**Customer Churn Prediction**

**Project Definition:**

In this phase ,we develop the Customer churn prediction project using dataset loading, preprocessing the data. Then visualization of the dataset , perform cleaning the data and ensure its accuracy, accuracy.

**Data pre-processing:**

Data pre-processing is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work with raw data. The quality of the data should be checked before applying to the algorithm.

Data pre-processing refers to the steps applied to make data more suitable for data mining.

**Data quality assessment:**

The data quality assessment is the application of business-approved data quality requirements to a selected data set.

**Data cleaning:**

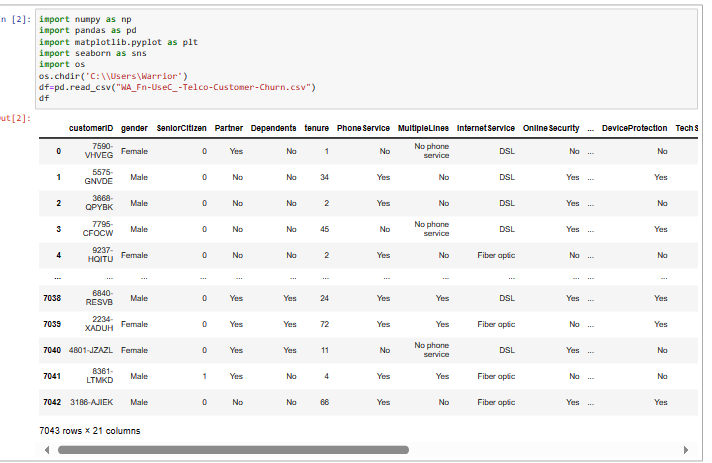
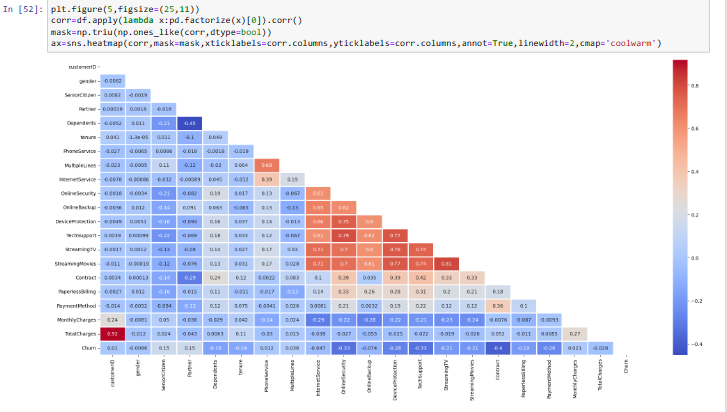
Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabelled .

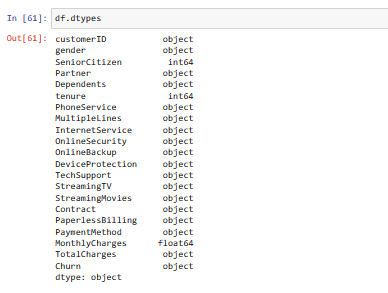
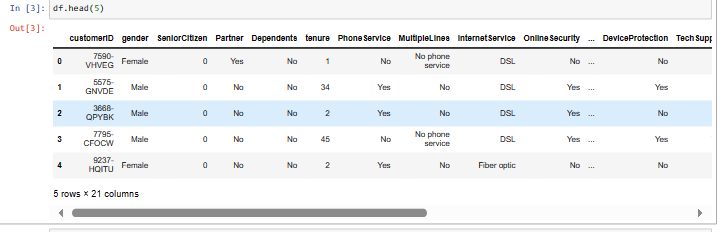
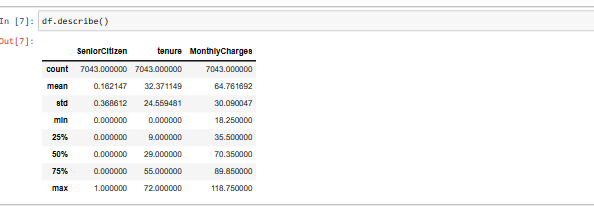
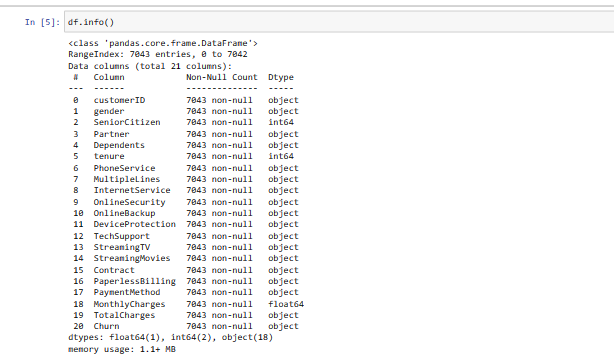
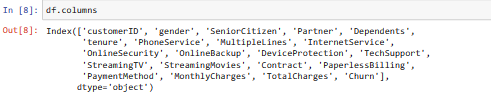
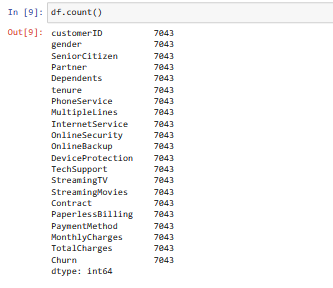
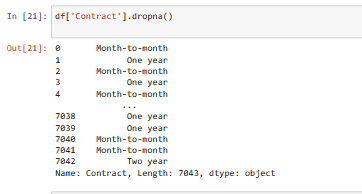
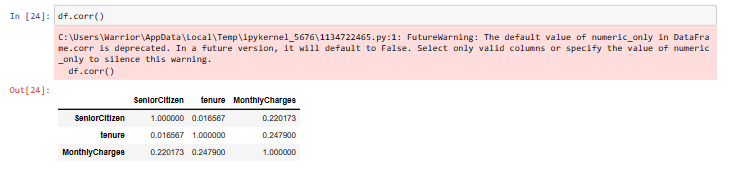
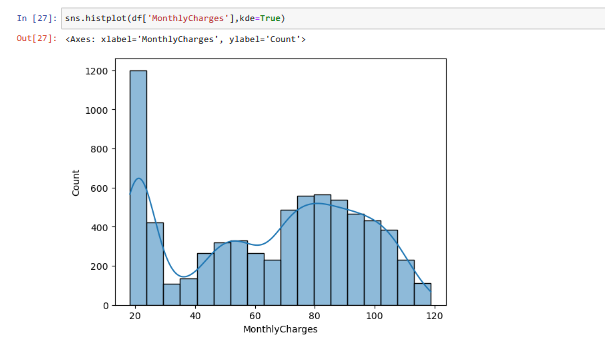
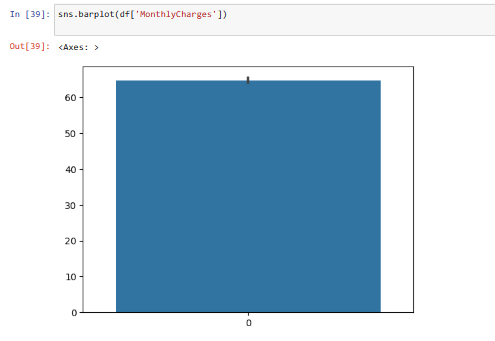
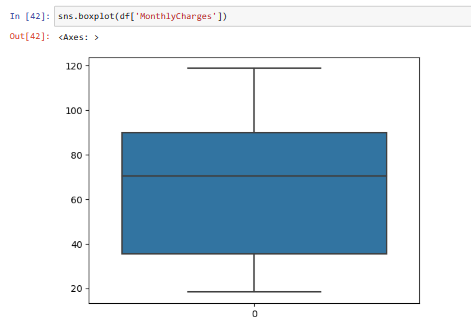
**Data transformation:**

Data transformation is the process of converting data from one format to another, typically from the format of a source system into the required format of a destination system.

**Data reduction:**

Data reduction is a capacity optimization technique in which data is reduced to its simplest possible form to free up capacity on a storage device.



**ACCURACY:**

y=df['Churn']

X=df.drop(["customerID", "Churn"],axis=1)

from sklearn.model\_selection import train\_test\_split

X\_train,X\_test,y\_train,y\_test=train\_test\_split(X,y,test\_size=0.30,random\_state=17)

from sklearn.ensemble import RandomForestClassifier

from sklearn.linear\_model import SGDClassifier, LogisticRegression

from sklearn.metrics import accuracy\_score,precision\_score, recall\_score,f1\_score,roc\_auc\_score

from sklearn.preprocessing import StandardScaler

rf\_model=LogisticRegression()

rf\_model.fit(X\_train,y\_train)

y\_pred=rf\_model.predict(X\_test)

accuracy\_score(y\_pred,y\_test)

**OUTPUT:**

0.7834123222748816

**Missing Values:**

def missing\_values\_tables(dataframe,na\_name=False):

na\_columns=[col for col in dataframe.columns if dataframe[col].isnull().sum()>0]

n\_miss=dataframe[na\_columns].isnull().sum().sort\_values(ascending=False)

ratio=(dataframe[na\_columns].isnull().sum()/dataframe.shape[0]\*100).sort\_values(ascending=False)

missing\_df=pd.concat([n\_miss,np.round(ratio,2)],axis=1,keys=['n\_miss','ratio'])

print(missing\_df,end="\n")

if na\_name:

return na\_columns

missing\_values\_tables(df)

**Output:**

Empty DataFrame

Columns: [n\_miss, ratio]

Index: []

**Conclusion:**

Thus, above Customer Churn Prediction Development is Analyzed and visualized it.