customer-churn-data

September 26, 2024

1 Exploratory Data Analysis for Telecom Customer Churn

By-Aneesh Kumar

2 Importing all the important Libraries for Exploratory Data Analysis

[29]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

3 Importing Dataset

[35]:

```
df = pd.read_csv('customer_churn_data.csv')
```

[22]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   CustomerID       1000 non-null   int64
 1   Age              1000 non-null   int64
 2   Gender           1000 non-null   object
 3   Tenure           1000 non-null   int64
 4   MonthlyCharges   1000 non-null   float64
 5   ContractType     1000 non-null   object
 6   InternetService  703 non-null    object
 7   TotalCharges     1000 non-null   float64
 8   TechSupport      1000 non-null   object
 9   Churn            1000 non-null   object
dtypes: float64(2), int64(3), object(5)
memory usage: 78.3+ KB
```

4 Replacing blanks with 0 as tenure is 0 and no totalcharges are recorded

```
[ ]: df["TotalCharges"] = df["TotalCharges"].replace("","0")
     df["TotalCharges"] = df["TotalCharges"].astype("float")
```

5 Changing Datatype for InternetService

```
[60]: df['InternetService'].fillna('NaN', inplace=True)
```

```
[62]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   CustomerID       1000 non-null   int64
 1   Age              1000 non-null   int64
 2   Gender           1000 non-null   object
 3   Tenure           1000 non-null   int64
 4   MonthlyCharges   1000 non-null   float64
 5   ContractType     1000 non-null   object
 6   InternetService  1000 non-null   object
 7   TotalCharges     1000 non-null   float64
 8   TechSupport      1000 non-null   object
 9   Churn            1000 non-null   object
dtypes: float64(2), int64(3), object(5)
memory usage: 78.3+ KB
```

6 Check Null Values

```
[64]: df.isnull().sum()
```

```
[64]: CustomerID         0
      Age                0
      Gender             0
      Tenure             0
      MonthlyCharges     0
      ContractType       0
      InternetService    0
      TotalCharges       0
      TechSupport        0
      Churn              0
      dtype: int64
```

7 Descriptive Statics

[38]:

```
df.describe()
```

[38]:

```
        CustomerID          Age      Tenure  MonthlyCharges  TotalCharges
count  1000.000000  1000.000000  1000.00000     1000.000000   1000.000000
mean    500.500000    44.674000    18.97300       74.391290   1404.364060
std     288.819436     9.797741    18.89257       25.712083   1571.755048
min       1.000000    12.000000     0.00000       30.000000      0.000000
25%     250.750000    38.000000     5.00000       52.357500    345.217500
50%     500.500000    45.000000    13.00000       74.060000    872.870000
75%     750.250000    51.000000    26.00000       96.102500   1900.175000
max    1000.000000    83.000000   122.00000      119.960000  12416.250000
```

[74]:

```
df["CustomerID"].duplicated().sum()
```

[74]:

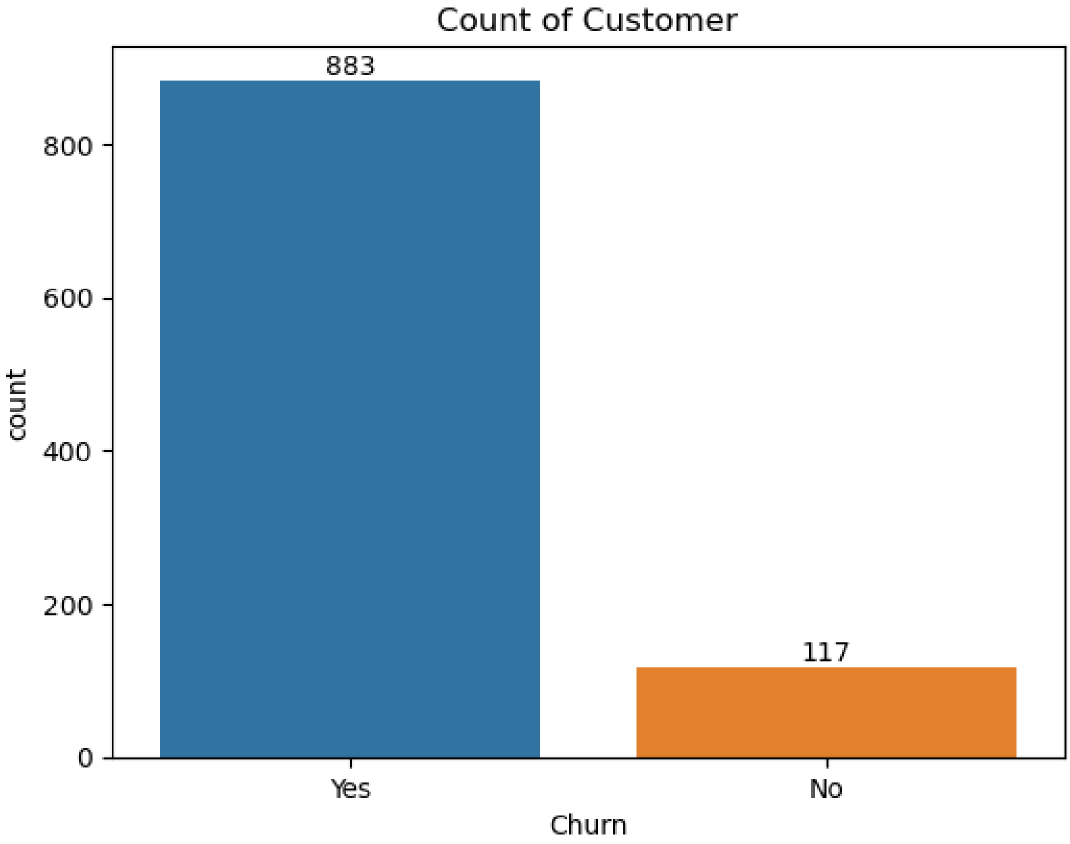
```
0
```

8 Number Of Customers Churned

[95]:

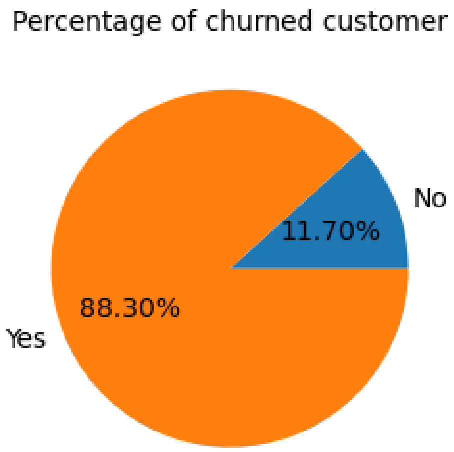
```
ax = sns.countplot(x = 'Churn', data =df)

ax.bar_label(ax.containers[0])
plt.title("Count of Customer")
plt.show()
```



This Bar chart depicts both the count of customers who have churned out and the retained ones

```
[97]: plt.figure(figsize = (3,4))
gb = df.groupby("Churn").agg({'Churn':"count"})
plt.pie(gb['Churn'], labels = gb.index, autopct = "%1.2f%%")
plt.title("Percentage of churned customer", fontsize =10)
plt.show()
```



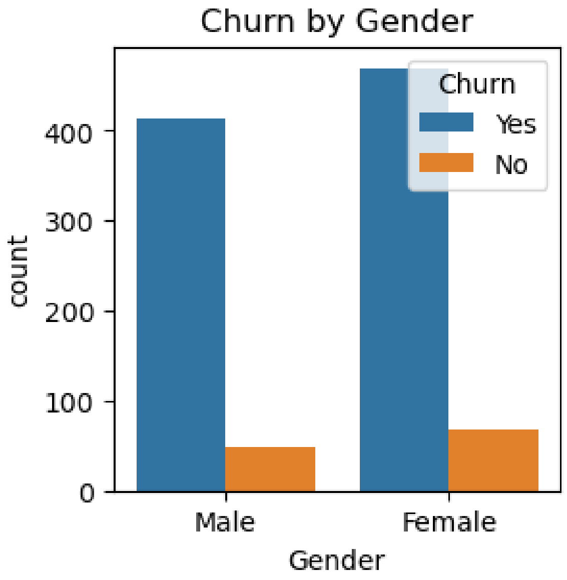
We can clearly observe that about 88.30% of the customers have churned out.

Now let's explore the possible reasons for this

9 Impact of Gender on Churn Rate

[49]:

```
plt.figure(figsize =(3,3))
sns.countplot(x = "Gender", data =df, hue = "Churn")
plt.title("Churn by Gender")
plt.show()
```

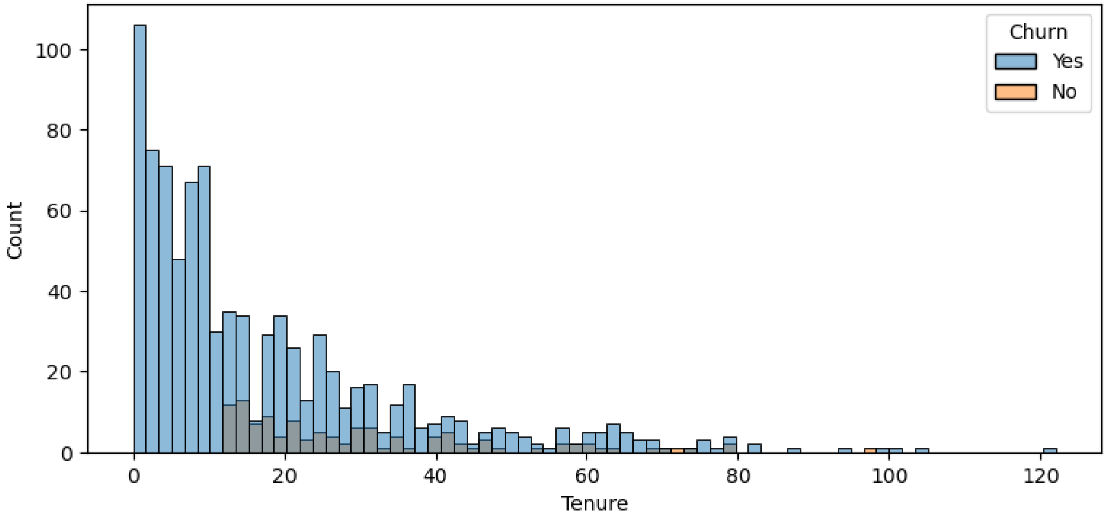


10 Based On Tenure

[68]:

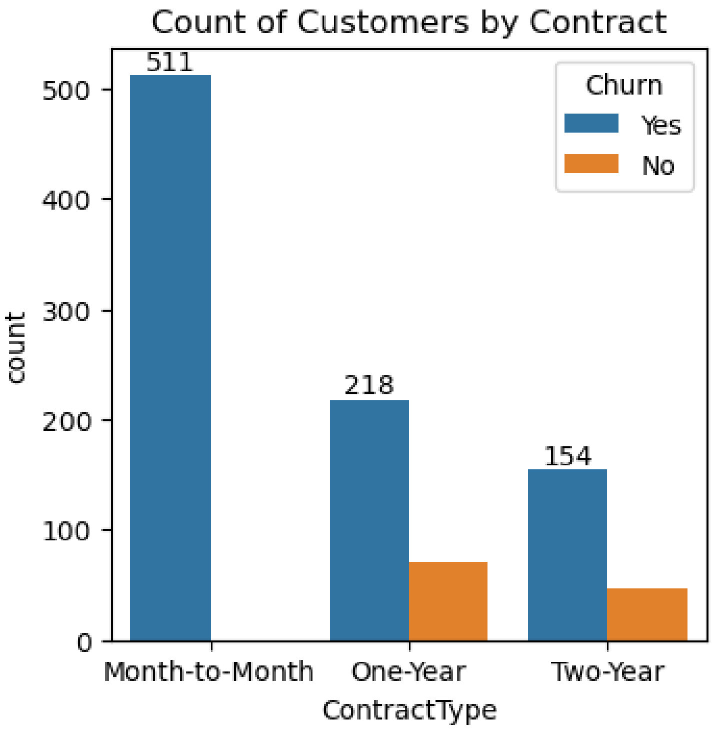
```
plt.figure(figsize =(9, 4))
sns.histplot(x = "Tenure", data = df, bins = 72, hue = "Churn")
plt.show()
```

```
D:\anaconda\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



[74]:

```
plt.figure(figsize = (4,4))
ax = sns.countplot(x = "ContractType", data = df,)
ax.bar_label(ax.containers[0])
plt.title("Count of Customers by Contract")
plt.show()
```



Insights:

Customers having month-to-month billing contract are more likely to churn out in comparison to one year and two year contract

[ ]: