

Churn Prediction – Milestone Report - 1

Objective: Increase customer retention by building predictive model (Churn) to push churn rate down closer to 0%.

Introduction:

Overview: This data set consists of Customer left within last month. Service that each customer has signed up for. Customer account information like payment method, billing type. Demographic info about customers like gender, age and dependents

Url/Link: <https://www.kaggle.com/blastchar/telco-customer-churn>

Approach:

Data Wrangling

EDA

Inferential Statistics

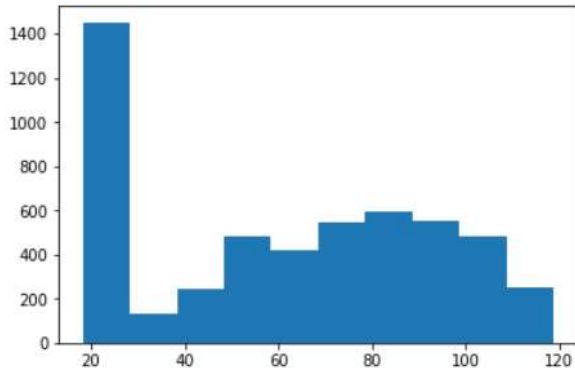
Data Wrangling & More:

- Fortunately, this is not a 'messy' data set
- The dataset has 11 missing values for total charges column, so replace it with 0 using “df.replace”
- We have more than 7000 rows and 21 attributes (columns)
- Some data that should be categorical are saved as number.
- If there are any missing values then 'df.fillna' can be used to fill the missing data
- Convert Total Charge to numeric
- There are some features that contains ambiguous information's, for example, Online Security contains 3 different labels, but the correct labels are Yes or No using “replace”
- This Data has 16 categorical features:

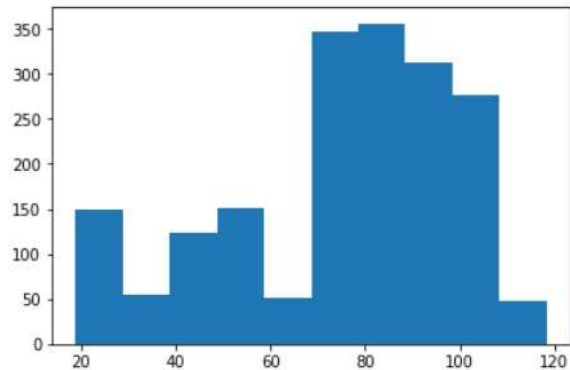
Data Visualization:

- Histogram for total charges (Churn Customers)

```
# Histogram (Total Charges for Not Churn Customers)  
plt.hist(df_not_Churn['MonthlyCharges'],bins= 10)  
plt.show()
```

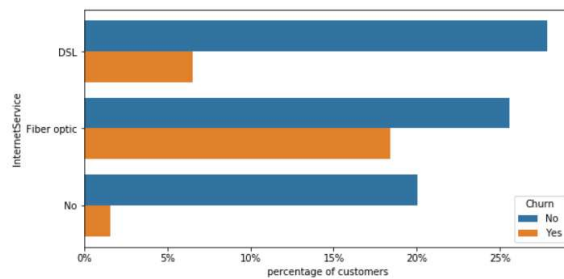
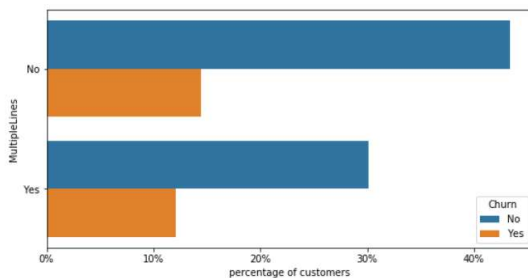


```
# Histogram (Total Charges for Churn Customers)  
plt.hist(df_churn['MonthlyCharges'],bins= 10)  
plt.show()
```

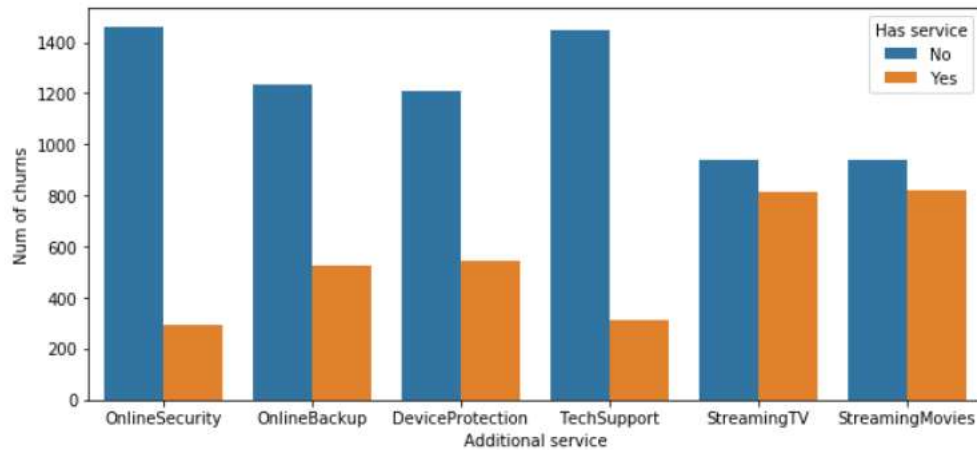


:

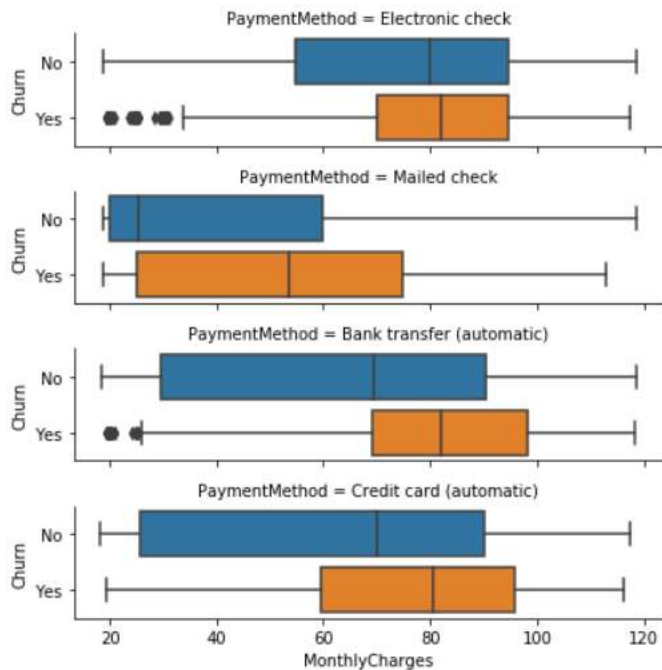
- Bar plot



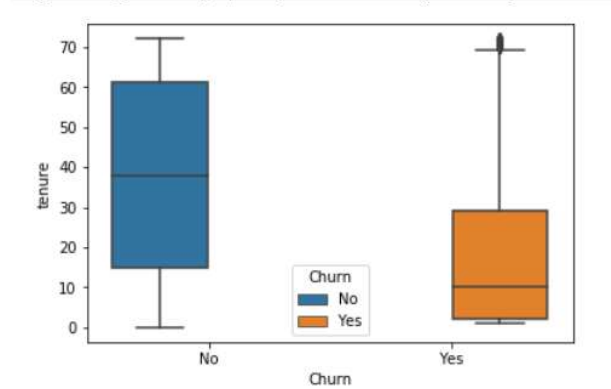
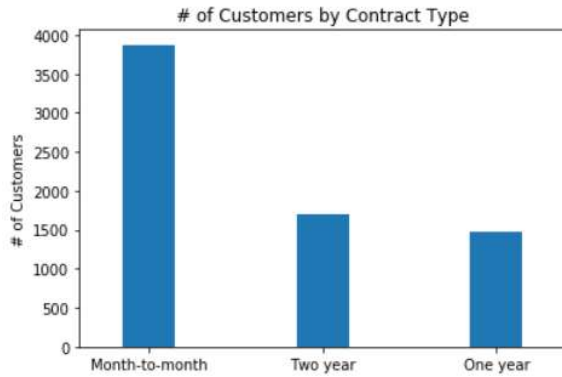
- Customers with multiple lines have a slightly higher churn rate
- Customers without internet have a very low churn rate
- Customers with fiber are more probable to churn than those with DSL connection



- Customers with online Security and tech support has a very low churn rate
- Customers with Streaming services are more likely to churn



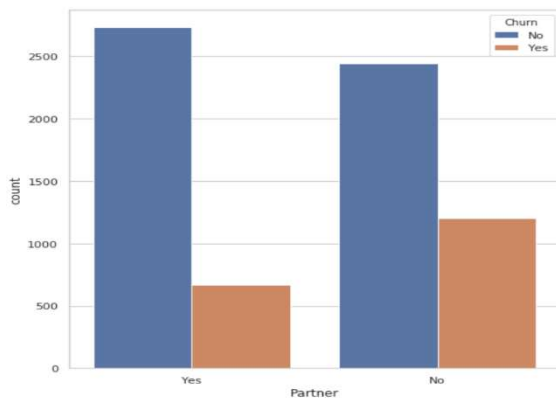
- Customers with automatic payments has high churn rate



- As we can see from this graph most of the customers are in the month to month contract.
- While there are equal number of customers in the 1 year and 2-year contracts.
- Churn rate is high for the customers who has less tenure/contract with Telco company.

Inferential Statistics

- There is no behavior difference between women and men.



- Customer with partners has a very low churn rate when compared with customer without partners.

Statistical summary of data set

	count	mean	std	min	25%	50%	75%	max
gender	7043.0	0.504756	0.500013	0.00	0.00	1.000	1.0000	1.00
SeniorCitizen	7043.0	0.162147	0.368612	0.00	0.00	0.000	0.0000	1.00
Partner	7043.0	0.483033	0.499748	0.00	0.00	0.000	1.0000	1.00
Dependents	7043.0	0.299588	0.458110	0.00	0.00	0.000	1.0000	1.00
tenure	7043.0	32.371149	24.559481	0.00	9.00	29.000	55.0000	72.00
PhoneService	7043.0	0.903166	0.295752	0.00	1.00	1.000	1.0000	1.00
PaperlessBilling	7043.0	0.592219	0.491457	0.00	0.00	1.000	1.0000	1.00
MonthlyCharges	7043.0	64.761692	30.090047	18.25	35.50	70.350	89.8500	118.75
TotalCharges	7032.0	2283.300441	2266.771362	18.80	401.45	1397.475	3794.7375	8684.80
Churn	7043.0	0.265370	0.441561	0.00	0.00	0.000	1.0000	1.00
customerID_0002-ORFBO	7043.0	0.000142	0.011916	0.00	0.00	0.000	0.0000	1.00
customerID_0003-MKNFE	7043.0	0.000142	0.011916	0.00	0.00	0.000	0.0000	1.00
customerID_0004-TLHLJ	7043.0	0.000142	0.011916	0.00	0.00	0.000	0.0000	1.00