



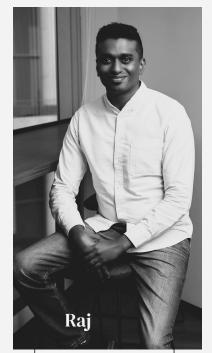
### Our company

Start-up launching an online platform that enables end-users to estimate HDB flats resale price online.

#### Our team









Yeow Wen

## Table of contents



04

Exploring Different Approaches 05

Data Modelling 06

Challenges

07

Recommendations & Conclusions

# 1 Problem Statement

Determining the **true**sale price of a HDB
resale flat is often
complicated due to the
confluence of external
and internal factors.



## Record \$1.5 million flat sold in **Tiong Bahru**

Michelle Ng Housing Correspondent ② Jul 06, 2023 03:16 pm







Prices of Housing Board resale flats were up 0.6 per cent in June, even as fewer units changed hands, which analysts said could be attributed to the school holiday seasonal lull and other factors.

An estimated 1.858 HDB resale flats were sold in June, down 17.8 per cent from the previous month and the lowest sales volume since February 2023. Compared with the same period a year ago, sales were down 13.1 per cent in the previous year, flash data released by real estate portals 99.co and SRX on Thursday showed.

June is typically a lull period for property transactions as it is the school holiday period and more families could have opted to travel averease this year as many countries no longer



## 'Getting out of hand': Netizens baulk at \$1.5m price tag for 5-room HDB flat in Queenstown



PHOTO: YouTube/PropertyLimBrothers

## '\$400K profit' — 5-room Sembawang HDB resale flat sold for \$745,000!



Photo: Google maps

#### **Problem**

In Singapore, HDB flat owners typically buy and sell property via agents.

Buyers and sellers are at the mercy of the final price quoted, before including the processing and commission fees.

Absence of pricing model can lead to uncertainty and unfair pricing.

#### Goal

Create a predictive model that estimates the resale price of HDB flats based on historical prices and property attributes.

This empowers users with a fair valuation alternative to prices quoted by property agents.

#### **Data Sources**

#### 1. Kaggle Dataset

For all the HDB resale prices and attributes

#### 2. Singstat

To get HDB Resale Price Index to account for inflation

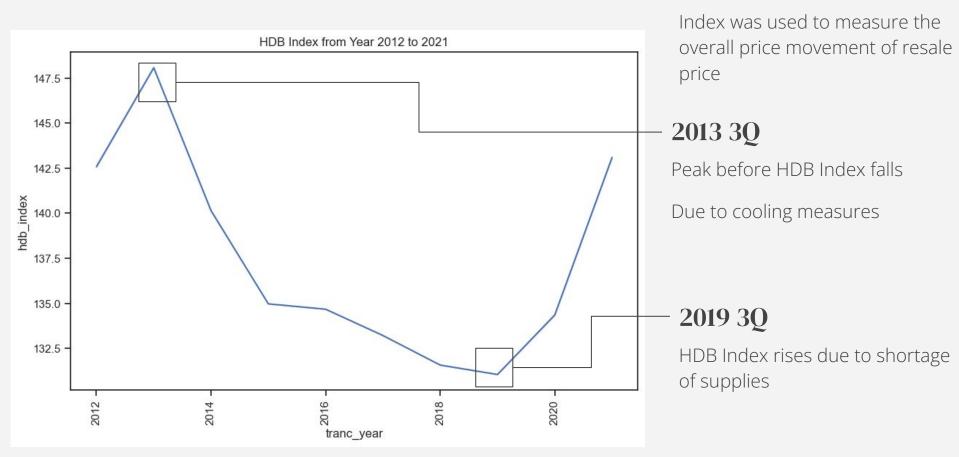
#### 3. Property Guru

Defining the CCR, RCR and OCR by the 28 Districts

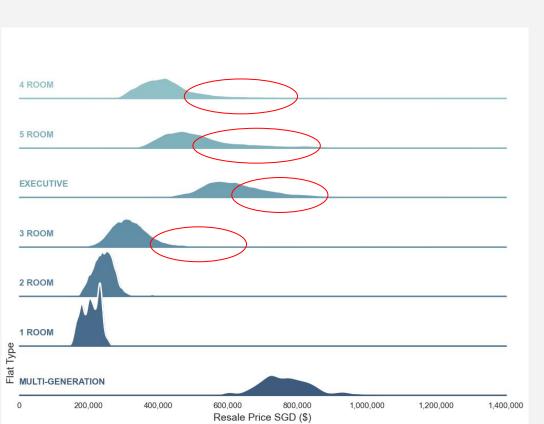




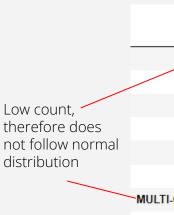
#### **HDB Price Index from 2012 1Q to 2021 2Q**



#### The higher the number of rooms, the higher the Resale Price. All flat types follow a normal distribution except 1 Room and Multi Gen due to low count

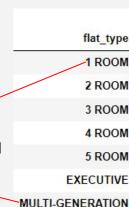






Low count

distribution



82
1896
39060

flat type

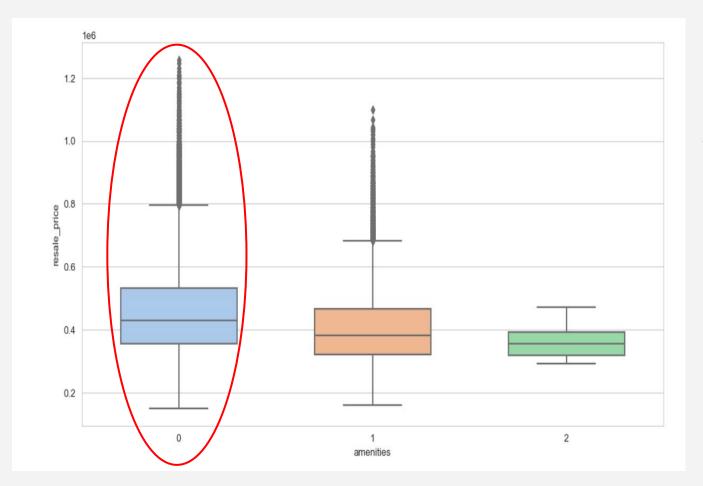
4 ROOM	61136
5 ROOM	36415

Count

56

5 ROOM	3641
EXECUTIVE	1198

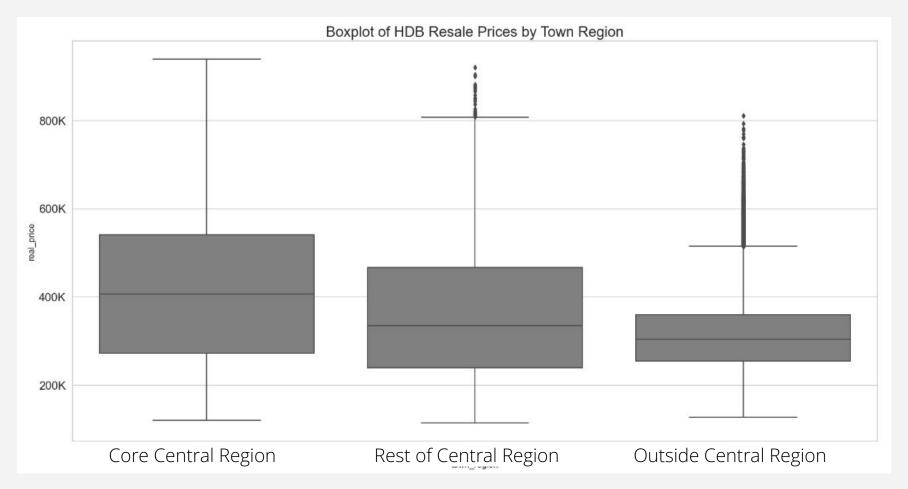
Contrary to popular belief, the lesser the no. of amenities within the same block, the higher the Resale Price.



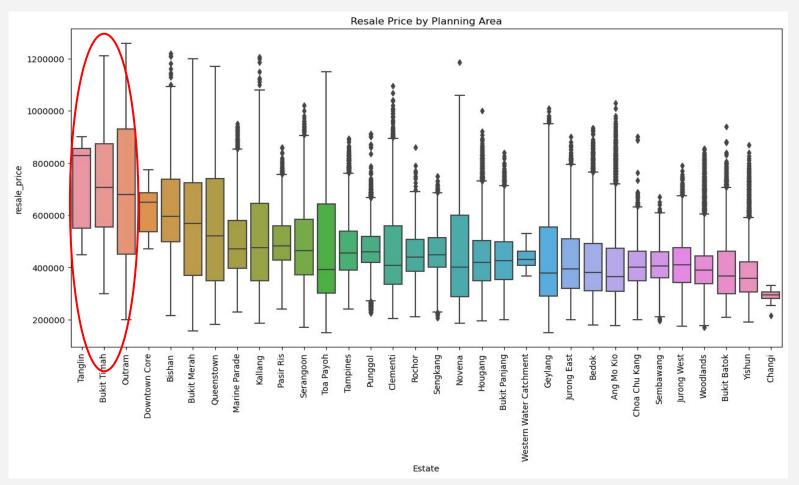
#### **Amenities:**

- Commercial
- Market / Hawker
- Multi-storey Carpark
- **Precinct Pavilion**

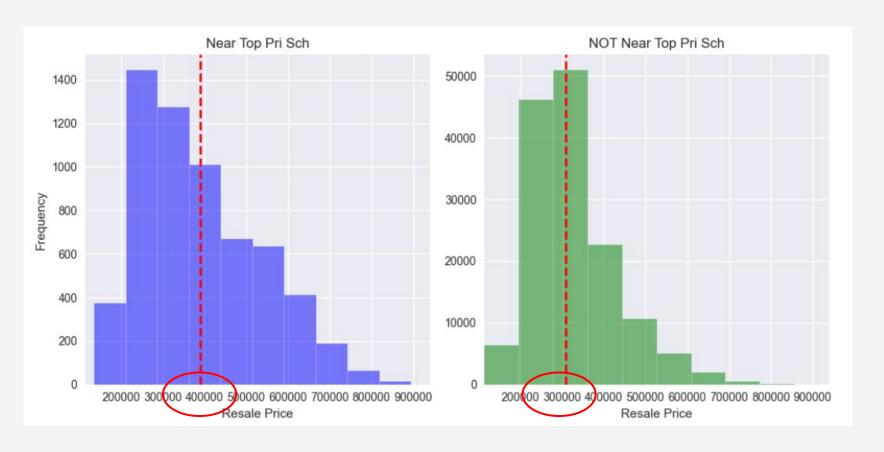
#### Houses in Central Catchment Region have higher median resale prices

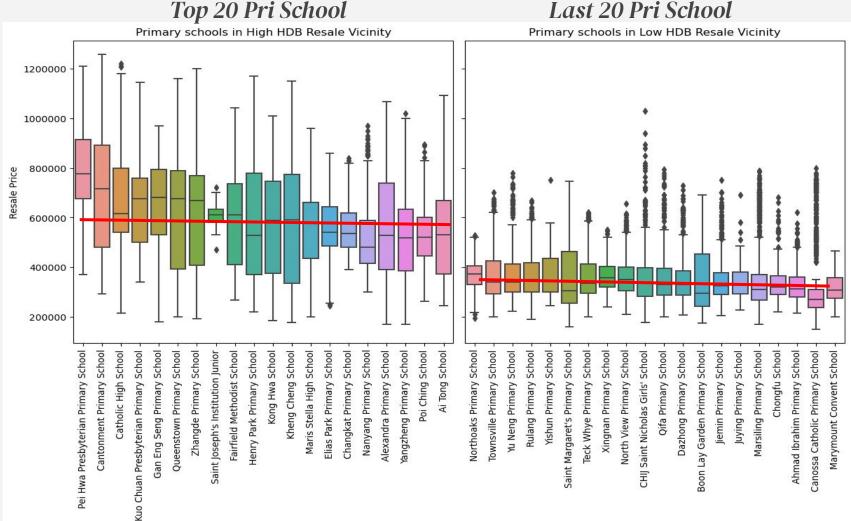


#### Tanglin, Bukit Timah, Outram have the highest median resale prices



#### HDB Flats located nearer to top primary schools\* have a higher median price







#### **Using Multilinear Regression**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon$$

- Y = Target Variable (HDB Resale Price)
- X = Features
- $\beta$  = Weights / Coefficients

<u>Goal</u>: To sieve out the best features X, so to predict the House price Y accurately. Success of the model will be evaluated by the *Root Mean Square Error (RMSE)* score.

#### Summary of Different Combination of Features explored by the Team

	Model 1	Model 2	Model 3	Model 4
HDB Price Index	X		X	X
Amenities (Merge Features)	X			
Region Mapping	X		X	
Primary school Feature interaction*		X	X	
RMSE SCORE	50,100	43,300	43,400	43,800



<sup>\*</sup>pri\_sch\_nearest\_distance' X 'pri\_sch\_affiliation' X 'cutoff\_point',

#### Model 2 had the best RMSE score, using the Ridge model.

	Model 2
Null	142926
Linear	43 034
Lasso	43120
Ridge	42965



4	Features	coef	abs_coef
0	remainderfloor_area_sqm	78835.353739	78835.353739
1	remainderremaining_lease	58124.249822	58124.249822
2	remainder_mid	18652.788900	18652.788900
3	onehotencoderpri_sch_name_Pei Tong Primary S	16965.479670	16965.479670
4	onehotencoderplanning_area_Bedok	15501.421469	15501.421469
5	onehotencoder_town_BEDOK	12332.300607	12332.300607
6	onehotencoderpri_sch_name_Qifa Primary School	10928.339089	10928.339089
7	onehotencoderflat_model_Terrace	10774.355402	10774.355402
8	onehotencoderflat_type_EXECUTIVE	10673.607506	10673.607506
9	remaindermax_floor_lvl	10567.197256	10567.197256
10	onehotencoderflat_model_DBSS	9092.540529	9092.540529



43,300

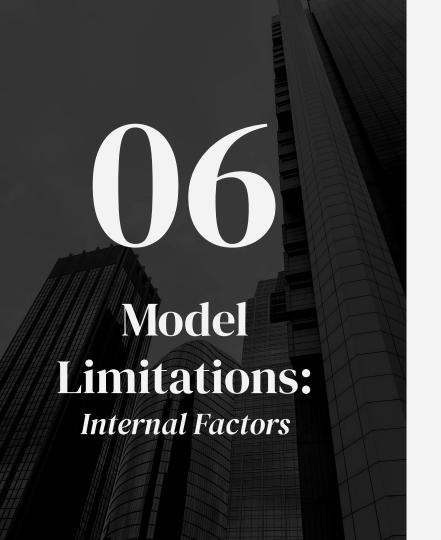
RMSE

Our model can predict a resale price that has a  $\pm$ -\$40,000 deviation

90%

Adjusted R2

Our model explains up to 90% (Adjusted R2 Score) of variability in HDB resale prices

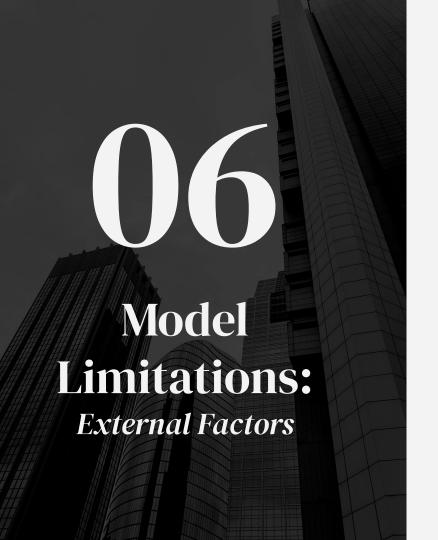


#### 1. Trade off between No. of Features and Accuracy

With an increase in number of features, this increases model accuracy and reduce RMSE score. However, this also might result in overfitting & more processing power involved.

## 2. Scope of Data Science Model Selection: Linear Regression

Can explore other models that might be more accurate



#### 3. Subject to future changes in government policies

- Classification of HDB into standard, prime and plus housing
- Change in Minimum Occupancy Period from
   5 to 10 Years for Plus Houses

#### 4. Subject to future changes in town developments

- Future Town plans
- Future Transport Network

## Conclusion

Important features for predicting HDB resale prices:

- 1. Floor area
- 2. Remaining lease
- 3. Unit level

Model could account for 90% of variability in resale prices.

Using the model, we are able to evaluate property price fairly.