

PREDICTING HOUSING (HDB) RESALE PRICES

**Project 2 - Singapore Housing Data and
Kaggle Challenge**

Gloria Neo
27 July 2023



TABLE OF CONTENTS

01

PROBLEM
STATEMENT

02

DATA CLEANING,
FEATURE SELECTION

03

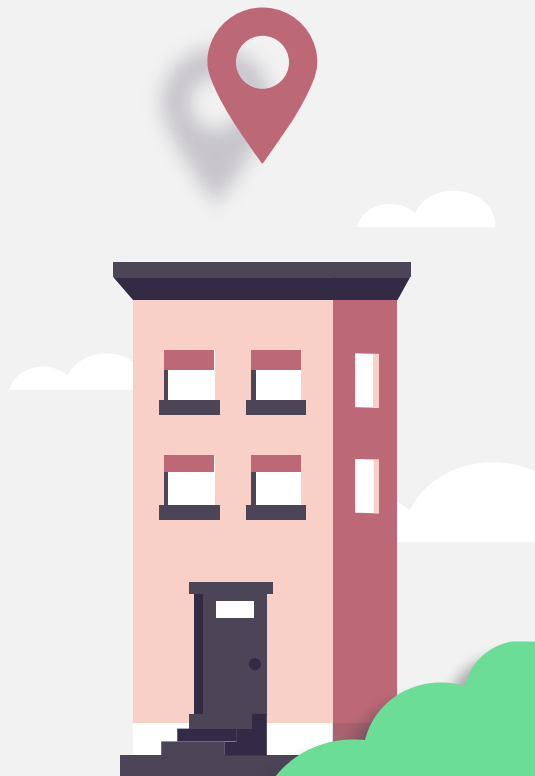
EDA,
VISUALTNS

04

SELECTED
MODEL

05

CONCLUSION



01

PROBLEM STATEMENT



CONTEXT

You have been hired by **Carousell Property**,
a Singapore-based property platform that

**allows property owners in Singapore to sell or
rent out property independently, without a real
estate agent.**

"For Sale By
Owner"
(FSBO)



**Carousell
Property**

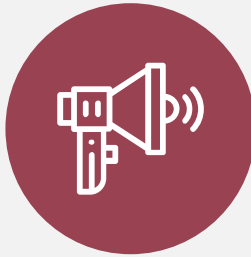


CONTEXT

provides a user-friendly and accessible platform that **alleviates the challenges a layperson would face**



**Comparative Market
Analysis (CMA)**



**Marketing the
property**



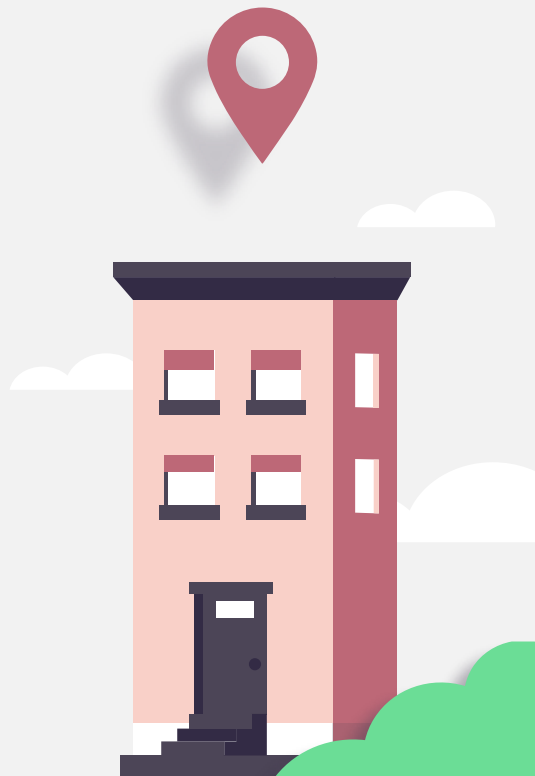
**Price
Negotiation**

while allowing users to reap the main benefit of **cost savings through omission of agent commission fees**

PROBLEM STATEMENT

Carousell Property seeks to implement a new feature on their platform that provides the layperson with data-driven insights, to aid in their property pricing strategy.

You are tasked to implement a regression model that would assist property owners in **identifying potential key selling features in their properties** and **predict potential resale price of their property.**

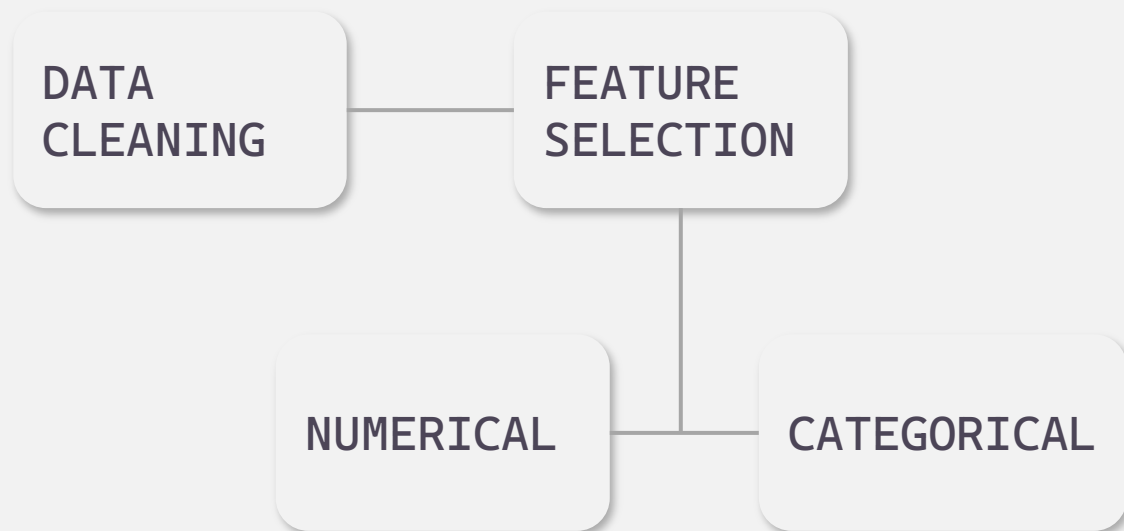


02

DATA CLEANING, FEATURE SELECTION



PROCESS



FEATURE SELECTION

NUMERICAL

CATEGORICAL

27

'floor_area_sqm'	'studio_apartment_sold'
'resale_price'	'2room_rental'
'tranc_year'	'mall_nearest_distance'
'tranc_month'	'mall_within_2km'
'mid_storey'	'hawker_nearest_distance'
'hdb_age'	'hawker_within_2km'
'max_floor_lv'	'hawker_food_stalls'
'total_dwelling_units'	'hawker_market_stalls'
'1room_sold'	'mrt_nearest_distance'
'2room_sold'	'bus_stop_nearest_distance'
'3room_sold'	'pri_sch_nearest_distance'
'4room_sold'	'vacancy'
'5room_sold'	'sec_sch_nearest_dist'
'exec_sold'	'cutoff_point'

15

'town'
'flat_type'
'block'
'street_name'
'storey_range'
'flat_model'
'full_flat_type'
'address'
'postal'
'planning_area'
'mrt_name'
'bus_stop_name'
'pri_sch_name'
'sec_sch_name'
'resale_price'



03

EDA, VISUALISATIONS



EDA, VISUALISATIONS



RESALE
PRICE



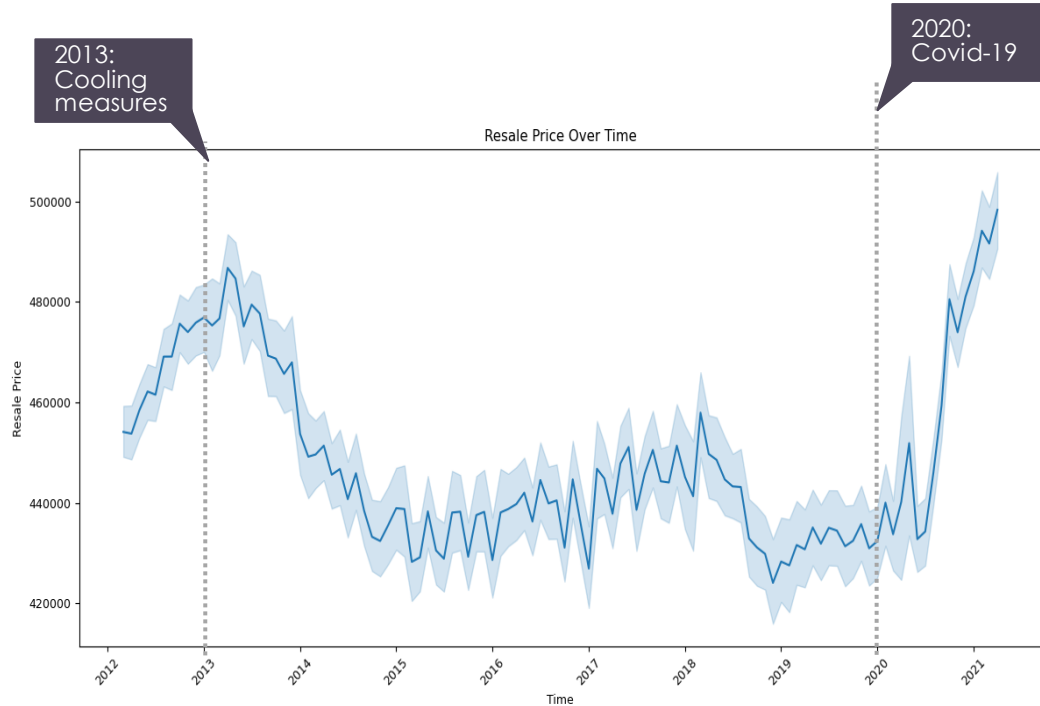
NUMERICAL
FEATURES



CATEGORICAL
FEATURES



RESALE PRICE OVER TIME



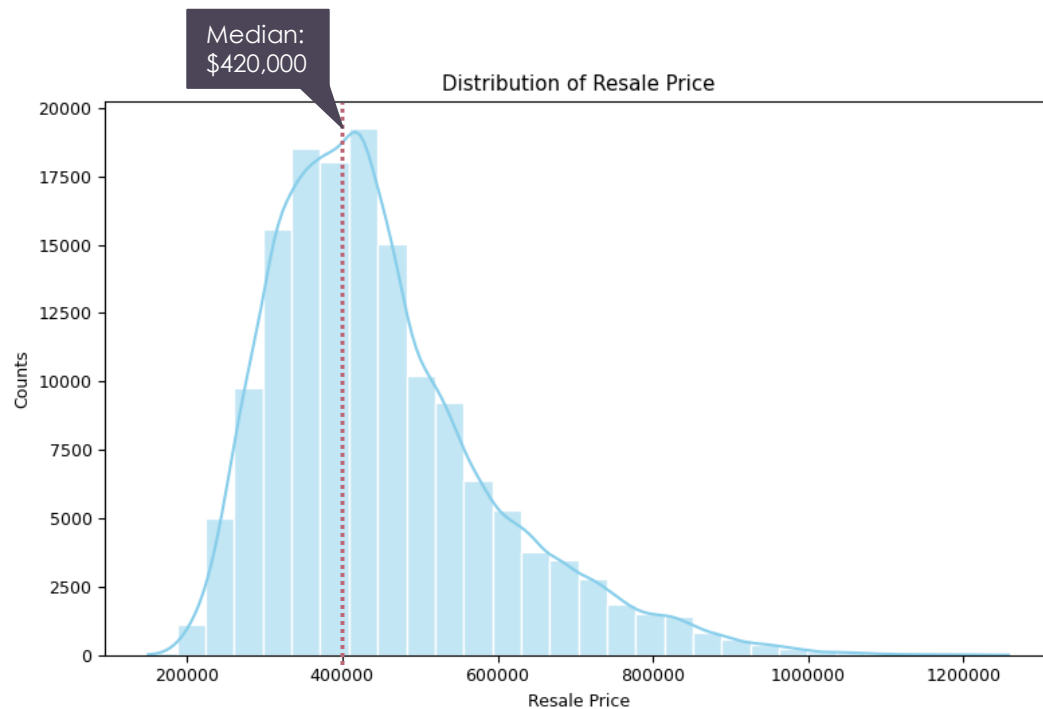
- Mean resale prices falling from mid 2013 to 2015.

From 2015 to mid 2020, prices remained quite stagnant before drastically increasing in mid 2020.

Possible reasons:

- 2013 Fall in prices: Cooling measures implemented in Jan 2013
- 2020 spike: Covid-19 - changes in lifestyle and demand

RESALE PRICE – DISTRIBUTION



- Few expensive homes contribute to the higher end of the price range, while most homes are priced below the median value.
- The median resale price of \$420,000 represents the middle value in the distribution

EDA, VISUALISATIONS



RESALE
PRICE



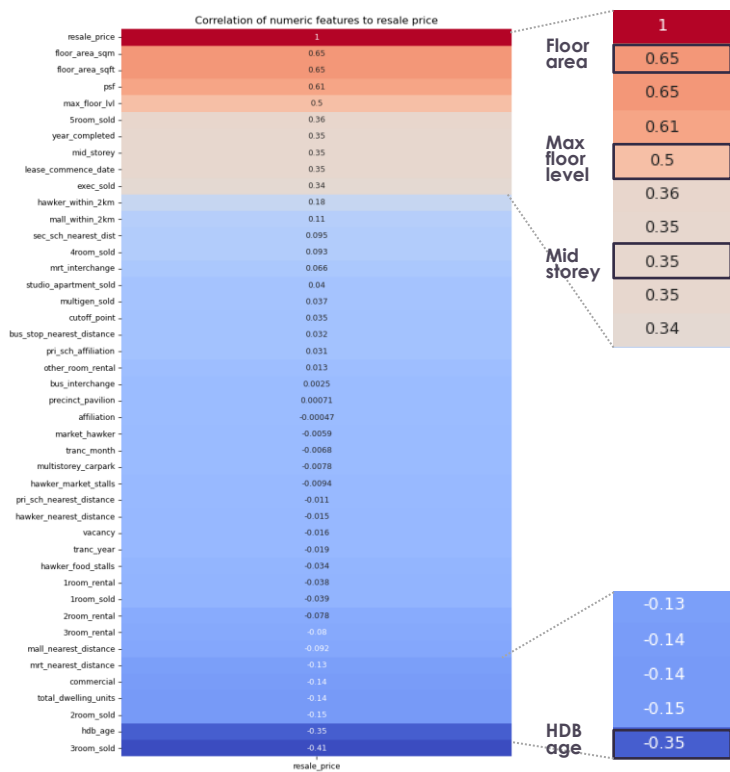
NUMERICAL
FEATURES



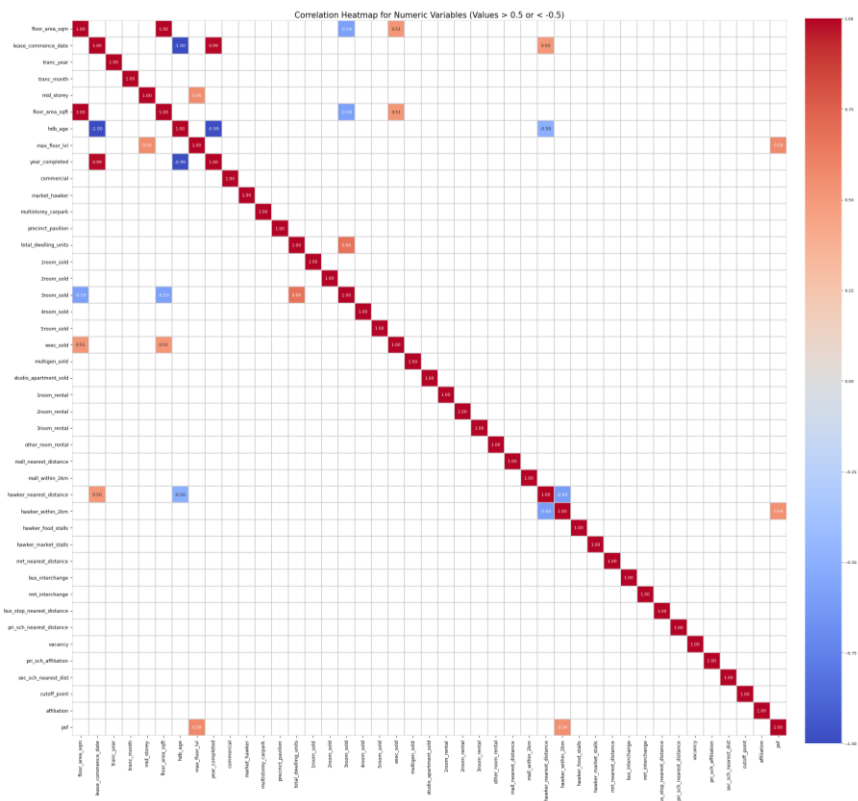
CATEGORICAL
FEATURES

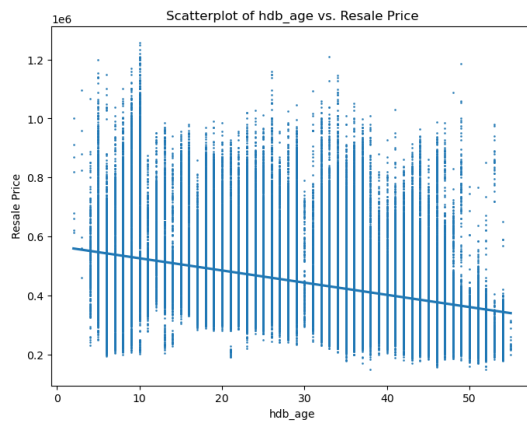
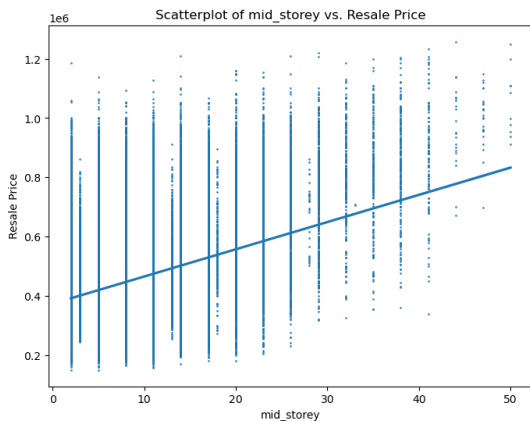
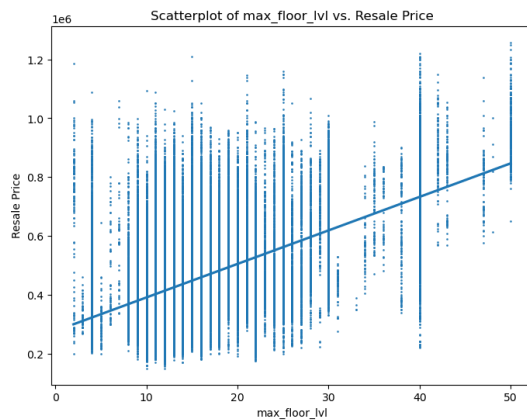
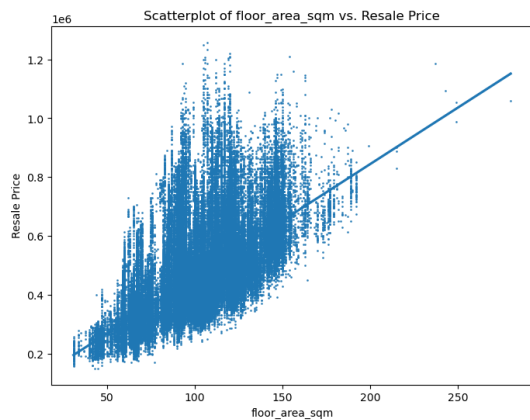


CORRELATION BETWEEN NUMERICAL FEATURES AND RESALE PRICE



CORRELATION BETWEEN NUMERICAL FEATURES





NUMERICAL FEATURES WITH STRONGEST CORRELATION TO RESALE PRICE

- Floor Area (sqm)
- Max Floor Level
- Mid storey
- HDB age

EDA, VISUALISATIONS



RESALE
PRICE



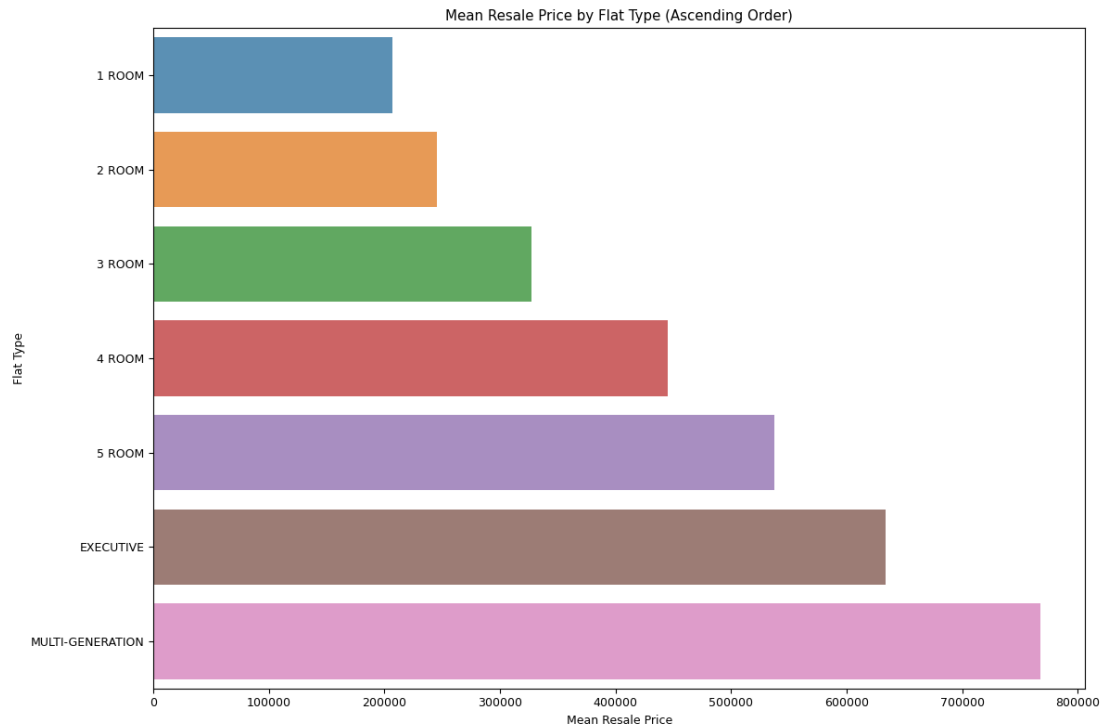
NUMERICAL
FEATURES



CATEGORICAL
FEATURES



MEAN RESALE PRICE BY FLAT TYPE



- The larger the flat type (more rooms), the higher the resale price

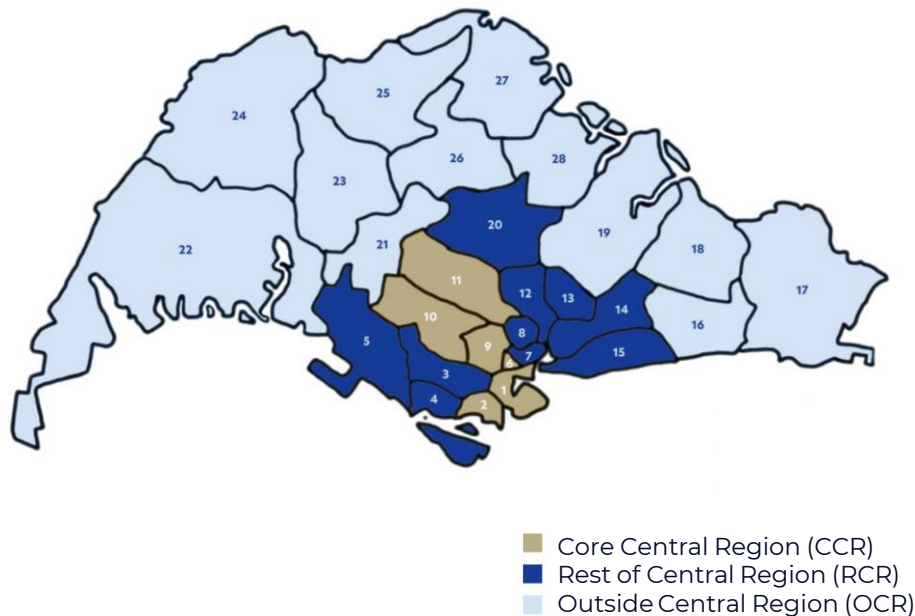
MEAN RESALE PRICE BY FLAT TYPE & MODEL



- Smaller flats of special models could outperform larger flats in terms of resale price

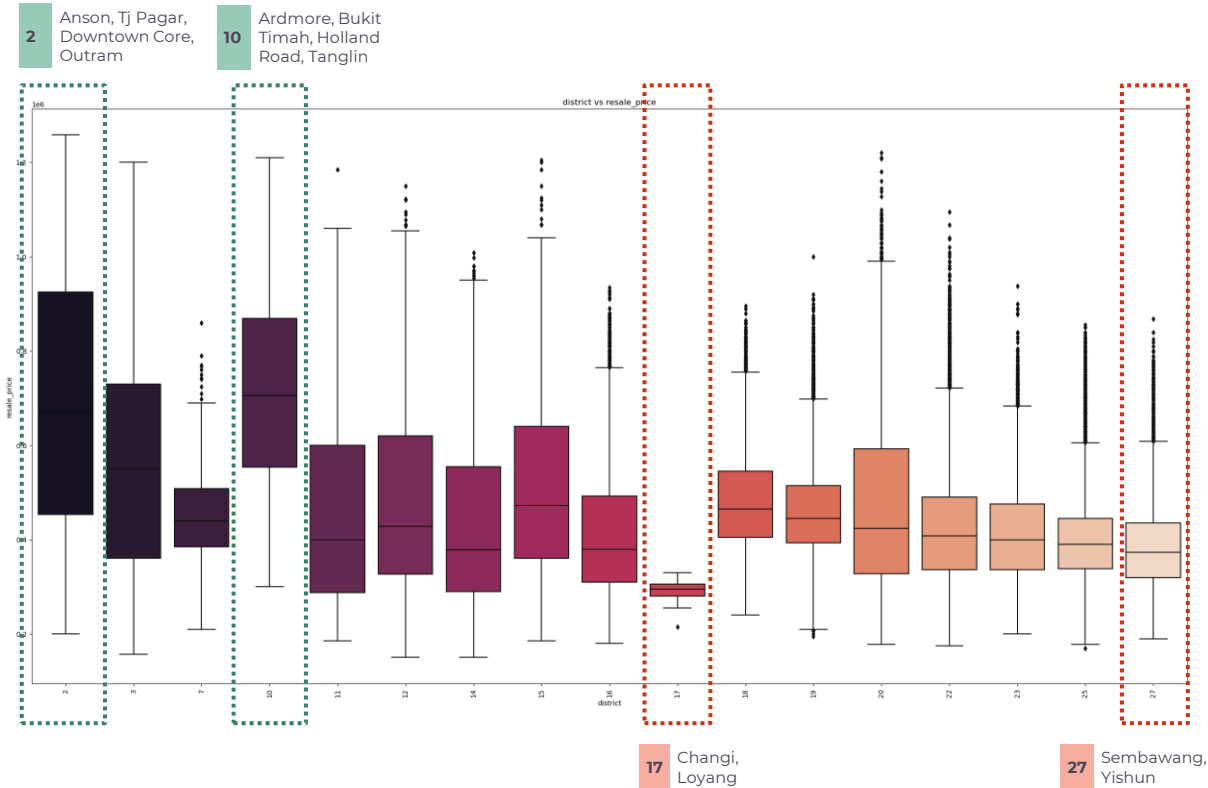
THE 28 SINGAPORE DISTRICTS

1	Raffles Place, Marina, Cecil, People's Park	15	Joo Chiat, Amber Road, Katong
2	Anson, Tanjong Pagar, Downtown Core, Outram	16	Bedok, Upper East Coast, Eastwood, Kew Drive
3	Tiong Bahru, Queenstown, Bukit Merah	17	Changi, Loyang
4	Telok Blangah, Harbourfront	18	Tampines, Pasir Ris
5	Pasir Panjang, Hong Leong Garden, Clementi New Town	19	Punggol, Hougang, Serangoon Gardens
6	High Street, Beach Road, City Hall	20	Ang Mo Kio, Bishan
7	Middle Road, Golden Mile	21	Upper Bukit Timah, Ulu Pandan, Clementi Park
8	Little India, Farrer Park, Serangoon Road	22	Jurong, Boonlay, Tuas, Lakeside, Jurong East, Jurong West, Clementi
9	Orchard, Cairnhill, River Valley	23	Choa Chu Kang, Dairy Farm, Hillview, Bukit Panjang, Choa Chu Kang, Bukit Batok, Western Water Catchment
10	Ardmore, Bukit Timah, Holland Road, Tanglin	24	Lim Chu Kang, Tengah
11	Watten Estate, Novena, Thomson	25	Kranji, Woodgrove
12	Toa Payoh, Serangoon, Balestier	26	Upper Thomson, Springleaf
13	Macpherson, Bradell	27	Sembawang, Yishun
14	Geylang, Eunos	28	Seletar



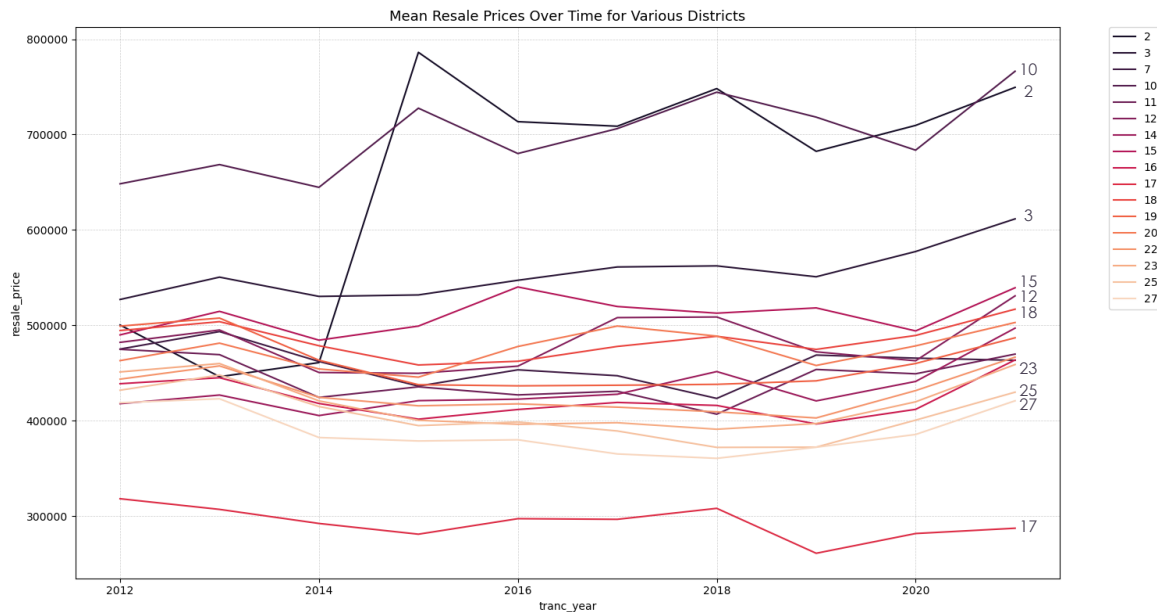
Source: <https://www.fraxtor.com/singapore-districts-and-their-origins/>

DISTRICTS AGAINST RESALE PRICE



- Districts with highest median resale price: district 2 and district 10
- Districts with the lowest median resale price: district 17 and 27

RESALE PRICE AGAINST TIME FOR VARIOUS DISTRICTS



- In general, the districts with the smaller numbers (mature estates) seem to perform better
- Besides districts 2 and 10 which have larger fluctuations, all other districts seem to have a stable mean resale price over the years.

04

SELECTED MODEL



SUMMARY OF METRICS

Model	R-square Train score	R-square Test Score	Cross Val Score	Train RMSE	Test RMSE
Linear Regression	0.9280	0.9262	$3.327 \times 1e14$	38539.1887	38700.8151
Ridge	0.9289	0.9261	0.9262	38565.9087	38718.0444
Lasso	0.8559	0.8533	0.9263	54528.7702	54554.5655

- Selected model manages to explain 92.6% of variance in resale price.
- Test RMSE tells us that there could be a potential error amount of \$38,718.

05

CONCLUSION



CONCLUSION



GAINED INSIGHTS ON KEY FEATURES

To assist homeowners in marketing their properties on the platform



PREDICTION OF RESALE PRICE

To assist home owners in pricing and negotiation when selling their home



LIMITATIONS



LIMITATIONS OF DATA COLLECTED

Limitations in terms of its scope and recency as the data is only collected up to April 2011



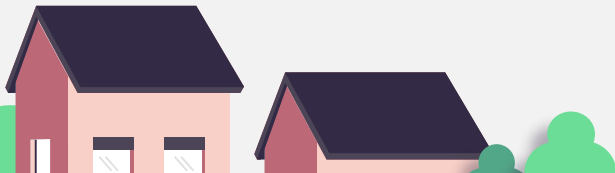
LIMITED FEATURES CONSIDERED

Dataset could not encompass all relevant variables. Data also not document characteristics within the unit itself



CHANGING MARKET DYNAMICS, GOVT POLICIES

Such unexpected changes would affect the relevance of historical data



THANK YOU

PPT Template Credits: Slidesgo, Flaticon, Freepik

