**ANL252**

**Python and Data Analytics**

**ECA**

**Submitted by:**

|  |  |
| --- | --- |
| **Name** | **PI No.** |
| Gan Ke Yi | B1972108 |

**Tutorial Group: ­­­­­­­­­­­­** T09

**Submission Date:** 4/09/2022

Q1

Table

Description automatically generated

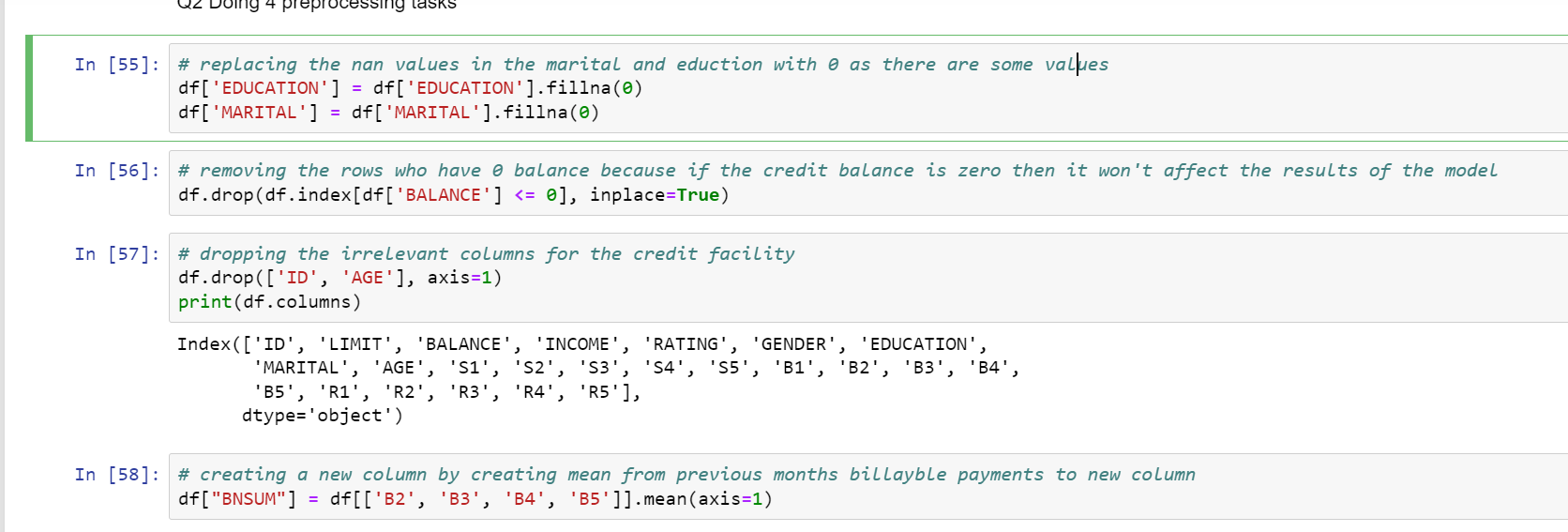
Graphical user interface, text, application

Description automatically generated

Categorical variables are: ['RATING', 'GENDER', 'EDUCATION', 'MARITAL', 'S1', 'S2', 'S3', 'S4', 'S5']

Numeric variables are: ['R2', 'LIMIT', 'R4', 'R5', 'R1', 'B2', 'B4', 'B5', 'ID', 'BALANCE', 'R3', 'INCOME', 'B3', 'B1', 'AGE']

Q2)



The 4 preprocessing data task that I did was to first replace the “nan” values into 0 for education and marital columns. As the values “0” refers to “Others”.

The second pre-processing task would be to remove the records that have $0 Balance as this is will not affect the model.

Next is to remove the columns that will not help in our analysis.

Lastly is to create a new column to find the average billable amount from 2nd month to the 5th month. This will give a gauge on how much a customer tend to spend in a month.

Q3)

Chart

Description automatically generated

Insights: the first insight I can get from this graph is that it shows me the income of our customers. From this graph we can tell that the majority earns below 600000. Secondly, from this chart we can also see some anomalies that lies in the age of 0 and 200, these data can be ignore or dropped.

Chart, scatter chart

Description automatically generated

Insights: In this graph we can see a relationship between positive correlations between billing and balance. This is logical because the amount the balance a person have is correlated to the amount of credit a customer can have.

Chart

Description automatically generated

Insights: From the plot we can tell that the billable amount for married couples is the highest among the single and others category. This is a logical conclusion as married couple may tend to have more expenses than people that are tagged as single or others.

Chart, scatter chart

Description automatically generated

Insights: From this chart we can also spot a positive correlation the average billing amount and the current billing month. This shows that the customer spending habits are consistent.

Q4:

A picture containing graphical user interface

Description automatically generated

Q5:

Graphical user interface, text, application, email

Description automatically generated

The linear regression equation is 0.33329575499920483\*AGE + 435.4944356915702\*MARITAL + 0.37966578932264383\*BNSUM + 3.6334611154226195\*BALANCE + -408.7635395433972.

The main key insights for the results obtained are the value of B1 is highly dependent on the Balance of the Customer and also the Customers billable amounts in the last 4 months along with its Age and Marital Status as unmarried person tends to show less billable amount from the insights aboves.