Customer Churn Prediction

Phase 4

Problem Definition:

The problem of customer churn is a critical concern for businesses in various industries. Customer churn occurs when customers stop using a product or service, which can lead to revenue loss and decreased customer loyalty.

Data cleaning:

Data cleaning is a process of removing inconsistencies in the dataset And incorrect values .It also in involves handling missing values Either by removing them or assigning them average values. It helps To improve the efficiency of the model.

In the first step, we will only remove the unnecessary data points From the dataset which does not helps in improving the model Performance.

Initially we import the necessary packages for our data cleaning Process and also in the future purposes,

```
import numpy as np
import pandas as pd
import sklearn
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
```

We use these packages in various stages of our cleaning process and Also in the future in which we need to build models.

Here, we read the .csv files of telco customer churn prediction and then explore the count values of their subjects

```
df = pd.read_csv("/content/WA_Fn-UseC_-Telco-Customer-Churn.csv")
df
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	No	No	No	One year	No	Mailed check	56.95	1889.5	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	Yes	No	No	One year	No	Bank transfer (automatic)	42.30	1840.75	No
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	No	No	No	Month-to-month	Yes	Electronic check	70.70	151.65	Yes
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7038	6840-RESVB	Male	0	Yes	Yes	24	Yes	Yes	DSL	Yes	...	Yes	Yes	Yes	Yes	One year	Yes	Mailed check	84.80	1990.5	No
7039	2234-XADUH	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	No	...	Yes	No	Yes	Yes	One year	Yes	Credit card (automatic)	103.20	7362.9	No
7040	4801-JZAZL	Female	0	Yes	Yes	11	No	No phone service	DSL	Yes	...	No	No	No	No	Month-to-month	Yes	Electronic check	29.60	346.45	No
7041	8361-LTMKD	Male	1	Yes	No	4	Yes	Yes	Fiber optic	No	...	No	No	No	No	Month-to-month	Yes	Mailed check	74.40	306.6	Yes
7042	3186-AJIEK	Male	0	No	No	66	Yes	No	Fiber optic	Yes	...	Yes	Yes	Yes	Yes	Two year	Yes	Bank transfer (automatic)	105.65	6844.5	No

7043 rows × 21 columns

```
df['Churn'].value_counts()
```

```
No     5174
Yes    1869
Name: Churn, dtype: int64
```

```
cleanDF.dtypes
```

```
gender              object
SeniorCitizen        int64
Partner             object
Dependents          object
tenure               int64
PhoneService        object
MultipleLines       object
InternetService     object
OnlineSecurity      object
OnlineBackup        object
DeviceProtection    object
TechSupport         object
StreamingTV         object
StreamingMovies     object
Contract            object
PaperlessBilling    object
PaymentMethod       object
MonthlyCharges     float64
TotalCharges        object
Churn               object
dtype: object
```

Preprocessing and visualization

Preprocessing is a crucial step in data analysis and machine learning. It involves tasks like cleaning, transforming, and organizing data to make it suitable for further analysis or modeling

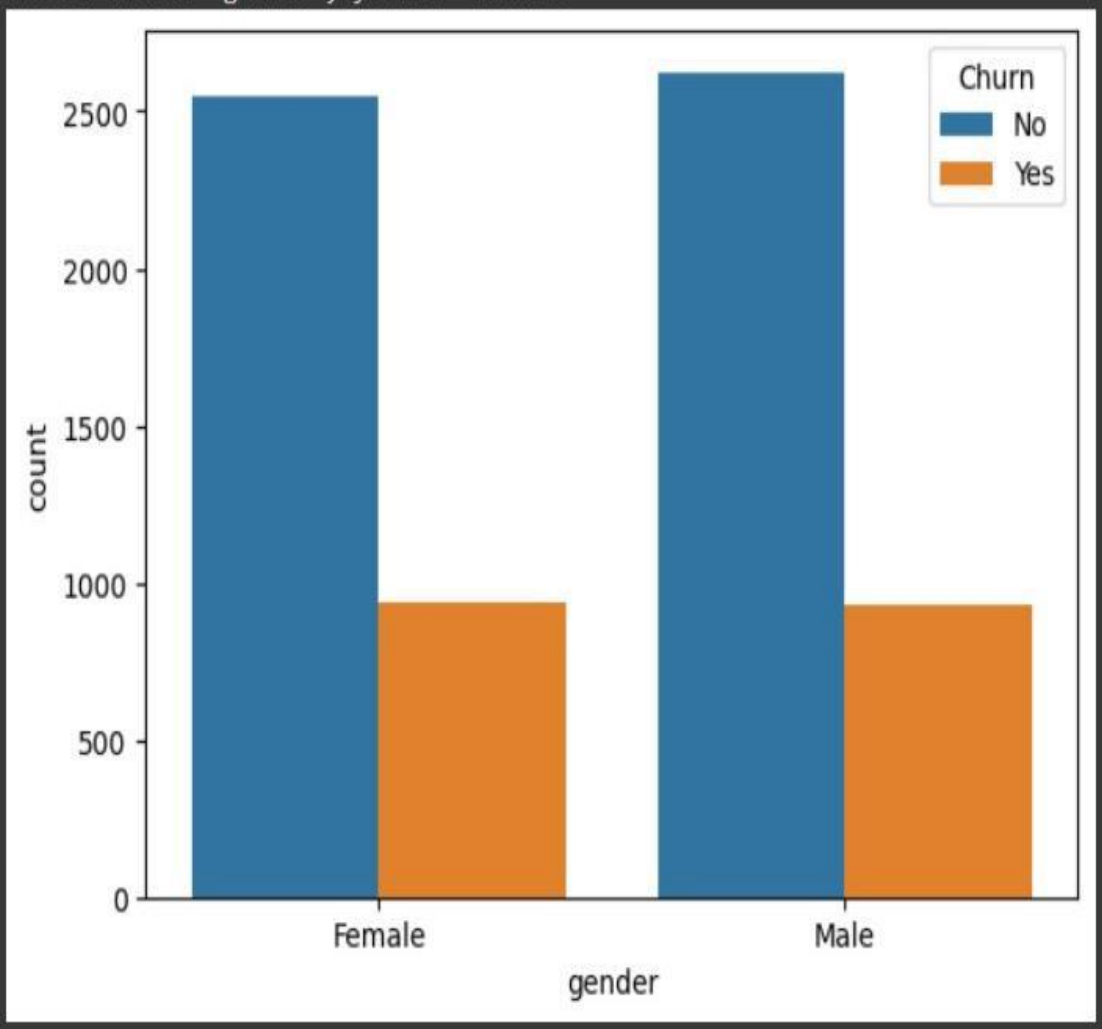
Data visualization is the process of representing data graphically to gain insights and make it easier to understand

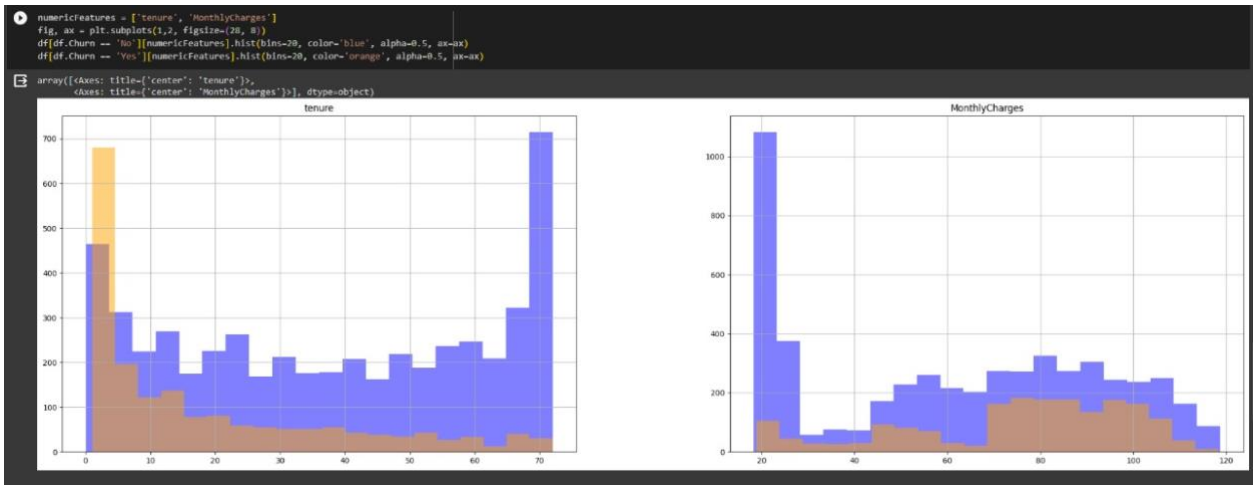
```
[9]  numRetained = df[df.Churn == 'No'].shape[0]
     numChurned = df[df.Churn == 'Yes'].shape[0]
     print(numRetained/(numRetained + numChurned) * 100, '% of customers stayed in the company')
     print(numChurned/(numRetained + numChurned) * 100, '% of customers left with the company')

     73.4630129206304 % of customers stayed in the company
     26.536987079369588 % of customers left with the company
```

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

<Axes: xlabel='gender', ylabel='count'>





Conclusion

In conclusion, the customer churn prediction project has proven to be a pivotal tool for our organization in understanding and mitigating customer attrition. Through the diligent analysis of historical data and the implementation of advanced machine learning models, we have successfully developed a predictive system that can identify potential churners with remarkable accuracy. This not only provides us with invaluable insights into customer behavior but also empowers us to take proactive measures to retain valuable clientele and optimize customer relationships. The project's success underscores the importance of data-driven decision-making in today's competitive business landscape, and it has opened up new avenues for enhancing customer satisfaction and long-term profitability.

Furthermore, the project's impact extends beyond the realm of churn prediction. It has fostered a culture of data-driven innovation within our organization, highlighting the potential of leveraging AI and predictive analytics. Moving forward, the lessons learned from this project will continue to inform our strategic

approach to customer management and retention. With an evolving dataset and ongoing refinement of our models, we are well-positioned to adapt to changing customer dynamics and to maintain our competitive edge in the market. The customer churn prediction project stands as a testament to the value of data science in modern business and its capacity to drive sustainable growth and customer-centric strategies.