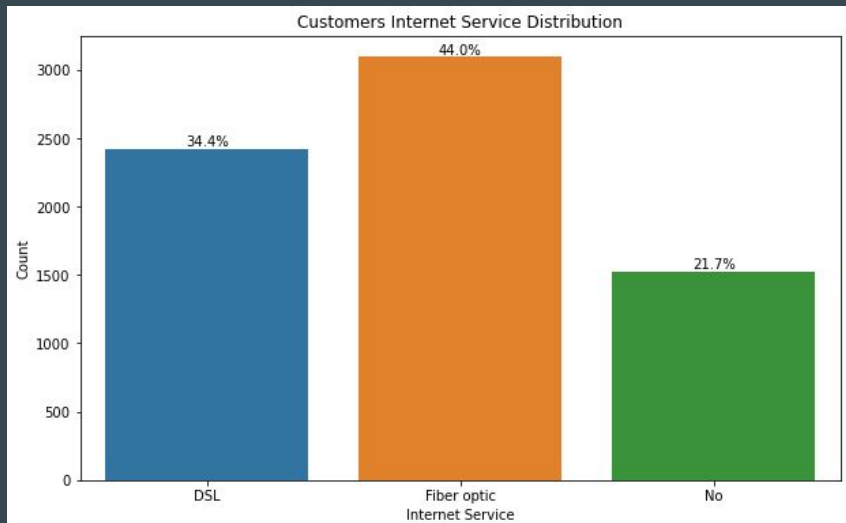
*pardon 0.1% rounding error

- Most customers (78.3%) use internet service, with more customer using Fiber Optic (43.9%) than DSL (34.4%)

- Customers with no internet service are less likely to churn

- Customers with Fiber optic are more likely to churn than DSL customers
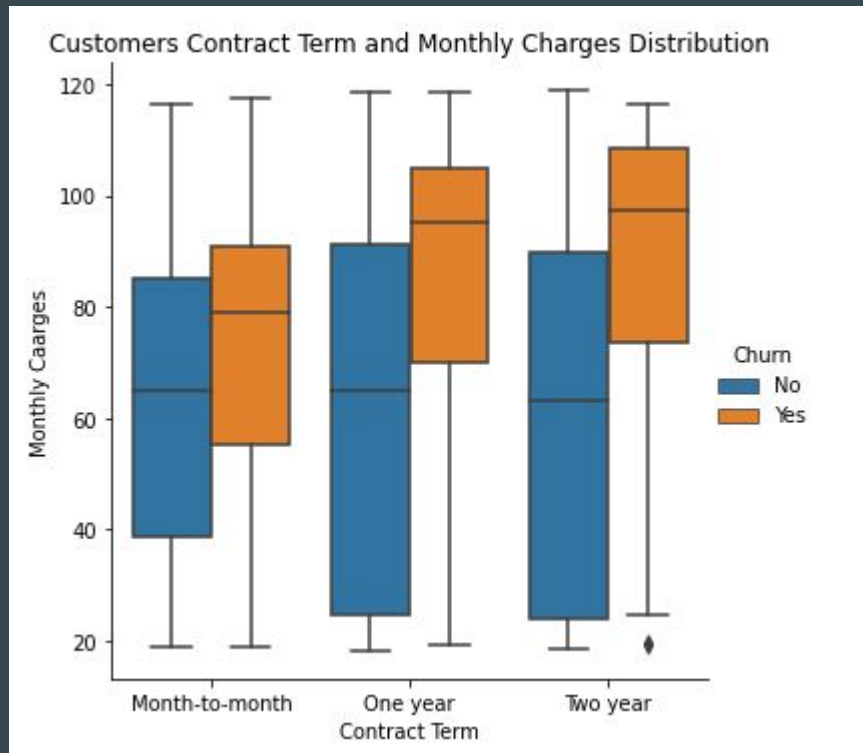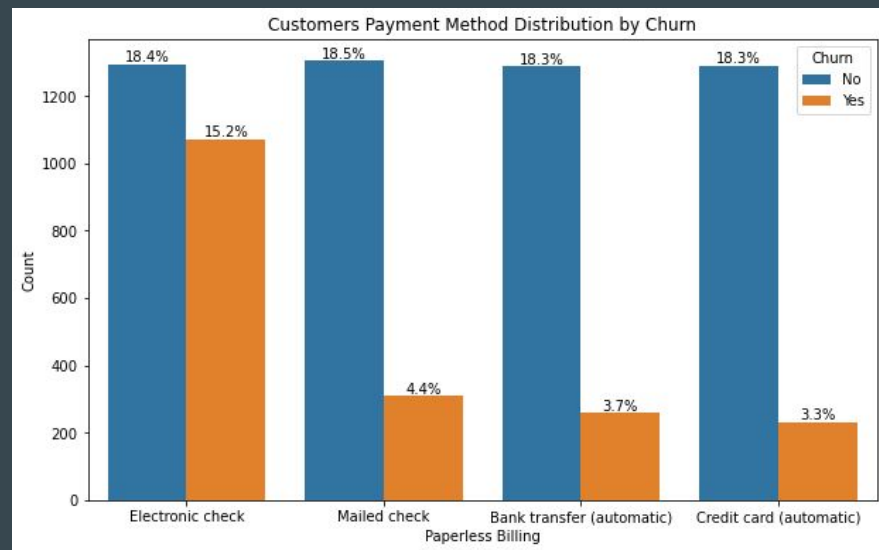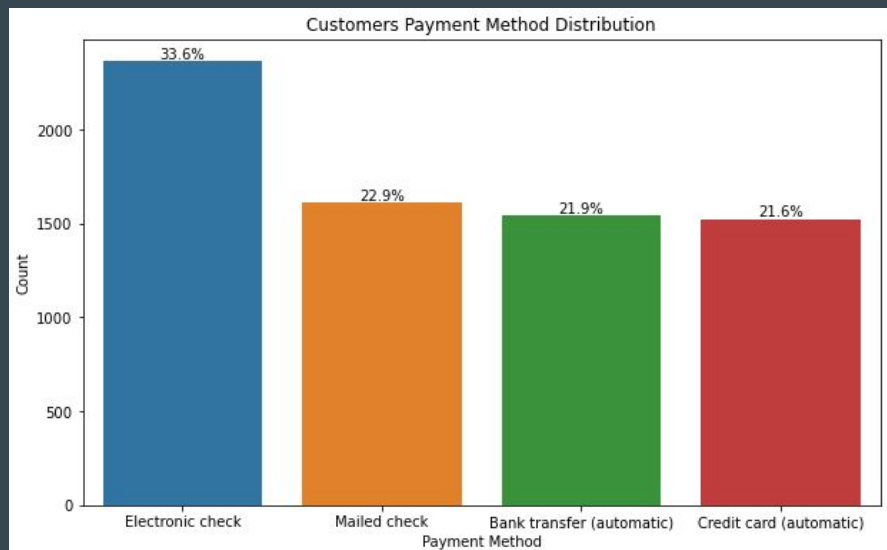
Contract Term

*pardon 0.1% rounding error

- Around half of total customers (55%) signed month-to-month, two year contract customers (24.1%) slightly higher than one year contract customers (20.9%)

- Customers with month-to-month contracts are more likely to churn

- Customers with two year contracts are much less likely to churn compared to other contracts

Contracts, Monthly Charges, and Churn

- Churn customers has higher median of Monthly Charges compared to non-churn over all terms
- Longer contracts churns are more affected by higher monthly charges

Customers Contract Term and Monthly Charges Distribution

Payment Method

Customers Payment Method Distribution

Customers Payment Method Distribution by Churn

*pardon 0.1% rounding error

- 4171(59.2%) customers receives paperless billing, while 2872(40.8%) customers don't, 60:40 relatively balanced spread
- 33.6 % of customers use Electronic check, this method also has much higher churn rate compared to other payment methods

Paperless Billing
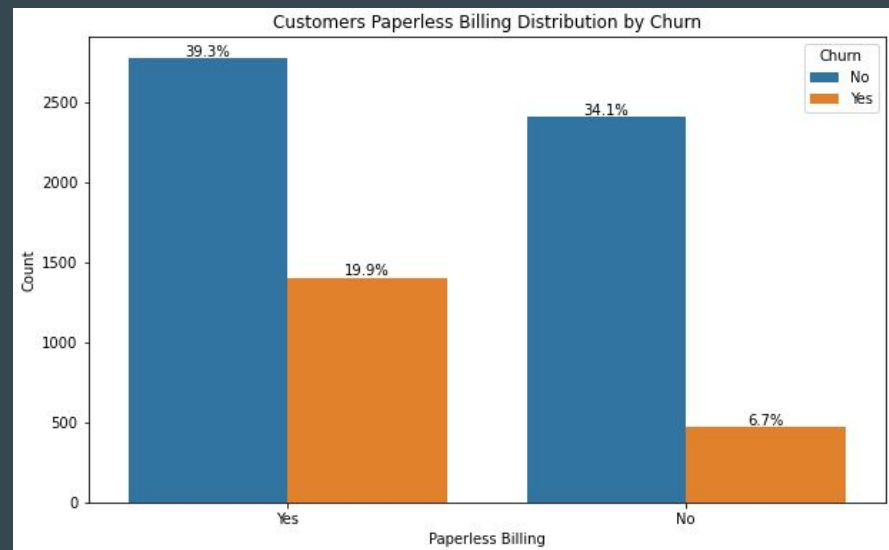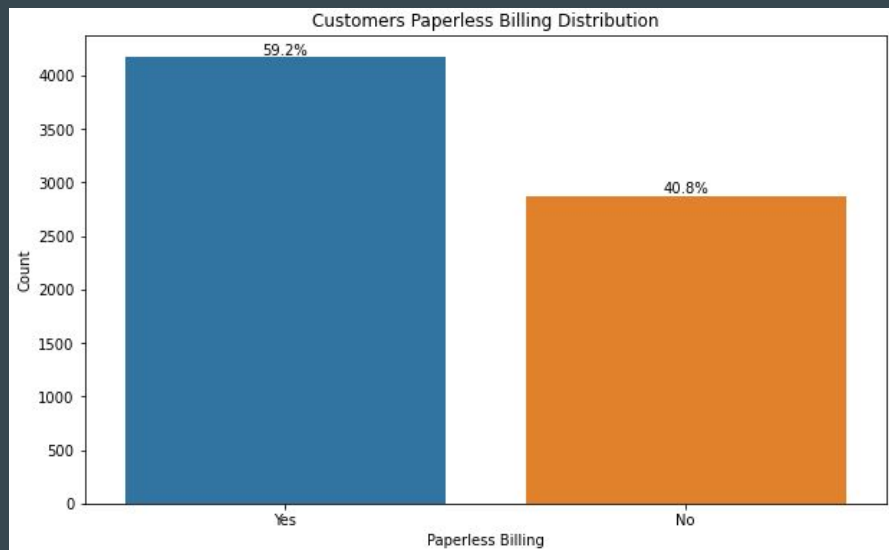
- 4171(59.2%) customers receives paperless billing, while 2872(40.8%) customers don't, 60:40 relatively balanced spread
- Customers with paperless billing is more likely to churn compared to the others
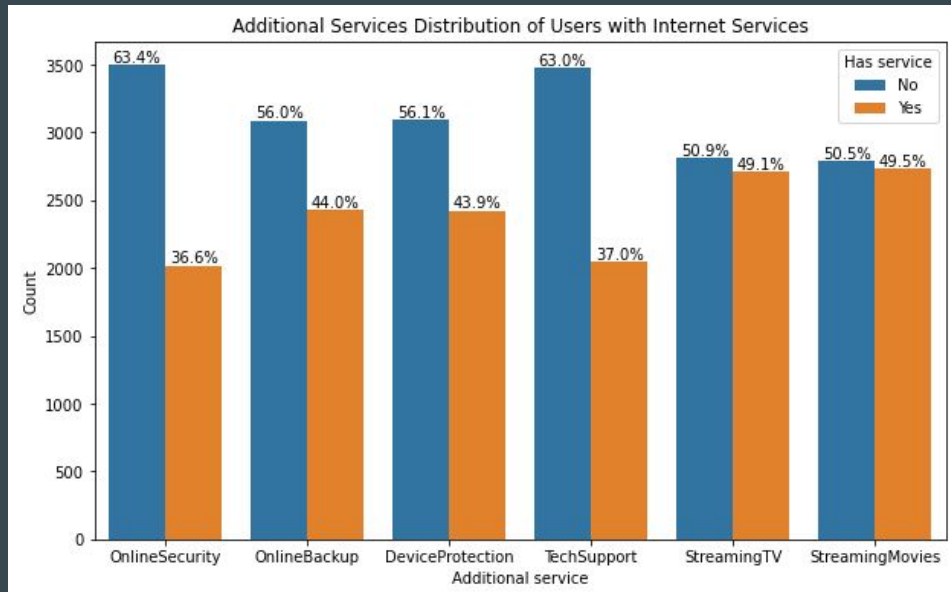
Additional Services

6 additional features for customers with Internet Service:
- Online Security,
- Online Backup,
- Device Protection,
- Tech Support,
- Streaming TV,
- Streaming Movies

Additional Services Distribution

- These services are available to customers with Internet Service, 5517 of 7043 customers, about 78.3% of total customers
- For each additional service, there's at least 50% of customers who doesn't subscribe,
- With the exception of slightly balanced distribution for Streaming TV and Streaming Movies services

Additional Services Distribution of Users with Internet Services

Additional Services and Churn

- Number of Services shows the churn rate among ALL customers according to number of additional services
- Customers with high number of services (4-6) are less likely to churn
- High churn is also apparent on customers with 1 additional service

Customers Additional Services Distribution by Churn
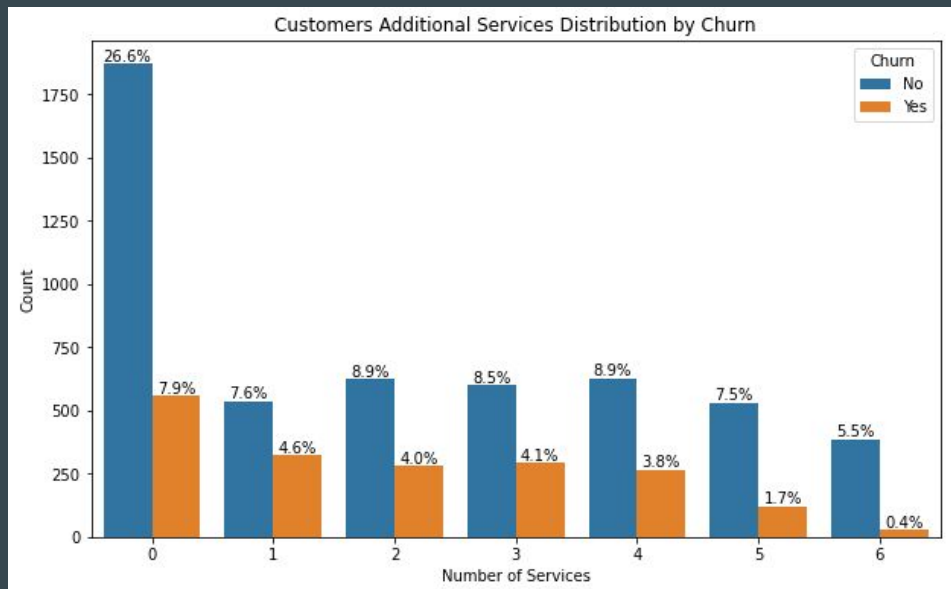
Correlation Matrix

Correlation Matrix

- Correlation matrix is made after encoding the dataset and running it through correlation analysis in Python
- The closer the value to 1 or (-1) the stronger the 2 variable correlates positively or negatively

Pearson Correlation Heatmap

Churn Correlations
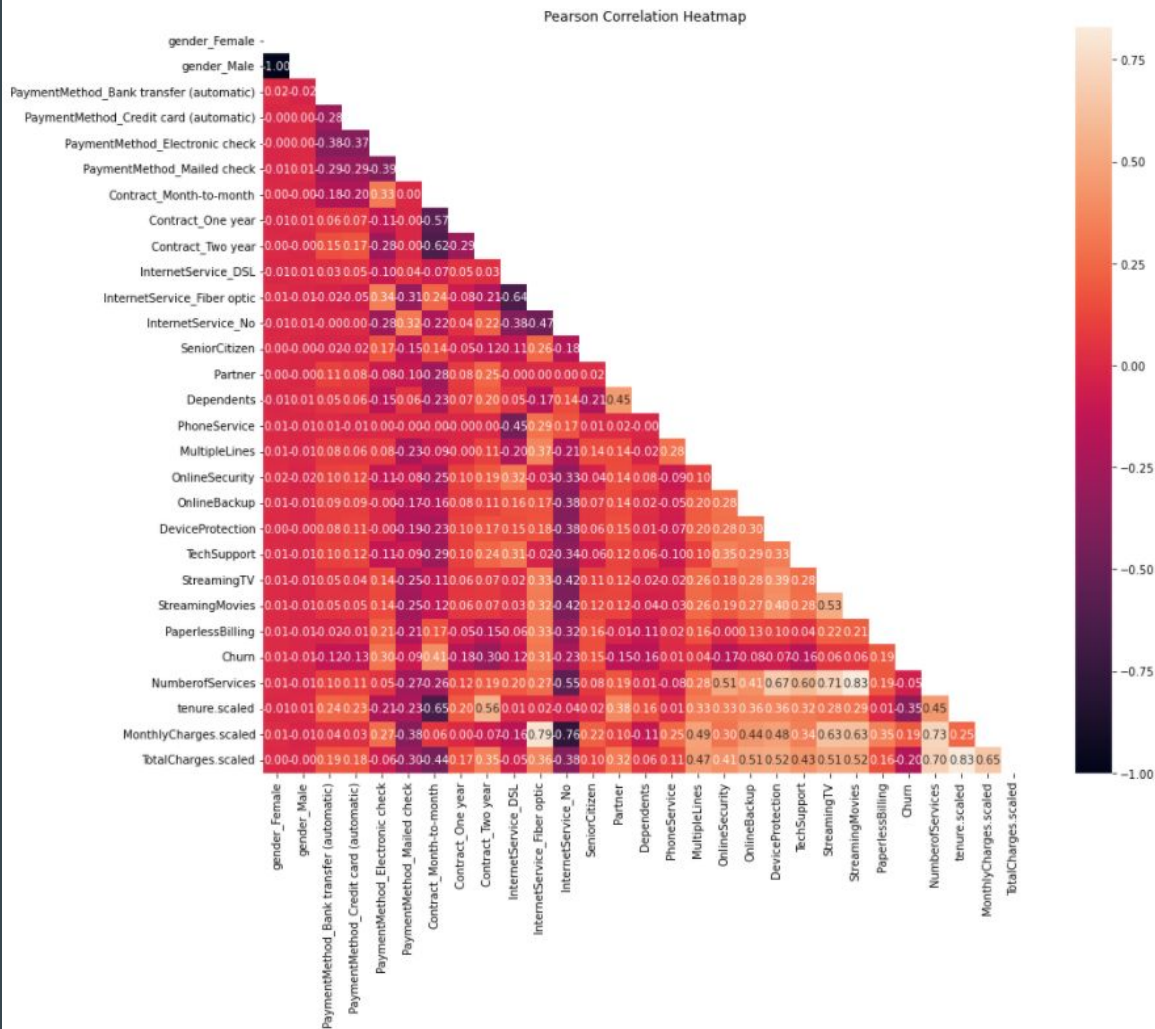
Churn Correlation Analysis
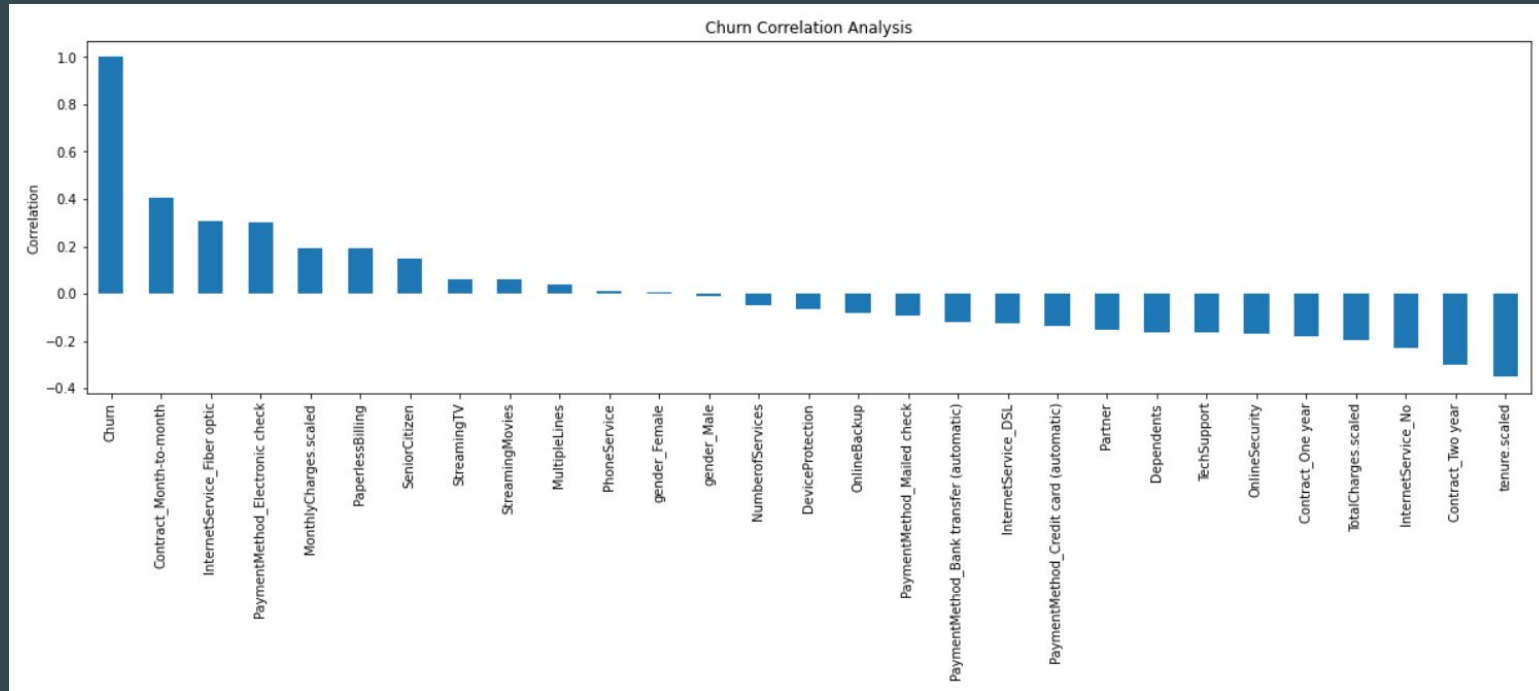
- The correlations are in line as our previous observations, we can see how strong each variable correlates to churn.
- Since correlation is based on linearity relationship between variables, other models might have different mapping of feature importance

Summary

Based on our hypothesis and observations of the dataset in relation to predicting churn, we can form the following summary:

- Generally, tenure and contract term seem to be the most important variable to predict churn
- Monthly Charges and Total Charges are still relatively relevant in predicting churn
- From customers demographic info, gender is relatively irrelevant in predicting churn
- Some certain categories of customers contract and services can help in predicting churn (Fiber Optic Internet Service, Electronic Check Payment Method, No Internet Service, etc.)
- Additional services/Number of services seem to be a weak predictor of churn and have multicollinearity with each other

Thank you.