Quantitative Strategy Case Study Interview

Section 1

Please pick ONE of the hypothetical scenarios in this section, and prepare a deck to brief the relevant hypothetical senior management about your findings.

Points to note in the hypothetical scenario:

1. The deck will be used for a face-to-face briefing with management.

2. The deck will be circulated to the management for reading beforehand.

3. The briefing is expected to take 30 minutes, excluding Q&A.

4. Management includes a mix of directors from technical as well as non-technical backgrounds.

5. Management is not available for further clarification. If you find the scenario to be coached in overly broad terms, you may make reasonable assumptions to narrow things down, but be prepared to explain / justify them.

6. Management is interested to hear what can be done to further improve the analysis, given more time / resources / etc., if the problem described in the scenario surfaces again in the future.

A list of suggested data sources is included below. You may use them (or any other publicly available data source) for your analysis.

Suggested data sources for Section 1

Data.gov.sg

Resale flat prices	https://data.gov.sg/dataset/resale-flat-prices
NEA licensed eating establishments	https://data.gov.sg/dataset/list-of-nea-licensed-eating-establishments-with-grades-demerit-points-and-suspension-history
National map line	https://data.gov.sg/dataset/national-map-line
Singapore Master Plan 2019	https://data.gov.sg/dataset/master-plan-2019-region-boundary-no-sea
	https://data.gov.sg/dataset/master-plan-2019-planning-area-boundary-no-sea
	https://data.gov.sg/dataset/master-plan-2019-subzone-boundary-no-sea

LTA DataMall

Bus stop locations / train station locations	https://www.mytransport.sg/content/mytransport/home/dataMall/static-data.html#Whole%20Island
Bus routes	https://www.mytransport.sg/content/mytransport/home/dataMall/dynamic-data.html#Public%20Transport

OneMap

Geocode API	https://docs.onemap.sg/#onemap-rest-apis

Parliament

Official reports – parliamentary debates (Hansard)	https://sprs.parl.gov.sg/search/home

Scenario 1

Some forum posters have complained that the value of their HDB flats suffer because they are near expressways, which are very noisy. Others say expressway proximity is good, due to the unblocked view (at least for higher floors).

The Housing and Development Board has tasked your team to analyse whether there is merit to either view, based on transaction prices for resale HDB flats in recent years.

~

Scenario 2

There is concern that the experience of the current public transport infrastructure varies significantly across HDB towns.

The Land Transport Authority has tasked your team to assess the current infrastructure, and review if there is merit to accelerate the construction of any specific phase of any MRT line that is due to open before 2030.

~

Scenario 3

The Singapore Food Agency is planning to conduct a 12-hour inspection operation on food establishments, using 5 teams of inspectors.

It has tasked your team to design an inspection schedule for each inspection team. Combined, the schedules should enable the inspectors to have a good geographical coverage of food establishments island-wide, with focus on those who have performed poorly in the past[1]. You may assume that each inspection takes 30 minutes on average, and that all food establishments are open for business during the inspection period.

(For the purpose of this question, you may limit your analysis to mainland food establishments that belong to one of the following two categories: 1) located within food centres, or 2) include postcodes in their addresses. I.e. excluding temporary canteens at construction sites, offshore hawker centre on Jurong Island, etc.)

~

Scenario 4

The National Archives of Singapore is preparing an exhibition about the evolution of parliamentary debate in Singapore and wants to do a feature on Committee of Supply (COS) debates[2] in recent years.

It has tasked your team to summarise common topics from COS debates that took place within the last Parliament (13th), as well as identify MPs who asked questions frequently about each topic.

(Sample codes written in R / Python are provided in another document to demonstrate the process of scraping text from the Parliament website. You may use them—with any necessary modification for your use case—or write your own code in your preferred language for the task.)

End of Section 1

---

[1] Based on inspection records from Jun 2015 to Sep 2016 available here: https://data.gov.sg/dataset/list-of-nea-licensed-eating-establishments-with-grades-demerit-points-and-suspension-history.

[2] A brief explanation on what the COS debates are can be found here: https://www.singaporebudget.gov.sg/budget_2019/about-budget/budget-features/how-singapore-s-finances-are-managed

Section 2

Please answer ALL the questions in this section.

Question 1 (Association)

Download the list of transactions by each property agent from https://data.gov.sg/dataset/cea-salesperson-residential-transaction-record, and focus on HDB resale flat transactions where a property agent represented the seller.

Task 1: Examine the distribution for number of sales closed by an agent in a year & suggest a probability distribution that may be suitable for modelling this set of values. What are some ways in which your suggested distribution is appropriate? What are some of its limitations?

~

Question 2 (Classification)

Download the Wireless@SG hotspots file from https://data.gov.sg/dataset/wireless-hotspots (in either KML or geoJSON format), and extract the data associated with it. You should obtain a table with over 1600 rows and several columns, where each row corresponding to a different WiFi hotspot in Singapore.

Task 1: From the table, what are some of the information you can deduce for each hotspot?

Task 2: Due to a system error, the location type column for the last 200 rows of the dataset has become garbled. Using all earlier rows as well as all other columns in the dataset, build a classification model to predict the location type for these hotspots. You may treat the three rarest location types as one category.
(*Note: you may wish to create some additional features based on available ones*.)

Task 3: The information has now been recovered from a backup copy of the file. Compared to the true location types, how good was your model? Be prepared to explain the metrics you use to evaluate your model.

~

Question 3 (Data Visualisation)

A colleague is working with a salary dataset based on recent poly graduates in a specific course of study highly subsidised by the government, to compare whether the career choices made by students from Group X are different from those from Group Y in any manner. She has already produced the following summary table and listed out the main insight she wishes to highlight, as well as pertinent observations on the dataset's characteristics, but is struggling to come up with a good way to communicate the insight to her audience though visualisations while also accurately reflecting the dataset's characteristics.

Summary table:

Job Nature	Industry	Student Group X Median Salary	Student Group X Count	Student Group Y Median Salary	Student Group Y Count
Closely related to course of study	A	3150	83	3000	23
	B	3300	53	3100	9
	C	2650	47	2600	32
	D	2400	12	2400	15
Somewhat related to course of study	E	4100	30	3900	3
	F	3400	23	3150	7
	G	2800	12	2600	22
	H	2300	8	2200	11
Unrelated to course of study	Others	2900	21	1900	28

Main insight:

It may be worth reviewing the policy behind subsidising this course of study, as a considerable proportion of students from each group do not go on to work in industries closely related to it. For Group X, this may be partially due to higher / comparable salaries offered by other industries. For Group Y, non-salary factors may play a more prominent role.

Data characteristics:

1. There are considerably more students from Group X than Group Y in this course of study.
2. Proportionately more students from Group Y are in jobs unrelated to their course of study.
3. The distribution of students among various industries is considerably different between the two student groups.
4. Students from Group X tend to command higher salaries, for the same type of job & industry.
5. The salary differential between the two student groups differs by job nature and industry.

Task: Help your colleague present the insight in an intuitive manner that is easily understood by a non-technical audience, and that reflects as many characteristics in the list as possible. Be prepared to justify any and every aspect of your visualisation (e.g. chart choice, colour palette, labels, orientation, etc.).