

ANL252

End-of-Course Assessment - July Semester 2022

Python for Data Analytics

INSTRUCTIONS TO STUDENTS:

1. This End-of-Course Assessment paper comprises **6** pages (including the cover page).
2. You are to include the following particulars in your submission: Course Code, Title of the ECA, SUSS PI No., Your Name, and Submission Date.
3. Late submission will be subjected to the marks deduction scheme. Please refer to the Student Handbook for details.

IMPORTANT NOTE

ECA Submission Deadline: Monday, 05 September 2022 12:00 pm

ECA Submission Guidelines

Please follow the submission instructions stated below:

A - What Must Be Submitted

You are required to submit the following item for marking and grading:

- *A Report*

Please verify your submissions after you have submitted the above item.

B - Submission Deadline

- *The report is to be submitted **by 12 noon** on the submission deadline.*
- *You are allowed multiple submissions till the cut-off date for the report.*
- *Late submission of the report **will be subjected to mark-deduction scheme** by the University. Please refer to Section 5.2 Para 2.4 of the Student Handbook.*

C - How the report Should Be Submitted

- *The Report: submit online to Canvas via TurnItIn (for plagiarism detection)*
 - *please ensure that your Microsoft Word document is generated by Microsoft Word 2007 or higher.*
 - *the report must be saved in .docx format.*
- *Avoid using a public WiFi connection for submitting large video files. If you are using public wireless (WiFi) connection (e.g. SG Wireless at public areas), you might encounter a break in the connection when sending large files.*

D – Please be Aware of the Following:

Submission in hardcopy or any other means not given in the above guidelines will not be accepted. You do not need to submit any other forms or cover sheets (e.g. form ET3) with your ECA.

*You are reminded that electronic transmission is not immediate. The network traffic may be particularly heavy on the date of submission deadline and connections to the system cannot be guaranteed. Hence, you are advised to submit your work early. **Canvas will allow you to submit your work late but your work will be subjected to the***

mark-deduction scheme. You should therefore not jeopardise your course result by submitting your ECA at the last minute.

It is your responsibility to check and ensure that your files are successfully submitted to Canvas.

E - Plagiarism and Collusion

Plagiarism and collusion are forms of cheating and are not acceptable in any form in a student's work, including this ECA. Plagiarism and collusion are taking work done by others or work done together with others respectively and passing it off as your own. You can avoid plagiarism by giving appropriate references when you use other people's ideas, words or pictures (including diagrams). Refer to the APA Manual if you need reminding about quoting and referencing. You can avoid collusion by ensuring that your submission is based on your own individual effort.

The electronic submission of your ECA will be screened by plagiarism detection software. For more information about plagiarism and collusion, you should refer to the Student Handbook (Section 5.2.1.3). You are reminded that SUSS takes a tough stance against plagiarism or collusion. Serious cases will normally result in the student being referred to SUSS's Student Disciplinary Group. For other cases, significant mark penalties or expulsion from the course will be imposed.

(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 accessed 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 -----