

Assignment-2

Student Name : ENIYAN T S

Student Roll Number: 111519106033

```
#1 load the data set
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt #charts
import seaborn as sns

df = pd.read_csv('Churn_Modelling.csv')
df.head()
df['NumOfProducts'].unique()
plt.bar(df['Gender'], df['Tenure'])

sns.countplot(df['Gender'])
sns.boxplot(df['Tenure'])

#2 (i)univariate analysis
plt.plot(df['Balance'], np.zeros_like(df['Balance']), 'o')

#2 (ii)bivariate analysis

sns.FacetGrid(df, hue="Gender", size=5).map(plt.scatter, "Gender", "Balance").
add_legend()
plt.show()

#2 (iii)multivariate analysis
sns.pairplot(df, hue="Gender", height=2)

#3 descriptive analysis
```

```

print(df.describe())

#4 Handle the Missing values
df.isnull().sum()

#5 Find the outliers and replace the outliers
df['Balance'].median(), df['Age'].mean()
df['Balance']=df['Balance'].fillna(df['Balance'].median())
df

#6 Check for Categorical columns and perform encoding
obj_df = df.select_dtypes(include=['object']).copy()
obj_df

from sklearn.preprocessing import LabelEncoder
from collections import Counter as count
le=LabelEncoder()
count(df['Geography'])

df['Exited']=df['Exited'].replace([1, 0], ('Yes', 'No'))
df

df['Exited']=df['Exited'].replace([1, 0], ('Yes', 'No'))
df

df1=pd.read_csv('/content/Churn_Modelling.csv')
df1['Geography']

df1=pd.read_csv('/content/Churn_Modelling.csv')
df1['Geography']

#7 Split the data into dependent and independent variables
x=df.iloc[:,0:3]
y=df['HasCrCard']
x
y

#8 training the data
from sklearn.model_selection import train_test_split

```

```
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.6,random_state=
11)
```

```
xtrain
```

```
xtest
```

```
ytrain
```

```
ytest
```