​​**Division / Team: Government Digital Services, CC4.0**

As Singaporeans continue to face an uncertain future, the government has been quick to act to ensure that the country is future ready. One of the most important initiatives is the Smart Nation Programme, which was launched in November 2014.

The programme aims to harness Singapore's infocomm and technology resources to create a more connected and smarter society.

One of the key initiatives of the Smart Nation Programme is Digital Economy and helps to create new jobs and opportunities for Singaporeans.

The Career Coach 4.0 (CC4.0) product is a collaboration between WSG and GovTech which aims to help Singaporeans improve their employability and gain access to job and training opportunities.

The platform also assists coaches to help applicants make informed job decisions and enhance their career journey.

Our role is to build and improve career matching services and ingest data from various sources so that we can find out what both employers and job seekers need and tailor those needs in a unique way.

**Assessment Title:** Take-home Assignment for Data Engineer Intern

**Overview/Background:** This take-home assignment is used to demonstrate your ability to design, build, collect and process data according to a specific set of requirements.

**Assessment Instructions / Case Study Scenarios:**

Steven is a travel blogger that intends to create a travel food series. He is looking at data from Zomato for inspiration. He wants to find restaurants that have good user ratings and interesting past events.

Steven obtained the following data:

· [List of restaurants](https://raw.githubusercontent.com/Papagoat/brain-assessment/main/restaurant_data.json)

· [Country Code (in excel format)](https://github.com/Papagoat/brain-assessment/blob/main/Country-Code.xlsx?raw=true)

1. Extract the following fields and store the data as **restaurants.csv**.

◦ Restaurant Id

◦ Restaurant Name

◦ Country

◦ City

◦ User Rating Votes

◦ User Aggregate Rating (in float)

◦ Cuisines

2. Extract the list of restaurants that have past event in the month of **April 2019** and store the data as **restaurant\_events.csv**.

◦ Event Id

◦ Restaurant Id

◦ Restaurant Name

◦ Photo URL

◦ Event Title

◦ Event Start Date

◦ Event End Date

*Note: Populate empty values with "NA".*

3. From the dataset (**restaurant\_data.json**), determine the threshold for the different rating text based on aggregate rating. Return aggregates for the following ratings only:

◦ Excellent

◦ Very Good

◦ Good

◦ Average

◦ Poor

**How to submit your answers:**

* Your code must be hosted on GitHub, or any other similar service, in a publicly accessible repository (e.g., GitHub / Bitbucket / GitLab).
* You may include a section with the assumptions, interpretations you have made about the requirements above or notes on your architecture decisions.
* Do show the progress of your work with atomic git commits.
* You may choose to host your solution on a server, container, or serverless framework.
* During your submission you may provide the URL where your app is hosted, and link to Google Video or YouTube of the video screen capture, but do not share the URL publicly.
* Please send us a link to your repository when you complete your assessment to Xue Ling ([xue\_ling\_ang\_from.tp@tech.gov.sg](mailto:xue_ling_ang_from.tp@tech.gov.sg)). Do title the subject of your email as CC4 Data Engineer Internship Tech Test Submission - Your Name (Eg: CC4 Data Engineer Internship Tech Test Submission – Ang Xue Ling).

**Deadline:**

1. Please complete the assessment **within 5 calendar days** upon receiving this email. Thereafter, you might be invited for a chat with our tech assessors to discuss your submission.

**Grading Criteria / Matrix:**

Besides source code submission, we are looking out for documentation as markdown file with:

1. Instructions on how to run your source code locally on our laptop in the README file.

2. A short summary on how you would design/deploy this using cloud services and any decisions and considerations you made when designing the solution. Max 500 words.

3. Architecture diagram describing the infrastructure components.

**Additional Notes to Candidates:**

· Preferred language: Python

· Try to limit Time and Space Complexity to O(n^2) or better

· Inclusion of unit or end-to-end tests is a plus

· For questions relating to the assessment, please contact terence\_yap@tech.gov.sg