**SCHOOL OF DIGITAL MEDIA AND INFOCOMM TECHNOLOGY  
ST0248 Programming for Data Science**

**Self Reflection (CA2)**

|  |
| --- |
| **Instructions:**   1. Submit this together with your other deliverables at Blackboard “Assignments->CA2” folder 2. Name your file “YourModuleLecturerName-YourStudentID-YourName.docx” |

|  |  |
| --- | --- |
| **Your Lecturer’s Name** | Mrs Dora Chua |
| **Your Name** | Lewis Lye Cheng See |
| **Your Student ID** | 1727571 |
| **Your Class** | DIT/2B/23 |

# QUESTION 1: CHALLENGES - SELF-REFLECTION FOR CA2

Provide a brief reflection of the challenges you have faced in this assignment (CA2) with a comparison of those you faced in CA2. Did you manage to apply any learning experiences from the previous assignment to improve the way you did this assignment?

|  |
| --- |
| **General Problems**   * Had issues figuring out what datasets to use * Could not think of a hypothesis to work on * After finding datasets, I had issues figuring out how to plot them as well * I had to also search for additional datasets as I did not have suitable datasets to plot both box plot and histogram   **Reading and Storing Data**   * Did not realise that my data was added to mongoDB multiple times, so I had to delete my database a few times * For the market share dataset, I originally only intended to use data from 2015. But realized that it was too little, so I tried to get data from 2009 (earliest available). Upon doing that, I realized that I had to repeat the same code multiple times to append all the data. * The crime rates dataframe would not save to mongoDB, unless I changed the year column to string. And I still don’t know what caused the error.   **Plotting**   * Original plan was to use bokeh, however, the lack of documentation made it difficult for me to use it. Furthermore, the sample code and codes I tried to use would not work. * Attempted to use pyplot but my version of Python was not supported. Lack of time lead me to avoid using Dash, I settled with using matplotlib again as it is still used in the real world * Attempting to annotate multiple graphs of mine, the annotations at time would cut into the plot or go out of the axes. Had to settle with the most aesthetically pleasing position for some, which might not have been the best * Original idea for scatter was to make it a bubble plot, but the bubbles ended up covering the entire plot so I decided to put that plan aside * For my life still cannot figure out how to create dataframes dynamically within a loop |

# QUESTION 2: ACHIEVEMENTS - SELF-REFLECTION FOR CA2

Provide a brief reflection of what you think you have personally achieved in this assignment. Do not be shy to highlight extra efforts you have put in for this assignment. This is the place for you to impress upon your lecturer on your competency in this area and ‘fight’ for extra bonus points! 😊

|  |
| --- |
| * Created a method to loop the reading of all market share datasets, and concatenating them into one dataframe * Shortened my code by creating a function to add dataframes to mongoDB. The function allowed me to do it in a single line * Attempted to create a pie chart * Managed to make my plots more pleasing to the eye * As a focus, I decided to work on my analysis. I think I managed to do a better analysis this time round, compared to CA1. This time round I had a hypothesis that my datasets related to and a proper flow to my analysis, ending off with a proper conclusion with suggestions. * Managed to display values of bars within bar chart * I also hosted my Jupyter Notebook on myBinder, which will allow people to view my codes and plots. However, they are not able to run my code. |

# QUESTION 3: EVALUATE YOUR CA2

What grade do you think you deserve for CA2? Justify the grade you gave yourself with reasons.

|  |
| --- |
| **What grade do you think you deserve for CA2?**  A  **State your justifications here**  For this CA, I focused on increasing my analytical skills. I realized my mistake in CA1, which was a lack of hypothesis to justify why I used the datasets I used.  Therefore, this time round, I focused on ensuring that I had a proper hypothesis and analysed my datasets accordingly. I properly observed the data that I had and came up with a flow to how I want to analyse them. I feel that there was a great improvement in my analysis of data.  With regards to plotting, I used matplotlib again as I felt it was sufficient in assisting me in my data analysis. However, I tried to up to looks of my plots by adding annotations and using more pleasing colours as well etc. This allowed to gain a better understanding of matplotlib, instead of trying another plotting library. I chose not to use interactive elements as I felt they would not help me better understand my data.  Overall, I think the effort and improvements I made this time round, warrant me an A because I have met the assignment requirements well. |

# QUESTION 4: FUTURE PLANS

How do you rate your programming competency with data analysis tasks after completing this assignment? Give yourself a rating from 0 to 10. Would you be confident to take on a job that requires you to write a Python program to analyse real-world data? What are your future plans in terms of sustaining the knowledge you have gained from this module?

|  |
| --- |
| 6.5/10. There is still much for me to learn. I have yet to fully understand Python as well as matplotlib. However, I think I will venture into other plotting libraries.  As of now, I would be apprehensive to do that, however, I would like to challenge myself and give it a shot.  With regards to the future, I will hopefully be taking the AI and Machine Learning module next semester, so the knowledge I have accumulated in this module will be of use. Furthermore, if time permits, I will try to learn AI and Machine Learning first through online courses. Additionally, I am interested in challenging myself by trying to use online data science projects that I found and complete them with the use of Dash. |

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