Take home test - Data science/ analytics/ engineering

Hi there! If you have made it this far, I would first like to congratulate you! We have a very rigorous selection process here in the AI/Machine Learning team.

The following take home challenge is meant to gauge your technical and analytical ability. Don't worry if you are unable to complete them all (we do not expect you to, this is a diversified challenge designed to find out which roles fit you best). That being said, we do encourage you to at least attempt them all. Each section can be answered relatively independently from one another, and so, if you find yourself stuck with one section, try to answer all the other sections to the best of your ability.

Please be prepared to provide a clear, step-by-step explanation to your answers. We are also trying to see how well you present your ideas and logic.

Do feel free to use all resources available to you (this is an open book challenge!). If you think you can provide a much better set of answers given extra time, please contact us, and do reach out to us if you have any questions or clarifications.

If you feel like you have better suggestions to expand this challenge, we are open to feedback, and to changing it to better suit you. Good luck, and most importantly, have fun coding!

Challenge 1: Data Cleaning, Transformations and ETL pipeline architecture

Using the datasets given in the link:

Take_home_test - Google Drive

1. By using the datasets provided, clean the data and find the correlation between both datasets.
2. After cleaning and linking:-

- Explain 5 reasons, why did you join the datasets in that way?
- Provide a business use case from the datasets provided, that would likely bring benefits to a business.
- Design a cloud/local-machine based pipeline architecture encompassing the business use case provided by you above. Architecture must take into consideration scalability, performance and simplicity. Data processing tools are up to your discretion. (Google cloud platform preferred)
- How do you validate your analytics?

Challenge 2 : Customer engagement

Based on the dataset and analytics derived from the dataset above:

- How would you present to a customer regarding the insights derived from their dataset, in your own words?
- Come up with a few proposal's, justify your answer

Challenge 3: Scrapping , Datasourcing and Enriching data

Enrich the dataset:

- What is data enrichment, in your own words.
- Come up with 1 new use-cases with the core dataset as above by enriching with public data
- Please included your scrapping or datasourcing methods in Github

Submission Instructions

1. Write down the answers in a document. Share the document in pdf format with these emails:

- arthurftl@sunway.com.my
- noellethw@sunway.com.my
- henrylmr@sunway.com.my
- khairulhka@sunway.com.my

CC:

- tingky@sunway.com.my
- elaynehel@sunway.com.my

when you are done.

1. Create a GitHub account if you don't have one already.
2. Create a GitHub public repository for this test and paste the link in the take home test email.
3. Upload the answers to your GitHub public repository by doing Git Push.
4. A short (less than 5 slides) presentation deck to explain your work would be great (but not strictly neccesary)

Some resources for video footages:

https://www.pexels.com/search/videos/people%20walking/