DATA SCIENCE CASE STUDY INTERVIEW

Data Strategy and Analytics Division

INSTRUCTIONS FOR COMPLETING THE CASE STUDY

This case study is divided into two parts:
1. Part 1 aims to assess your exploratory data analysis and data visualisation competency.
2. Part 2 aims to assess your ability to develop an analytical use case in depth. This includes scoping an analytics project from a broadly defined problem, applying relevant data science techniques to derive meaningful insights, and communicating your findings in a coherent and impactful narrative to stakeholders with appropriate visuals.

The entire case interview is to be completed within 24 hours from the time you receive it. You are advised to allocate more time to Part 2 than Part 1.

Please submit the following within the allocated time:
1. A deck of slides for each part, preferably in *.pptx or *.pdf form.
2. The complete code used in your analysis, such that we should be able to run the code ourselves with appropriate languages/tools and reach conclusions comparable to what you include in your deck.

Instructions for interview

You will meet with the human resource team, business user team, and current members of the data team for around an hour, where you will be invited to deliver your findings for both parts. You can make reasonable assumptions for these case studies wherever necessary.

You will be assessed on the quality of your delivery, as well as the validity of your answers.

Part 1: Answer the Question

1. As the data scientist from the data planning division, you are tasked to discover insights and trends in data science from the Kaggle data science survey from 2017 to 2021 and present them during the next management meeting. Discover 5-8 insights and trends in data science during the dataset period and present them with good narratives and visualisations for the general audience to understand.

   Dataset:
   https://www.kaggle.com/datasets/andradaolteanu/kaggle-data-science-survey-20172021

Part 2: Choose **ONE** Case to answer from 2 or 3

2. Management is interested in the skills trends in data science from the United States to take reference for Singapore's skills funding policy.

   Datasets available:
   1. Dictionary of standardised skills from SkillsFuture (Link Here)
   2. Singapore Standard Occupational Classification (SSOC): https://www.singstat.gov.sg/standards/standards-and-classifications/ssoc
   3. Careerbuilder Job Listing 2020 https://www.kaggle.com/datasets/promptcloud/careerbuilder-job-listing-2020

   As the data science from the skills insight team, you are tasked to find out the jobs and skills insights and trends from the given data sets and draft actionable proposals for the management for Q3 2020. (Imagine the current date is around the end of June 2020)

3. As the only data scientist of the newly formed risk management division, you are tasked to develop a fraud detection model to flag potentially fraudulent transactions for early intervention. In addition to the model development, you need also deliver the model implementation and monitoring plan for operations.

   Due to resource limitations, only 1000 manual investigations can be conducted monthly for ground truth establishment.

   Dataset:
   https://www.kaggle.com/datasets/dermisfit/fraud-transactions-dataset