**SCHOOL OF COMPUTING (SOC)**

**IT8701 Introduction to Programming for Data Science**

**Self Reflection (CA1)**

|  |
| --- |
| **Instructions:**   1. Submit this at Polymall “Assignments->CA1->Self-Reflection” folder 2. Name your file “YourModuleLecturerName-YourStudentID-YourName.docx” |

|  |  |
| --- | --- |
| **Your Lecturer’s Name** | Chia Yao An, Kendrick |
| **Your Name** | Sng Tian Hao, Keith |
| **Your Student ID** | 1273381F |
| **Your Class** | C |

# QUESTION 1: CHALLENGES - SELF-REFLECTION FOR CA1

Provide a brief reflection of the challenges you have faced in this assignment.

|  |
| --- |
| Onerous to use numpy and matplotlib to manipulate and visualise data. Also challenging to create new features and customise plots by having to specify exact arrays/values for various plot arguments (x, y, colour, etc.).  Making graphs visually appealing was also difficult as there were many arguments to adjust. Many of which are usually hidden in the library documentation as kwargs. Often had to search on forums such as stackoverflow for clues on where to begin. A lot of things that I didn’t know existed. |

# QUESTION 2: ACHIEVEMENTS - SELF-REFLECTION FOR CA1

Provide a brief reflection of what you think you have personally achieved in this assignment or the knolwedge or skills you have found satisfaction in learning / acquiring.

|  |
| --- |
| Customisation of matplotlib graphs was challenging but useful as I internalised the thought process/workflow for creating visuals (e.g., examine variables and determine the type of chart to create > arrange data into relevant arrays for plotting > format title, axes, labels > size graphs appropriately  Utilised API calls to feature engineer and reduced run time by optimising code (e.g., get only distinct addresses for API call → reduced initial run time for 179,586 records to just 9,561 within 20+ mins) |

# QUESTION 3: SELF-EVALUATION

Grade yourself using the marking rubrics below.

### **How well did I meet the BASIC assignment requirements?**

For each criteria, place a tick ☑ in the column that best matches what you have done for the assignment.

State the evidence in the “Evidence” so that your lecturer can verify, otherwise you will be asked to show evidence during your assignment interview

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Criteria | **Fully met** | **Partially met (at least 50%)** | **Below requirements** | **Evidence** |
| My CA1 submission uses at least 3 different datasets from HDB at data.gov.sg | ☑ |  |  | Used 5 datasets |
| My CA1 submission included ALL the 4 compulsory charts | ☑ |  |  | Created 6 charts |
| My CA1 submission purely used the Numpy library in my Python codes to perform data manipulation only (i.e. I did not resort to easier ways to achieve the requirements using other libraries such as pandas etc) | ☑ |  |  | Numpy to manipulate raw data. However, json\_normalize function from pandas library was used only to help parse results from API calls (pandas not used for manipulating raw datasets) |
| My CA1 submission purely used the Matplotlib library in my Python codes to perform data visualization only (i.e. I did not resort to easier ways to achieve the requirements using other libraries such as seaborn, pygal etc) | ☑ |  |  | Charts created using matplotlib |
| My CA1 submission includes a deck of Powerpoint slides that explain the datasets I used, what was done to process these datasets and summarizes the insights gained from the analysis of the data | ☑ |  |  | See attached in zip file |
| My CA1 submission includes a self-reflection document that outlines my challenges and achievements doing this assignment | ☑ |  |  | See attached in zip file |

### **How high is the quality of my CA1 assignment?**

For each criteria, place a tick ☑ in the column that best matches what you have done for the assignment.

Justify your answer in the “Evidence” column so that your lecturer can verify, otherwise you will be asked to show evidence during your assignment interview.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Criteria | **Above Average** | **Average** | **Below average** | **Evidence** |
| I evaluate the **technical complexity** of my assignment as:  \*A technically complex assignment should include many advanced features that are not taught in the class and are not trivial to code | ☑ |  |  | Defined helper function in a separate py module for API calls, used dictionaries, list comprehension and loops to help with manipulation and extracting data |
| I evaluate the **code quality** of my assignment as:  \*An assignment with high code quality often includes high usage of reusable functions, demonstrates code efficiency through use of appropriate language constructs (e.g. for loops) and is well-documented. | ☑ |  |  | Heavily utilised for loops to minimise redundancy, created a module for API extraction with docstring for reusability |
| I evaluate the **user-friendliness** of my assignment as:  \*A user-friendly application is typically one that provides an easy-to-use user interface (UI) that novice users can understand and navigate with ease. For the purpose of CA1, since there is no /limited UI, please evaluate user-friendliness of your assignment as “How organised is your code and how easily and smoothly another person like your lecturer can run the code on his computer” | ☑ |  |  | Added annotations using markdown, code and datasets stored in properly organised folders which should allow the next user to run the code correctly |
| I evaluate the **aesthetics** of my assignment as:  \*An assignment which has a high level of aesthetics for this module’s CA1, should show effort by the student to enhance their graphical outputs with attractive and pleasant layouts and color combinations | ☑ |  |  | Use of appropriate visuals (colours, positioning, formatting, size) |
| I evaluate the **creativity** of my assignment as:  \*An assignment which demonstrates creativity includes ideas that are novel and not implemented by other students | ☑ |  |  | Created a map using scatter plots |

### **How in-depth and insightful is my data analysis?**

For each criteria, place a tick ☑ in the column that best matches what you have done for the assignment.

Justify your answer in the “Evidence” column so that your lecturer can verify, otherwise you will be asked to show evidence during your assignment interview.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Criteria | **Above Average** | **Average** | **Below average** | **Evidence** |
| I evaluate the **completeness** of my data analysis as:  \*A data analysis that has a high level of completeness requires the analyst to perform a lot of drilling-in/ cross-analysis of the data. If you think you performed above average here, you should show evidence that you went ‘very deep” in digging out details and made effort to explore related datasets etc | ☑ |  |  | Derived insights from various housing options in the market, allowing potential home owners to make an informed decision depending on their preferences/financial capability. Further went to derive more granular insights by performing aggregations (e.g., by estate or room type) for descriptive statistics and by engineering features to plot a map of resale and rental units, something that is not achievable using the raw dataset alone |
| I evaluate the **quality** of my analysis as:  \*A data analysis report that is above average is usually prepared by a student who shows high clarity about the goals he wants to achieve through analysing the data. This includes knowing the specific target audience he wants to present the report to and the insights he wants to derive from the analysis. The analyst is able to present a convincing argument and conclusion to what he sets up to find. If you think your data analysis is of high quality, ask yourself if you are you confident that the target party for your analysis would think your analysis is quite interesting and useful to him/her? |  | ☑ |  | Set context of analysis with identified target audience for better flow. However, analysis is only basic and to some extent haphazard as graphs for each dataset were created based on what’s available and to fulfil assignment criteria (i.e., 4 charts, ideally with little repeats). For completeness, insights for individual components and be better conceived as there are many more variables to explore |

### **How much effort did I put in for my self-reflection?**

For each criteria, place a tick ☑ in the column that best matches what you have done for the assignment.

Justify your answer in the “Evidence” column so that your lecturer can verify, otherwise you will be asked to show evidence during your assignment interview.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Criteria | **Above Average** | **Average** | **Below average** | **Evidence** |
| I evaluate the effort that I put in to explain the **challenges** that I faced in this self-reflection assignment as: | ☑ |  |  | See above |
| I evaluate the effort that I put in to explain the **achievements** that I faced in this self-reflection assignment as: | ☑ |  |  | See above |

**-- End of Self-Reflection --**