# Assignment -2

Data Visualization and Preprocessing

|  |  |
| --- | --- |
| Assignment Date | 19 September 2022 |
| Student Name | KOULURU NANDA KISHORE REDDY |
| Student Roll Number | 211419104142 |
| Maximum Marks | 2 Marks |

# Question-1:

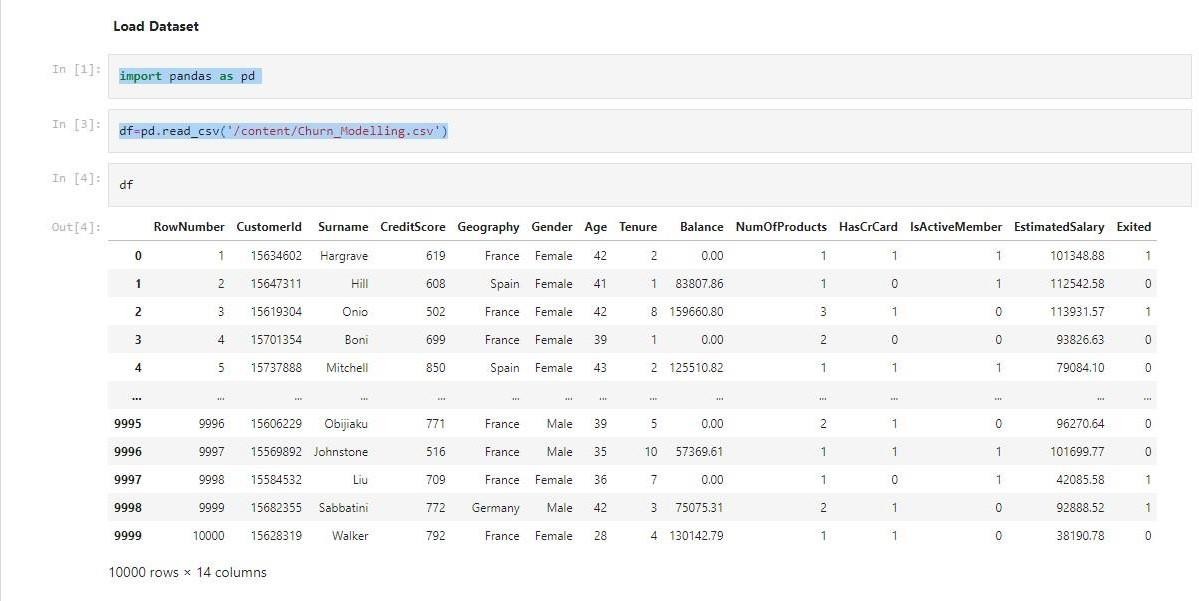
Download the dataset:

# Question-2:

Load the dataset.

# Solution:

import pandas **as** pd df**=**pd**.**read\_csv('/content/Churn\_Modelling.csv')



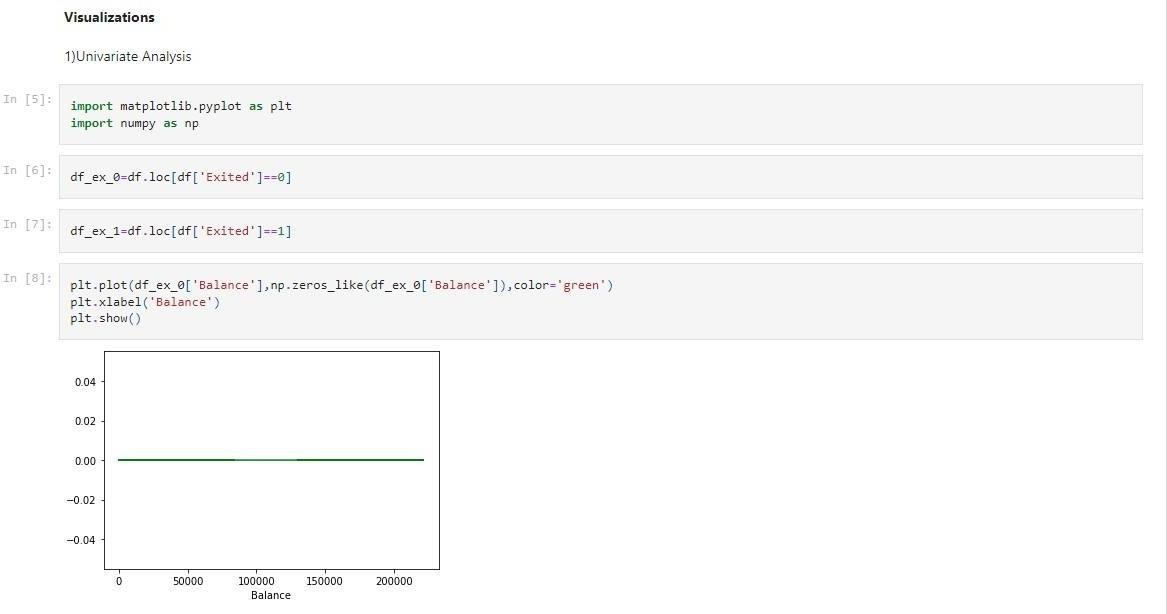
# Question-3:

Perform Below Visualizations. 1)Univariate Analysis

# Solution:

import matplotlib.pyplot **as** plt import numpy **as** np df\_ex\_0**=**df**.**loc[df['Exited']**==**0] df\_ex\_1**=**df**.**loc[df['Exited']**==**1]

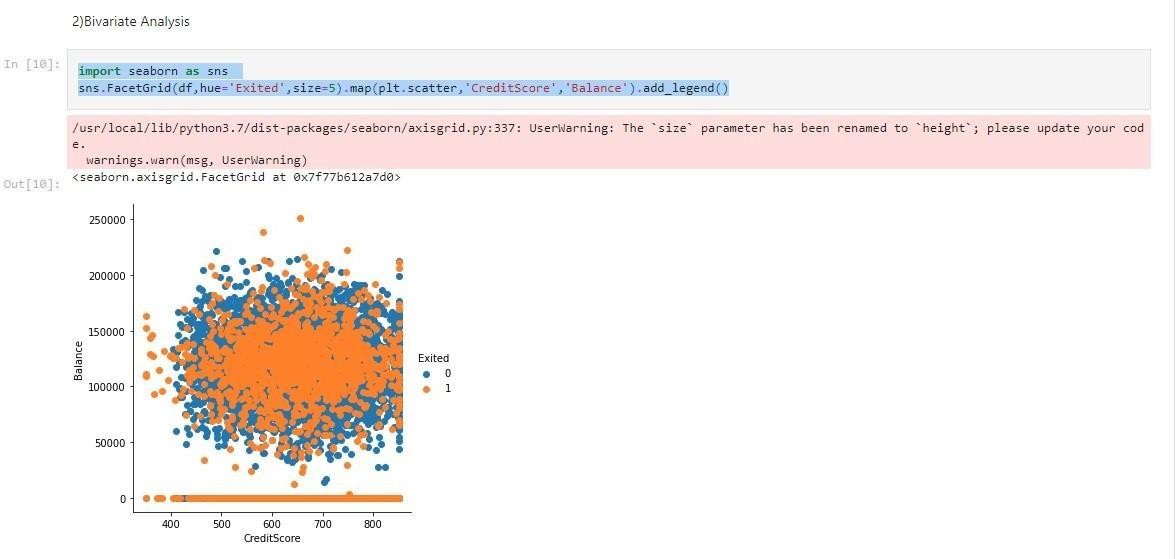
plt**.**plot(df\_ex\_0['Balance'],np**.**zeros\_like(df\_ex\_0['Balance']),color**=**'green') plt**.**xlabel('Balance') plt**.**show()



2)Bi - Variate Analysis

# Solution:

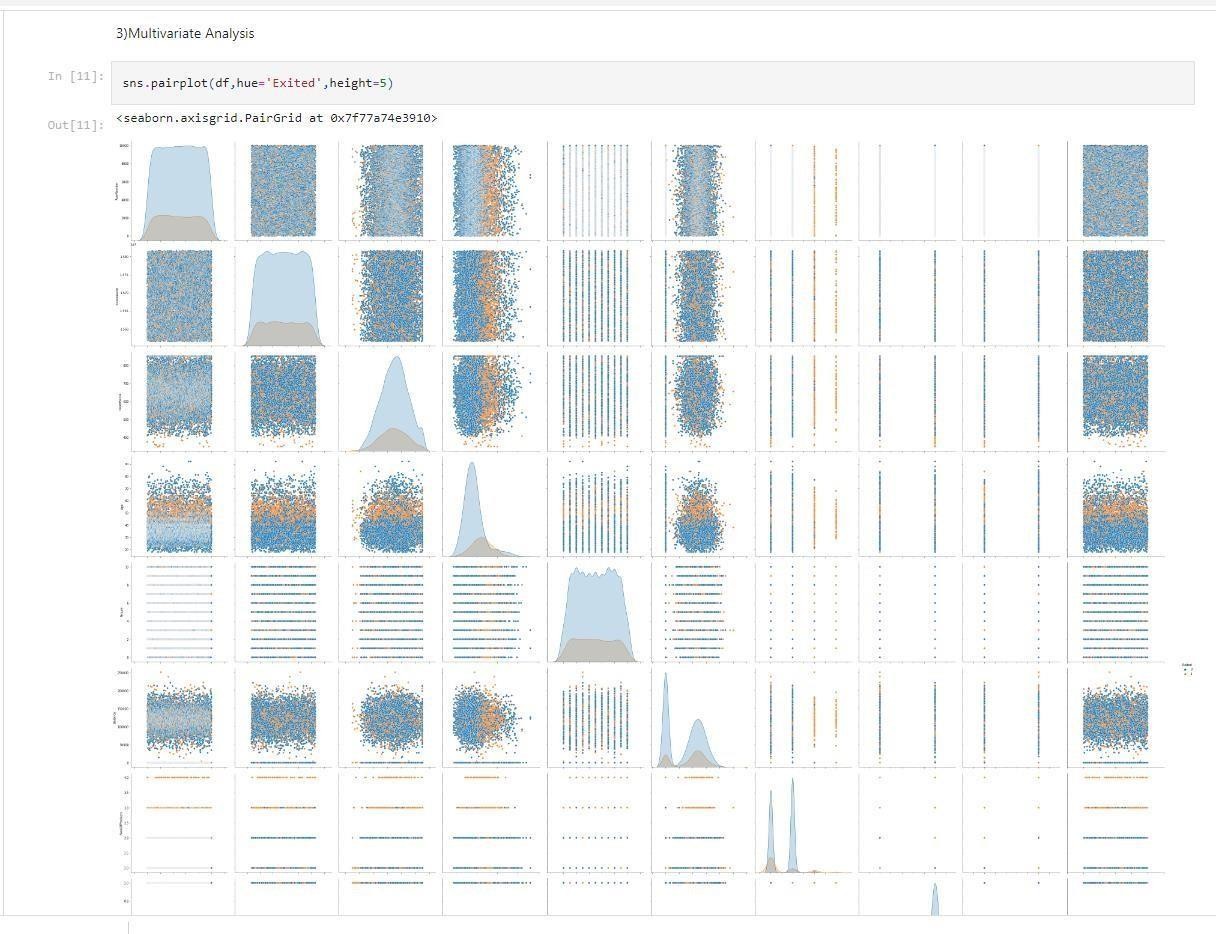
import seaborn as sns sns**.**FacetGrid(df,hue**=**'Exited',size**=**5)**.**map(plt**.**scatter,'CreditScore','Balance')**.**add\_legend()



1)Multivariate Analysis

# Solution:

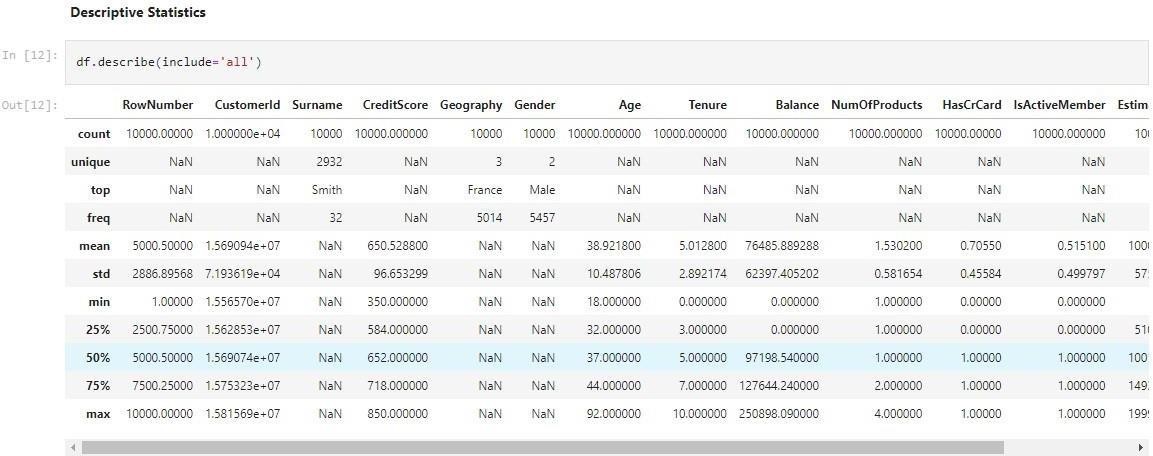
sns**.**pairplot(df,hue**=**'Exited',height**=**5)



# Question-4:

Perform descriptive statistics on the dataset.

**Solution:** df**.**describe(include**=**'all')

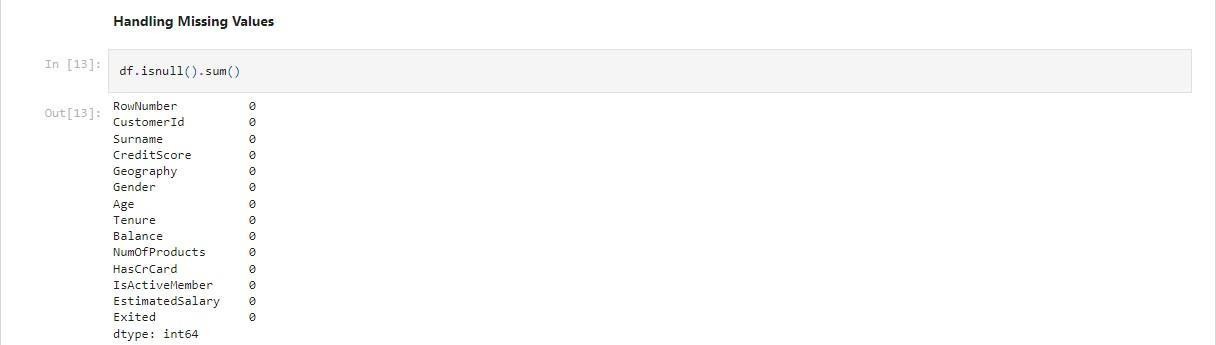


# Question-5:

Handle the Missing values.

# Solution:

df**.**isnull()**.**sum()



# Question-6:

Find the outliers and replace the outliers

# Solution:

**import** seaborn **as** sns sns**.**boxplot(df['Balance'])

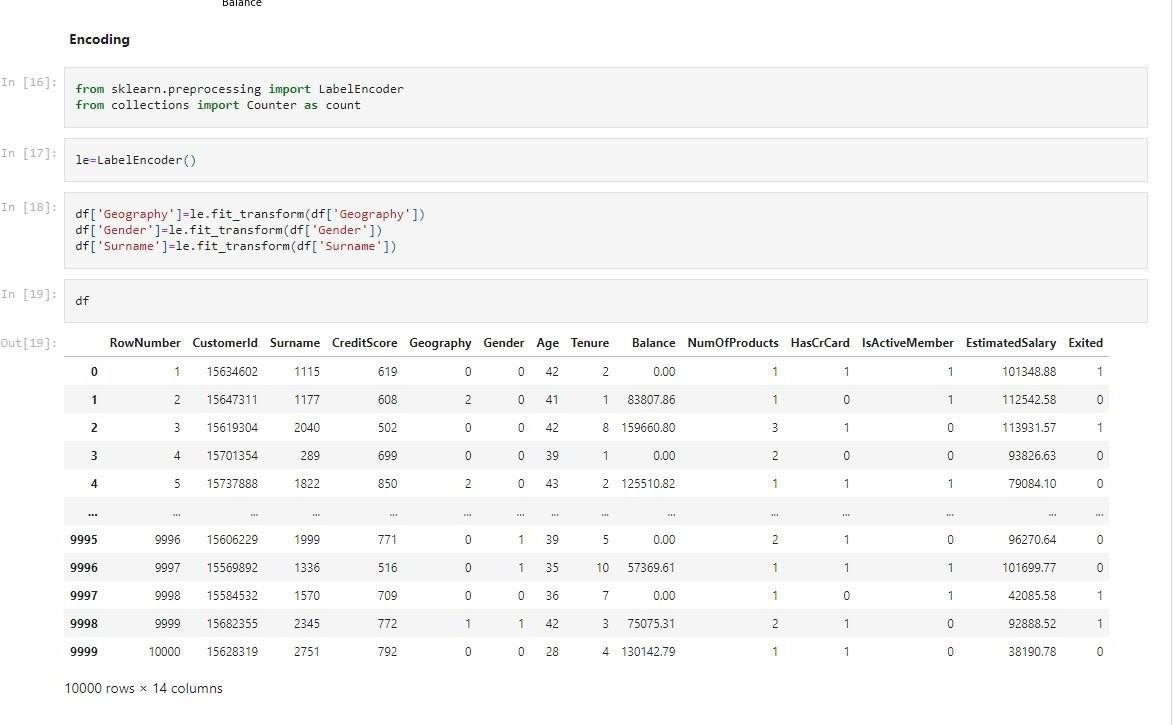


# Question-7:

Check for Categorical columns and perform encoding.

# Solution:

**from** sklearn.preprocessing **import** LabelEncoder **from** collections **import** Counter **as** count le**=**LabelEncoder() df['Geography']**=**le**.**fit\_transform(df['Geography']) df['Gender']**=**le**.**fit\_transform(df['Gender']) df['Surname']**=**le**.**fit\_transform(df['Surname'])



# Question-8:

Split the data into dependent and independent variables.

# Solution:

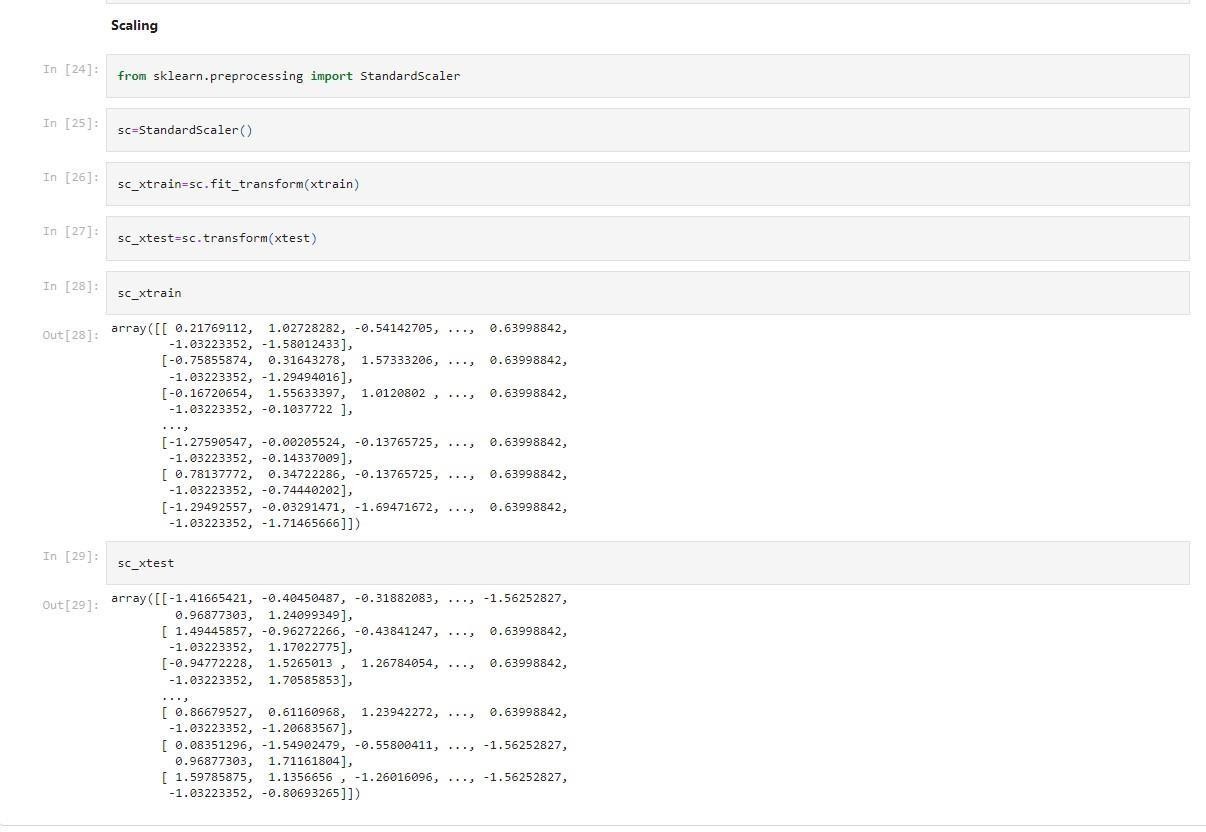
x**=**df**.**iloc[:,0:13] y**=**df['Exited']



# Question-9:

Scale the independent variables

**Solution:** from sklearn.preprocessing import StandardScaler sc**=**StandardScaler() sc\_xtrain**=**sc**.**fit\_transform(xtrain) sc\_xtest**=**sc**.**transform(xtest)



# Question-10:

Testing and training data

**Solution:** from sklearn.model\_selection import train\_test\_split

xtrain,xtest,ytrain,ytest=train\_test\_split(x,y,test\_size=0.3,random\_state=10)

