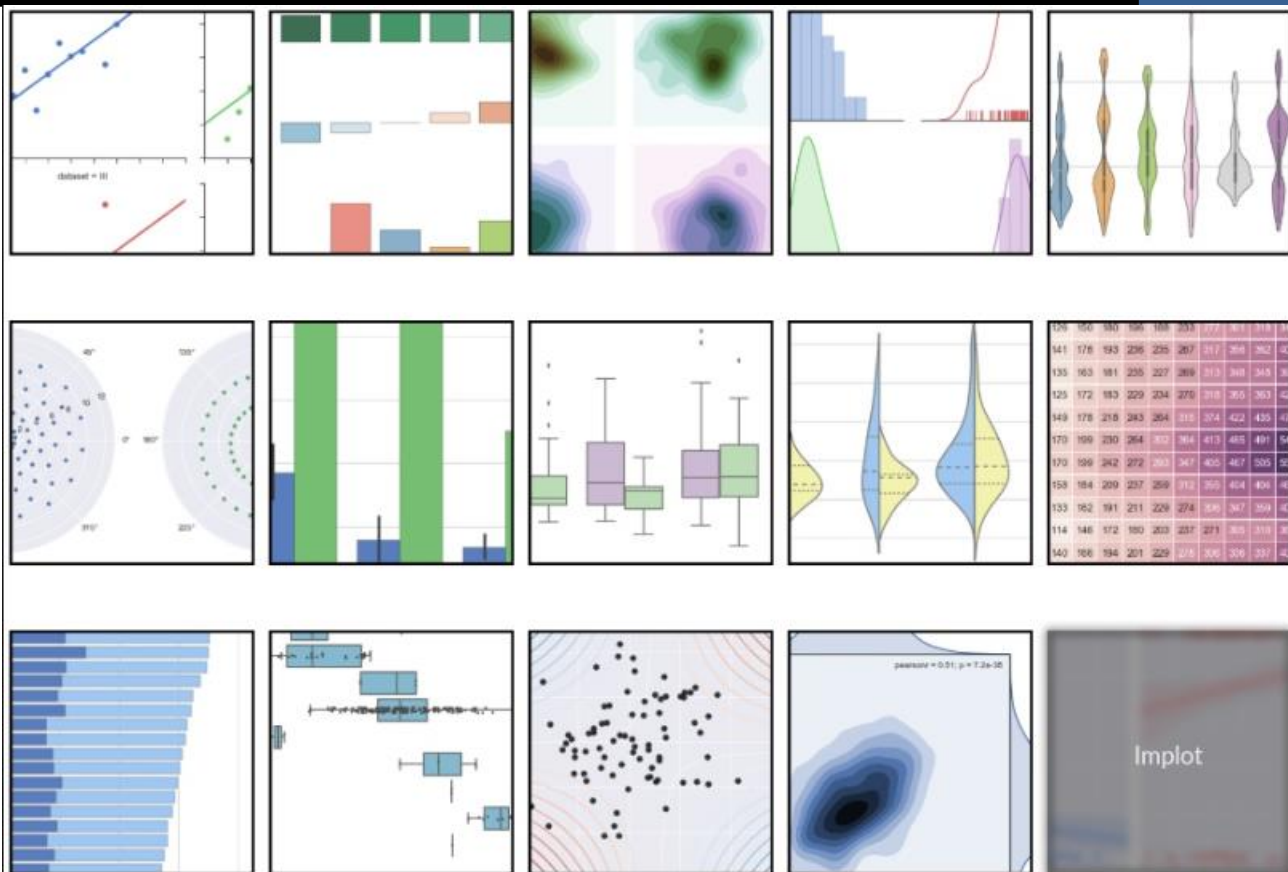


2021

DAVI Assignment 2 (CA2)



School Name: School of Computing

Semester: AY2021 semester 2

Course Name: DAAA

Module Code: ST1502

Module Name: Data Visualization

Deadline: Mon, 15 Feb 2020 by 8:00 am

2/15/2021

Table of Contents

Section 1 Instructions and Guidelines	2
Section 2 Scope of the assignment	3
Basic Requirements.....	3
Section 3 Marking Scheme	4
Section 4 Sample outputs expected	5
Example 1: Swarm Plot (Bivariate).....	5
Example 2: Violin Plot (Trivariate)	6

Section 1

Instructions and Guidelines

1. This is an **INDIVIDUAL** assignment which requires the student to code a Python application using seaborn library. Please note that DAVI Assignment 2 and PDAS Assignment 2 are integrated assignment. The final output dataset from PDAS CA2 module is the dataset you will use to perform data visualization and data analysis.
2. The requirements of this assignment are outlined in Section 2 of this document.
3. Submissions should be made via the Blackboard CA2 Assignment Submission link by the stated deadline.
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. You will need to give a 10 mins **presentation** cum interview using the Powerpoint slides you have prepared. Your module tutor may ask you questions related to Python code and Data visualization during this interview / presentation session.
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 DAVI (ST1502) Assignment 2 is “Singapore’s Transport System”. This topic is the same as PDAS (ST1510) assignment 2. For PDAS assignment 2, you are required to source for 3 data sets and work on data wrangling. The output dataset from PDAS assignment 2 is then used in DAVI module where you are required to use seaborn to plot charts and provide compelling insights about Singapore’s Transport system.
2. For DAVI Assignment 2, you may use one or 2 datasets to plot charts; one for each objective. You may use the Pandas package to load the dataset(s).
3. State **2 objectives**. For each objective, plot at least 3 charts using seaborn package. In total, you are required to plot 6 charts. 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 graph
4. Write **python program** and use seaborn package to plot charts.
 - a. Create attractive and aesthetically pleasing charts using Seaborn.
 - b. You will need to plot at least one univariate chart (e.g. bar chart), one bivariate chart (e.g. pairplot) and one multivariate chart (e.g. heatmap)
 - c. Use plotly package to create maps (optional)
5. Explain the insights of each chart. For each chart, you may explain using at least three points
6. 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 peculiarities about it you wish to highlight.
- For each dataset, the **insights** you have gained from analysing the data and provide 2 recommendations.

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 or Udemy to learn more about Data visualization using seaborn.

Component	Weightage
1. Clarity of project objectives The topic is on Singapore's Transport System. Public transport system includes mass rapid transport (MRT), bus and taxi. The final dataset is the output from PDAS Assignment 2 module. <ul style="list-style-type: none">• You will need to state 2 objectives.• Provide a data dictionary of the final dataset• Summarise key insights gained from the analysis of the data• Provide 2 recommendations	30%
2. Quality of application <ul style="list-style-type: none">• Technical complexity• Code quality• User-friendliness• Aesthetics• Creativity	30% (6 charts * 5 marks each)
3. Data analysis (Powerpoint Slides) <ul style="list-style-type: none">• Coherent and Completeness in the analysis of data• Quality of a deck of slides	30% (6 charts * 5 marks each)
4. Presentation, Interview and Q and A session	10%

Section 4

Sample outputs expected

This section contains sample plots and insights.

Example 1: Swarm Plot (Bivariate)

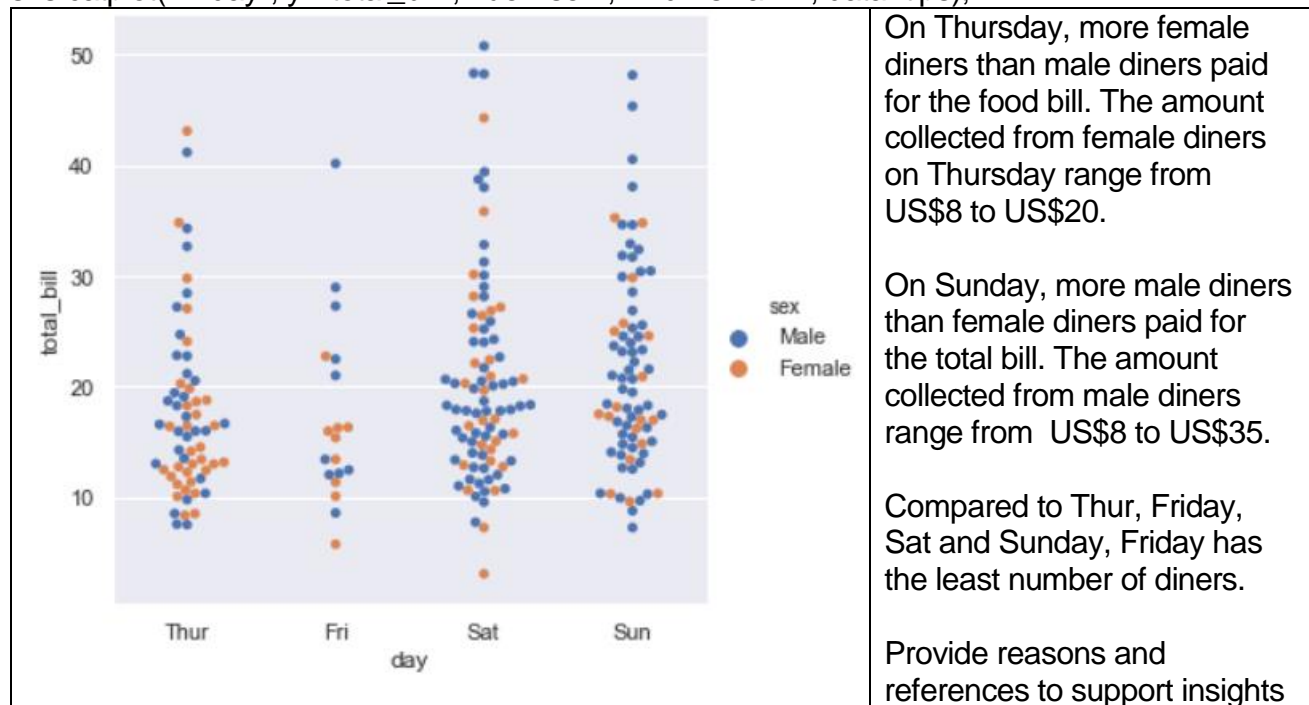
To determine the total bill collected from the different gender in a Swiss Café restaurant in USA.

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:

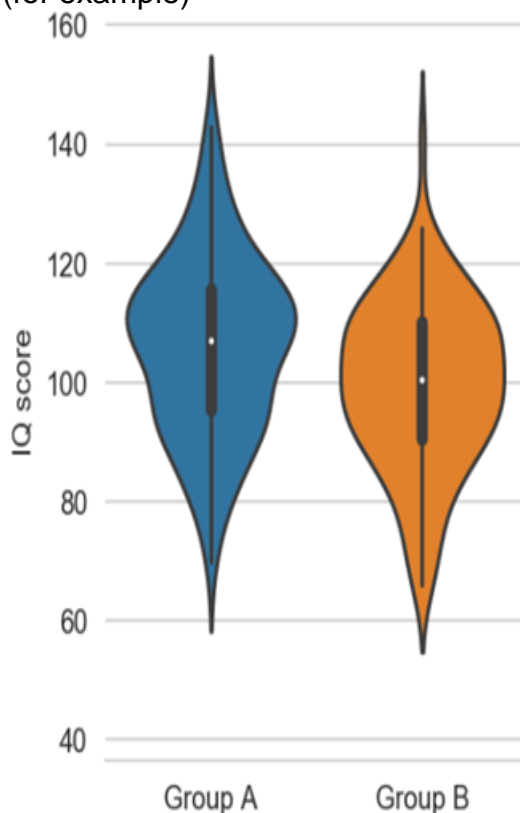
```
sns.catplot(x="day", y="total_bill", hue="sex", kind="swarm", data=tips);
```



**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 Group 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 --