**ASSIGNMENT-02**

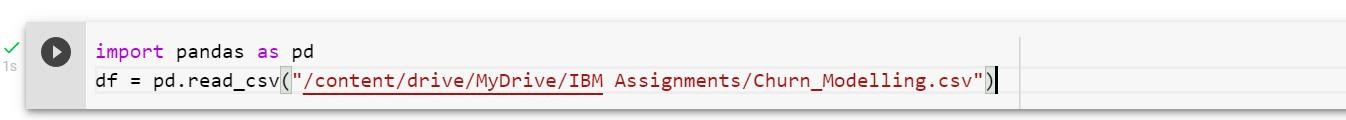
**DATA VISUALIZATION AND PRE PROCESSING**

|  |  |
| --- | --- |
| **Assignment Date** | 22 September 2022 |
| **Student Name** | Bhuvaneshwari M |
| **Student Roll Number** | 113219071004 |
| **Maximum Marks** | 2 Marks |

1. Download the dataset: Dataset Dataset downloaded in csv form.
2. Load the dataset.

import pandas as pd

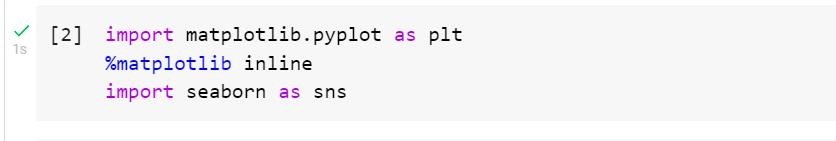
df = pd.read\_csv("/content/drive/MyDrive/IBM Assignments/Churn\_Modellin g.csv")

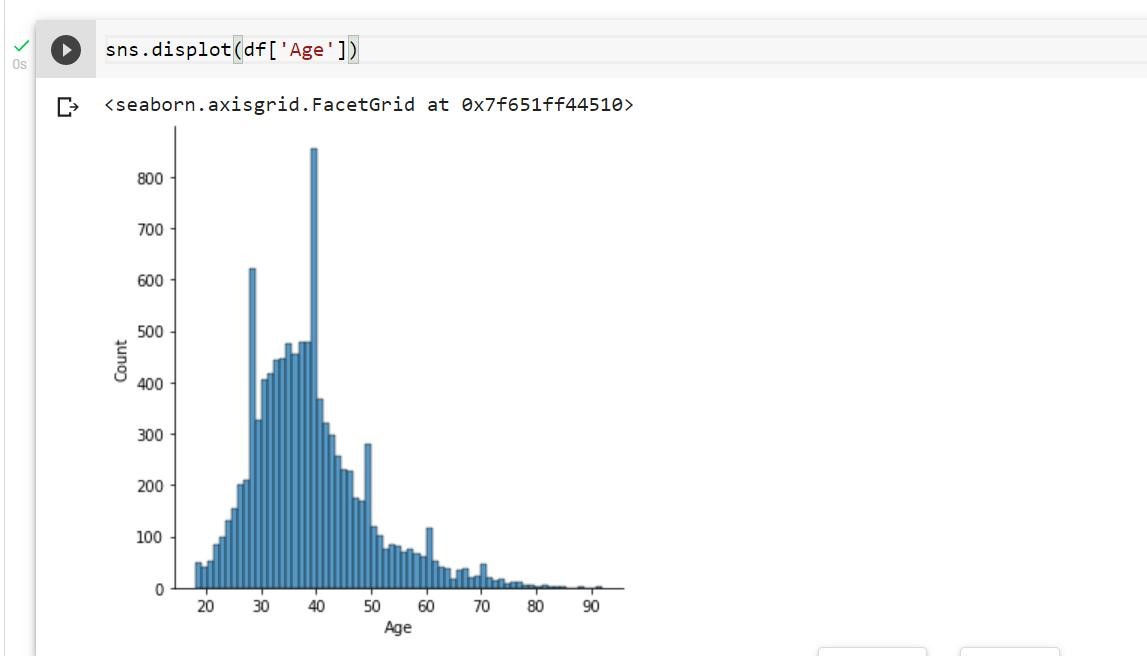


1. Perform Below Visualizations.

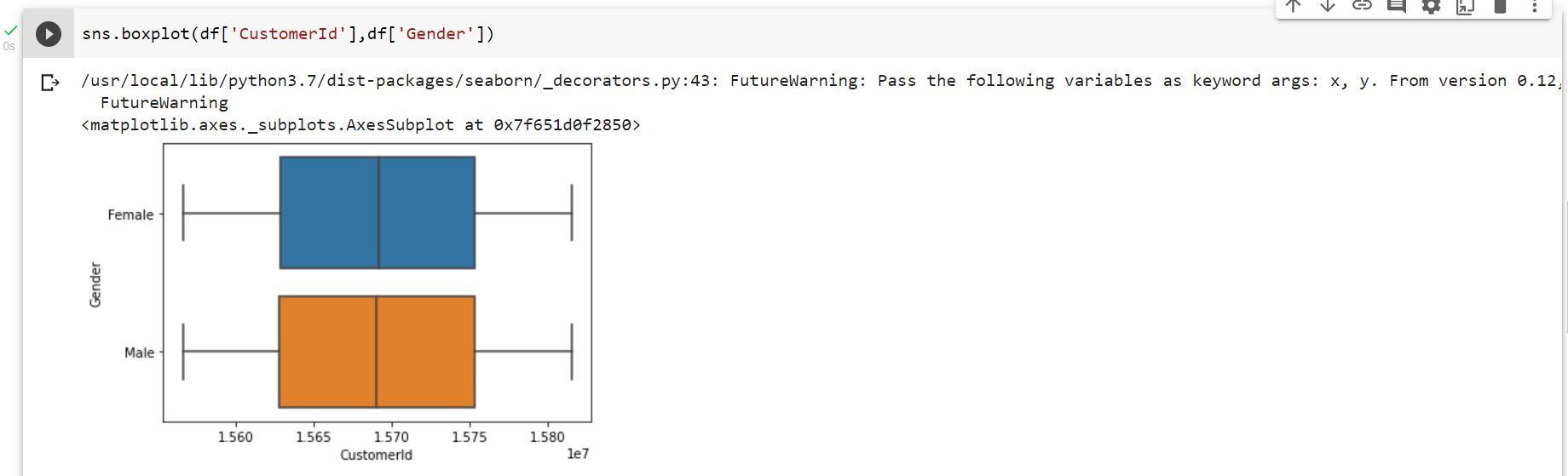
* Univariate Analysis

sns.displot(df['Age'])

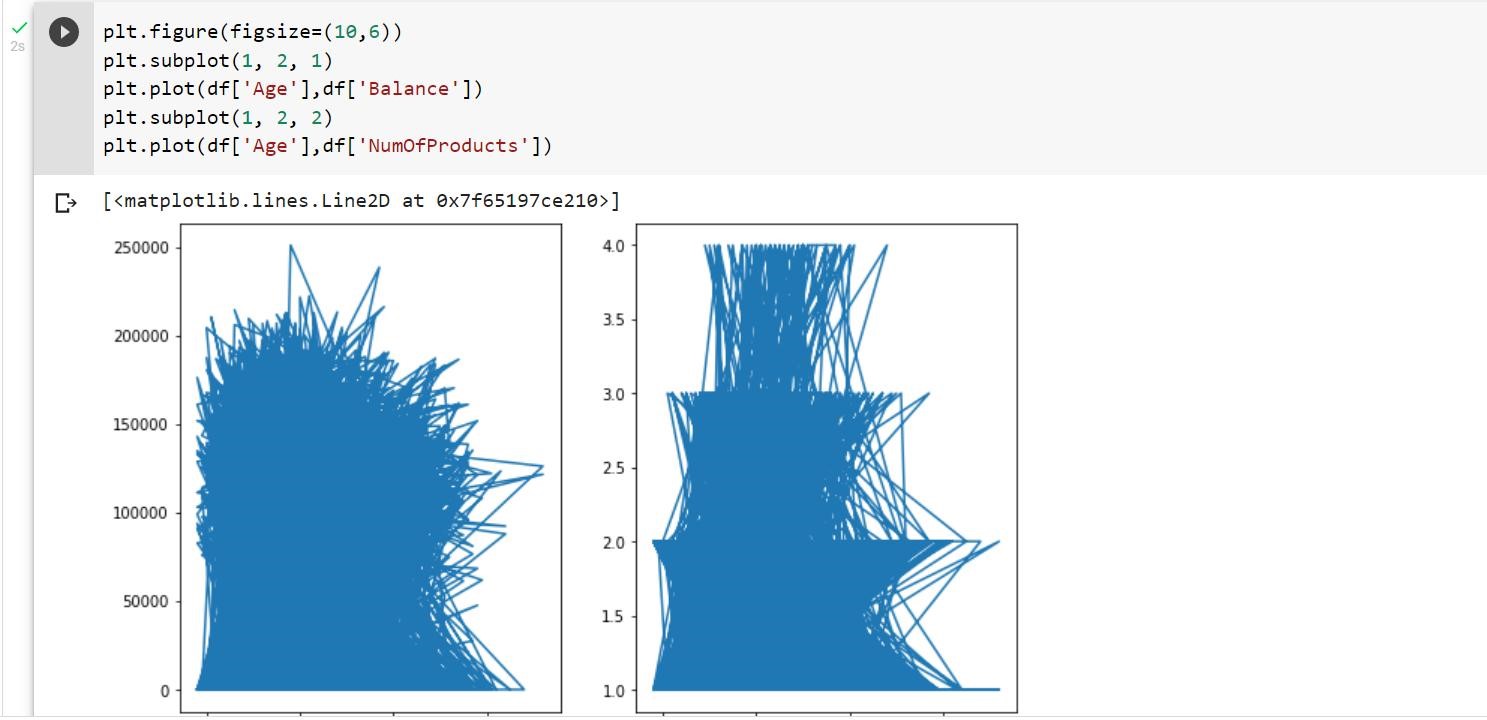




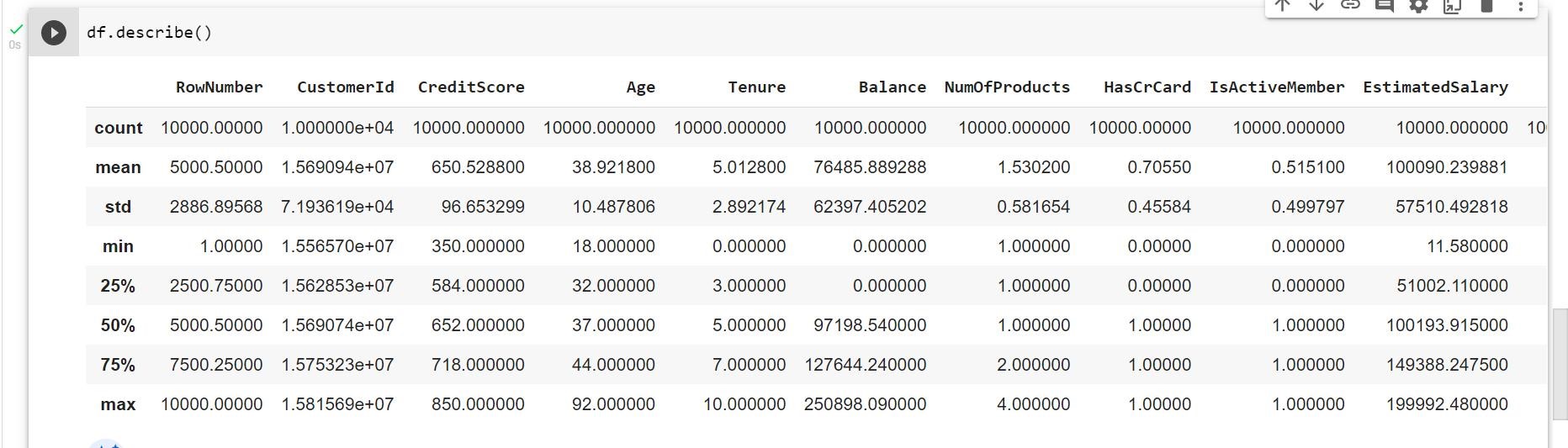
* Bi - Variate Analysis sns.boxplot(df['CustomerId'],df['Gender'])



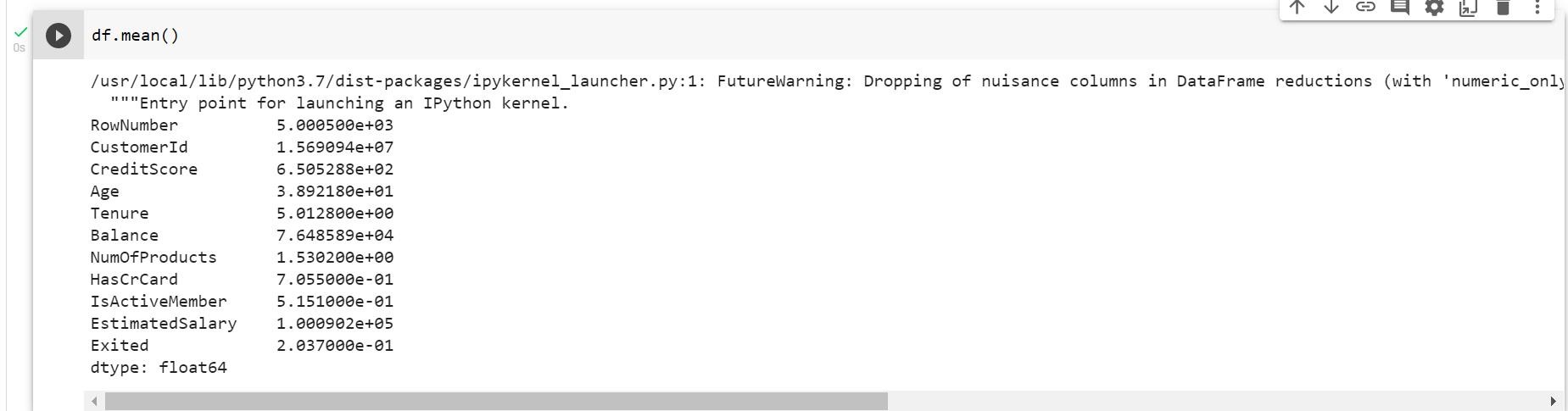
* Multi - Variate Analysis



1. Perform descriptive statistics on the dataset.



Mean:



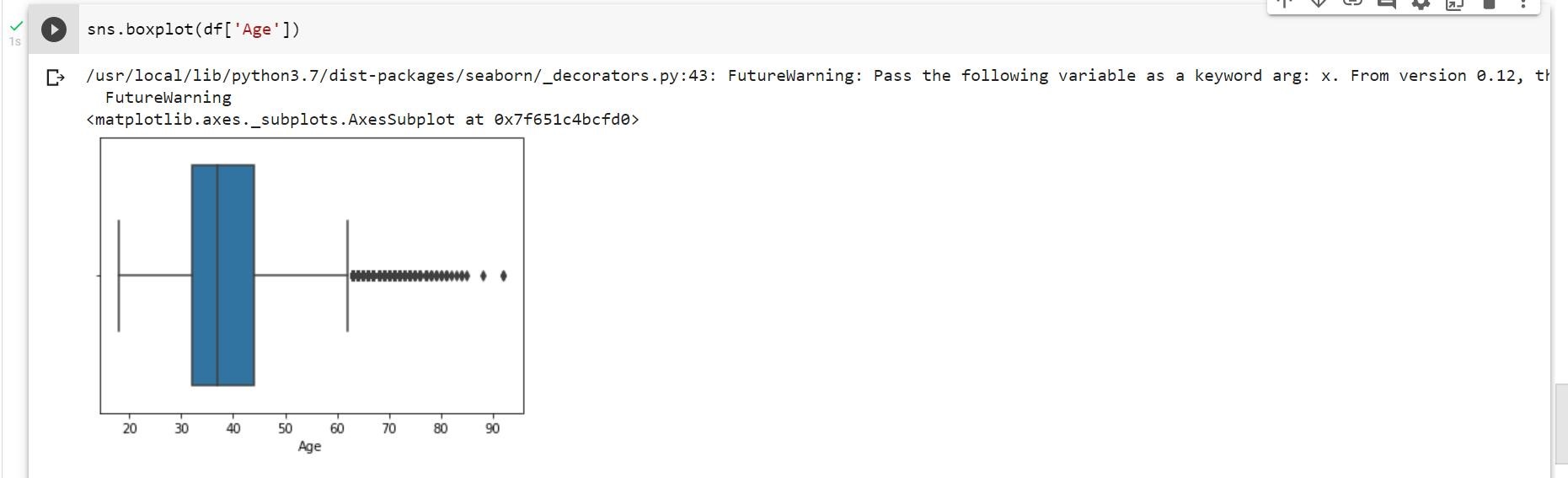
1. Handle the Missing values.



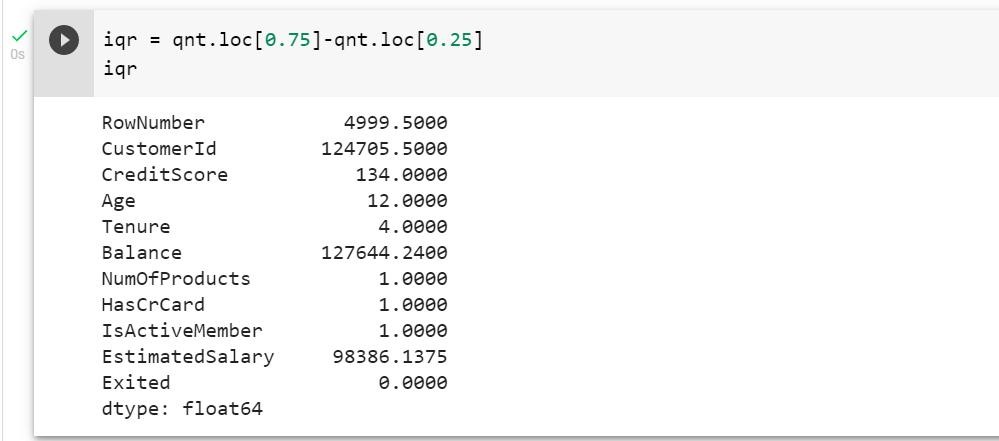
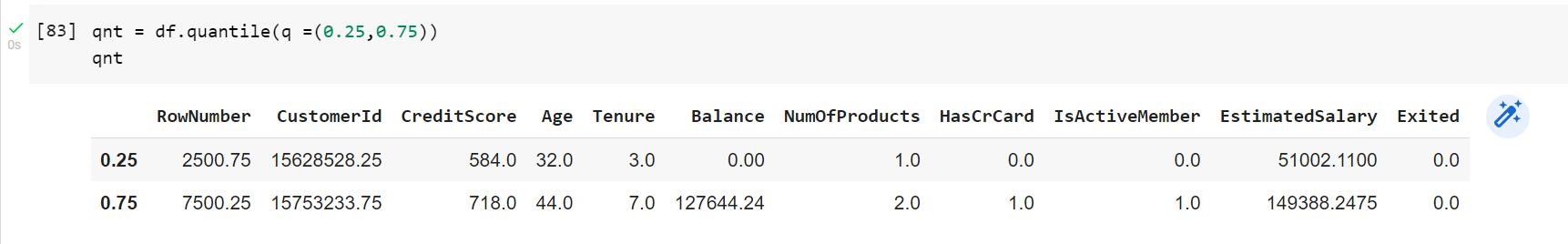
1. Find the outliers and replace the outliers

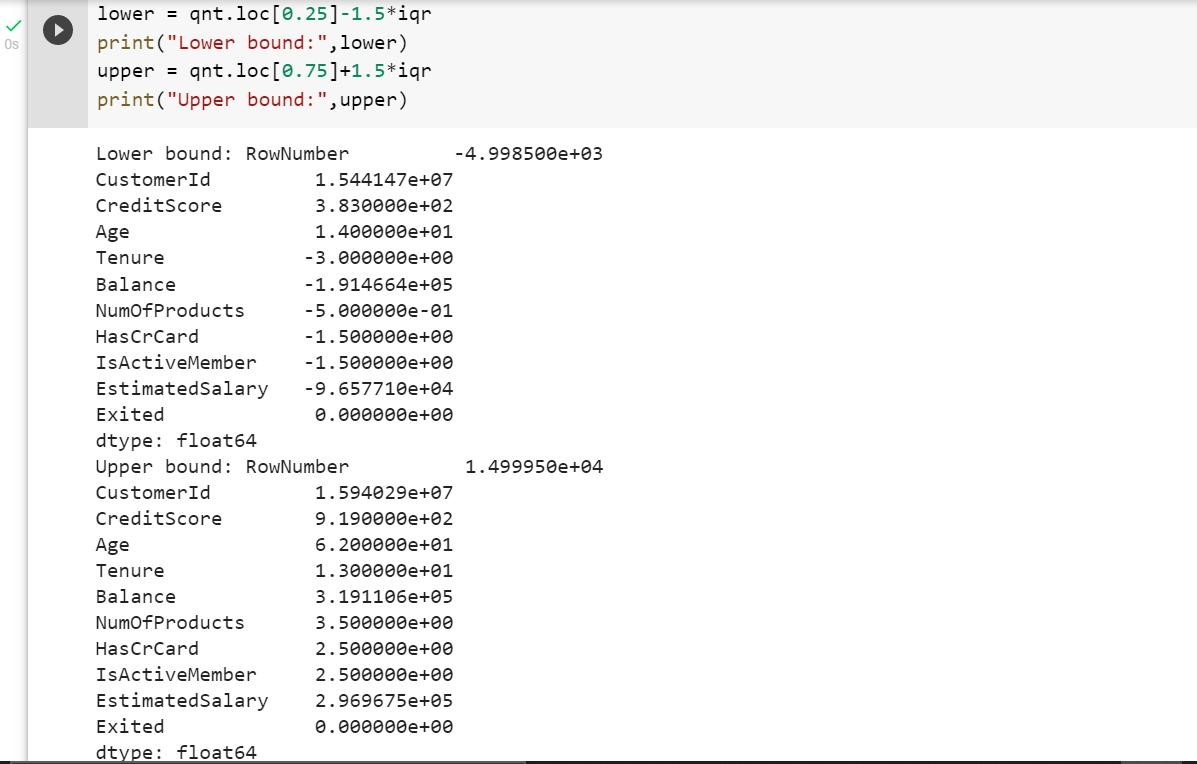
Finding Outliers:

Using Boxplot



Using method



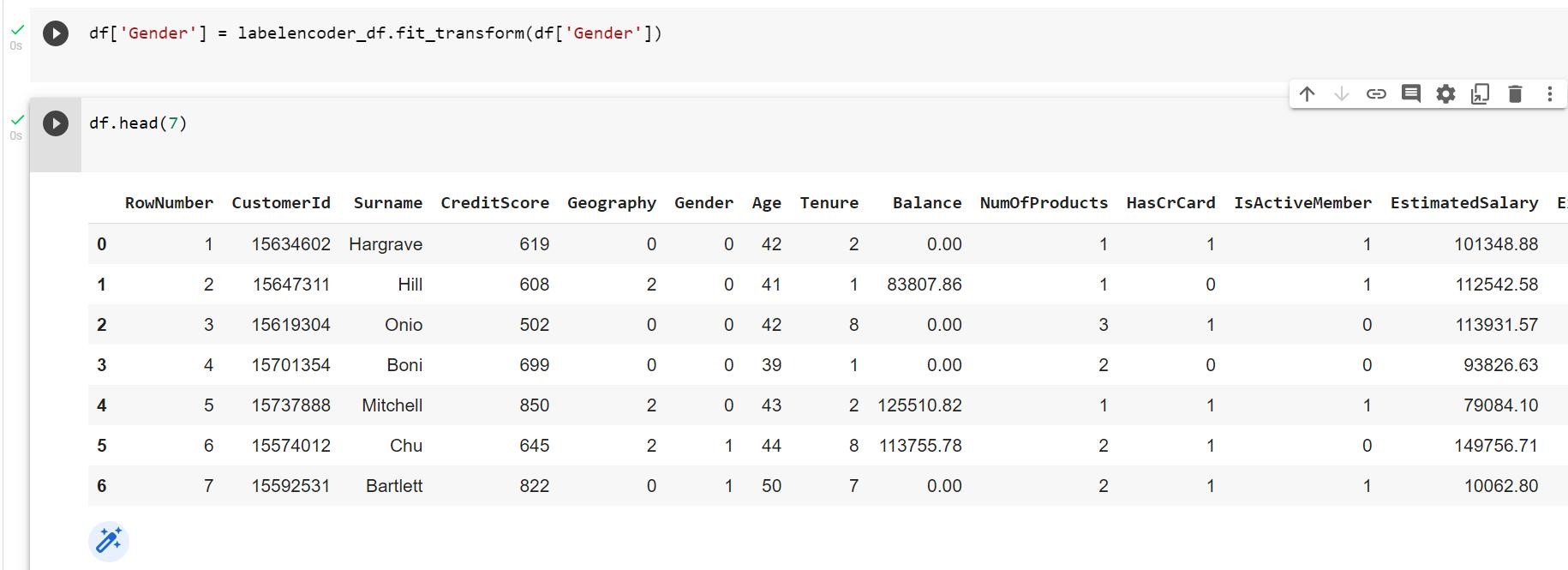
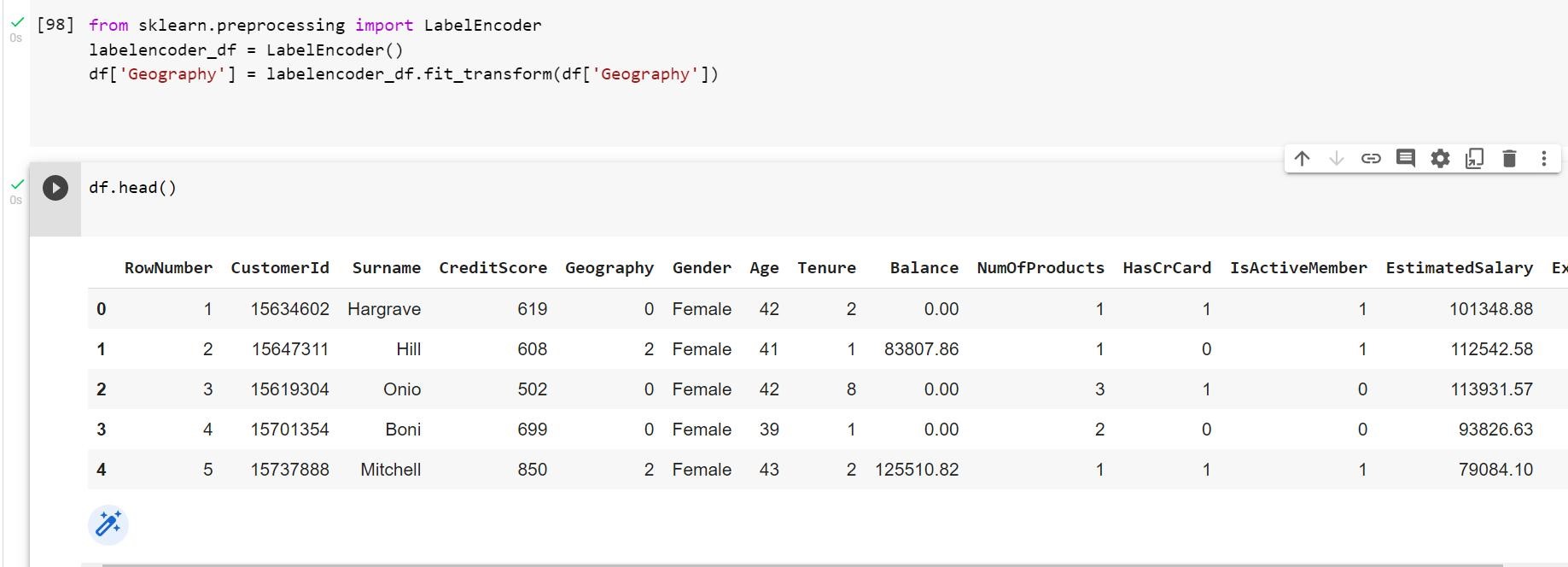


Replacing Outliers:



1. Check for Categorical columns and perform encoding.

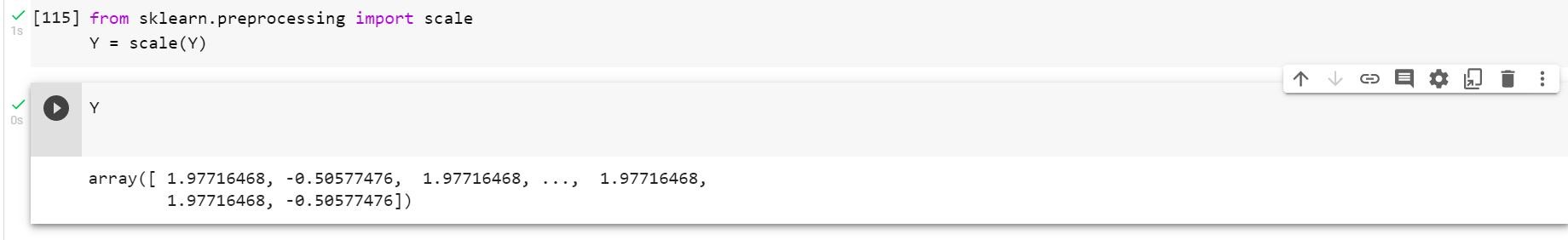
Categorical columns: Geography,Gender



1. Split the data into dependent and independent variables.



1. Scale the independent variables



1. Split the data into training and testing

