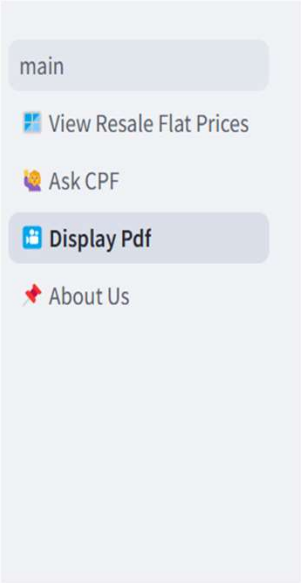
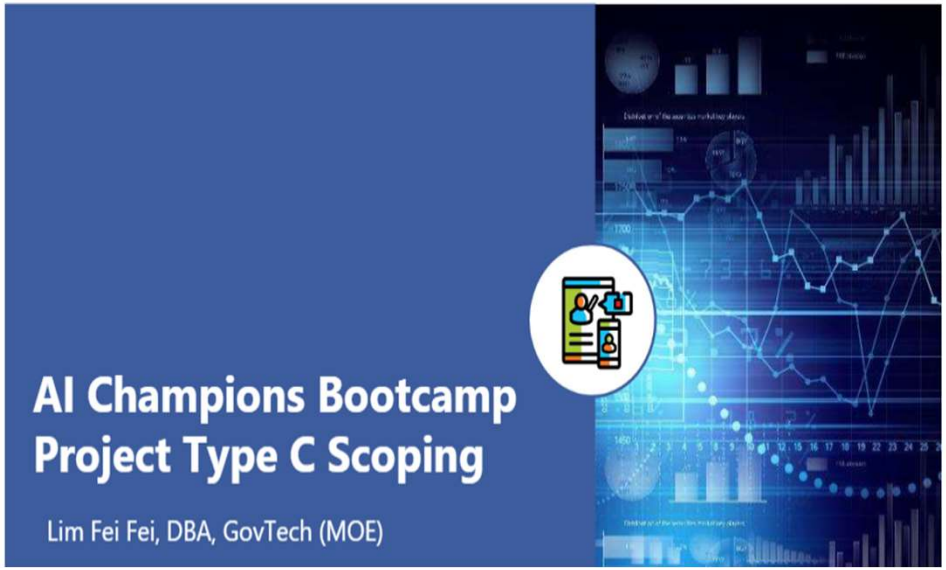
AI Champions Bootcamp Project Type C Scoping

Lim Fei Fei, DBA, GovTech (MOE)

Background





Project URL
https://project-typec-feifei.streamlit.app/

Display Pdf

Upon clicking the sidebar menu item > Display pdf, you will see this deck in pdf format.

This deck provide details on the 2 use cases:

1. Use Case 1 – Query on Resale flat getting text answers or data visualization answers.

2. Use Case 2 – Query on how to buy resale flat with details from CPF website.

Use Case 1 –
Query on HDB Resale Flat Price

Problem Statement: The Problem

1. Describe the problem you are trying to solve

A couple intends to purchase a resale HDB flat to accommodate their living needs, considering factors like affordability, location, and space.

2. At the end, sharpen it into a question

How can we help couples who face challenges in accurately finding the right resale price and the remaining lease of an HDB flat, which are critical factors in their purchasing decision of a resale HDB flat.

Problem Statement: Buying Resale Flats

1. What is wrong with the current situation?

(e.g. How does this problem affect your agency or involved individuals?)

The couple need to make informed decision on getting the right price of resale flats based on their financial position, they do not where to get the right information to make the right assessments.

2. What is the magnitude of this problem?

(e.g. number officers/citizens are affected, financial costs incurred, or productivity hours affected)

Resale flats often have a wide range of prices, influenced by the neighborhood, condition of the property, and market demand. This variability can make it difficult to determine a fair price.

Proposed Solution

1. How would you try to solve this problem?

Researching the available HDB datasets in data.gov.sg with Resale flat price and remaining lease.

2. How would you think Large Language Model(s) can be used to support your solution?

LLMs can be trained to analyze vast amounts of property transaction data, both historical and real-time, to generate price estimations for resale HDB flats based on various factors such as location, size and type of flats and resale price.

We have learned to deploy streamlit projects, therefore we can build a web-based solution that combines the power of LLMs for data analysis and conversational guidance with Streamlit for an interactive and user-friendly interface – using python programming language.

3. Are there any alternative solutions you have considered?

HDB Helpline or HDB mobile apps as the alternatives solutions.

Impact (part 1 of 2)

1. How would things be improved after the solution is implemented?

Users will have an easy-to-use interface for querying the resale flat dataset, making data accessible to non-technical users who would otherwise struggle with raw CSV files or SQL databases. They can search by criteria such as town, flat type, storey range, remaining lease, and month to get precise results quickly. The users can also request the result in data visualisation format (for e.g. bar chart)

2. How often the solution will be used?

It depends on the target audience and their needs. Potential home buyers could use the solution regularly, especially when they are in the market for purchasing a resale flat. Users are likely to query the platform multiple times during their research process to compare prices by location, flat type, or size, and to monitor trends over time.

The frequency could range from several times per week to daily during the home-buying journey, especially when market prices fluctuate or new data is released.

Impact (part 2 of 2)

1. What is the number of potential users can be benefited by the solution in long-run?

The primary audience for this tool is prospective HDB resale flat buyers. Assuming an average of **2-3 people** per transaction (including family members), this gives an estimate of **50,000 to 90,000 individuals** who may directly benefit from using the tool annually.

2. What is the estimated time-saving or financial-saving per year (from the agency's or WOG perspective)?

Time saving - an agency could spend an estimated **5,000 to 10,000 hours annually** on manual data analysis and respond to public queries.

Manual Data Compilation: Without such a tool, government agencies like HDB and URA typically rely on manual processes, such as compiling and analyzing datasets, preparing market reports, and responding to public inquiries. Automating these tasks can save a substantial amount of time.

Stakeholders & Users

1. Is this project's impact big enough to get the attention of the senior management?

As of now, this is a POC and we would get users to test to obtain their feedback on this tool.

2. Who are the users of the LLM applications?

We would submit this POC for trial testing and get users who are interested in this tool such as home buyers and real estate agents.

Available Data and Samples

1. What are relevant data do you currently collect and already have?

Using the resale flat prices dataset from data.gov.sg from HDB from Jan 2017 to Oct 2024

https://data.gov.sg/datasets?query=resale+flats+prices&page=1&resultId=d_8b84c4ee58e3cfc0ece0d773c8ca6abc

2. Is there any data that require approval to be used in the project?

No, these datasets are available at data.gov.sg

3. Please provide a data dictionary and a sample set of the data*.

Please see next slide.

Data Dictionary & Sample Data

Data dictionary

Field name	Data type	Description
month	Month (YYYY-MM)	Month of sale
town	Text	Designated residential area with its own amenities, infrastructure, and community facilities
flat_type	Text	Classification of units by room size. They range from 2 to 5 rooms, 3Gen units, and Executive units.
block	Text	Classification of units by room size. They range from 2 to 5 rooms, 3Gen units, and Executive units.
street_name	Text	A HDB building comprising multiple flats or apartments
storey_range	Text	Estimated range of floors the unit sold was located on
floor_area_sqm	Numeric	Total interior space within the unit, measured in square meters
flat_model	Text	Classification of units by generation of which the flat was made, ranging from New Generation, DBSS, Improved, Apartment
lease_commence_date	Year (YYYY)	Starting point of a lease agreement, marking the beginning of the lease term during which the tenant has the right to use and occupy the leased property
remaining_lease	Text	Remaining amount of time left on the lease.
resale_price	Numeric	Cost of the flat sold

Sample data

	month	town	flat_type	block	street_name	storey_range
589	2017-01	JURONG WES	4 ROOM	681C	JURONG WEST CTRL	10 TO 12
590	2017-01	JURONG WES	4 ROOM	987C	JURONG WEST ST 93	16 TO 18
591	2017-01	JURONG WES	4 ROOM	337A	TAH CHING RD	19 TO 21
592	2017-01	JURONG WES	5 ROOM	934	JURONG WEST ST 91	07 TO 09

	ge	floor_area_sqm	flat_model	lease_commence_date	remaining_lease	resale_price
589		96	Premium Apartme	2,000	82 years 06 months	445,000
590		93	Premium Apartme	2,006	88 years 04 months	446,000
591		92	Model A	2,010	92 years 03 months	470,000
592		122	Improved	1,988	70 years 09 months	350,000

***please share a mockup (dummy) data samples**

Data Classification

What is the Classification/Sensitivity of Data?

Official open. Reference from HDB datasets - https://data.gov.sg/datasets?query=resale+flats+prices&page=1&resultId=d_8b84c4ee58e3cfc0ece0d773c8ca6abc

About this dataset

Contact

lau_pei_wen@hdb.gov.sg

Created on

28 Jul 2021

Licence

Free forever for personal or commercial use, under the Open Data Licence.

Agency

HDB (Housing and Development Board)

Lines Of Inquiry

What are some of your hypothesis, or assumptions which you would like to be validated?

Hypothesis: Users such as home buyers, sellers, real estate agents who prefer to use self-service tools to access HDB resale flat data rather than contacting agencies or consultants.

Assumption to Validate: Users are tech-savvy enough to navigate and use a web application that allows them to query data without needing extensive guidance or customer support.

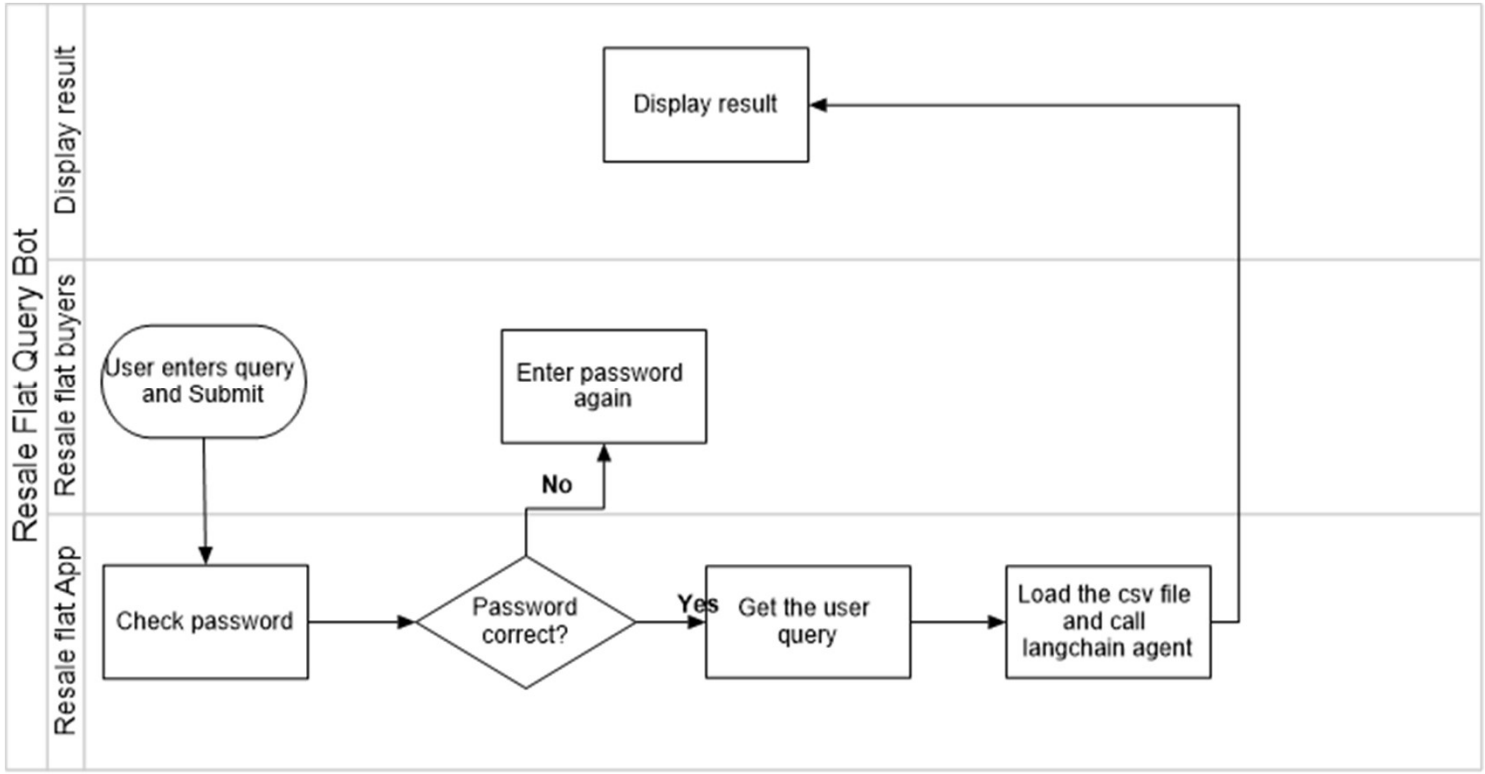
Lines Of Inquiry

How will you test your hypotheses or validate your assumptions?

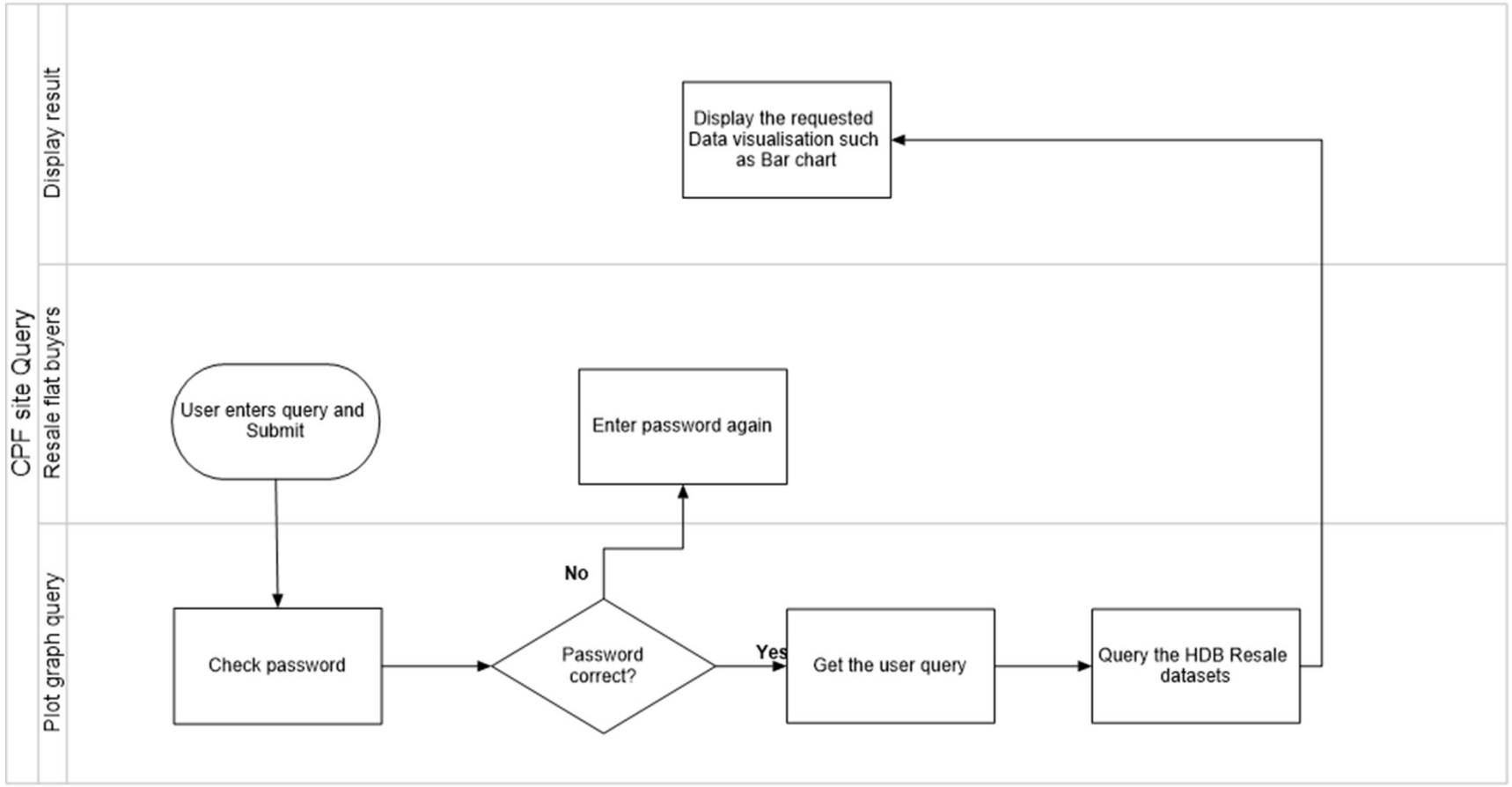
Assumptions related to resale flat buyers' self-service preferences, tech-savviness, desire for deeper insights, and privacy concerns which they can self-help.

We can validate these assumptions through surveys for selected home buyers and real estate agents.

Methodology (Flow Chart) – main



Methodology (Flow Chart) – Plot graph



Methodology (data source, API) – main

a. We use the dataset from HDB (reference slide 9)
b. We have below programs.

Main.py	customer_query_resale.py	utility.py	Llm.py
1. Call check password in utility.py to ensure user enter the correct password	process_user_messag function a. Load the hdb csv dataset b. Get the OpenAPI key from llm.py c. Call the langchain API with the OpenAPI key- create_csv_agent to process the question d. Return the result	check_password function – checking if user has entered the correct password as secrets.toml	1. Call load_dotenv from dotenv 2. Load the API key by checking it is in local development environment or Streamlit
2. If password is correct, call the function in customer_query_resale.py by passing in the user questions after user click "Submit" button.			

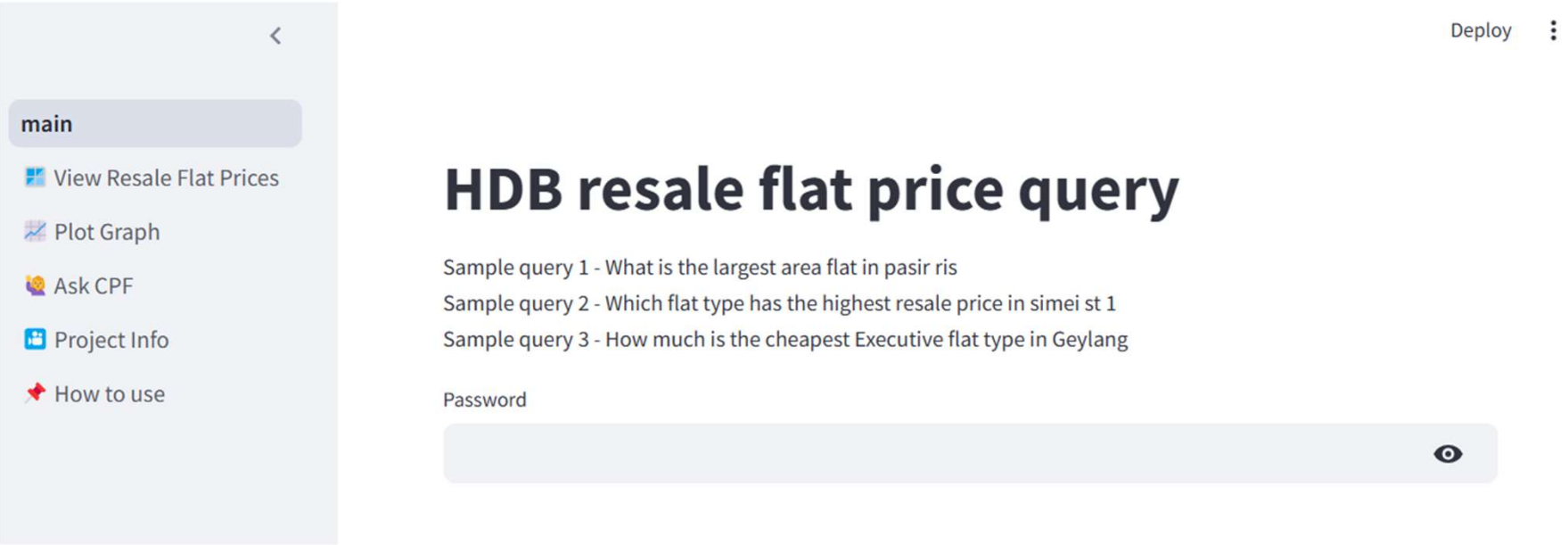
Methodology (data source, API) – Plot graph

a. We use the dataset from HDB (reference slide 9)
b. We have below programs.

Main.py	customer_plot_graph.py	utility.py	Llm.py
1. Call check password in utility.py to ensure user enter the correct password	process_user_messag function a. Load the hdb csv dataset b. Get the OpenAPI key from llm.py c. Call the langchain API with the OpenAPI key- create_pandas_dataframe_agent to process the question d. Return the result	check_password function – checking if user has entered the correct password as secrets.toml	1. Call load_dotenv from dotenv 2. Load the API key by checking it is in local development environment or Streamlit
2. If password is correct, call the function in customer_query_resale.py by passing in the user questions after user click "Submit" button.			

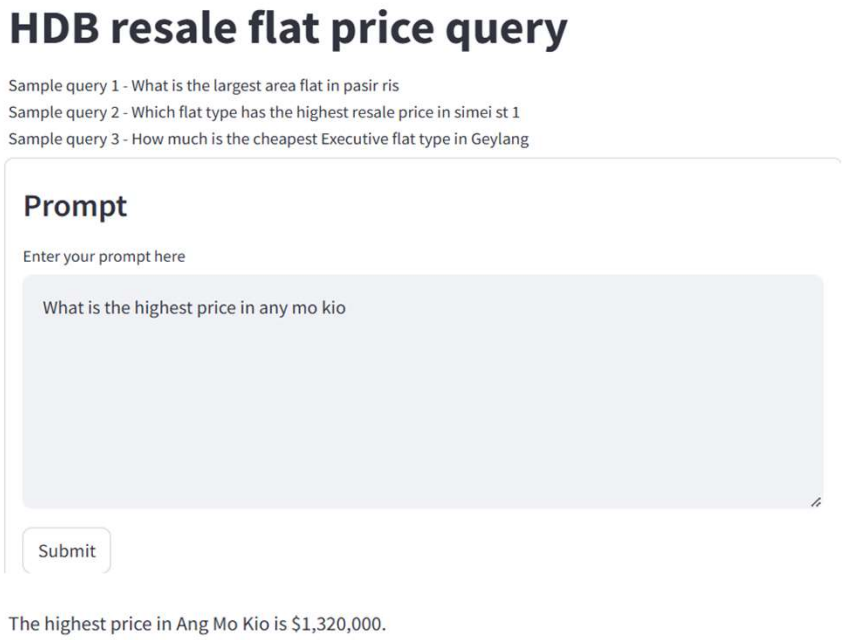
Screen flow – 1 - main

1. User enters the password



Screen flow – 2 - main

2. User enters the correct password and submits the question

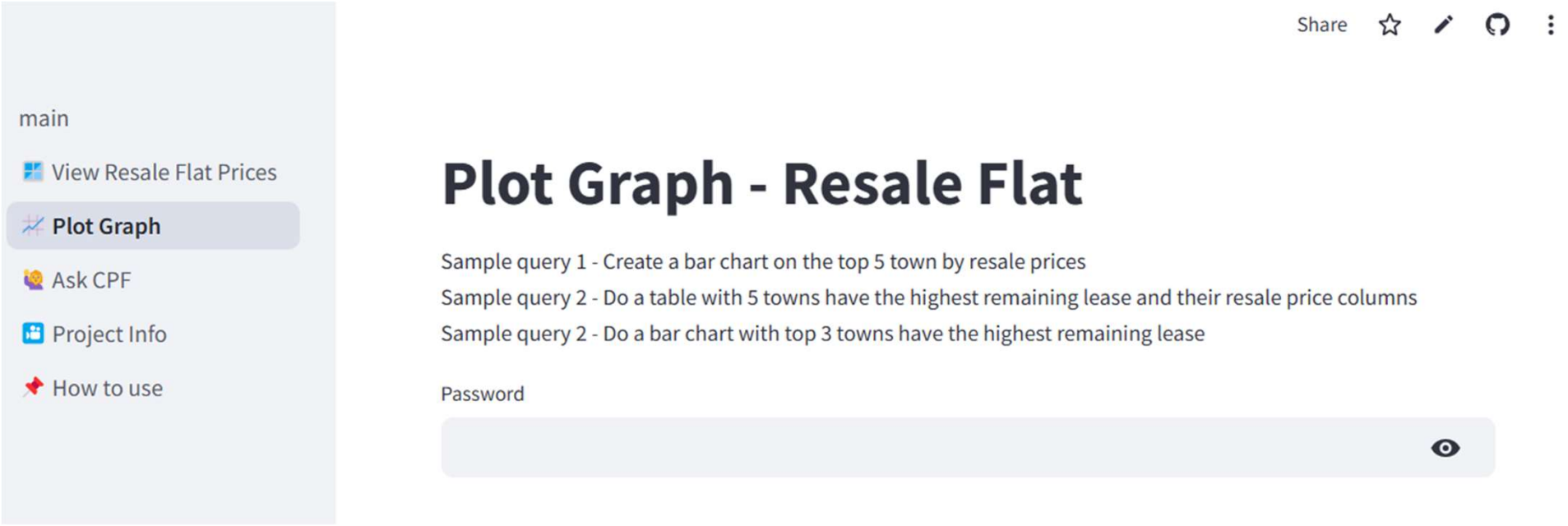


Questions such as:

- What is the highest price in any mo kio
- What is the largest area flat in pasir ris
- which flat type has the highest resale price in the ang mo kio
  Which flat type has the highest resale price in simei st 1
- What is the highest resale price for 4 room in simei st 1
- What is the highest resale price

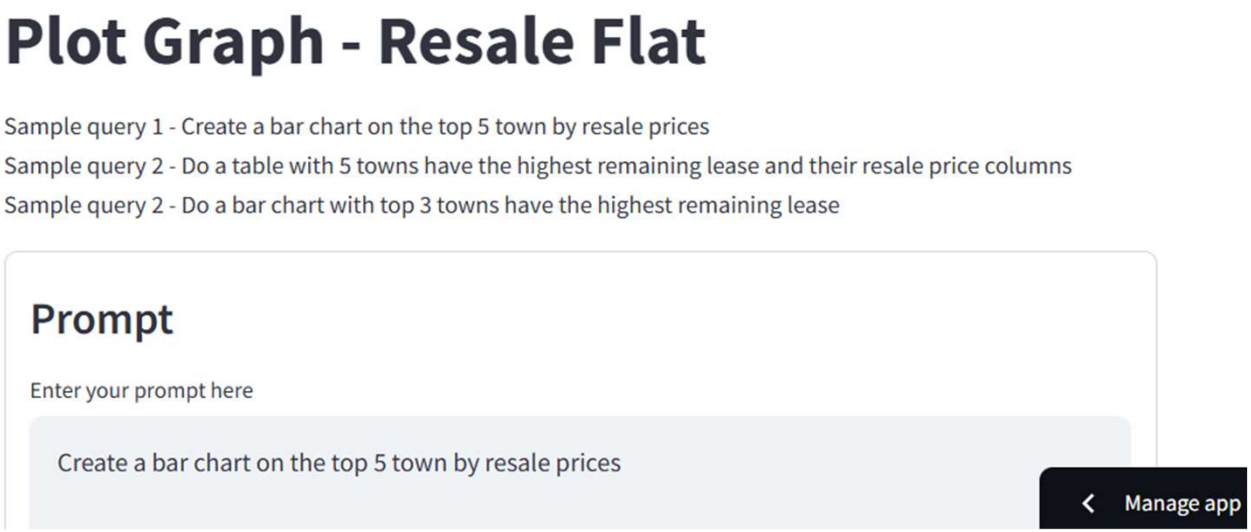
Screen flow – 3 – Plot graph

2. User enters the correct password and submits the question



Screen flow – 4 – Plot graph

2. User enters the correct password and submits the question

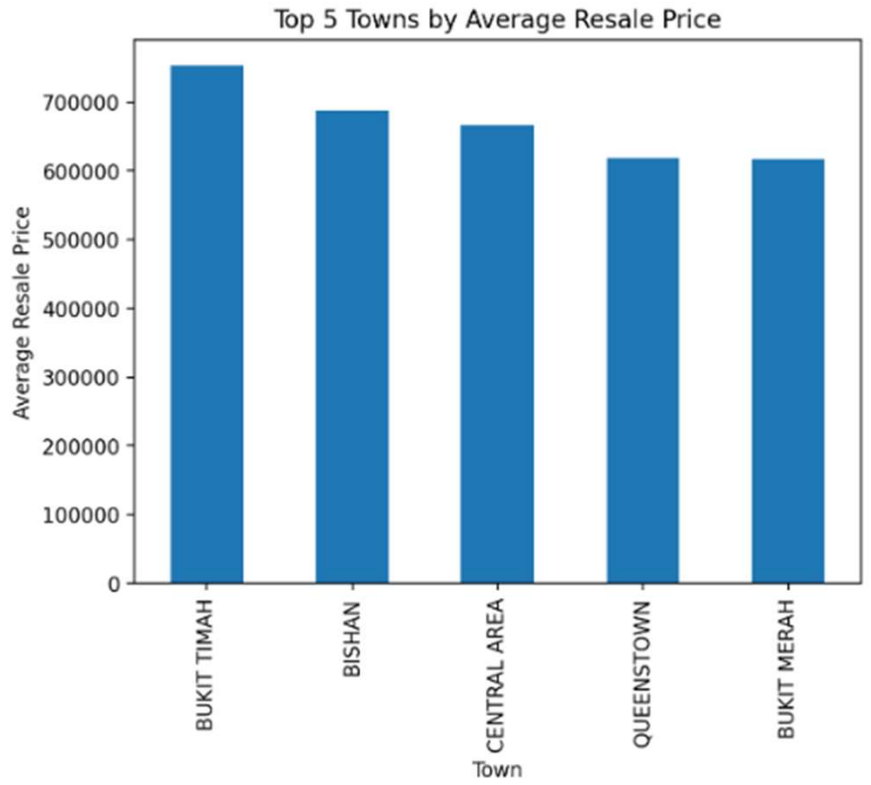


Questions such as:

- Create a bar chart on the top 5 town by resale prices
- Do a table with 5 towns have the highest remaining lease and their resale price columns
- Do a bar chart with top 3 towns have the highest remaining lease

Screen flow – 5 – Plot graph

3. The graph displays below the prompt



Obstacles (Applies for Use Case 1 and 2)

1. Do you foresee obstacles in implementing the solutions?

To get the updated and refreshed resale flat datasets and the CPF website information needs to be updated if there is a change in policies.

2. What resources (financial, human, technical) will be needed to overcome these obstacles?

The app needs to be promoted so that potential resale flat buyers are aware of its existence. We might need some marketing funds to promote this app.

Use Case 2 –
Guide on Buying First Time Resale Flat

Problem Statement: The Problem

1. Describe the problem you are trying to solve

The problem being addressed is guiding first-time buyers through the process of purchasing a resale flat in Singapore.

2. At the end, sharpen it into a question.

How can first-time buyers in Singapore effectively navigate the complex process of purchasing a resale flat, ensuring they manage financial planning, loan options, and the legal steps involved?

Problem Statement: How to buy Resale Flats

1. What is wrong with the current situation?

(e.g. How does this problem affect your agency or involved individuals?)

The current issue is that many first-time resale flat buyers in Singapore face confusion due to the complexity of the process, including financial planning, loan options, eligibility criteria, and the various steps involved in buying a resale flat.

2. What is the magnitude of this problem?

(e.g. number officers/citizens are affected, financial costs incurred, or productivity hours affected)

Buyers often face challenges in understanding financial eligibility, CPF savings usage, available grants, loan options, and legal procedures.

Proposed Solution

1. How would you try to solve this problem?

We search through CPF website to get the right guide for the buyers.
https://www.cpf.gov.sg/member/infohub/educational-resources/guide-for-first-time-resale-flat-buyers

2. How would you think Large Language Model(s) can be used to support your solution?

LLMs can act as a virtual assistant, answering detailed questions about the resale process, legal requirements, and timeline.

3. Are there any alternative solutions you have considered?

Government mobile apps or CPF helpline to address the public queries.

Impact (part 1 of 2)

1. How would things be improved after the solution is implemented?

First-time resale flat buyers will experience greater clarity and confidence throughout the process. They will be able to access personalized, step-by-step guidance, reducing confusion about financial planning, grants, and legal steps.

2. How often the solution will be used?

The solution would likely be used frequently, particularly by first-time homebuyers, which constitutes a significant group in the resale flat market.

Impact (part 2 of 2)

1. What is the number of potential users can be benefited by the solution in long-run?

The primary audience for this tool is prospective HDB resale flat buyers. Refer to use case 1.

2. What is the estimated time-saving or financial-saving per year (from the agency's or WOG perspective)?

Refer to use case 1

Stakeholders & Users

1. Is this project's impact big enough to get the attention of the senior management?

As of now, this is a POC and we would get users to test to obtain their feedback on this tool.

2. Who are the users of the LLM applications?

We would submit this POC for trial testing and get users who are interested in this tool such as first time resale flat buyers.

Available Data and Samples

1. What are relevant data do you currently collect and already have?

We refer to this CPF website which is publicly available: https://www.cpf.gov.sg/member/infohub/educational-resources/guide-for-first-time-resale-flat-buyers

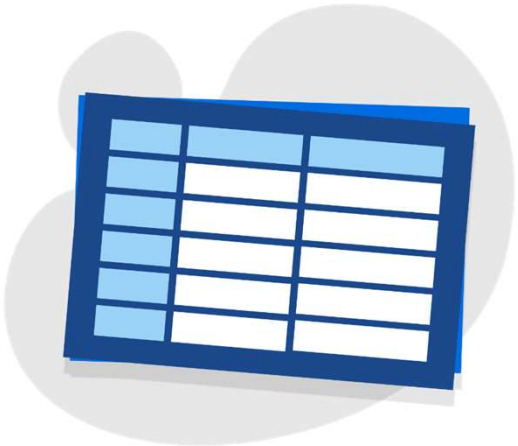
2. Is there any data that require approval to be used in the project?

No, this CPF site offers publicly available information.

3. Please provide a data dictionary and a sample set of the data*.

Not applicable as I am referencing a website for crawling on the information to provide the guide for the buyers.

Data Classification



What is the Classification/Sensitivity of Data?

Official open. This CPF website is publicly available: https://www.cpf.gov.sg/member/infohub/educational-resources/guide-for-first-time-resale-flat-buyers

Lines Of Inquiry

What are some of your hypothesis, or assumptions which you would like to be validated?

Hypothesis: First time resale flat buyers prefer to use self-service tools to for self help rather than contacting CPF or HDB.

Assumption to Validate: Users are tech-savvy enough to navigate and use a web application that allows them to query data without needing extensive guidance or customer support.

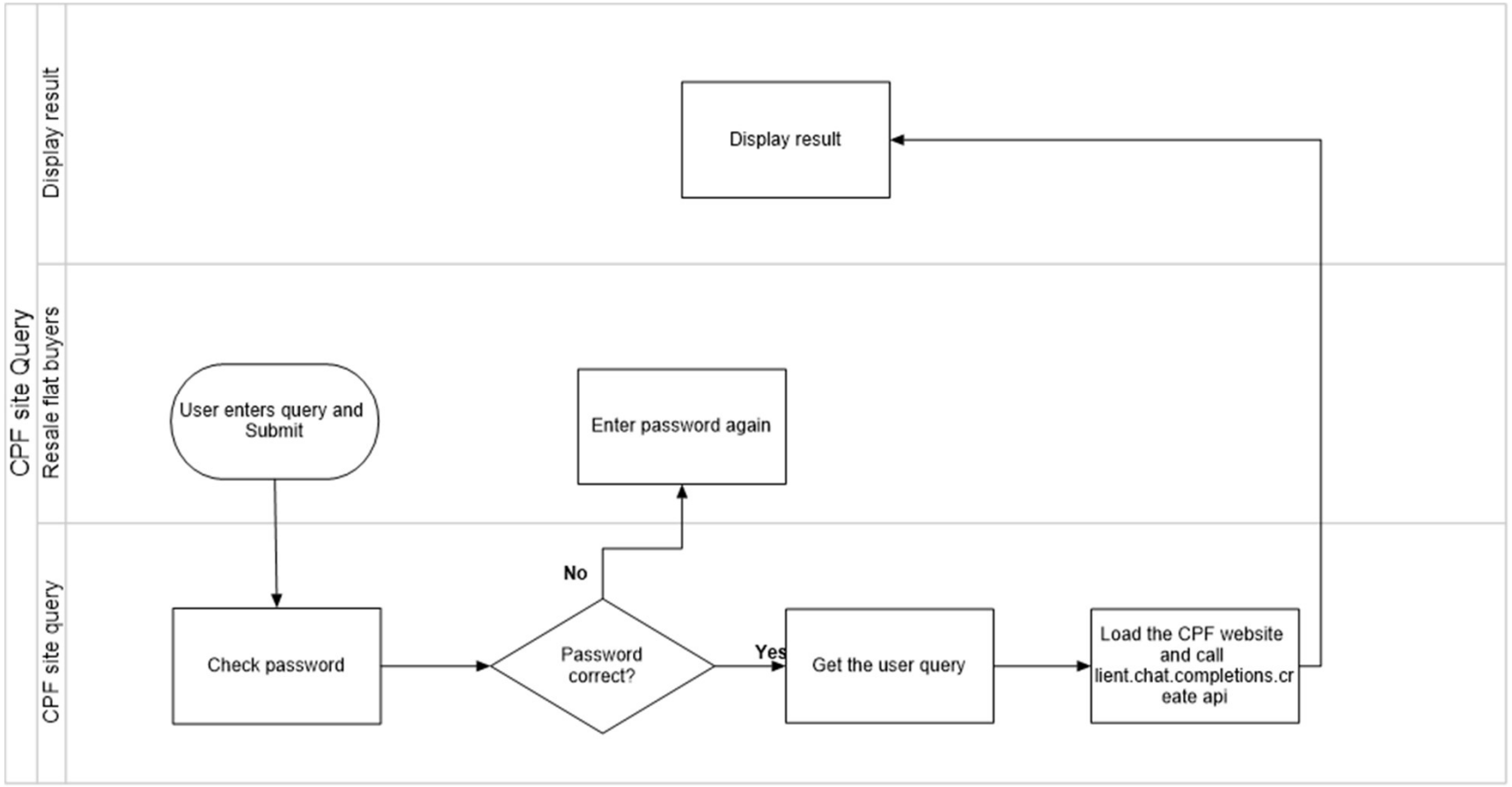
Lines Of Inquiry

How will you test your hypotheses or validate your assumptions?

Assumptions related to self-service preferences, tech-savviness, desire for deeper insights, and privacy concerns

Validate through conduct surveys

Methodology (Flow Chart)



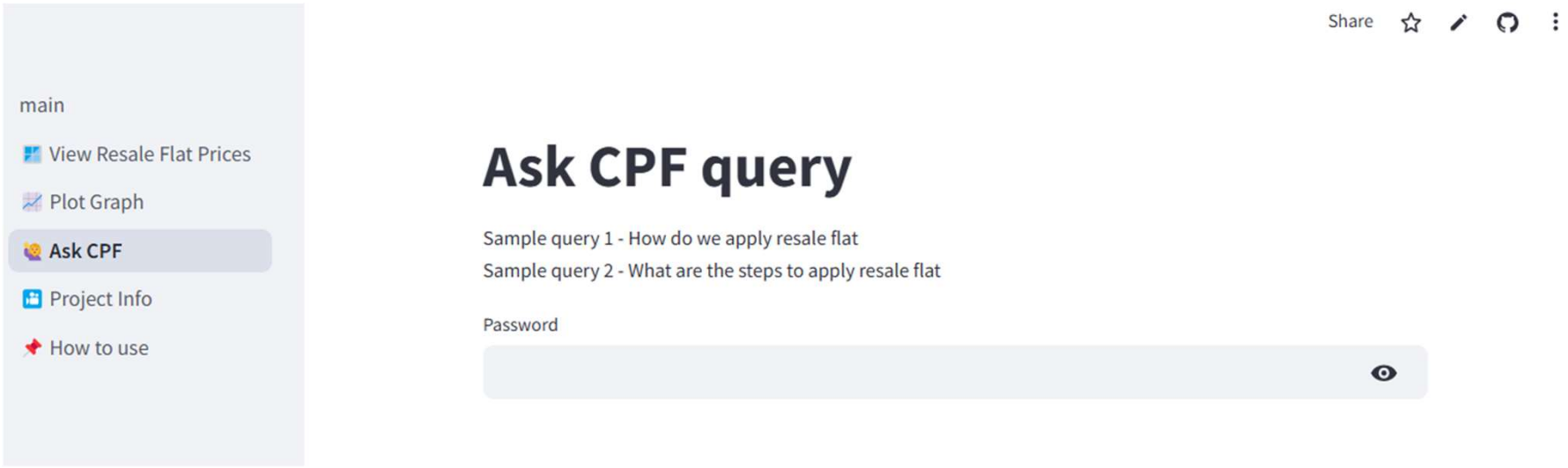
Methodology (data source, API)

a. We crawl the CPF website (reference slide 28)
b. We have below programs.

Ask_cpf.py	customer_query_cpf.py	utility.py	Llm.py
1. Call check password in utility.py to ensure user enter the correct password	process_user_messag function a. Load the Cpf url b. Get the OpenAPI key from llm.py c. Call the OpenAPI API with the OpenAPI key-client.chat.completions.create to process the question d. Return the result	check_password function – checking if user has entered the correct password as secrets.toml	1. Call load_dotenv from dotenv 2. Load the API key by checking it is in local development environment or Streamlit
2. If password is correct, call the function in customer_query_cpf.py by passing in the user questions after user click "Submit" button.			

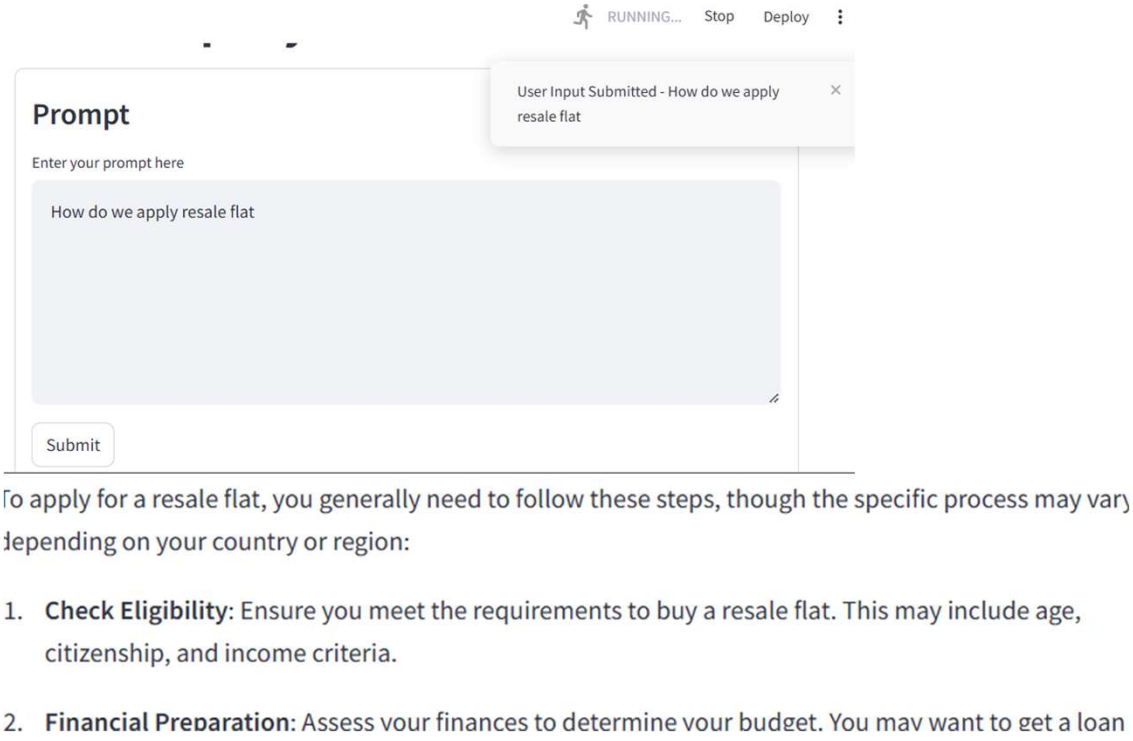
Screen flow - 1

1. User enters the password



Screen flow - 2

2. User enters the correct password and submits the question



Questions such as:

- How do we apply resale flat
- What are the steps to apply resale flat
- What do we need to know before applying resale flat