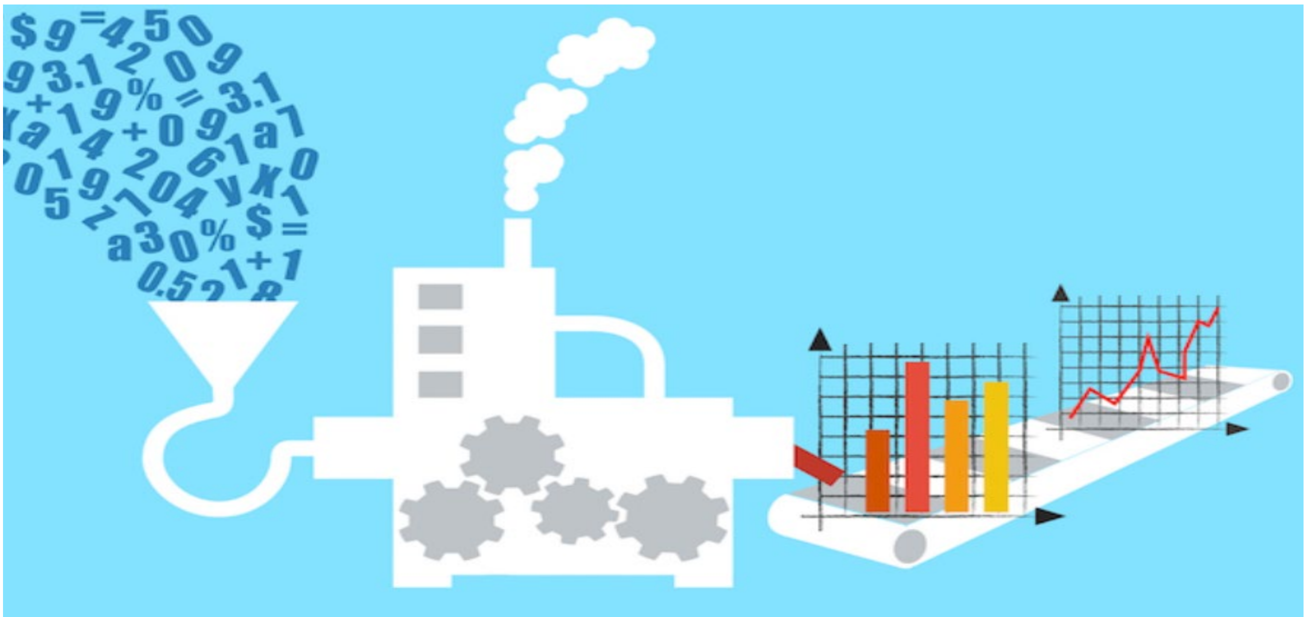


School Name:	School of Computing
Academic Year:	AY2223 Semester 2
Course Name:	DAAA
Module Code:	ST1502
Module Name:	Data Visualization
Assignment:	CA2 (Individual)
Deadline:	Mon, 6 Feb 2023 by 2359

Table of Contents

Section 1
Instructions and Guidelines

1. This is an **INDIVIDUAL** assignment which requires you to use Python Seaborn and Plotly to plot charts and visuals and gather data insights.

2. The requirements of this assignment are outlined in Section 2 of this document.

3. Submit CA2 Assignment in Brightspace.

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

5. Zip file should include the following items:
    - **One Jupyter notebook** that accomplishes the given tasks using the Python programming language
    - **A set of Powerpoint slides** that summarizes the data insights that you have gained through the Python code you have written

6. Each student is allocated 7 mins to present using the slides and python codes to their tutor. There is a 3 mins Q and A session. Your module tutor may ask you questions relating to Python code and Data visualization during the Q and A session. The total number of powerpiont sliges is 15.

7. This assignment will account for **40%** of the **module grade**.

8. No marks will be awarded, if the work is copied or you have allowed others to copy your work.

9. 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.

Section 2
Scope of the assignment

In this individual 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. The topic for the Assignment 2 is "Singapore's Land Public Transport". You are required to use seaborn and plotly to plot charts and provide compelling insights. You may use the three datasets provided (note: datasets are from kaggle): Bus_routes, Bus_services and Bus_stops. Alternatively, you may use other datasets.

2. State **2 objectives.** For each objective, plot 3 charts. You are required to plot 6 different charts (note: 4 charts, use seaborn package and 2 charts use python plotly). Please label each chart. 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 program** and use seaborn and python plotly 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. For each chart, you may explain using three points

5. Your Python programs should help you to gain deeper insights into the chosen dataset(s) 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**

   Your Powerpoint slides should include the following sections:

   - A cover page that lists your name and the title of your data analysis
   - A slide that lists the URLs of all the datasets you have used
   - 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.
   - For each dataset, the **insights** you have gained from analysing the data and provide 2 recommendations.

6. You may create interactive data visualisation using plotly dash. (optional).

Section 3
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 such as datacamp to learn more about Data visualization using seaborn and plotly.

Component	Weightage
1. Clarity of project objectives and data wrangling The topic is on Singapore's Land Public Transport. • You will need to state 2 objectives you wish to work on in the context of Singapore's Land Public Transpot. • 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 2 recommendations	20%
2. **Quality of application** • Technical complexity • User-friendliness • Aesthetics **&** Creativity	36% (6 charts * 6 marks each)
3. **Data analysis (<u>Powerpoint Slides</u>)** • Quality of Presentation & Slides • Coherent and Completeness in the analysis of data	24% (6 charts * 4 marks each)
4. Presentation, Interview and Q and A session	10%
5. Additional • You may also explore dash python(optional) • Advanced techniques	10%

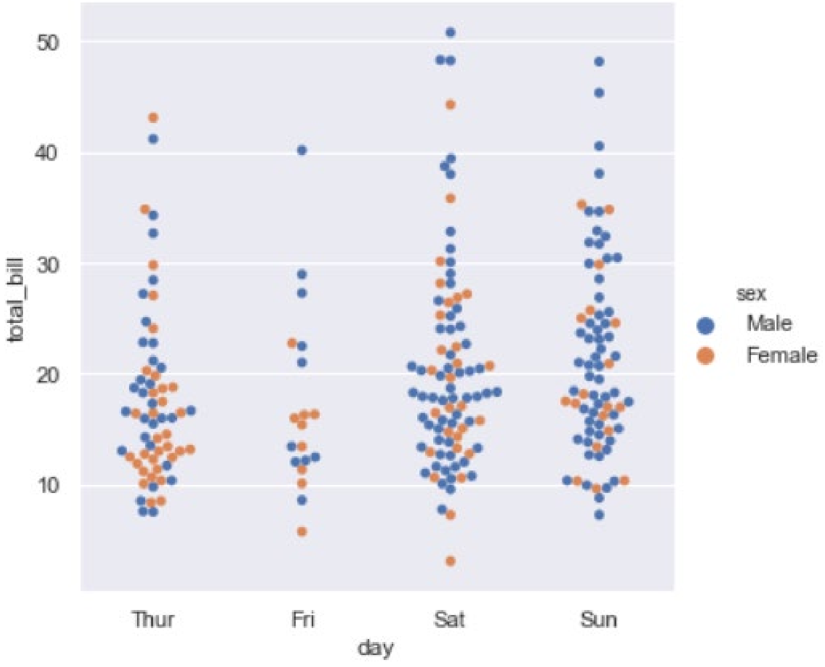
Section 4
Sample outputs expected

This section contains sample plots and insights.

Example 1: Swarm Plot

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);



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.

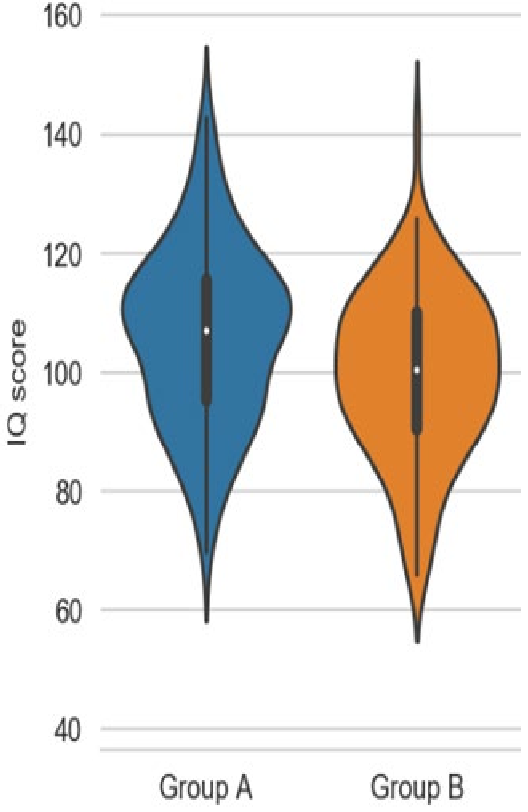
Compared to Thur, Friday, Sat and Sunday, Friday has the least number of diners.

Provide reasons and references to support insights

Example 2:
Violin Plot

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

-- End of Assignment Specifications --