DAVI: CA2 Assignment Specification
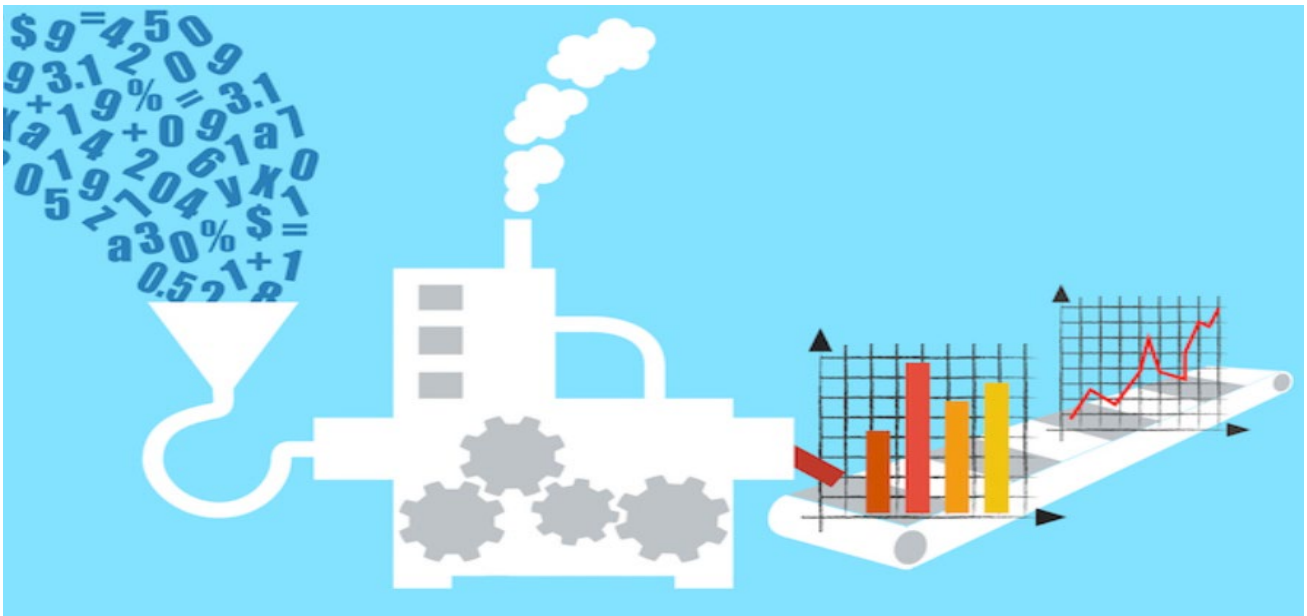
School Name:	School of Computing
Academic Year:	AY2425 Semester 1
Course Name:	DAAA
Module Code:	ST1502
Module Name:	Data Visualization
Assignment:	CA2 (Individual)
Deadline:	**Part 1:** Monday, 15 July 2024 by 0800 hrs
	Part 2: Friday, 2 Aug 2024 by 2359 hrs

Table of Contents

Section 1.1
Seaborn: Instructions and Guidelines

1. This is an **INDIVIDUAL** assignment that contains TWO parts. They add up to **40%** of the module grade. Both parts aim to plot charts and visuals and gather insights using Python.

   a. Part 1: 40% of CA2 (16 marks of DAVI module) requires you to use Python seaborn
   b. Part 2: 60% of CA2 (24 marks of DAVI module) is to use Plotly

2. Section 1 of this document outlines the instructions and requirements of Part 1 of the assignment, whereas Section 2 is about Part 2.

3. Each part has its own submission deadline, as stated on the cover page of this document. Submit all deliverables to Brightspace.

4. Deliverable for Part 1 should be a zip file with the following file-naming convention **"YourStudentID-YourName-Part1.zip"**

5. Zip file should include the following items:
   - **One Jupyter notebook** that accomplishes the given tasks using the Python programming language. Your programs should include remarks or comments.
   - **Datasets** that you have cleansed and can be analyzed by the Python code that you have written
   - **A set of Powerpoint slides** that summarizes the data insights that you have gained through your Python code

6. No presentation or interview is required for Part 1. Thus, your Powerpoint slides must be self-explanatory.

7. A penalty will be imposed on late submission. 50% of the marks will be deducted for assignments that are received within ONE (1) calendar day after the submission deadline. No marks will be given thereafter.
   Exceptions to this policy will be given to students with valid LOA on medical or compassionate grounds. Students in such cases will need to inform the lecturer as soon as reasonably possible. Students are not to assume on their own that their deadline has been extended.

8. Plagiarism is a serious offence. No marks will be awarded, if the work is copied or you have allowed/enabled others to copy your work. If you are found to have committed, aided, and/or abetted the offence of plagiarism, disciplinary action will be taken against you.

   Warning: Plagiarism means passing off as one's own ideas, works, writings, etc., which belong to another person. In accordance with this definition, you are committing plagiarism if you copy the work of another person and turning it in as your own, even if you have the permission of that person.

Section 1.2
Seaborn: Scope of the assignment

In this part of the assignment, you are required to write Python program and produce a data analysis presentation for a dataset based on the requirements stated below.

Basic Requirements

1. Assignment 2 is about analysing the Student Profiles of a business training institute. You are required to use seaborn to plot charts and provide compelling insights about the student profiles.

2. For Part 1, state **1 objective of your own choice.** For this objective, plot 3 charts. Label each chart clearly. Include variable for x-axis, variable for y-axis, add scale to the x-axis, add scale for the y-axis, units of measurement, title to the chart.

3. Write **python programs** to cleanse or wrangle the dataset. Use **seaborn** package to plot charts.
   a. Create attractive and aesthetically pleasing charts.
   b. You will need to plot univariate, bivariate and multivariabe charts.

4. Explain the insights of each chart by explaining **not** more than three points, as demonstrated in the example in Section 1.4 of this document.

5. Your Python programs should help you to gain deeper insights into the given dataset such that you are able to craft a 'storyline' or produce an interesting data analysis on it.

   Compile your findings into a deck of **Powerpoint slides** (not more than 10 slides).

   Your Powerpoint slides should include the following sections:

   - A cover page that lists your name and the title of your data analysis
   - A slide to briefly explain the **nature of that dataset** (i.e. what is in that dataset) or any pecularities about it you wish to highlight.
   - Explain the **process** you went through to cleanse or wrangle the dataset.
   - A screenshot of each chart
   - For each chart, the **insights** you have gained from analysing the data
   - Provide at least 1 recommendation

Section 1.3
Seaborn: Marking Scheme

Marks will be awarded to each student based on the following rubrics.

To score higher marks, you are encouraged to explore and experiment beyond the syllabus and demonstrate your independently-acquired skills via your deliverables. You may access to the online learning platforms such as datacamp to learn more about Data visualization using seaborn.

Component	Weightage
1. **Clarity of project objective and data wrangling** • State the objective on analyzing the given dataset. • Explain the process you went through from getting the raw data to working on data wrangling and finally working on the final set of data • Summarise key insights gained from the analysis of the data • Provide at least 1 recommendation	10%
2. **Quality of application** • Technical complexity • User-friendliness • Aesthetics **&** Creativity	18% (3 charts * 6 marks each)
3. **Data analysis (<u>Powerpoint Slides</u>)** • Quality of Presentation & Slides • Coherent and Completeness in the analysis of data	12% (3 charts * 4 marks each)

Section 1.4
Seaborn: Sample outputs

This section contains sample plots and insights using seaborn.

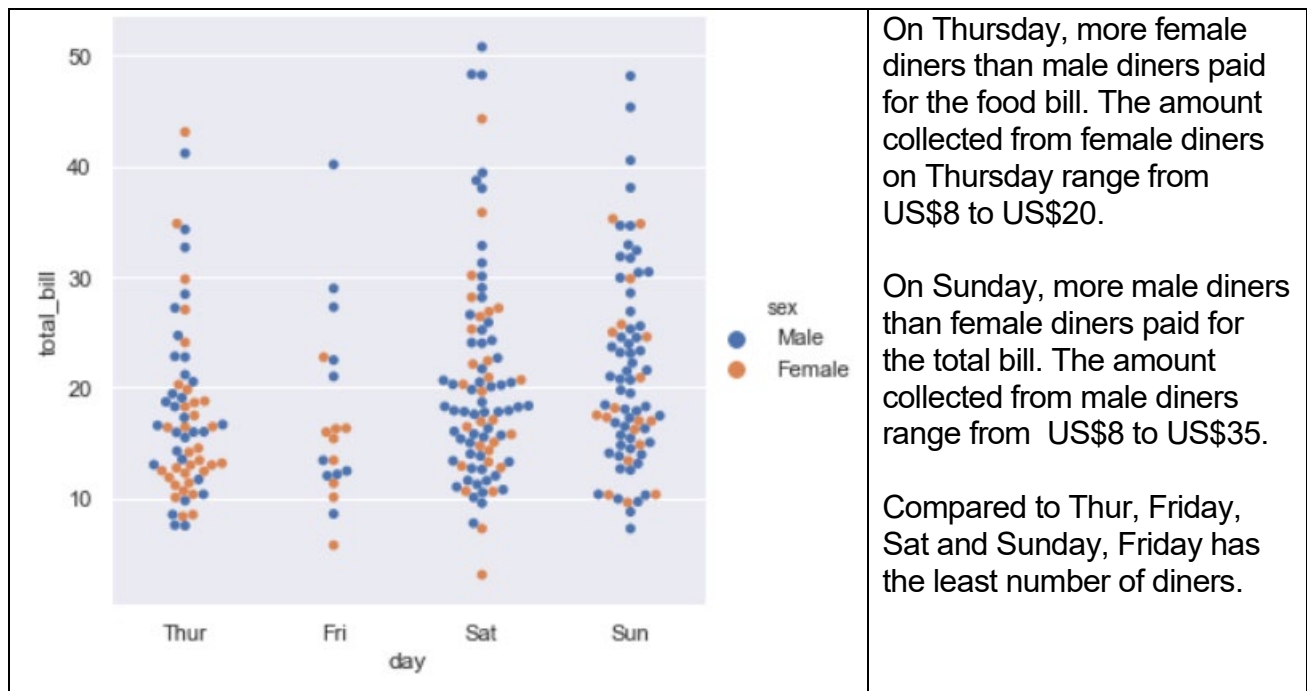| Example 1: Swarm Plot (Bivariate) |
|---|

To determine the total bill collected from the different gender in a Swiss Café restaurant in USA.
The restaurant is open from 5:00 pm till 11:00 pm on Thursday till Sunday.

You may add another dimension to a categorical plot by using a hue semantic.
Code:
sns.catplot(x="day", y="total_bill", hue="sex", kind="swarm", data=tips);

<table>
<tr><td>

</td><td>

On Thursday, more female diners than male diners paid for the food bill. The amount collected from female diners on Thursday range from US$8 to US$20.

On Sunday, more male diners than female diners paid for the total bill. The amount collected from male diners range from US$8 to US$35.

Compared to Thur, Friday, Sat and Sunday, Friday has the least number of diners.

</td></tr>
</table>

Example 2:
Violin Plot (Trivariate)

This sample output uses the [seaborn](#) library to plot a static violin chart visualization showing the IQ scores of different test groups. Seaborn produces much more aesthetically-pleasing charts than Matplotlib.

IQ Scores for different test groups (Group A and Gropu B). Group B mean IQ score is 100 (for example)
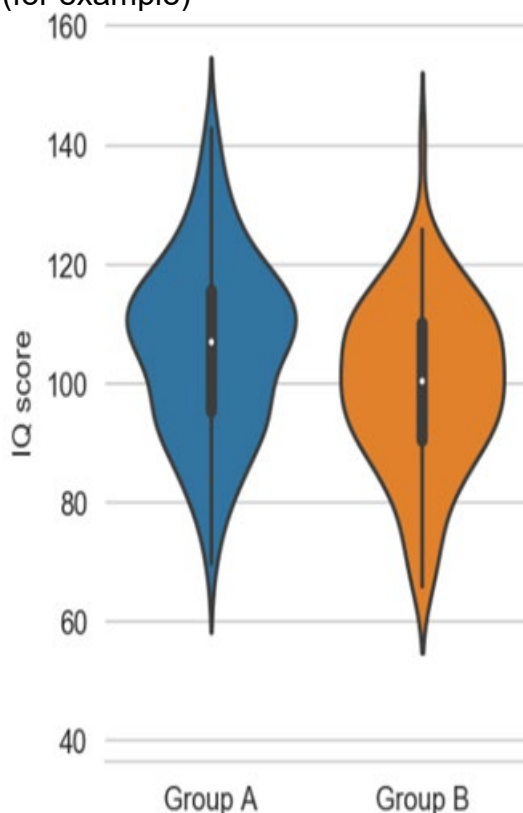
Sample of Analysis:

- The median for Group B is lower than as compared to Group A. The median IQ score for Group B is 100. There were 50% participants who had IQ score between 90 and 110. The mode IQ score is 100. For Group B, the shape of the distribution is wide in the middle indicating the IQ score are highly concentrated around the median. Group B exhibited a normal distribution where the average IQ score in the United States is about 100.

- The median for Group A is slightly higher than Group B, the median IQ score is 110. There were 50% of the participants in Group A that has IQ score of between 95 and 115. There is bi-modal for group A, the IQ scores are 95 and 115. For Group A, there is a group of participants with IQ score of 115 and another group of participants with IQ score of 90.

- Reference: Available at https://www.healthline.com/health/what-is-considered-a-high-iq

Section 2.1
Plotly: Instructions and Guidelines

1. This outlines the instructions and requirements of Part 2 of the assignment which takes up 60% of CA2.

2. Observe Part 2 submission deadline stated on the first page of this document. Submit all deliverables to Brightspace.

3. Deliverable for Part 2 should be a zip file with the following file-naming convention **"YourStudentID-YourName-Part2.zip"**

4. Zip file should include the following items:
   - **One Jupyter notebook** that accomplishes the given tasks using the Python programming language. Your programs should include remarks or comments.
   - **Datasets** that you have cleansed and can be analyzed by the Python code that you have written
   - **A set of Powerpoint slides** that summarizes the data insights that you have gained through your Python code

5. You will need to give a 8-minute **presentation** cum interview using the Powerpoint slides and python codes you have prepared. Your module tutor may ask you questions related to Python code and Data visualization during this interview / presentation session.

6. Like in Part 1, a penalty will be imposed on late submission. 50% of the marks will be deducted for assignments that are received within ONE (1) calendar day after the submission deadline. No marks will be given thereafter.

   Refer to Part 1 on the exceptions to this policy.

7. Plagiarism is a serious offence. No marks will be awarded, if the work is copied or you have allowed/enabled others to copy your work. If you are found to have committed, aided, and/or abetted the offence of plagiarism, disciplinary action will be taken against you.

   Warning: Plagiarism means passing off as one's own ideas, works, writings, etc., which belong to another person. In accordance with this definition, you are committing plagiarism if you copy the work of another person and turning it in as your own, even if you have the permission of that person.

Section 2.2
Plotly: Scope of the assignment

In this part of the assignment, you are required to write Python program and produce a data analysis presentation for a dataset based on the requirements stated below.

Basic Requirements

1. Part 2 is based on the same datasets given in Part 1. It is also about analysing the Student Profiles of a business training institute, but this time you are required to use plotly to plot charts and provide compelling insights about the student profiles.

2. For Part 2, state **1 objective of your own choice**. The objective should be different from that of Part 1.

3. If necessary, you may wish to perform further data wrangling on the dataset cleansed in Part 1 so as to meet the objective for Part 2.

4. Plot charts that use python **plotly** package and meet the following requirements:

   a. Two charts using plotly express or graph objects
   b. One plotly dash that may contain one or more charts and include interactive elements such as dropdown menu.

5. Explain the insights of each chart by explaining **not** more than three points, as demonstrated in the example in Section 1.4 of this document (but using plotly).

6. Your Python programs should help you to gain deeper insights into the given dataset such that you are able to craft a 'storyline' or produce an interesting data analysis on it.

   Compile your findings into a deck of **Powerpoint slides** (not more than 10 slides).

   Your Powerpoint slides should include the following sections:

   - A cover page that lists your name and the title of your data analysis
   - A slide to briefly explain the **nature of that dataset** (i.e. what is in that dataset) or any pecularities about it you wish to highlight.
   - Explain the **process** you went through to cleanse or wrangle the dataset.
   - A screenshot of each chart
   - For each chart, the **insights** you have gained from analysing the data
   - Provide at least 1 recommendation

7. You may explore features in plotly dash that are not taught in the lesson, e.g. other interative elements in dash such as checkbox.

Section 2.3
Plotly: Marking Scheme

Marks will be awarded to each student based on the following rubrics.

To score higher marks, you are encouraged to explore and experiment beyond the syllabus and demonstrate your independently-acquired skills via your deliverables / interview. You may access to the online learning platforms to learn more about Data visualization using plotly.

Component	Weightage
1. **Clarity of project objective and, if necessary, data wrangling** • State the objective on analyzing the given dataset. • Explain the process you went through for performing further data wrangling, if necessary, and finally working on the final set of data • Summarise key insights gained from the analysis of the data • Provide at least 1 recommendation	6%
2. **Quality of application** • Technical complexity • User-friendliness • Aesthetics **&** Creativity (a) Two charts using plotly express or graph objects (b) One plotly dash	22% (a) 2 charts * 6 marks (b) 1 plotly dash * 10 marks
3. **Data analysis (<u>Powerpoint Slides</u>)** • Quality of Presentation & Slides • Coherent and Completeness in the analysis of data	12% (3 charts * 4 marks each)
4. Presentation, Interview and Q and A session	10%
5. Additional • Features of plotly dash not demonstrated in the lesson • Advanced techniques	10%

- **End of assignment specifications -**