

(Full marks: 100)

Section A (90 marks)

Answer all questions in this section.

Question 1

This credit facility dataset to be analyzed comprises records of customers' demographics, amount owed, repayment history/status etc. The data dictionary of this dataset is depicted in Appendix 1.

List the categorical and numeric variables in this dataset.

(5 marks)

Question 2

Conduct **four (4)** data pre-processing tasks for the analysis of the data, explaining results obtained.

(20 marks)

Question 3

Articulate **five (5)** relevant insights of the data, with supporting visualization for **each** insight.

(25 marks)

Question 4

Perform linear regression modelling to predict the variable, B1, explaining the approach taken, including any further data pre-processing.

(25 marks)

Question 5

State the linear regression equation and explain key insights from the results obtained in Question 4.

(15 marks)

Section B (10 marks)

Answer all questions in this section.

Question 6

Organization of Code

The submitted Jupyter notebook will be assessed based on the following:-

- Readability, Consistency and Efficiency
- Well-documented

(10 marks)

Appendix:

APPENDIX 1 – DATA DICTIONARY

Variable	Description
ID	Customer unique identifier
LIMIT	Customer total limit
BALANCE	Customer current credit balance (snapshot in time)
INCOME	Customer current income
GENDER	Customer gender (0: Male, 1: Female)
EDUCATION	Customer highest education attained (0: Others, 1: Postgraduate, 2: Tertiary, 3: High School)
MARITAL	Customer marital status (0: Others, 1: Single, 2: Married)
AGE	Customer age in years
S(n)	Customer repayment reflected status in nth month. (-1; Prompt payment, 0: Minimum sum payment, x = Delayed payment for x month(s))
B(n)	Customer billable amount in nth month
R(n)	Customer previous repayment amount, paid in nth month
RATING	Customer rating (0: Good, 1: Bad)

Note:

n=1 signifies the most recent month, while n=5 signifies the previous 4th month.
If n=1 is the month of May 2022, then n=5 is the month of January 2022.

----- **END OF ECA PAPER** -----