**Assignment -2**

Python Programming

| Assignment Date |  |
| --- | --- |
| Student Name | Ruthvika bojjala |
| Student Roll Number | 111519104013 |
| Maximum Marks | 2 Marks |

# Question-1:

Download the dataset: Dataset

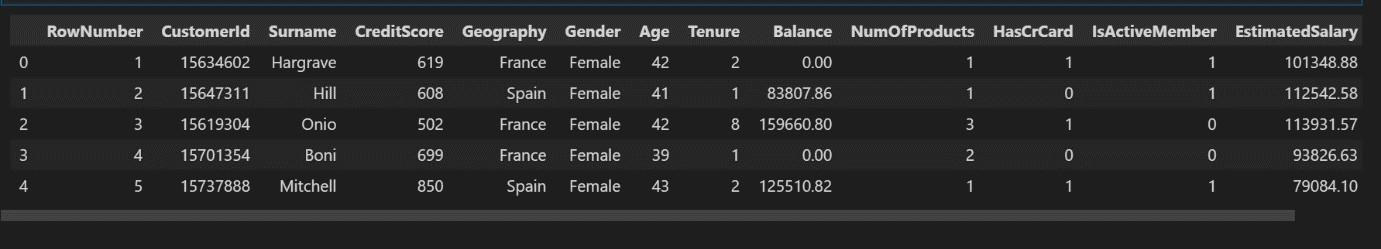
| **Solution:** |
| --- |
| Downloaded successfully |

# Question-2:

Load the dataset.

| **Solution:** |
| --- |
| import pandas as pd  importnumpyas np |



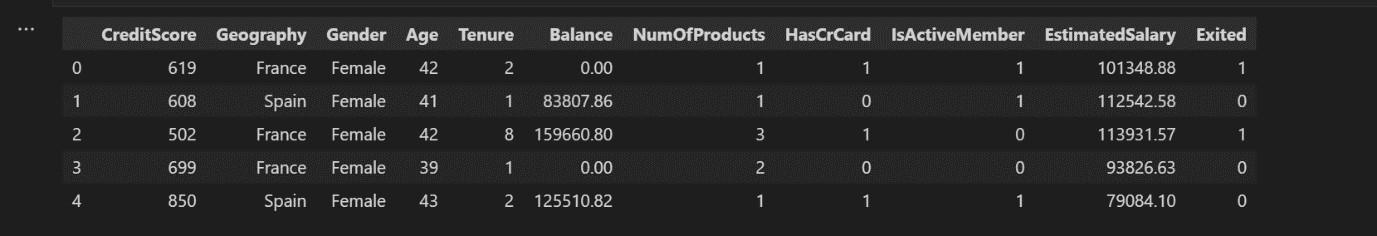


df['HasCrCard'] = df['HasCrCard'].astype('category')



df = df.drop(columns=['RowNumber', 'CustomerId', 'Surname'])

df.head()



# Question 3:

Perform Below Visualizations:

Univariate Analysis, Bi - Variate Analysis, Multi - Variate Analysis

# Solution:



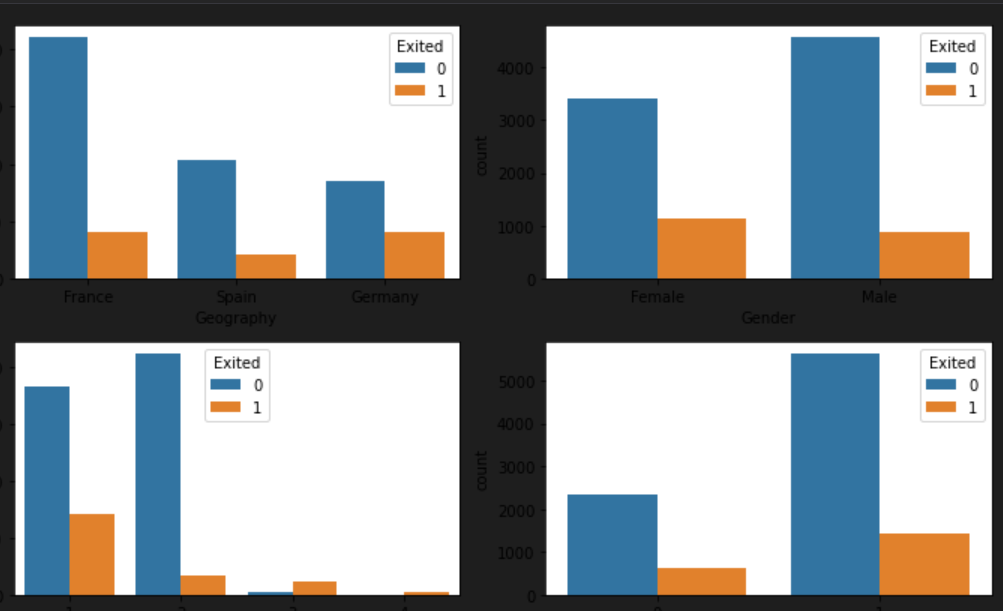
|  | index | Exited |
| --- | --- | --- |
| 0 | 0 | 0.7963 |
| 1 | 1 | 0.2037 |



The data is significantly imbalanced







# Question 4:

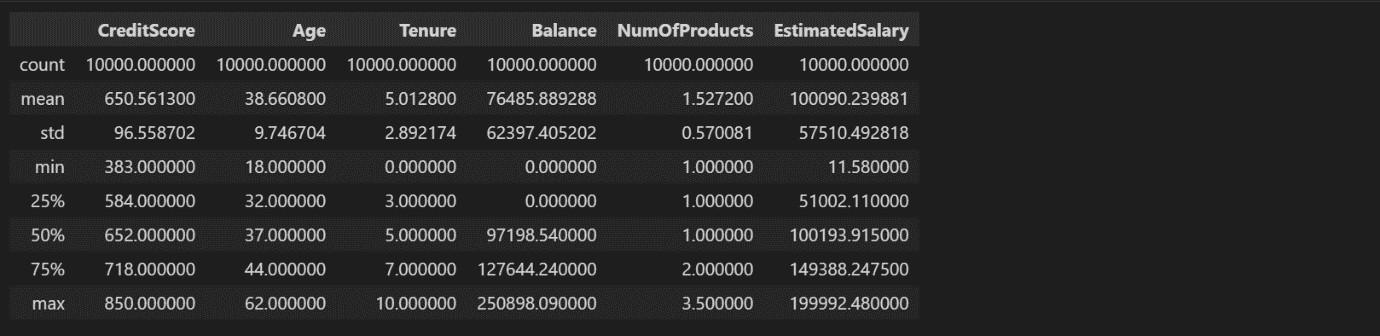
Perform descriptive statistics on the dataset.

# Solution:

df.info()



df.describe()



# Question 5:

Handle the Missing values.

# Solution:

df.isna().sum()



There is no missing values in dataset

foriindf:



unique of Geography is 3 they are {‘France’,’Germany’,’Spain’} unique of Gender is 2 they are {‘Male’,’Female’}

unique of Has CrCard is 2 they are {0,1} unique of Is Active Member is 2 they are {0,1} unique of Exited is 2 they are {0,1}

# Question 6:

Find the outliers and replace the outliers.

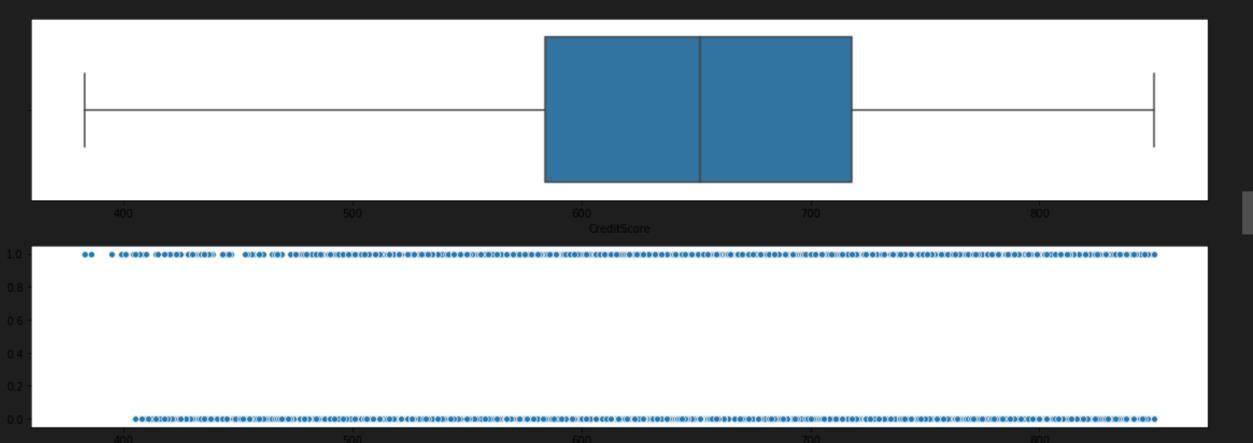
# Solution:

Checking for outliers



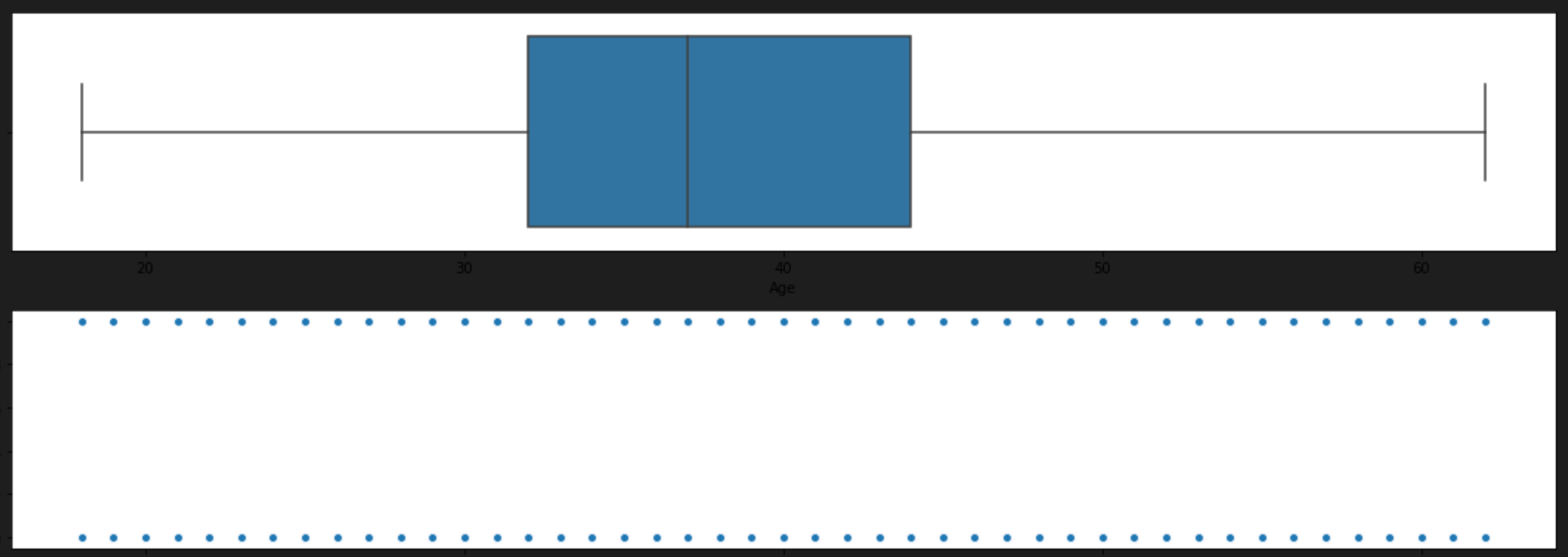


# of bivariate Outliers:19



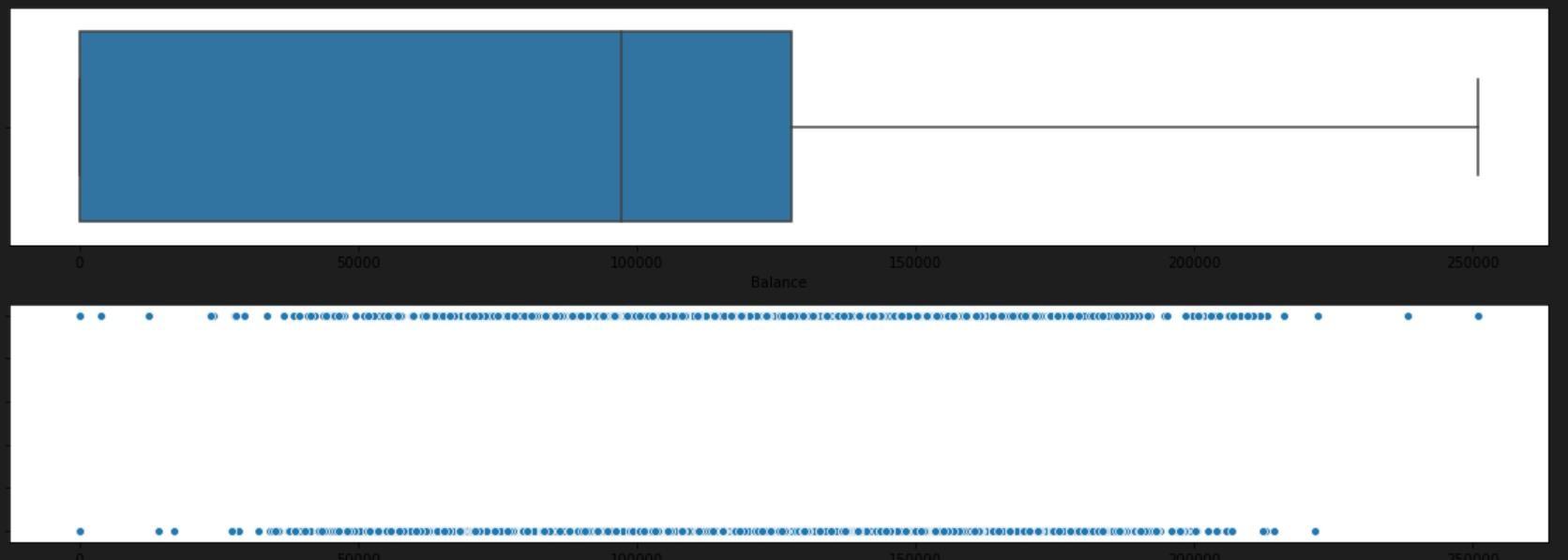


# of bivariate Outliers:0

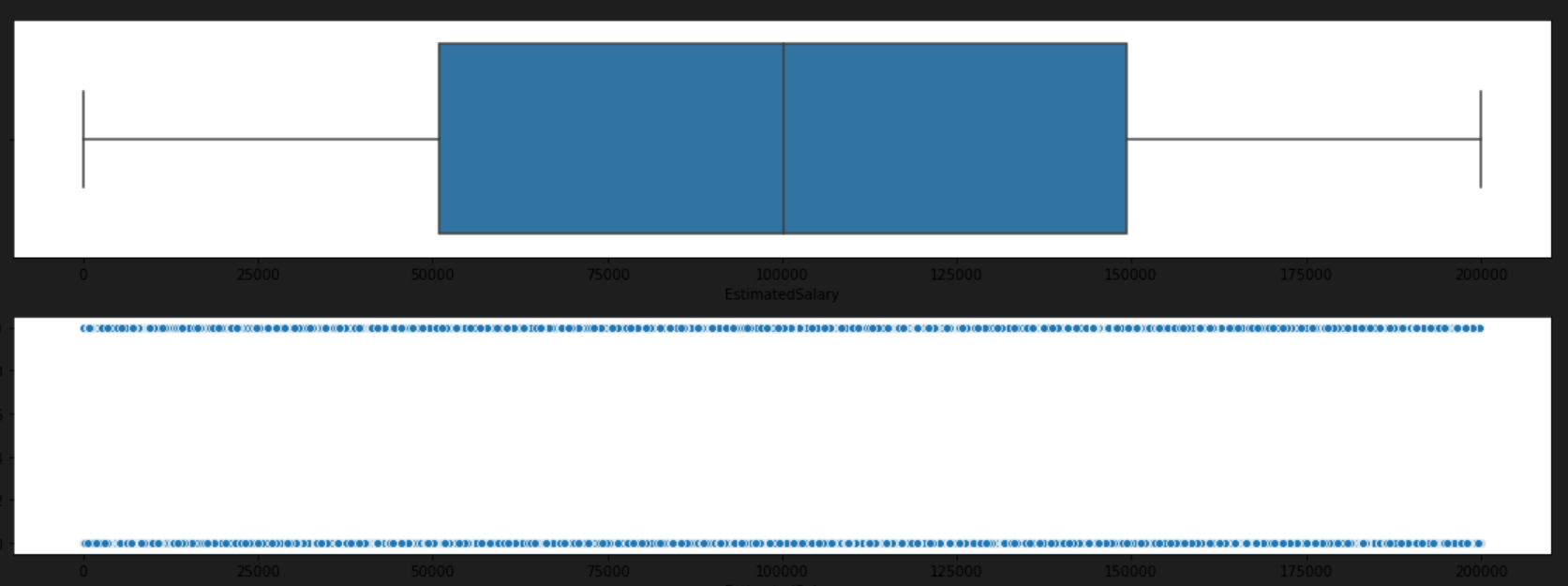




# of bivariate Outliers:4







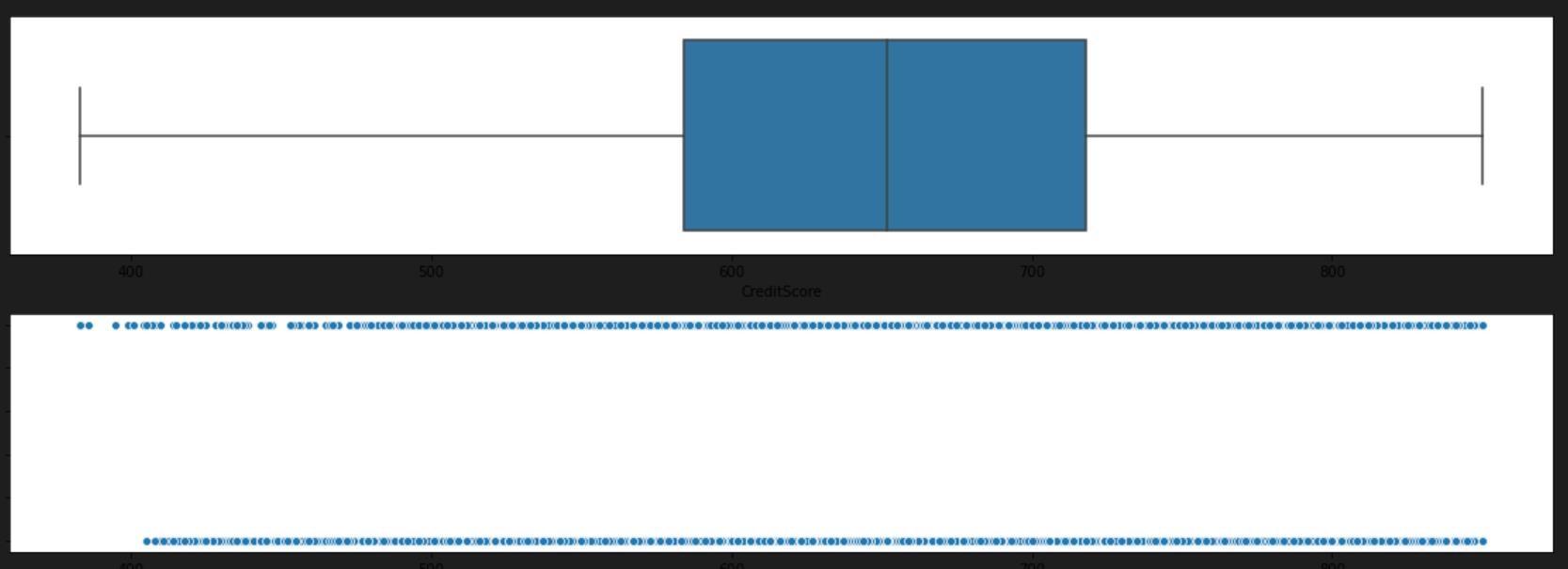
Removing Outliers



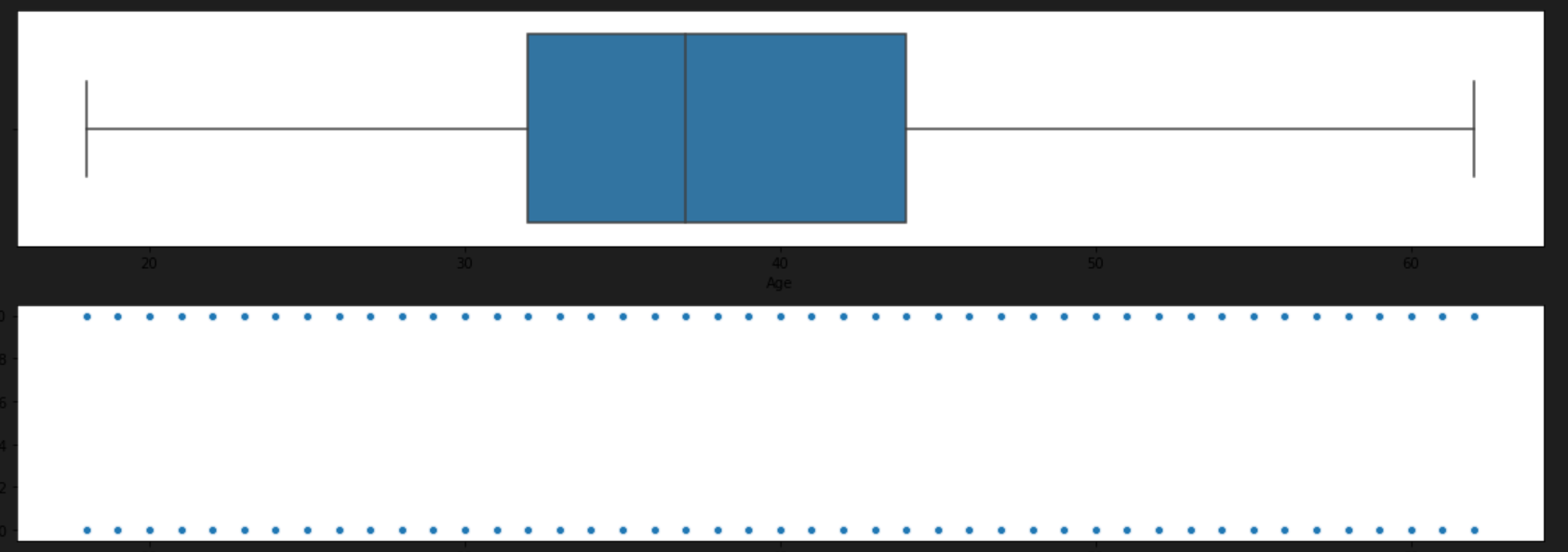
After removing outliers,boxplot will be like



# of bivariate Outliers:19

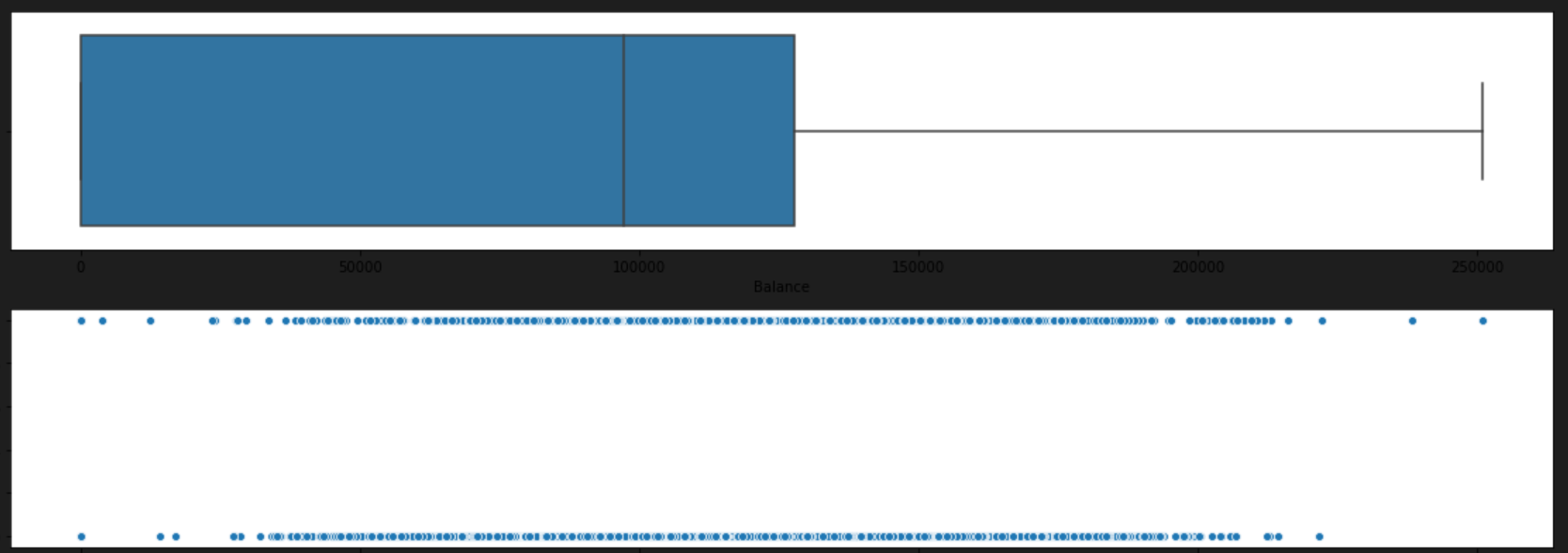


# of bivariate Outliers:0





# of bivariate Outliers:4



# Question 7:

Check for Categorical columns and perform encoding.

# Solution:

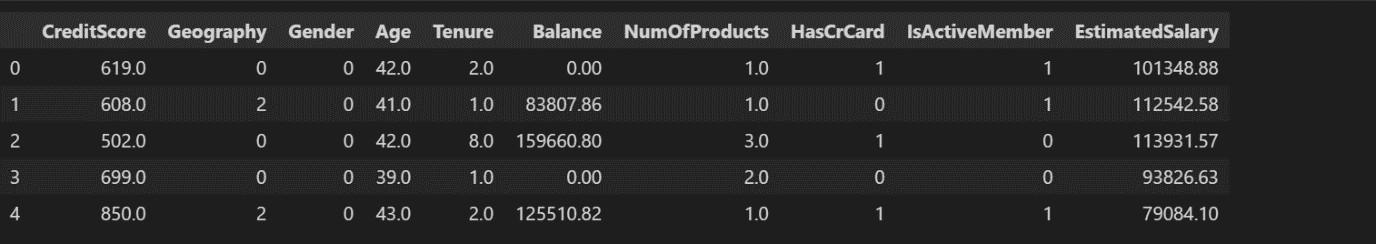


**Question 8:**

Split the data into dependent and independent variables.

# Solution:









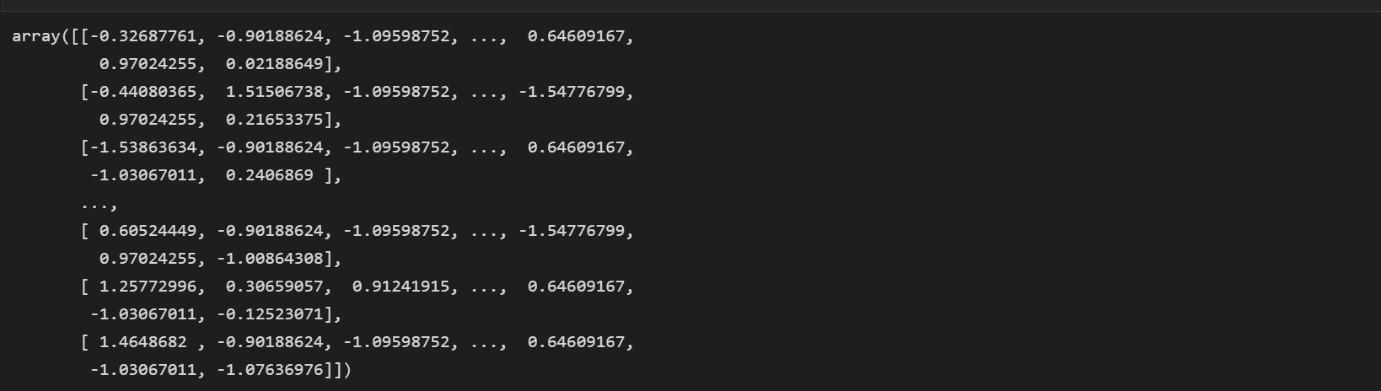
**Question 9:**

Scale the independent variables

# Solution:



x



# Question 10:

Split the data into training and testing

# Solution:



