Assignment Date	19 September 2022
Student Name	R deepa
Student Roll Number	962719106007
Maximum Mark	2 Marks

Question –1

1. Download the dataset:

Question-2

2. Load the dataset.

Solution:

import numpy as np

import pandas as pd
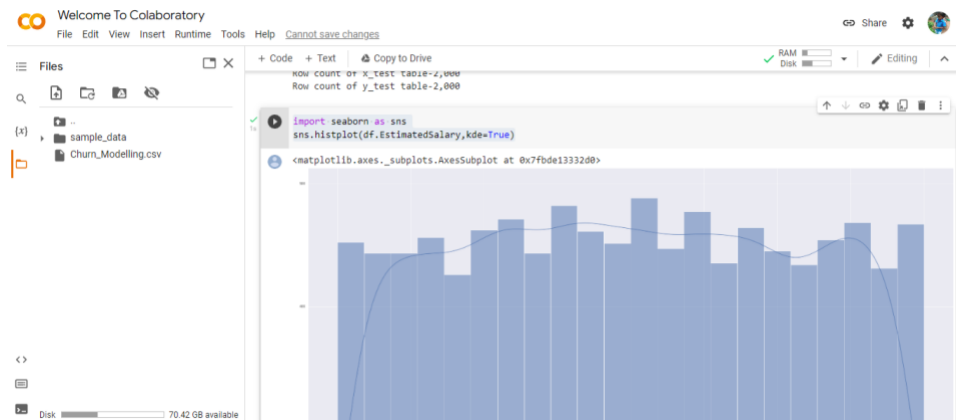
df = pd.read_csv("/content/Churn_Modelling.csv")

Question_3

3. Perform Below Visualizations.

● Univariate Analysis

Solution:

```python
import seaborn as sns

sns.histplot(df.EstimatedSalary,kde=True)
```


● Bi - Variate Analysis

Solution:

```python
import seaborn as sns

sns.scatterplot(df.Balance,df.EstimatedSalary)

plt.ylim(0,15000)
```
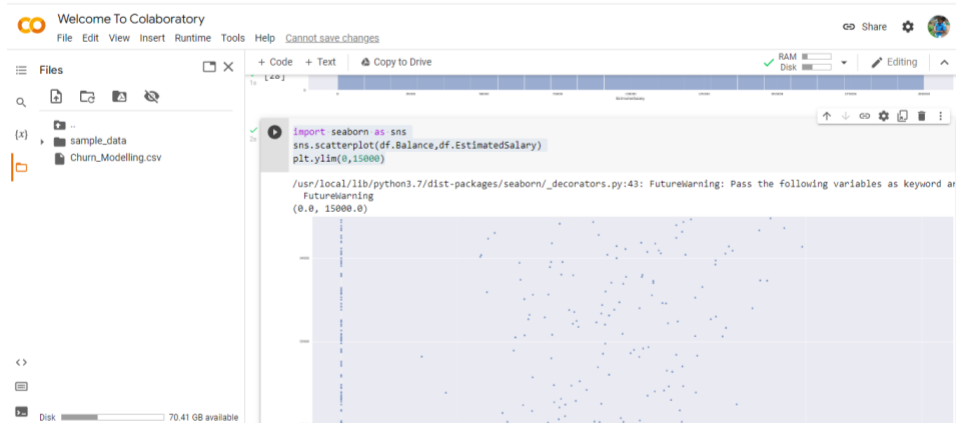
● Multi - Variate Analysis

Solution:

import seaborn as sns

df=pd.read_csv("/content/Churn_Modelling.csv")

sns.pairplot(df)

Question_4

4. Perform descriptive statistics on the data set

Solution:

df=pd.read_csv("/content/Churn_Modelling.csv")

df.describe(include='all')

Question_5

5. Handle the Missing values.

Solution:

```python
from ast import increment_lineno
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(color_codes=True)
df=pd.read_csv("/content/Churn_Modelling.csv")
df.head()
```
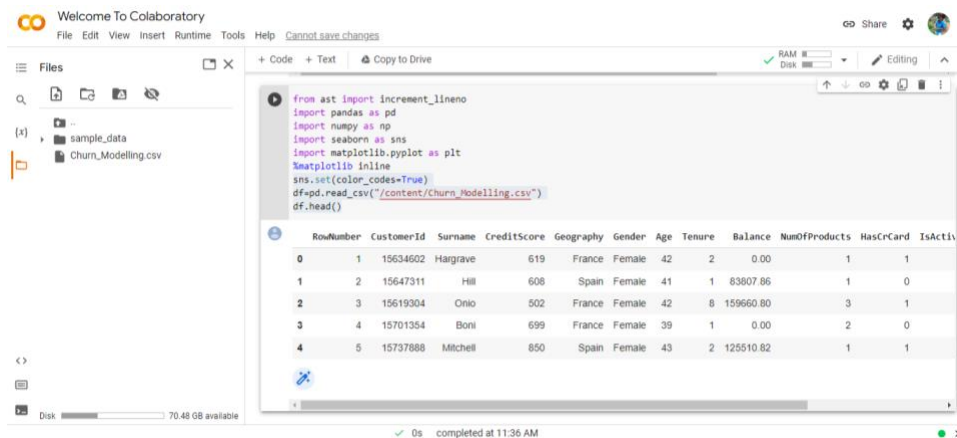
Question_6

6. Find the outliers and replace the outliers

Solution:

import pandas as pd

import matplotlib

from matplotlib import pyplot as pyplot

%matplotlib inline

matplotlib.rcParams['figure.figsize']=(10,6)

df=pd.read_csv("/content/Churn_Modelling.csv")

df.sample(5)

Question_7

7. Check for Categorical columns and perform encoding.

Solution:

df=pd.read_csv("/content/Churn_Modelling.csv")

df.columns

import pandas as pd

import numpy as np

headers=['RowNumber','CustomerID','Surname','CreditScore','Geography',

'Gender','Age','Tenure','Balance','NumofProducts','HasCard'

'IsActiveMember','EstimatedSalary','Exited']

import seaborn as sns

df.head()

Questioh_8

8. Split the data into dependent and independent variables.

Solution:

x=df.iloc[:,:-1].values

print(x)

y=df.iloc[:,-1]._values

print(y)

Question_9

9. Scale the independent variables

Solution:

import seaborn as sns

df=pd.read_csv("/content/Churn_Modelling.csv")

dff=df[['Balance','Age']]

```
sns.heatmap(dff.corr(), annot=True)

sns.set(rc={'figure.figsize':(40,40)})
```

Question_10

10. Split the data into training and testing

Solution:

```
from scipy.sparse.construct import random

x=df.iloc[:, 1:2].values

y=df.iloc[:,2].values

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,random_state=0)

print('Row count of x_train table'+'-'+str(f"{len(x_train):,}"))

print('Row count of y_train table'+'-'+str(f"{len(y_train):,}"))

print('Row count of x_test table'+'-'+str(f"{len(x_test):,}"))

print('Row count of y_test table'+'-'+str(f"{len(y_test):,}"))
```