School of Computing
EP0302 Programming for Data Science PDS
Assignment Brief: CA2 AY2526S2

Assignment Rubrics

By completing this assignment, you will have been able to:

- Demonstrate basic competency in writing a Python program in a notebook environment
- Demonstrate basic competency in using the Python Pandas, Matplotlib and Seaborn packages for data visualization
- Demonstrate basic competency in applying the insights gained from the outputs of your Python programs to deliver a useful data visualization report for your stakeholders

Assignment Instructions and Guidelines

1) This is an **individual** assignment which requires you to write a simple Python program that retrieves data from a CSV data source and perform basic operations on the dataset such as, but not limited to, transformation, cleansing and visualization.
2) The requirements and deliverables are outlined in the following section of this document.
3) The deadline of this assignment is on 20th February 2026, 11.59PM.
4) Submissions should be made via the Brightspace CA2 assignment submission link by the stipulated dateline.
5) Your deliverables should be contained in a zip file with the following convention for its name:
   CA2-[Elective Class]-[Admin Number]-[Name].zip
   Example: CA2-01-23120303-JohnDoe.zip
6) Your zip file should include the following deliverables:
   - One Python notebook environment that accomplishes the given tasks in the assignment brief below using the Python programming language. The notebook will also document the insights into the data that you have gained through the Python code that you have written.
   - One HTML exported version of the Jupyter notebook.
   - All datasets (.csv files) used (including the recommended datasets).
   - One Declaration of Academic Integrity (SOC version).
7) As part of the assignment requirements, you will be interviewed based on the Jupyter notebook environment that you have submitted. Your module tutor may ask you to reproduce certain parts of your code during this interview session. You do not need to reproduce the same code in question, but the code should be able to perform the same task. Usage of Google will be allowed during the process. Usage of Gen AI tools and any other form of AI assisted tools will not be allowed during the interview process.
8) This assignment will account for 40% of the module grade.
9) No marks will be awarded upon meeting any of the following conditions:
   a. Your work is copied/plagiarized, or was allowed to have been copied/plagiarized.
   b. You are unable to reproduce code or are unable to answer most of the interview process.
   c. You use prohibited packages in your assignment.
   d. Your zip file is corrupted (Please double check by downloading your work after submitting it).
   e. You submit your assignment more than one day late.
10) 50% of the marks will be deducted for assignments that are received within one (1) calendar day after the submission dateline. No marks will be given thereafter.

Assignment Brief and Scope

Introduction and Brief

This individual assignment is a data science project to be written in a Python notebook environment. The scope of this data science project is centered around relevant data in the context of domains in Singapore. You are required to come up with your own topic statement relevant to an issue/domain in Singapore (e.g., housing, car prices, food prices, etc., etc.,) and write a Jupyter notebook investigating your topic statement.

Scope

1) Relevant datasets regarding Singapore can be found at https://data.gov.sg. You may use external datasets not from this website; however, your topic statement should still be relevant to issues in the context of Singapore.
   a. For this assignment, you may reuse your previous topic statement and make relevant improvements to your CA1 project.
2) Come up with a topic statement relevant to Singapore that you would like to investigate and source at least three relevant datasets from the link provided above to use.
   a. Your data science project should have a strong and clear goal. (e.g., evaluation of certain government policies)
   b. Avoid broad subject topic statements such as “Analysis of COE prices in Singapore”
3) Your Jupyter notebook should include the following:
   a. Your name and the title of your data analysis
   b. The questions you want to answer and gain deeper insights into with support from the datasets that you have chosen and provide insights and analysis on it.
   c. A list of URLs of all the datasets that you have selected (including the those from the link that we have provided)
   d. For each dataset, you should develop a Python notebook environment that extracts useful statistical and summary information regarding that dataset that you have chosen using any package of your choice (e.g., NumPy and Pandas are both available to you for this assignment). Explain the process throughout the data extraction and cleaning process.
   e. For each dataset, you should explain the nature of the dataset that you have chosen, and report primary inferences and conclusions that you can make.
   f. For each dataset that you have extracted data from, you should use appropriate data visualization techniques to represent said data.
      i) You should have a variety of data visualization techniques to represent your data, but do not force a visualization technique if the dataset does not call for it.
      ii) You should think about the inferences you want to make when thinking of what variety of visualization techniques to use.
   g. For each dataset, describe the insights you have gained from analysing the data and any conclusions that you have made from the analysis.
      i) You should be making recommendations and/or be able to describe the impact of your study.

Marking Scheme

Marks will be awarded to each student based on the following marking scheme:

Component	Weightage
Base Assignment Requirements • Choose at least 3 non-trivial (more than 10 rows) relevant datasets. You are encouraged to use more datasets, especially datasets that are inter-related to others, and you can utilize the inter-relationship in your analysis. • Python program that can extract useful insights from the datasets using the NumPy library • Python program that can produce useful data visualizations from the datasets using the Pandas and the Seaborn library • Explanation of the datasets and what was done to process said datasets • Report quality of the insights gained from the analysis of the data	30%
Quality of Notebook • Python program technical complexity	15%

• Python program code quality and cleanliness • Report formatting and quality • Usage of markdown cells for text	
Data Analysis • Completeness in the analysis of the data • Depth of the questions explored • Quality of the answers you provide	30%
Interview • Technical competency • Explanation of insights and thought process • Flow of the interview process and content	25%
Total	100%

-- End of Assignment Specifications --