SCHOOL OF COMPUTING (SOC)

CA2 Specification

DIPLOMA IN APPLIED AI & ANALYTICS

ST1510
Programming for Data Analytics

2024/2025 Semester 2

Assignment Rubrics

1. Demonstrate basic competency in writing Python programs.
2. Demonstrate basic competency in using Numpy, Pandas and Statsmodel packages for data analysis and data visualization.
3. Demonstrate basic competency in applying the insights gained from the outputs of your Python programs to deliver a useful data analysis presentation.

Section 1
Instructions and Guidelines

1. This is a **group** assignment requiring you to write Python code to retrieve data from files and perform basic data manipulation operations, including cleansing, transformation, and visualization.

2. **You will form pairs.** In classes with an odd number of students, there will be at most one group of three.

3. The deadline for this assignment is **10 February 2024 (Monday) at 8am.**

4. Submit your assignment via the BrightSpace CA2 Assignment Submission link by the stated deadline.

5. The deliverable should be a zip file named using the following convention: **"PDASCA2_YourClass-YourStudentID-YourName.zip" e.g. "PDASCA2_1B04-2388888-StevenLee.zip"** The Zip file should include the following items:
   - One or more Jupyter Notebook (.ipynb) files that accomplishes the given tasks using the Python.
   - A set of PowerPoint slides that summarize the data insights that you have gained through your Python code.
   - All datasets used in your Jupyter Notebook files.
   - One Declaration of Academic Integrity.

6. A compulsory presentation/interview will be conducted. During the session, you must present your work using the submitted PowerPoint slides and the Jupyter Notebook. Your module tutor will ask questions related to the submission and ask you to reproduce certain parts of your code during the session.

7. This assignment will account for **40%** of the **module grade**.

8. 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 module tutor as soon as reasonably possible. Students are not to assume on their own that their deadline has been extended.

9. No marks will be awarded, if the work is copied or you have allowed/enabled others to copy your work. Plagiarism is a serious offence, and if you are found to have committed, aided, and/or abetted the offence of plagiarism, disciplinary action will be taken against you.

Warning: Plagiarism means passing off as one's own the ideas, works, writings, etc., which belong to another person. In accordance with this definition, you are committing plagiarism if you copy the work of another person and turning it in as your own, even if you would have the permission of that person.

Section 2
Assignment Requirements

Background

Here are some key factors to consider when choosing a job in Singapore:

1. **Company Culture**: Ensure it aligns with your values and offers a positive work environment.
2. **Job Role**: Match your skills and interests with the job responsibilities.
3. **Salary and Benefits**: Check if the compensation package meets your financial needs.
4. **Work-Life Balance**: Look for flexible hours and good leave policies.
5. **Location**: Consider the job's location and your daily commute.
6. **Growth Opportunities**: Seek companies that offer career development and advancement.
7. **Job Stability**: Research the company's reputation and stability.
8. **Meaningfulness**: Choose a job that aligns with your values and lets you see the impact of your work.
9. **Job Satisfaction**: Look for a role that offers a positive work environment and fulfilling tasks.

These factors provide a solid starting point, but they're not exhaustive. Feel free to consider these and other factors to shape your research question. You do not need to use all these factors.

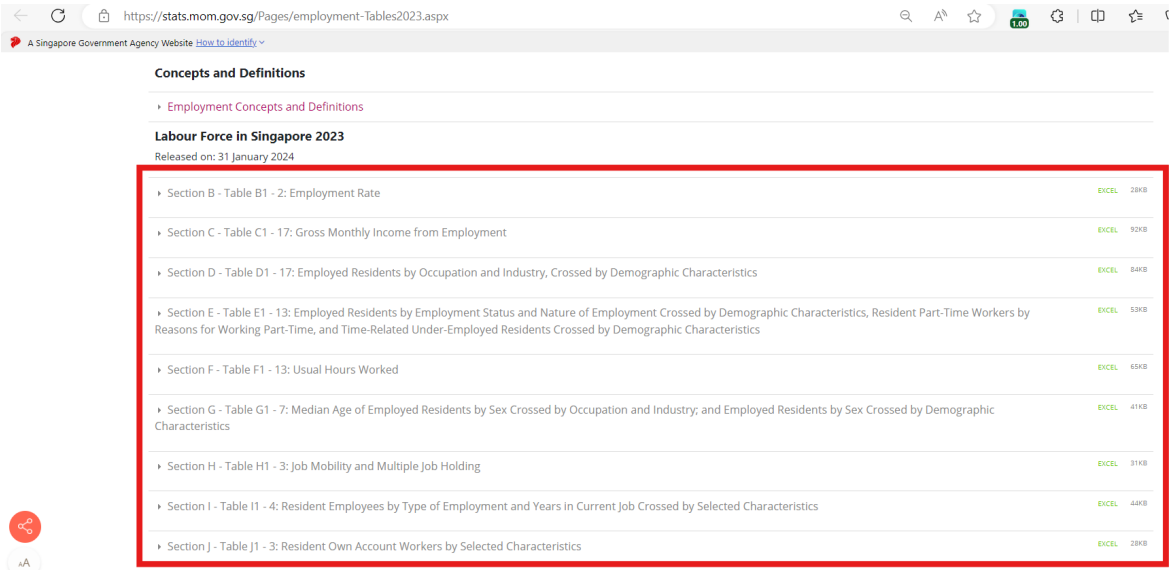
Requirements:

1. You must use **at least three** datasets related to employment. **Document the URLs of all datasets used** in both the Jupyter Notebook and PowerPoint slides. Clearly define your research question at the beginning of the Jupyter notebook and presentation slides. You are encouraged to select interrelated datasets that align with a research question. Address this research question using Pandas, Matplotlib, NumPy, and Statsmodels.

2. One of the datasets **must** be the "graduate employment survey modified.csv" file available on the Brightspace CA2 tab. The metadata is provided on the last page of this document.

3. One of the datasets **must** be from: https://stats.mom.gov.sg/Pages/employment-Tables2023.aspx

   Below is a screenshot illustrating how to access the documents from the link above. By clicking on any of these rows, you can download the Excel files and import the data into Python.



4. The final dataset(s) may be obtained online; just make sure to document the dataset's URL.

5. Select datasets that allows you to write Python code for data analysis and visualization, enabling you to address the criteria outlined in Section 3: Marking Scheme (detailed in the following pages). Focus on fulfilling the Marking Scheme in order to do well for this assignment.

6. Submit at least 5 visualizations but keep it within a maximum of 9.

7. Compile your findings into a deck of PowerPoint slides. The PowerPoint slides should include the following sections and to be presented within the stipulated time limit of the Lecturer. Changes of time limit may be done by your lecturer based on the situation.

    - Prepare your code for running. Your Lecturer may ask you to run some code.
    - A slide that lists your name and the research question of your data analysis.
    - A slide that lists the URLs of all the datasets you have used.
    - For each dataset, use at most three slide to explain the **process** you went through to analyse that dataset. Where possible, you should specifically mention **how you used NumPy, Pandas, Visualizations (Matplotlib, Seaborn, etc) and Statsmodels** to achieve certain outcome, for example, to identify and handling of missing values and outliners.
    - Maximum three sides on the **insights** obtained from analysing the datasets should use to answer the research question. You should support your analysis with visualizations libraries.
    - A slide detailing the responsibilities and tasks completed by each group member.

Section 3
Marking Scheme

Marks will be awarded based on the following rubrics:

Component	Description	Weightage
Use of Statsmodels	Apply linear regression, multiple linear regression, data distribution, QQ plot, and boxplot to address the research questionUse Markdown, comments, or slides to explain the rationale and methodology for answering the research question. Ensure that you primarily utilize tutorial and practical content, resorting to external materials only after thoroughly applying the tutorial and practical content.	20%
Use of NumPy	Apply NumPy Arrays, including but not limited to creating, subsetting, slicing, indexing, sorting, array manipulation, and array operations. Use Markdown, comments, or slides to explain the rationale and methodology behind addressing the research question. Ensure that you primarily utilize tutorial and practical content, resorting to external materials only after thoroughly applying the tutorial and practical content.	20%
Use of Visualization	Create at least 5 charts to interpret findings of the charts and to answer the research question. You may use other libraries other than Matplotlib.	10%
Use of Pandas	Utilize Pandas for tasks such as: Retrieving and Inspecting Data, Selecting Data, Reshaping Data (e.g., pivot, melt), Handling Missing Data, Groupby, Apply, or Lambda functions. Use Markdown, comments, or slides to explain the rationale and methodology behind addressing the research question. Ensure that you primarily utilize tutorial and practical content, resorting to external materials only after thoroughly applying the tutorial and practical content.	10%
Code and Notebook Quality	Code Notebooks should be documented with comments and markdown to explain the code. The research question is clear and addressed regularly.	10%
Presentation	The presentation should be well-rehearsed, keeping to the time limit and able to show evidence of work through code samples and graphics.	10%
Question and Answer	Answering all questions confidently and correctly.	10%
General Performance	Submit all your lab work for Practical 5 and 6. Participate in class to answer questions.	10%

Meta Data for "graduate employment survey modified.csv"

Title	Column name	Data type	Unit of measure	Description
Year	year	Year (YYYY)	-	-
University	university	Text	-	-
School	school	Text	-	-
Degree	degree	Text	-	-
Overall Employment Rate (%)	employment_rate_overall	Text	-	Overall employment rate refers to the number of graduates working in full-time permanent, part-time, temporary or freelance basis, as a proportion of graduates in the labour force (i.e. those who were working, or not working but actively looking and available for work) approximately 6 months after completing their final examinations.
Full-Time Permanent Employment Rate (%)	employment_rate_ft_perm	Text	-	Full-time permanent employment rate refers to the number of graduates in employment of at least 35 hours a week and where the employment is not temporary (including contracts of one year or more), as a proportion of graduates in the labour force (i.e. those who were working, or not working but actively looking and available for work) approximately 6 months after completing their final examinations.
Basic Monthly Salary - Mean (S$)	basic_monthly_mean	Text	-	Basic monthly salary pertains only to full-time permanently employed graduates. It comprises basic pay before deduction of the employee's CPF contributions and personal income tax. Employer's CPF contributions, bonuses, stock options, overtime payments, commissions, fixed allowances, other regular cash payments, lump sum payments, and payments-in-kind are excluded.
Basic Monthly Salary - Median (S$)	basic_monthly_median	Text	-	Basic monthly salary pertains only to full-time permanently employed graduates. It comprises basic pay before deduction of the employee's CPF contributions and personal income tax. Employer's CPF contributions, bonuses, stock options, overtime payments, commissions, fixed allowances, other regular cash payments, lump sum payments, and payments-in-kind are excluded.
Gross Monthly Salary - Mean (S$)	gross_monthly_mean	Text	-	Gross monthly salary pertains only to full-time permanently employed graduates. It comprises basic salary, overtime payments, commissions, fixed allowances, and other regular cash payments, before deductions of the employee's CPF contributions and personal income tax. Employer's CPF contributions, bonuses, stock options, lump sum payments, and payments-in-kind are excluded.
Gross Monthly Salary - Median (S$)	gross_monthly_median	Text	-	Gross monthly salary pertains only to full-time permanently employed graduates. It comprises basic salary, overtime payments, commissions, fixed allowances, and other regular cash payments, before deductions of the employee's CPF contributions and personal income tax. Employer's CPF contributions, bonuses, stock options, lump sum payments, and payments-in-kind are excluded.
Gross Monthly Salary - 25th Percentile (S$)	gross_mthly_25_percentile	Text	-	Gross monthly salary pertains only to full-time permanently employed graduates. It comprises basic salary, overtime payments, commissions, fixed allowances, and other regular cash payments, before deductions of the employee's CPF contributions and personal income tax. Employer's CPF contributions, bonuses, stock options, lump sum payments, and payments-in-kind are excluded.
Gross Monthly Salary - 75th Percentile (S$)	gross_mthly_75_percentile	Text	-	Gross monthly salary pertains only to full-time permanently employed graduates. It comprises basic salary, overtime payments, commissions, fixed allowances, and other regular cash payments, before deductions of the employee's CPF contributions and personal income tax. Employer's CPF contributions, bonuses, stock options, lump sum payments, and payments-in-kind are excluded.

~~ End of Assignment Specifications ~~