



# ARTIFICIAL INTELLIGENCE PRACTICE

## **Case Study Interview**

### **Instructions for completing the case study**

*This case study is divided into two sections:*

1. *Section 1 aims to assess your technical understanding in a variety of topics. Please answer **ALL** the questions in this section.*
2. *Section 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. Please pick **ONE** of the hypothetical scenarios in this section.*

*A list of suggested data sources for each section is included in the annex. You may use them (or any other publicly available data source) for each question.*

### **Submission Requirements**

*The entire case is to be completed within 36 hours from the time you receive it. You are advised to allocate more time to Section 2 than Section 1.*

*Create a GitHub repository containing:*

- *Your resume and/or CV.*
- *Your presentation slides in PDF format.*
- *Complete, executable code used in your analysis, with a readme file in each codes folder explaining how to use it.*

*using the following project structure*

```
section-1-question-1/
└── codes/
└── slides/
section-1-question-2/
└── codes/
└── slides/
section-2-case-study/
└── codes/
└── slides/
resume/
```

*Share your repository link with these GitHub users:*

- *victorong-github*
- *lishuoleo*
- *linzi-sg*
- *shaunkhoo*
- *johnsohh*

***Instructions for interview***

*You will meet two or more current members of the AI Practice Team for around an hour, where you will be invited to deliver your findings for Section 2 as though they were scenario-specific stakeholders. In addition, you may be asked questions relating to your work for both sections.*

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

---

---

## *Section 1*

*Please answer ALL the questions in this section.*

---

### **Question 1**

The HDB Resale Portal was officially launched in Jan 2018 to streamline transaction of HDB resale flats. This has made it easier for buyers and/or sellers to carry out transactions on their own, without utilising a property agent's professional services.

Using relevant data from 2017 onwards, quantify the business impact on agents. Present your findings in the form of a data story suitable for general public consumption (e.g.: a series of annotated charts and/or tables, a blog post-style write-up with accompanying visualizations, etc.).

~

### **Question 2**

There is growing concern over the volatility and sustained high prices of Certificate of Entitlement (COE) in Singapore. The COE quota, which determines the number of new vehicles allowed on Singapore's roads, is reviewed and adjusted by the Land Transport Authority (LTA) every three months. These quarterly quota announcements significantly influence COE bidding behaviour and subsequent prices across all vehicle categories.

The Land Transport Authority (LTA) needs to better understand the relationship between COE quotas and prices to make more informed policy decisions. Your team has been tasked to develop a model to predict COE prices for vehicle categories A & B. Additionally, the model should quantify the price elasticity of COE quotas by determining how incremental changes in quota (e.g., adding X number of certificates) would impact prices across the two categories.

---

*End of Section 1*

---

---

## Section 2

*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. *You are allocated 30 minutes for a face-to-face briefing, excluding Q&A.*
  2. *Your deck will be circulated to the management for reading beforehand.*
  3. *Management includes a mix of directors from technical as well as non-technical backgrounds.*
  4. *Management is not available for further clarification before the presentation. If you find the scenario coached in overly broad terms, you may make reasonable assumptions to narrow things down, but be prepared to explain / justify them.*
  5. *Management will be willing (within reason) to clarify during / after the presentation. You may include a list of questions you'd like to ask them at the end, if you so desire.*
  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.*
- 

### **Scenario 1**

There is concern over mismatch of demand and supply for preschools (see [The Big Read: The preschool conundrum — why shortage of places persists in some estates despite millions invested](#)).

The Early Childhood Development Agency (ECDA) has asked your team to forecast the subzone-level demand for preschool services, in particular those providing childcare (18 months to 6 years) programmes, over the next 5 years, to determine where they should prioritize building / relocating preschools. As a planning norm, you may assume each preschool can accommodate up to 100 children.

In addition to the one-off forecast, ECDA staff wants a tool that they can use to make subsequent decisions on a regular basis (e.g.: every time they receive updated population statistics, update newly opened / closed preschools, learn about upcoming BTO sites, etc.). While they do not expect your team to produce a full-fledged product within 2 days, they would like to have some idea of what the eventual tool could look like, as well as how it could be deployed.

~

### **Scenario 2**

In response to a [call for suggestions on ways to improve cycling experience in a neighbourhood](#), some netizens remarked that urban planners have designed Singapore's cycling routes<sup>1</sup> for two use cases (cycling for leisure within one's own neighbourhood, and cycling to the nearest transport node for transfer onto public transport) while neglecting a third use case: cycling *all* the way between home and work.

The Land Transport Authority (LTA) has asked your team to assess the validity of this statement and propose where the next tranche of cycling paths / cycling lanes should be built, to improve

---

<sup>1</sup> In this context, cycling routes include both park connectors and cycling paths.

connectivity for work commute. For simplicity, you may assume that it is possible to add these to all roads other than expressways, and up to 100km may be added in one tranche.

~

### **Scenario 3**

In 2018, the Ministry of Culture, Community & Youth (MCCY) announced a target to double Singapore's volunteerism rate from 35% in 2016 to 70% by 2024. While this is considered reachable, [especially with COVID-19 inspiring more people to step up to volunteer](#), more can be done to encourage volunteerism in all forms.

The National Council of Social Services (NCSS) is designing a new series of volunteer outreach campaigns to target people of different age groups and has asked your team to advise on 1) which age group(s) to prioritize; and 2) how to reach them effectively. The goal is to maximize long-term impact, and the available data is a recent survey on Singaporeans' moral foundations.

NCSS is also open to consider running an additional survey to collect more first-hand data, based on your preliminary findings. In your recommendations, please include a section on what to cover in the new survey, such as the specific profile(s) to target, possible questions to ask, etc.<sup>2</sup>

### **Scenario 4**

Design an LLM pipeline to generate AI-based analysis of the 2025 Geography examination paper, focusing on two key aspects: its alignment with syllabus objectives and the balance of topic weightage. Evaluate the AI outputs and assess how well these prompting techniques can be adapted to meet the specific requirements of different subjects.

You may use self-hosted language models, as we will not grade the quality of their outputs.

~

### **Annex**

---

<sup>2</sup> You *don't* have to cover the mechanics behind conducting a survey, such as the minimum sample size required for a representative sample, the appropriate number and style of questions for different medium (e.g.: face-to-face, phone, online).

Should any links not work, please visit data.gov.sg and search for the dataset directly. You are also encouraged to find other external source of data when required.

#### **Suggested data sources for Section 1**

[CEA Salespersons' Property Transaction Records \(residential\)](#)

[Resale Flat Prices](#)

[COE Quota](#)

[COE Price](#)

#### **Suggested data sources for Section 2**

[Singapore Residents by Planning Area / Subzone, Single Year of Age and Sex, June 2000-2020](#)

[Births and Fertility, Annual](#)

[Listing of Centres](#)

[Listing of BTO projects with subzone information](#)

[Master Plan 2019 Subzone Boundary \(No Sea\)](#)

[Master Plan 2019 SDCP Park Connector Line Layer](#)

[Master Plan 2019 SDCP Cycling Path Layer](#)

[Master Plan 2019 Road layer](#)

[Data on Moral Foundations Theory](#)

[2025 O-Level Geography examination paper](#)

[Upper Secondary Geography syllabus](#)