Assignment -2

Data Visualization and Pre-processing

|  |  |
| --- | --- |
| Assignment Date | 25 September 2022 |
| Student Name | Pranesh S |
| Student Roll Number | 19104050 |
| Maximum Marks | 2 Marks |

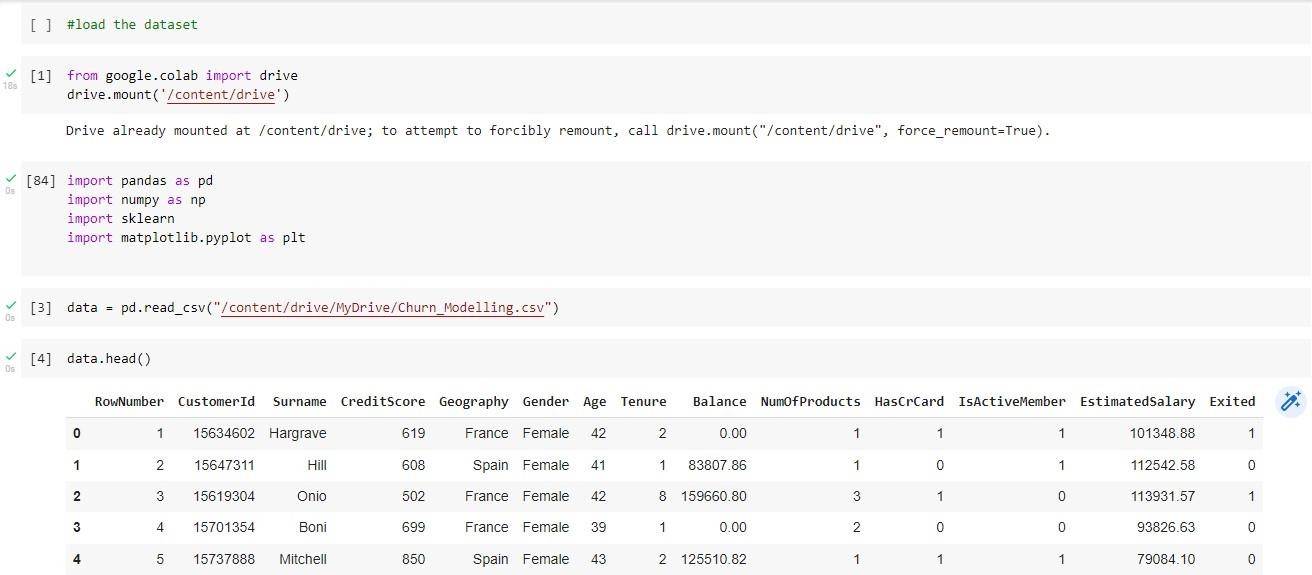
Question-1:

Download the dataset:

The dataset “Churn\_Modelling.csv” was downloaded Successfully

Question-2:

Load the Dataset:



Question-3:

Perform Below Visualization: Univariate Analysis



#Wnivariate Analysis for Categorical Data

[14] pBar Chant

df = pd . DataFrame (data)

X = is:(df.iloc[:,

s:(df.iloc[:, 1])

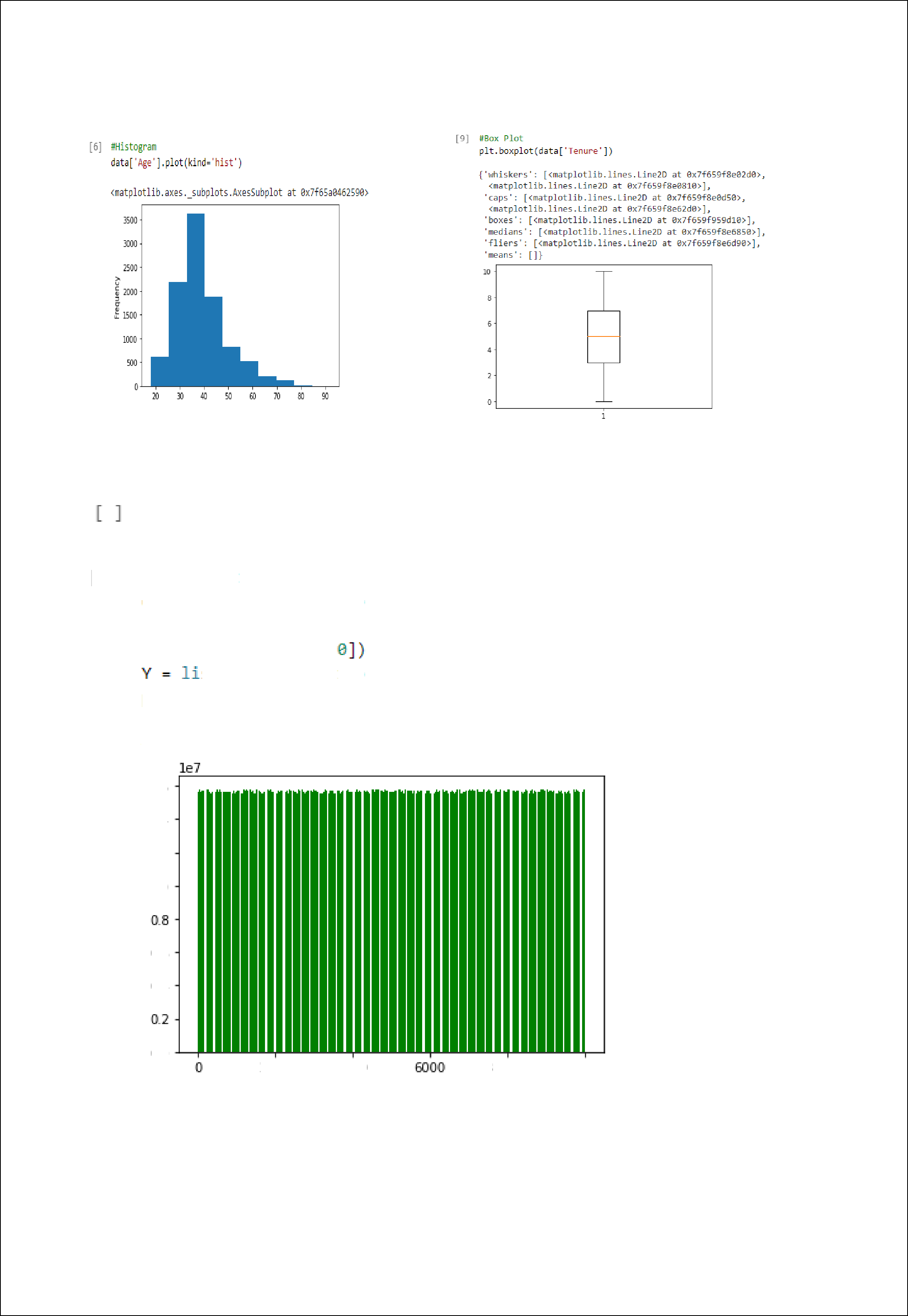
pit.bar(X, Y, color='g')

<BarContainer object of 10B00 artists›

L6 L4 L2 L0

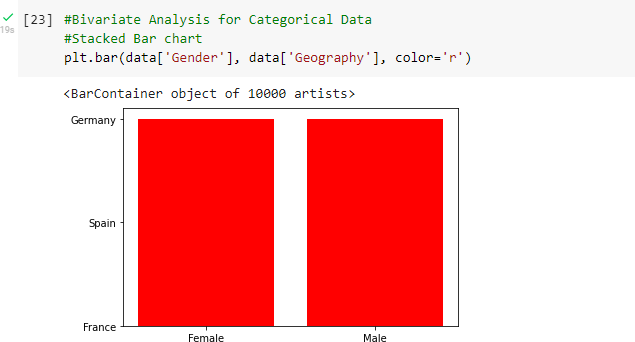
0.6

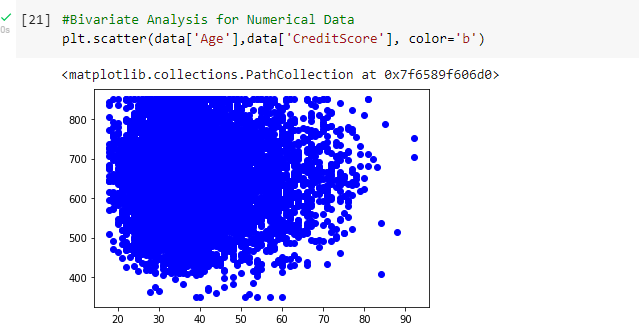
0.4

0.0

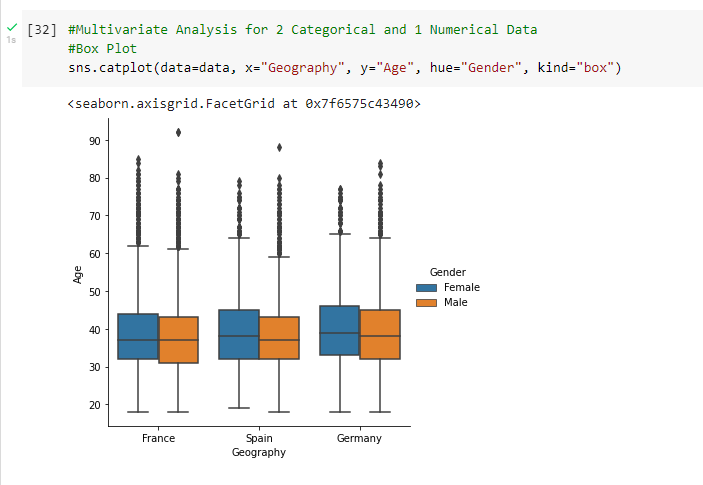
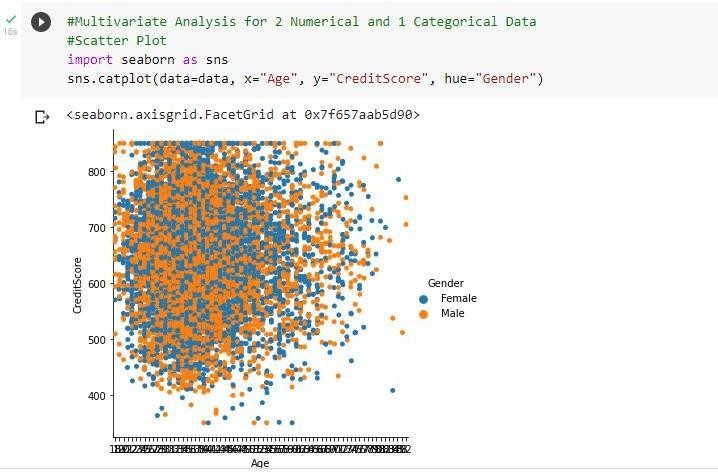
ZOOO 4000 BOOO IOOOO

Bivariate Analysis



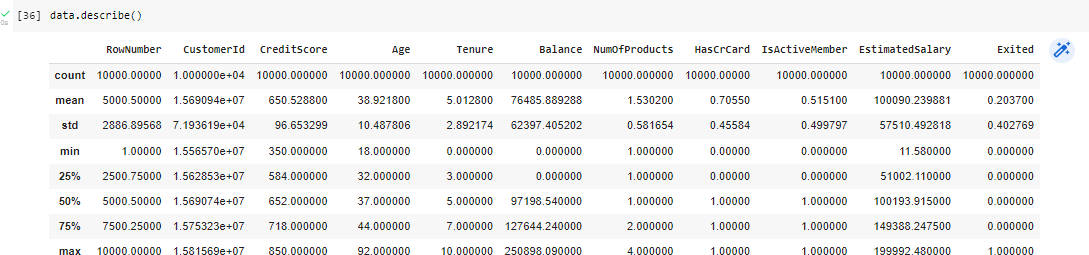
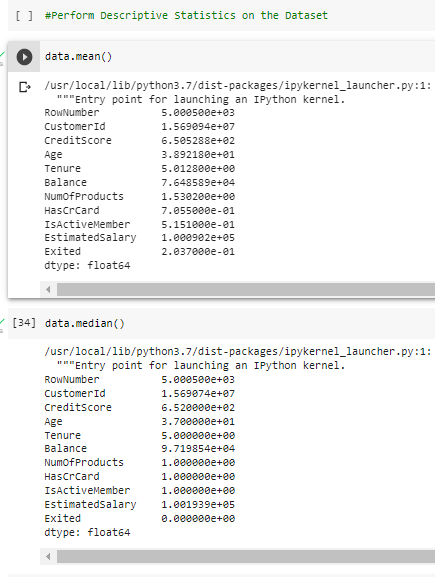


Multivariate Analysis



Question-4:

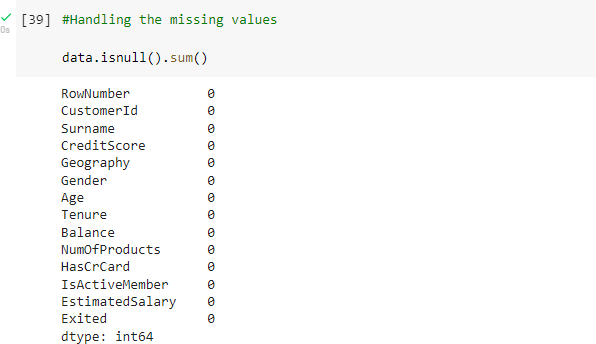
Perform Descriptive Statistics on the dataset:





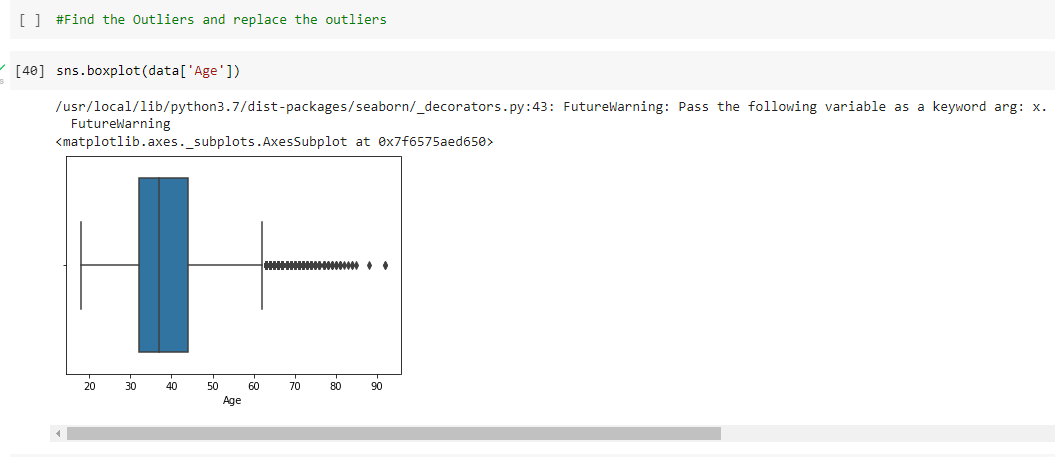
Question-5:

Handle the Missing values:



Question-6:

Find the outliers and replace the outliers:



[4t] qnt-data. quanti1e(q -[0.2„0.75])

RowHumber Customerld CreditScore Age Tenure Balance NumofProducts nascrCard IsActiveMember CstimatedSalary Exited ,}ñ

0.25 2500.75 15628528.25

0.75 7500.25 15753233.75

' {42] IQR = pnt.loc[0.73] - pnt.1oc[0.25]

584.0 32.0 3.0 0.00

718.0 44.0 7.0 127644.24

1 0

2.0

0.0

10

00 51002.1100 0.0

10 149388.2475 0.0

C u st omer' E d Cred i I Sco re Age





12470>.5000





1.0000

Es t\* maI edsaMary 98386. 137 5

E xited e. 00ee

 upper\_ext rerne = qnt . lac [0.75]+1.5\* IgR upper\_ext rerne

RowNumber 1.499950e+04

CustomerId 1.594029e+07

CreditScore 9.190000e+02

Age 6.200000e+01

Tenure 1.300000e+01

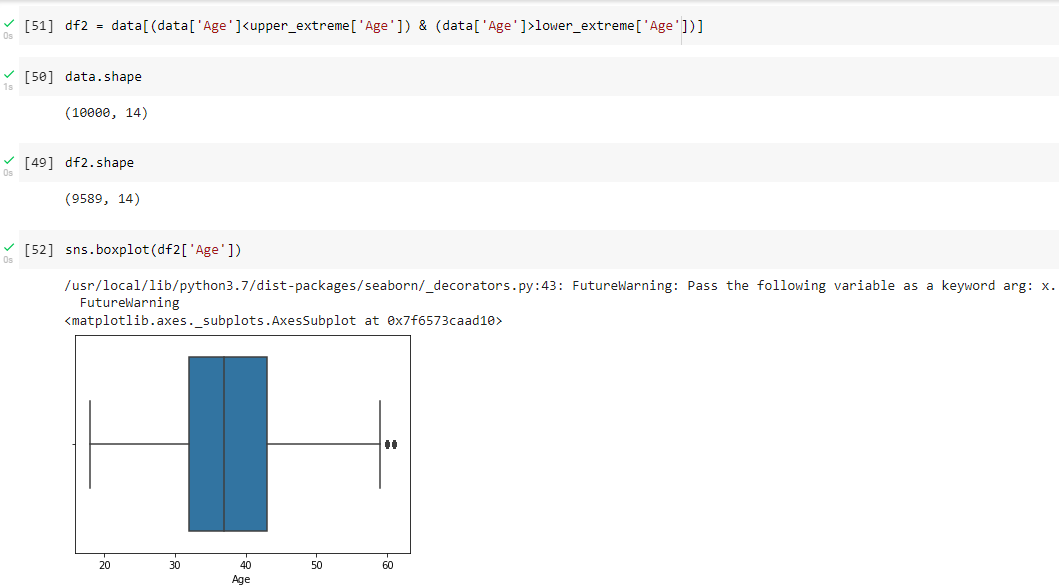
Balance 3.191106e+05

NumofProducts 3.500000e+00

HasCrCard 2.500000e+00 IsActiveMember 2.500000e+00 EstimatedSalary 2.9g9675e+0S Exited 0.000000e+00 dtype: floatd4

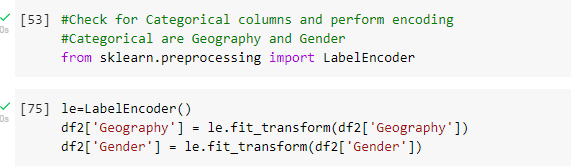
 lower\_extreme = qnt.loc[0.25]-1.5\*IQR lower\_extreme

|  |  |
| --- | --- |
| RowNumber | -4.998500e+03 |
| CustomerId | 1.544147e+07 |
| CreditScore | 3.830000e+02 |
| Age | 1.400000e+01 |
| Tenure | -3.000000e+00 |
| Balance | -1.914664e+05 |
| NumOfProducts | -5.000000e-01 |
| HasCrCard | -1.500000e+00 |
| IsActiveMember | -1.500000e+00 |
| EstimatedSalary | -9.657710e+04 |
| Exited | 0.000000e+00 |
| dtype: float64 |  |



Question-7:

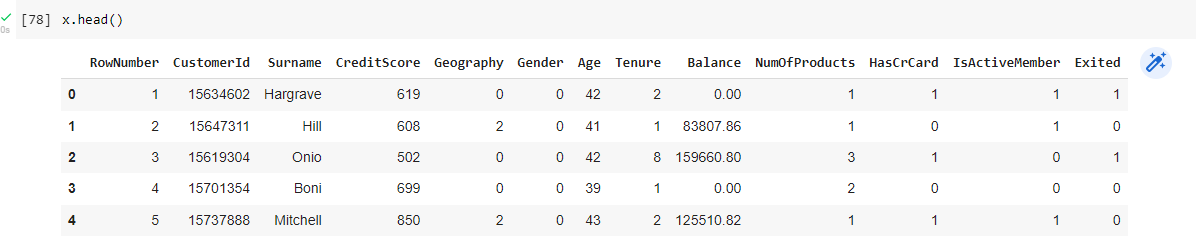
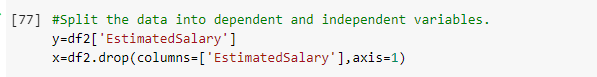
Check for Categorical columns and perform Encoding:





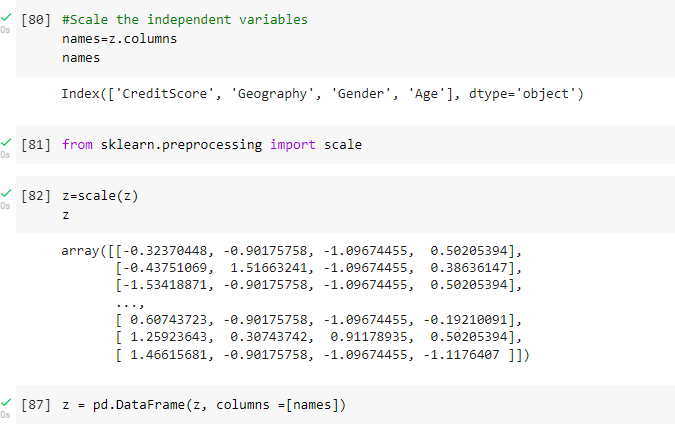
Question-8:

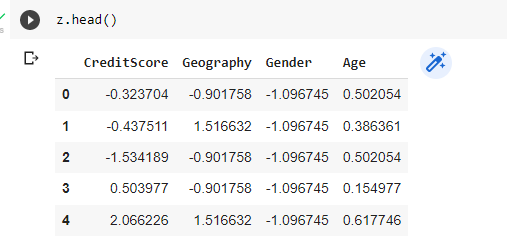
Split the data into dependent and independent variables:



Question-9:

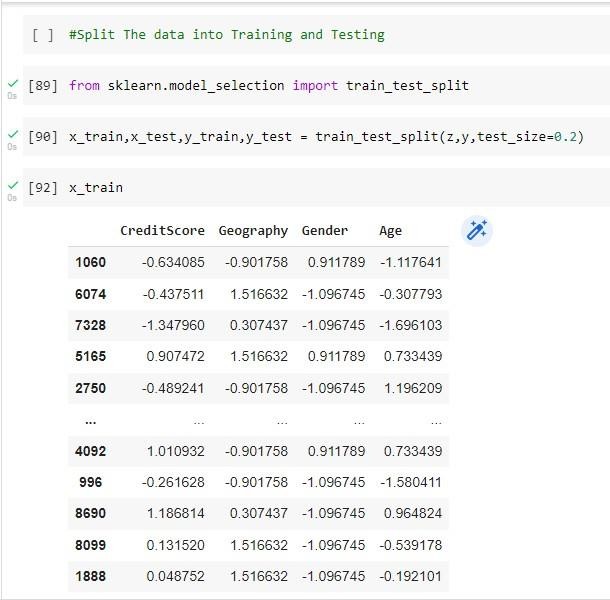
Scale the independent variables:





Question-10:

Split the data into training and testing:



 151005.96

6 334 143463. 2B

76J8 37577.66

5392 43€IS8.B2

2851 1£@478. 6B

4209 2048.55

1037 180969.55

9058 166896.Bl

8440 368d4.B5

196B 86013.96

Name: EstimatedSalary, Length: 7671. dtype: floatd4



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Cred1tScore | Geogpaptty | 6endez' | Age |
| 062 | 0.772974 | 0.307437 | 0.911789 | 0.154977 |
| 5d5T | 1.2488& | 1.516632 | -1.DW745 | 0.386361 |
| 7515 | —0.841005 | 0.307437 | —1.DW745 | —0.654871 |
| 6844 | 0.959202 | -0.90175B | -1.DW745 | -0.BB6256 |
| 4102 | -0.9961W | 1.516632 | -1.DW745 | 0.386361 |
| 60 | 0.379B25 | 0.307437 | —1.D@745 | —1.233333 |
| 5555 | 0.503977 | -0.90175B | 0.911789 | -0.D76408 |
| 5112 | 1.704115 | 1.516632 | —1.D 745 | 2.237441 |
| 138 | 0.131520 | —0.90175B | 0.911789 | —0.423486 |
| 4973 | 0.328095 | -0.90175B | -1.DW745 | 2.353134 |

191B rows x 4 columns



|  |  |
| --- | --- |
| 1002 | 1B4023.54 |
| 54B6 | 92914.67 |
| 7838 | 132B38.65 |
| 7133 | 1387B0.B9 |
| 42B1 | 36242.19 |
| 61 | 126404. B2 |
| 5797 | 83Z63.B4 |
| 5337 | 38941.44 |
| 141 | 18B427.24 |
| sisi | 7em.se |

Name: Estimated5alary, Length: 19\*8. dtype: floatd4