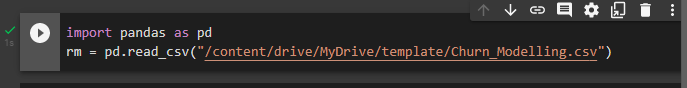
**ASSIGNMENT-02**

**DATA VISUALIZATION AND PRE PROCESSING**

|  |  |
| --- | --- |
| **Assignment Date** | 22 September 2022 |
| **Student Name** | Tharika Jayaraj |
| **Student Roll Number** | 113219071045 |
| **Maximum Marks** | 2 Marks |

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

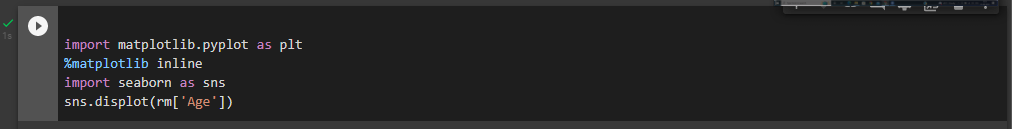
import pandas as pd rm = pd.read\_csv("/content/drive/MyDrive/template/Churn\_Modelling.csv")

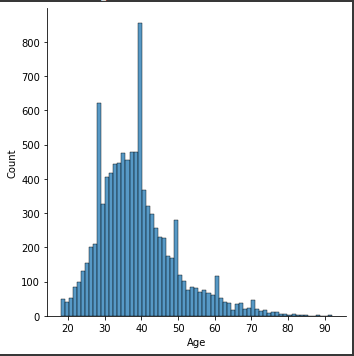


1. Perform Below Visualizations.

* Univariate Analysis

sns.displot(rm['Age'])

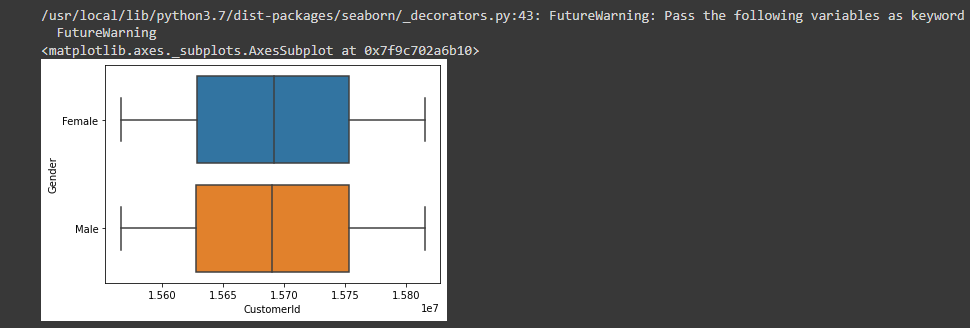




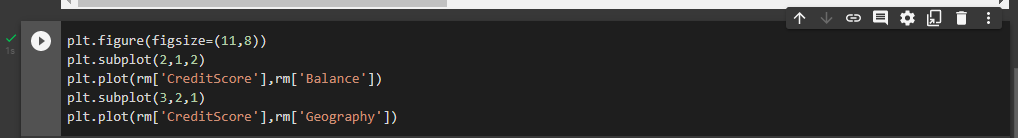
* Bi - Variate Analysis

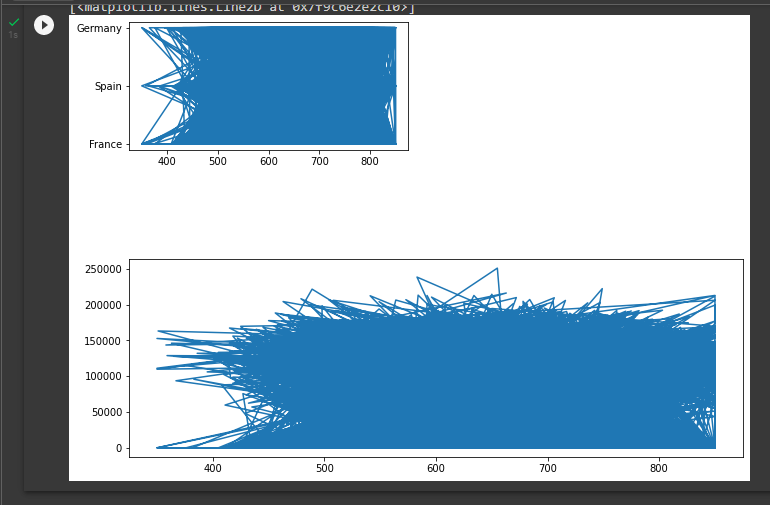
sns.boxplot(rm['CustomerId'],rm['Gender'])



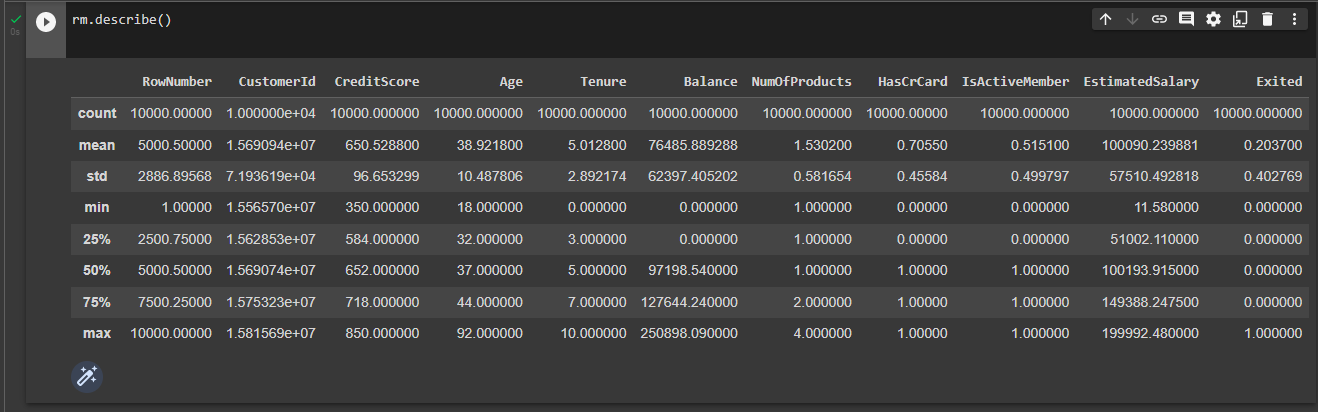


* Multi - Variate Analysis

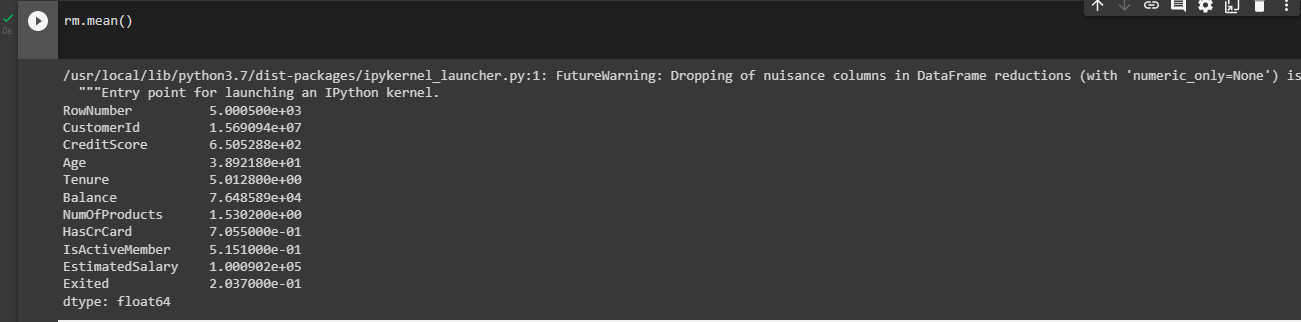




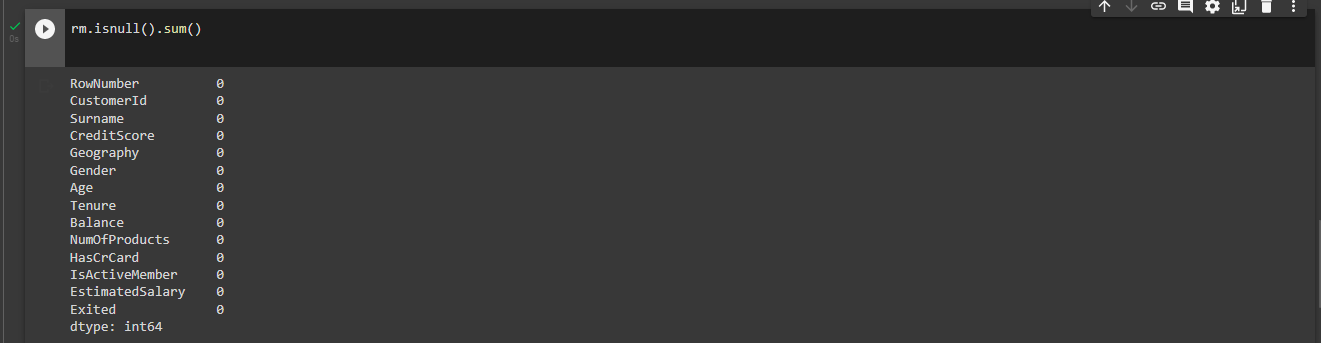
1. Perform descriptive statistics on the dataset.



Mean:



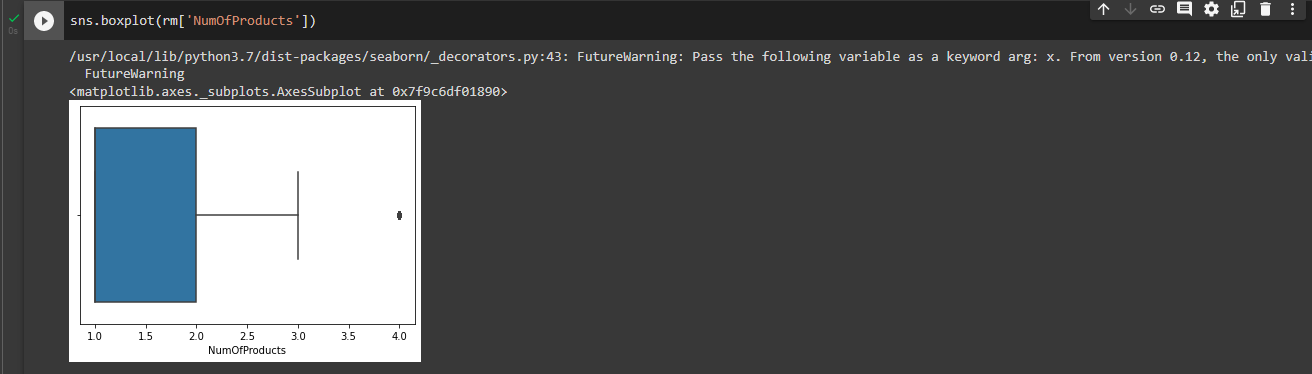
1. Handle the Missing values.



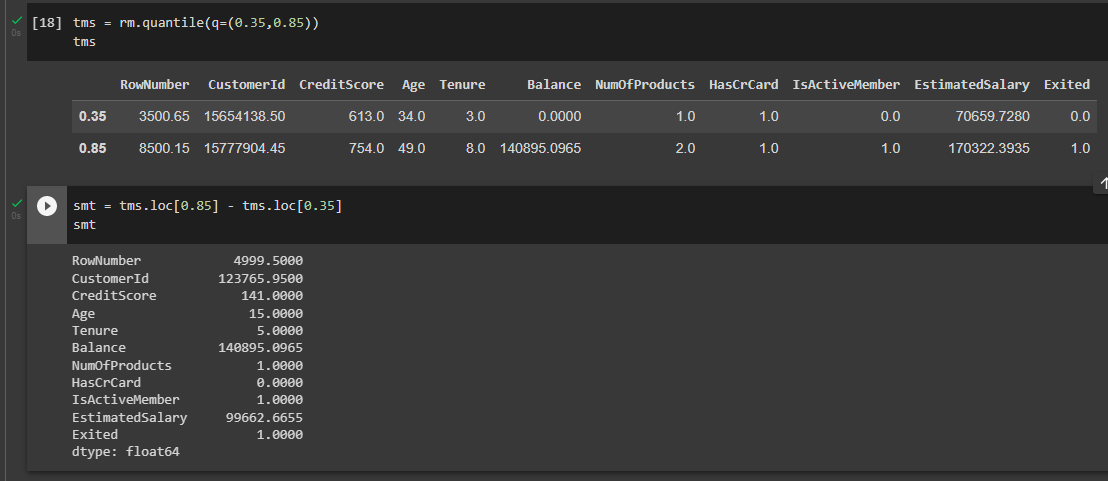
1. Find the outliers and replace the outliers

Finding Outliers:

Using Boxplot



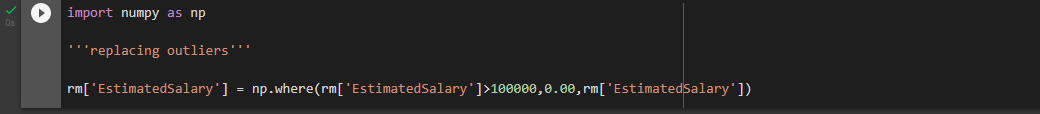
Using method





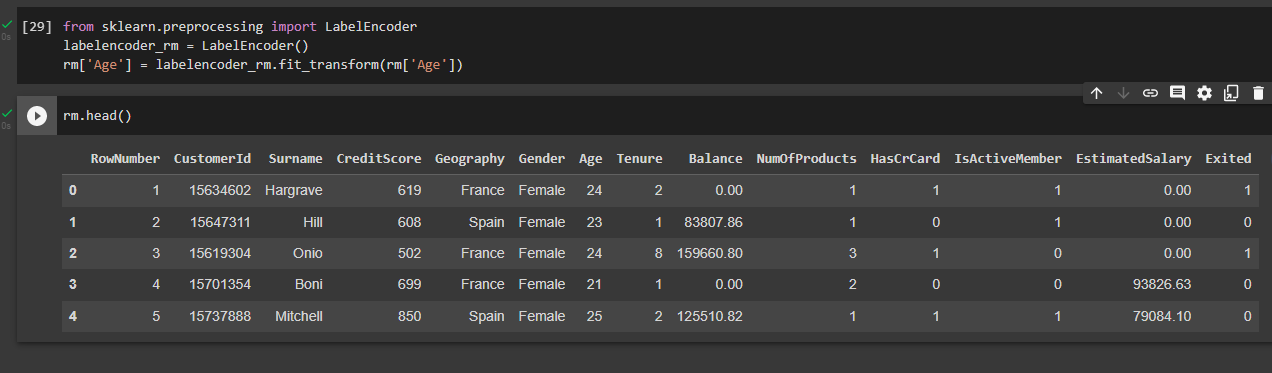


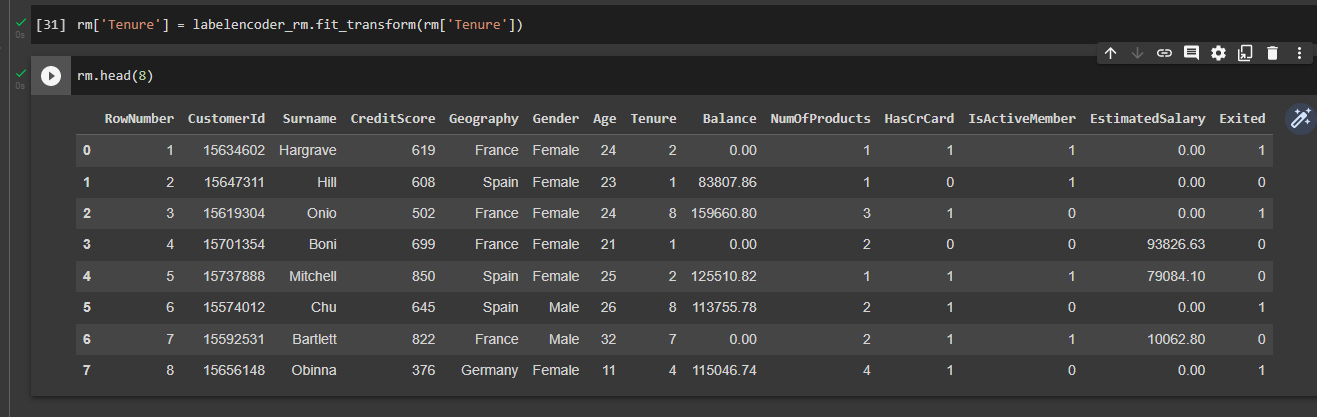
Replacing Outliers:



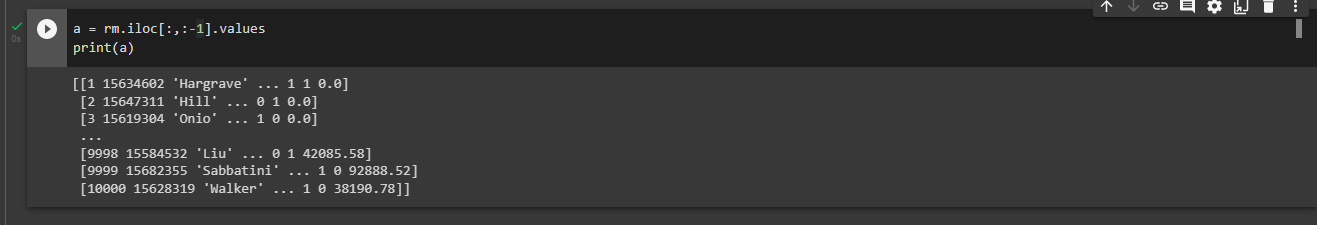
7.Check for Categorical columns and perform encoding.

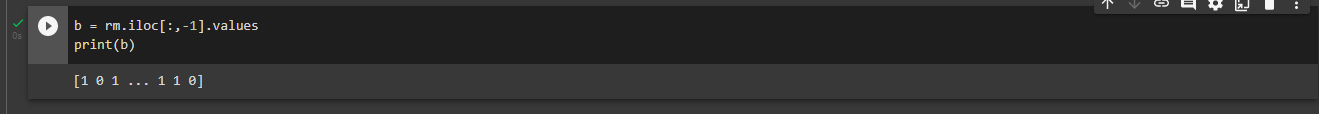
Categorical columns: Age,Tenure



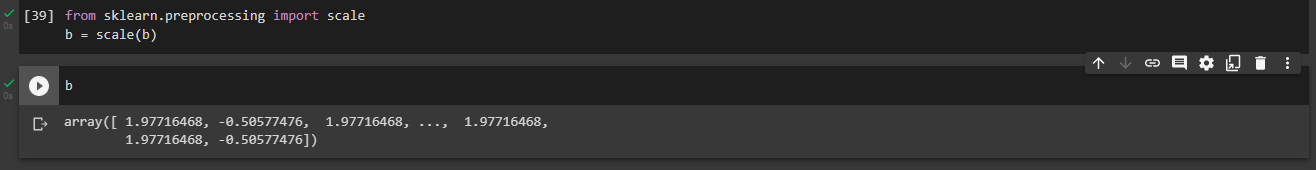


8.Split the data into dependent and independent variables.





9.Scale the independent variables



10.Split the data into training and testing

b\_train



b\_test

Text

Description automatically generated

a\_train

Text

Description automatically generated

a\_test

Text, letter

Description automatically generated