Assignment 2

Artificial Intelligence

Assignment Date	20-09-2022
Student Name	M Theepiga
Register Number	921319104212
Maximum Marks	2 marks

```
from google.colab import drive
drive.mount('/content/drive')
import pandas as pd
import numpy as np
import sklearn as sk
import seaborn as sns
data=pd.read_csv("/content/Churn_Modelling Dataset2.csv")
df=data.head(10)
```

1. **Univariate Analysis**

```
import matplotlib.pyplot as plt
plt.bar (df['Age'],4)
```

Bivariate Analysis

```
plt.scatter(df['Age'],df['CreditScore'])
```

Multivariate analysis

```
plt.scatter(df['Age'],df['CreditScore'],df['Tenure'])
```

Descriptive statistics on the dataset

```
data.describe()
```

2. **Handling Missing values**

```
data.isnull().sum()
```

3. **Find the outliers and replace the outliers**

Finding Outliners

```
sns.boxplot(data['Age'])
```

Replacing outliners

```
q=data.quantile(q=[0.75,0.5])
iqr=q.iloc[0]-q.iloc[1]
iqr
l=q.iloc[1]-(1.5*iqr)
l['Age']
u=q.iloc[1]+(1.5*iqr)
u['Age']
data['Age']=np.where(data['Age']>u['Age'],u['Age'],np.where(data['Age']
<l['Age'],l['Age'],data['Age']))
sns.boxplot(data['Age'])
```

4. **Check for Categorical columns and perform encoding**.

```
df.info()
from sklearn.preprocessing import LabelEncoder
from collections import Counter as count
le=LabelEncoder()
data['Surname']=le.fit_transform(data['Surname'])
data
data['Geography']=le.fit_transform(data['Geography'])
data['Gender']=data['Gender'].replace(['Male','Female'],[0,1])
data
```

5. **Split the data into dependent and independent variables.**

Independent Variables

```
x=data.iloc[:,0:13]
x
```

Dependent Variables

```
y=data['Exited']
y
```

Scale the independent variables

```
from sklearn.preprocessing import scale
scale(x)
```

Split the data into training and testing

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
x_train
x_train.shape
y_train
y_train.shape
x_test
x_test.shape
y_test
y_test.shape
```