SINGAPORE UNIVERSITY OF SOCIAL SCIENCES

# AIB503

**End-of-Course Assessment - January Semester 2025**

## Foundation to Python for AI

---

**INSTRUCTIONS TO STUDENTS:**

1. This End-of-Course Assessment paper comprises **7** 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: Friday, 28 February 2025 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 TWO (2) items for marking and grading:*

- *A Report*
- *A .Zip File that contains the dataset used for the report and code file used for data analysis*

*Please verify your submissions after you have submitted the above TWO (2) items.*

### *B - Submission Deadline*

- *The TWO (2) items of Report and .Zip File are to be submitted **by 12 noon** on the submission deadline.*
- *You are allowed multiple submissions till the cut-off date for each of the TWO (2) items.*
- *Late submission of any of the TWO (2) items **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 (2) Items Should Be Submitted*

- *The Report: submit online to Canvas via TurnItIn (for plagiarism detection) under the ECA submission link*
- *The .Zip File that contains the dataset and code file:*
  - *Zip the dataset and code file*
  - *Submit the .Zip file online to Canvas via the -ECA Zip File submission link*

### *D - Additional guidelines on file formatting are given as follows*

| | |
|---|---|
| ***1. Report*** | • *Please ensure that your Microsoft Word document is generated by Microsoft Word 2016 or higher.*<br>• *The report must be saved in **.docx** format.* |

| | |
|---|---|
| ***2. .Zip File*** | • *The dataset must be saved in .**xlsx** or .**xls** format.*<br>• *The dataset must be included in the **.zip** file.*<br>• *The code file must be saved in the required format.*<br>• *You are to include the following particulars in your submission: Course Code, Title of the ECA, SUSS PI No., Your Name, and Submission Date.* |

## *E - 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.*

## *F - 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.*

## *G - Use of Generative AI Tools (Allowed)*

*The use of generative AI tools is allowed for this assignment.*

- *You are expected to provide proper attribution if you use generative AI tools while completing the assignment, including appropriate and discipline-specific citation, a table detailing the name of the AI tool used, the approach to using the tool (e.g. what prompts were used), the full output provided by the tool, and which part of the output was adapted for the assignment;*

- *To take note of section 3, paragraph 3.2 and section 5.2, paragraph 2A.1 (Viva Voce) of the Student Handbook;*

- *The University has the right to exercise the viva voce option to determine the authorship of a student's submission should there be reasonable grounds to suspect that the submission may not be fully the student's own work.*

- *For more details on academic integrity and guidance on responsible use of generative AI tools in assignments, please refer to the TLC website for more details;*

- *The University will continue to review the use of generative AI tools based on feedback and in light of developments in AI and related technologies.*

(Full marks: 100)

**Section A (100 marks)**
Answer all questions in this section.

**Question 1**

**Clustering Analysis of Airline Passenger Satisfaction**

The objective of this assignment is to use clustering techniques in Python to group airline passengers based on the provided dataset. Clustering is an unsupervised machine learning method that identifies natural groupings in data by organizing similar data points into clusters. This analysis is particularly useful for the airline industry to gain insights into customer behavior and preferences, enabling more targeted strategies to improve passenger satisfaction and operational efficiency.

The Airline Passenger Satisfaction dataset from Kaggle provides detailed feedback from passengers, covering various aspects of their travel experience. Key features include gender, age, type of travel, class, flight distance, seat comfort, inflight entertainment, onboard service, cleanliness, departure delay, arrival delay, and overall satisfaction.

This assignment is designed to deepen your understanding of clustering algorithms and their real-world applications. Your submission must include a comprehensive report detailing your analysis and findings and a .zip file containing the codes for the following four sub-questions.

**Data Description**:
**Gender**: Gender of the passengers (Female, Male)
**Customer Type**: The customer type (Loyal customer, disloyal customer)
**Age**: The actual age of the passengers
**Type of Travel**: Purpose of the flight of the passengers (Personal Travel, Business Travel)
**Class**: Travel class in the plane of the passengers (Business, Economy, Economy Plus)
**Flight distance**: The flight distance of this journey (miles)
**Inflight wifi service**: Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)
**Departure/Arrival time convenient**: Satisfaction level of Departure/Arrival time convenient
**Ease of Online booking**: Satisfaction level of online booking
**Gate location**: Satisfaction level of Gate location
**Food and drink**: Satisfaction level of Food and drink
**Online boarding**: Satisfaction level of online boarding
**Seat comfort**: Satisfaction level of Seat comfort
**Inflight entertainment**: Satisfaction level of inflight entertainment
**On-board service**: Satisfaction level of On-board service
**Leg room service**: Satisfaction level of Leg room service
**Baggage handling**: Satisfaction level of baggage handling

**Check-in service**: Satisfaction level of Check-in service
**Inflight service**: Satisfaction level of inflight service
**Cleanliness**: Satisfaction level of Cleanliness
**Departure Delay in Minutes**: Minutes delayed when departure
**Arrival Delay in Minutes**: Minutes delayed when Arrival
**Satisfaction**: Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

**Question 1a**

- Data preprocessing:
  - Perform data preprocessing, including handling missing values (e.g., imputation or removal) and scaling numeric features, if necessary, to prepare the data for analysis.

- Explore and visualize the dataset:
  - Analyze the dataset to understand its features by examining distributions, relationships between variables, and correlations.
  - Use at least three visualization techniques, such as histograms, scatter plots, pair plots, or heatmaps, to gain deeper insights into the data.

(25 marks)

**Question 1b**

- Utilize Python to design and execute K-means clustering algorithm to segment airline passengers based on all the features.

- Use techniques such as the elbow method or silhouette score to identify the optimal number of clusters.

- Execute K-means clustering with the selected number of clusters and segment the airline passengers accordingly.

(30 marks)

**Question 1c**

- Appraise and interpret each airline passenger segments obtained from the clustering process based on their attributes.

- Provide insights and recommendations for the airline company based on the identified passenger segments. How can the company leverage these segments for passenger satisfaction improvements?

(25 marks)

**Question 1d**

- Create clear visualizations to represent the clusters derived from the clustering algorithm.

- Appraise the significant findings, discuss the challenges faced during the analysis, and potential next steps for further improvement of the clustering case study.

(20 marks)

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