1 . IMPORTING REQUIRED PACKAGE

```
import pandas as pd
import seaborn as sns
import numpy as np
from matplotlib import pyplot as plt
%matplotlib inline
```

2 . LOADING THE DATASET

```
df = pd.read_csv("notebooks/Churn_Modelling.csv")
df
```

3 . VISUALIZATIONS

```
sns.displot(df.Tenure)
```

3 . 1 . UNIVARIATE ANALYSIS

```
df.plot.line()
```

3 . 2 . BI-VARIATE ANALYSIS

```
df.plot.line()
```

3 . 3 . MULTI - VARIATE ANALYSIS

```
sns.lmplot("Age","NumOfProducts",df,hue="NumOfProducts",
fit_reg=False);
```

4 . PERFORM DESCRIPTIVE STATISTICS ON THE DATASET.

```
df.describe()
```

5 . HANDLE THE MISSING VALUES.

```
data = pd.read_csv("Churn_Modelling.csv")
pd.isnull(data["Gender"])
```

6 . FIND THE OUTLIERS AND REPLACE THE OUTLIERS.

```
df["Tenure"] = np.where(df["Tenure"] >10, np.median,df["Tenure"])
df["Tenure"]
```

7 . CHECK FOR CATEGORICAL COLUMNS AND PERFORM ENCODING.

```
pd.get_dummies(df, columns=["Gender", "Age"], prefix=["Age",
"Gender"]).head()
```

8 . SPLIT THE DATA INTO DEPENDENT AND INDEPENDENT VARIABLES

8 . 1 . SPLIT THE DATA INTO INDEPENDENT VARIABLES.

```
X = df.iloc[:, :-2].values
print(X)
```

8 . 2 . SPLIT THE DATA INTO DEPENDENT VARIABLES.

```
Y = df.iloc[:, -1].values
print(Y)
```

9 . SCALE THE INDEPENDENT VARIABLES

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
df[["RowNumber"]] = scaler.fit_transform(df[["RowNumber"]])
print(df)
```

10 . SPLIT THE DATA INTO TRAINING AND TESTING

```
from sklearn.model_selection import train_test_split
train_size=0.8
X = df.drop(columns = ['Tenure']).copy()
y = df['Tenure']
X_train, X_rem, y_train, y_rem = train_test_split(X,y, train_size=0.8)
test_size = 0.5
X_valid, X_test, y_valid, y_test = train_test_split(X_rem,y_rem,
test_size=0.5)
print(X_train.shape), print(y_train.shape)
print(X_valid.shape), print(y_valid.shape)
print(X_test.shape), print(y_test.shape)
```