**SCHOOL OF COMPUTING (SOC)**

**IT8701 Introduction to Programming for Data Science**

**Self Reflection (CA2)**

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

|  |  |
| --- | --- |
| **Your Lecturer’s Name** | MR. HUBERTUS |
| **Your Name** | WONG QI YUAN, JEFFREY |
| **Your Student ID** | P7359567 |
| **Your Class** | PA-01 |

# QUESTION 1: RATE THE EFFORTS AND COMPETENCY THAT IS DEMONSTRATED IN THIS ASSIGNMENT

Tick in the column that best describes the efforts, technical competency and depth of data analysis that is demonstrated in this assignment.

Justify your rating in the second and third questions below

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **WAY Above Average** | **Above Average** | **Average** | **Below Average** | **Way Below Average** |
| Coding |  | √ |  |  |  |
| Analysis |  | √ |  |  |  |

# QUESTION 2: JUSTIFICATION FOR RATING GIVEN FOR CODING

Please provide evidence that you have met the requirements (AVERAGE) or if you think your submission is above average or even above average, state details of what you have done here so that your lecturer does not miss out the efforts you have put in for this assignment. For CA2, the basic requirements are to produce 4 different graphs with at least 3 datasets from data.gov.sg, and write working code to store/retrieve at least 1 dataset to a relational or NoSQL database using Pandas.

|  |
| --- |
| * Use Matplotlib and Seaborn to produce great visualization in this CA2 * Written code to store and retrieve on more than one dataset to a relational SQL database using Pandas * Written code on more than 3 datasets from data.gov.sg in this CA2 with World Bank Data.   You may refer to my submission folder for evidence. |

# QUESTION 3: JUSTIFICATION FOR RATING GIVEN FOR DATA ANALYSIS

Please provide evidence that you have met the requirements (AVERAGE) or if you think your submission is above average or even above average, state details of what you have done here so that your lecturer does not miss out the efforts you have put in for this assignment.

|  |
| --- |
| Cleansing, transforming and modeling data were done with the goal of discovering useful information and derived with informing conclusions in this analysis. Please kindly refer to my submission folder for justification. |

# QUESTION 4: YOUR 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.

|  |
| --- |
| I believe there are still more room on improvement for both data coding and analysis, so I give myself a rating of 7.5. |

After finishing the PDC1 of your Specialist Diploma, which do you think you prefer or is stronger at? The Statistics or Programming portion? How has this realisation affected your mindset of a Data Science job? Do you enjoy a Data Science role that mainly involves application of lots of statistical concepts (improving predictive algorithms for instance) or one that requires a lot of programming (e.g. code to acquire or clean data) or perhaps both equally excite you? 😊

|  |
| --- |
| I believe I am still at best in Statistics portion. In PDC1, the concepts I received as part of the statistics has taught me how to think logically, and this is very important in my work in data management. I think patience was another very important skill I learned. I love what I do, and sometimes I take for granted that others have the same statistical training that I do because I am so entrenched in it. I gained a better sense of the challenges and difficulties when interpreting statistics and how to better communicate results and ideas to the people.  TO be honest, the most interesting part of the Data statistics module was seeing how many fields can grow and transform based on insights from data and statistics. In my case, I found it most interesting seeing how in integrates and interacts with computer science. Every time someone surfs the internet through Facebook, enters a Google search, or looks at an item on Amazon, data about what you are doing are collected and algorithms process this data to enrich the experience by saying, recommending books and offering specials deals on Amazon, recommending friends and showing relevant stories on Facebook, and the most groundbreaking of all; returning relevant search results. Although I loved Math, I sometimes had trouble connecting the theory to the application, and statistics is such an applied field. I look at it as a rite of passage and once I saw enough theory relevant to my interests, the learning process became easier. |

Are there any useful skills that you gained from this module? Share how you think the skills you learnt from this module can be applied in your current job or in a future career / job change.

|  |
| --- |
| It is ironic, but it is the more basic concepts of statistics and probability that I use everyday rather than the complicated models I learned. Concepts such as independence, confidence, power and accuracy, etc, are important building blocks for building my own models, or for choosing an existing one from those I learned in this module. I always start with some exploratory analysis such as computing some statistics and making plots that show relationships clearly via Python software. Then I set explicit guidelines for the input and output of the model I want to build and note any critical assumptions that are violated or that must be met. I then try several different methods and models and validate their results using common metrics before settling on a final model configuration. |

What was not taught in this module, but you wish to learn? How do you plan to learn these missing skills?

|  |
| --- |
| It was great to be able to learn some intermediate Python library in this module. However, it was unfortunate that not all Python library were taught in this module such as SciPy, Bokeh, Seaborn and Hypothesis Testing with Python. I am planning to learn all these missing skills during my free time on online platform such as Coursera or Udemy or even sign-up from other data science course. |

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