Singapore Housing Data and Kaggle Challenge

Project 2 - Team 1
(Johann, Anil, Kunhong, Shermine)

jaks

Agenda

- Introduction
- Problem Statement
- Dataset
- Methodology & Feature choices
  - Location / Planning area
  - Size of flat / Type
  - Transportation
  - Education
- Recommendations
- Summary

jaks

The HDB flat…

- Conceived and launched in 1964
- Cornerstone of Singapore’s public housing policy
- Home for 85-90% of all Singaporeans
- Asset class with near-lifetime appreciation

jaks

Scenario & Problem Statement

Scenario:

Singapore's URA and HDB is monitoring the rise in HDB flat resale prices over the past few years, and is understandably concerned that some units are being priced out for a lot of buyers. They also are about to launch the HDB Resale Portal, which will allow HDB unit owners to list their flats on their own.

We - **JAKS Pte Ltd** - are a specialist consultant firm and have been asked by URA and HDB to to help them identify the factors and levers affecting a HDB unit's price appreciation. This will help to:

1. Offer HDB unit owners pricing suggestions on the HDB Resale Portal
2. Plan for future developments as well as identify HDB estates for renewal and refurbishment plans in the future.

Problem Statement:

Based on given historical data of HDB resale prices and conditions prevalent at the time of resale, identify and present a data science model, which can accurately predict resale prices for HDB units, given the same criteria.

jaks

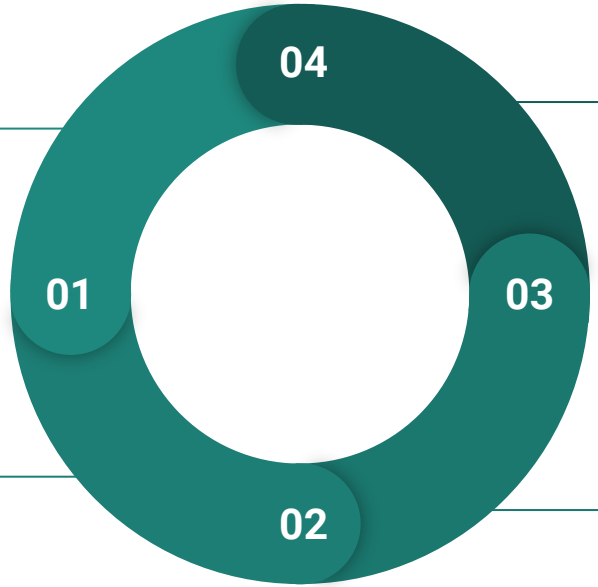
Dataset - features and columns

HDB Resale Dataset Categories

Transaction Details	Location Details	HDB Unit	HDB Block Utilities	HDB Units sold/rented	Nearby Amenities	Transportation	Education
Tranc_YearMonth	town	flat_type	residential	1room_sold	Mall_Nearest_Distance	mrt_nearest_distance	pri_sch_nearest_distance
Tranc_Year	block	flat_model	commercial	2room_sold	Mall_Within_500m	mrt_name	pri_sch_name
Tranc_Month	street_name	storey_range	market_hawker	3room_sold	Mall_Within_1km	bus_interchange	vacancy
resale_price	address	floor_area_sqm	multistorey_carpark	4room_sold	Mall_Within_2km	mrt_interchange	pri_sch_affiliation
	postal	floor_area_sqft	precinct_pavilion	5room_sold	Hawker_Nearest_Distance	mrt_latitude	pri_sch_latitude
	Latitude	full_flat_type	total_dwelling_units	exec_sold	Hawker_Within_500m	mrt_longitude	pri_sch_longitude
	Longitude	lease_commence_date		multigen_sold	Hawker_Within_1km	bus_stop_nearest_distance	sec_sch_nearest_dist
	planning_area	hdb_age		studio_apartment_sold	Hawker_Within_2km	bus_stop_name	sec_sch_name
		max_floor_lvl		1room_rental	hawker_food_stalls	bus_stop_latitude	cutoff_point
		year_completed		2room_rental	hawker_market_stalls	bus_stop_longitude	affiliation
		mid_storey		3room_rental			sec_sch_latitude
		lower		other_room_rental			sec_sch_longitude
		upper					
		mid					

jaks

Methodology / Features



Location

Education

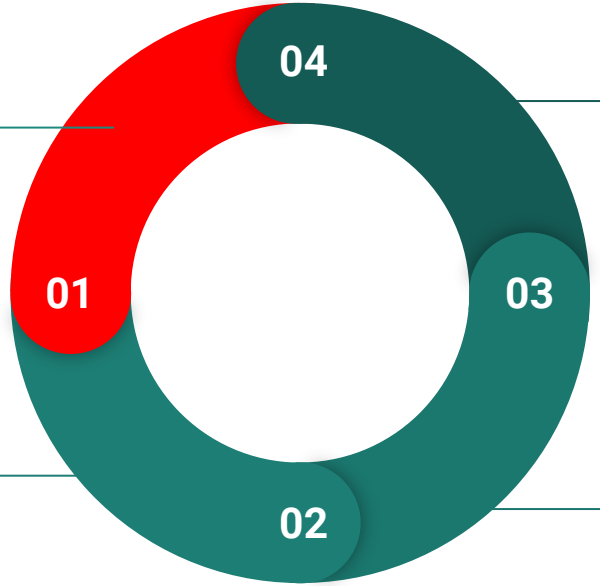
HDB Characteristic

Transport

jaks

Location

HDB Characteristic

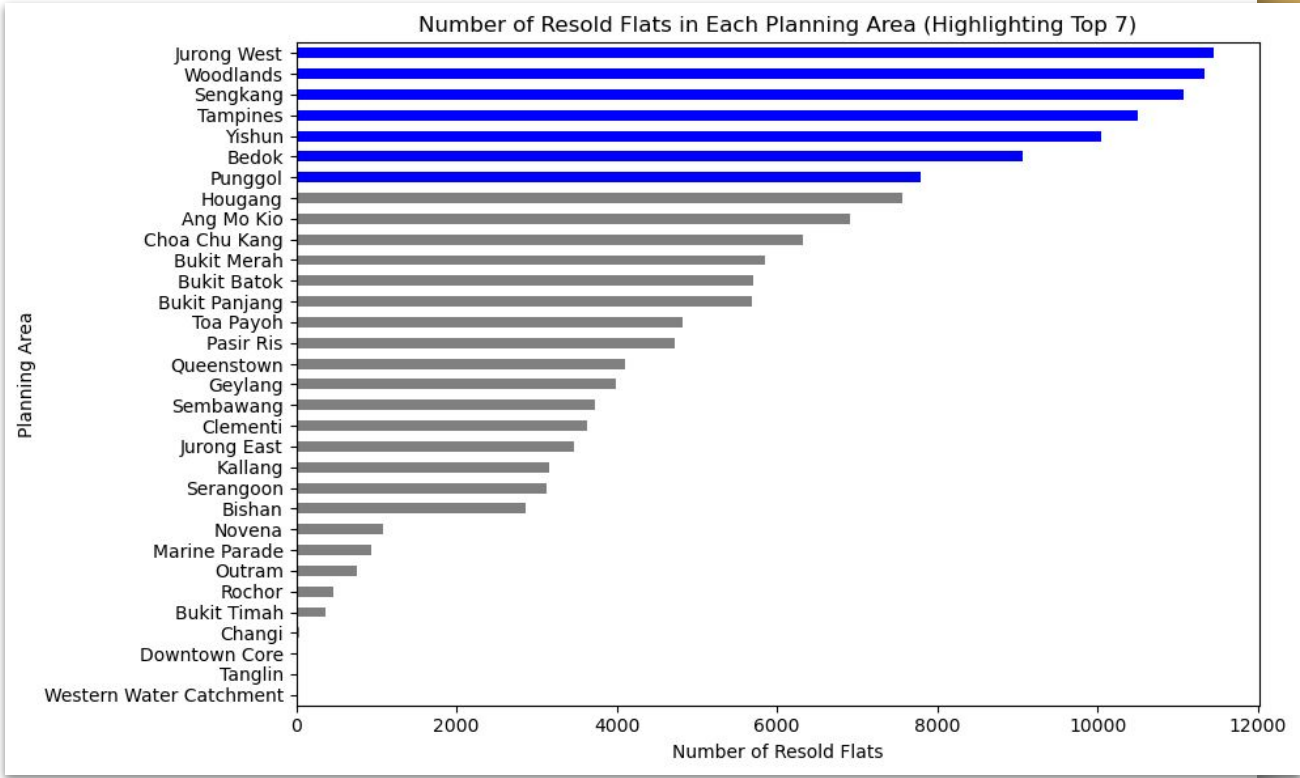


Education

Transport

Location

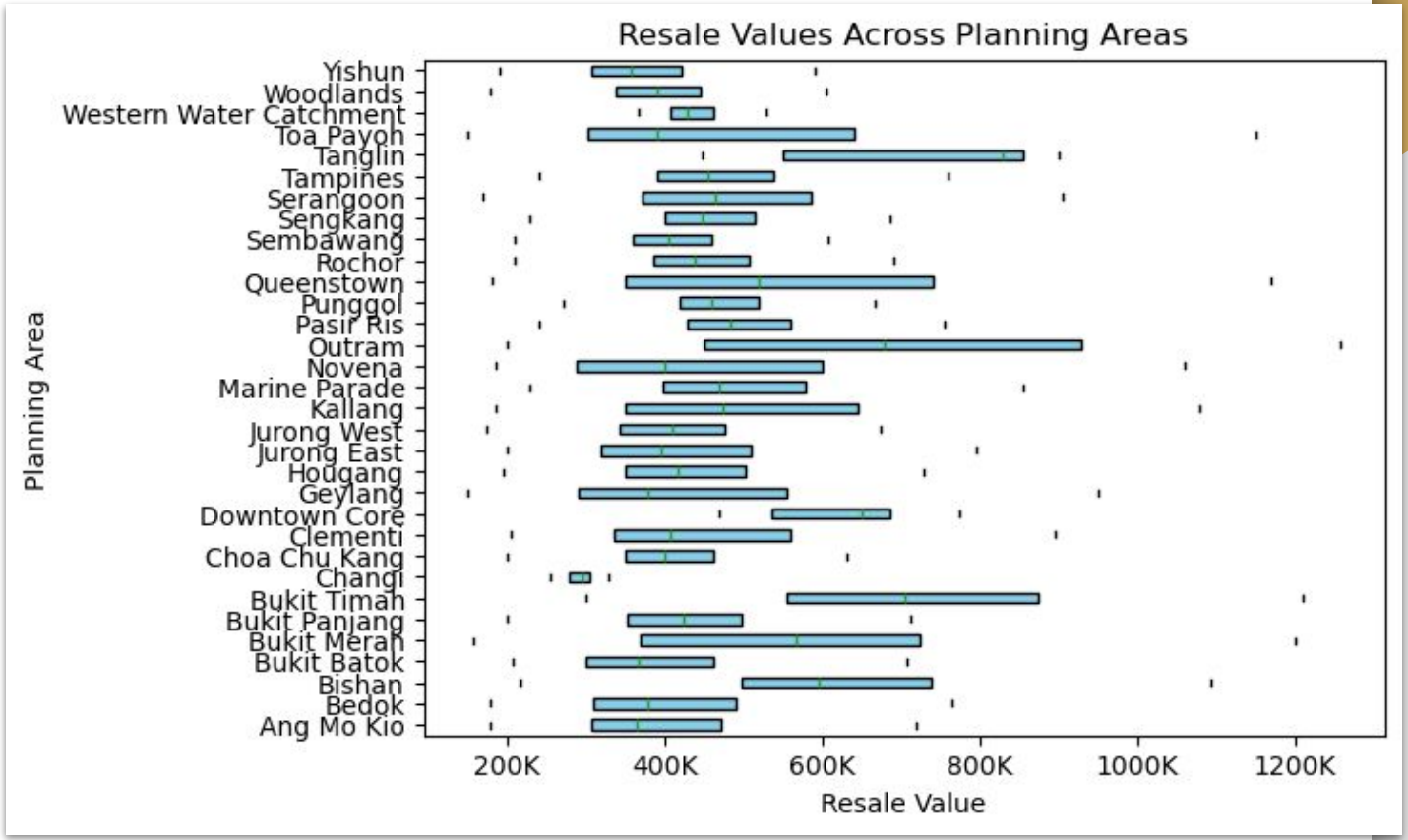
3 out of top 5 planning areas in terms of resold volume, were non-mature estates i.e. Jurong West, Woodlands, Sengkang etc.



jaks

Location (contd.)

But, these areas are not in the top 5 planning regions in terms of resale value.

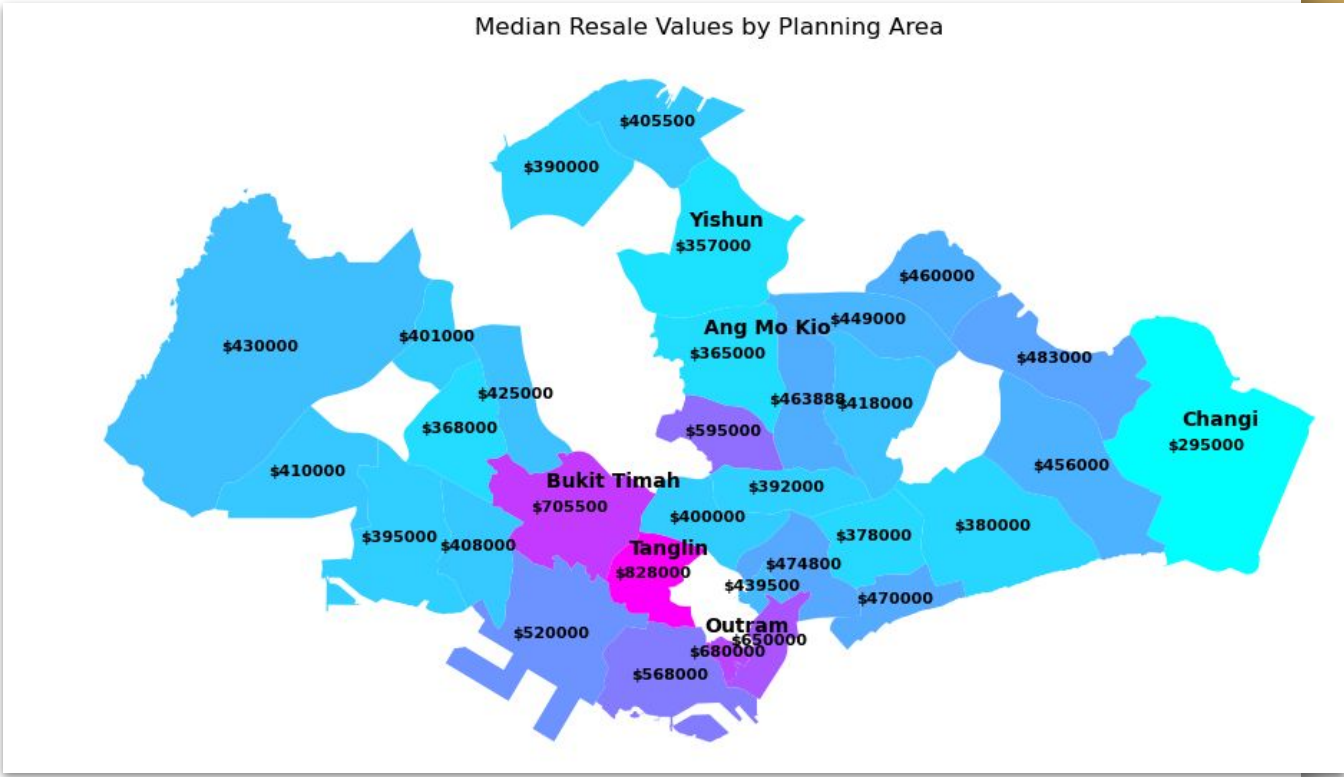


jaks

Location (contd.)

Mapping out planning_areas by median prices confirms that number of flats resold ≠ higher resale value

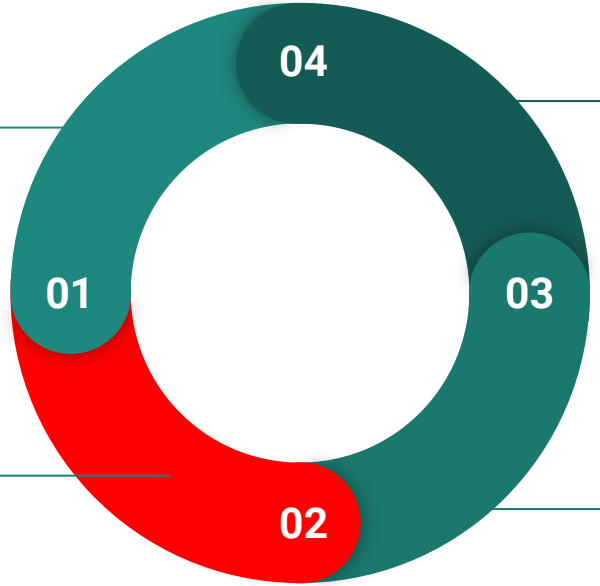
Instead reaffirms that location / proximity to work / education etc. drives demand and higher resale prices



jaks

Location

HDB Characteristic

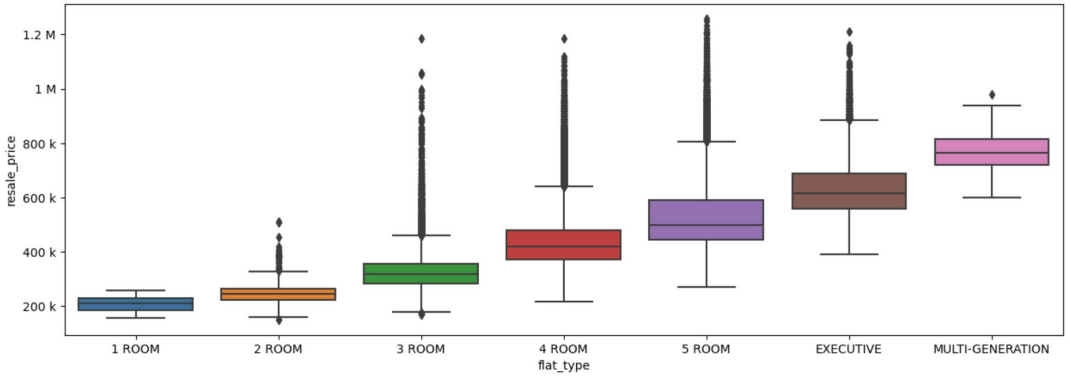


Education

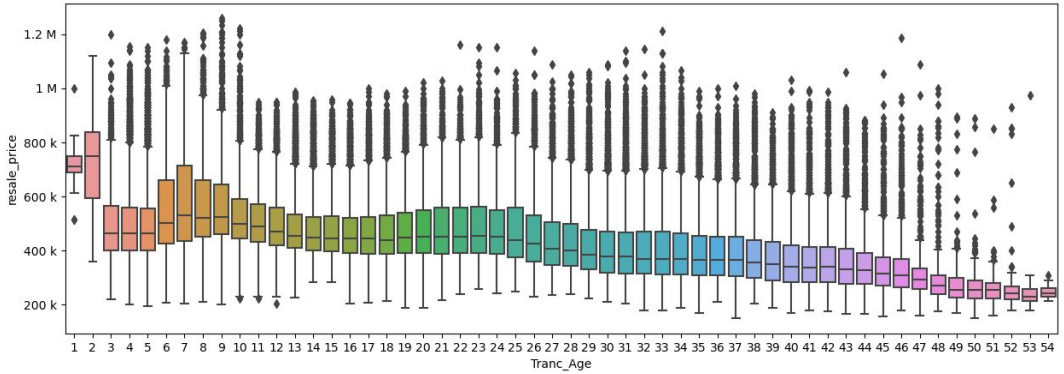
Transport

HDB Characteristics

- **Flat_type/ Floor_area_sqft**
  - Positively correlated with price



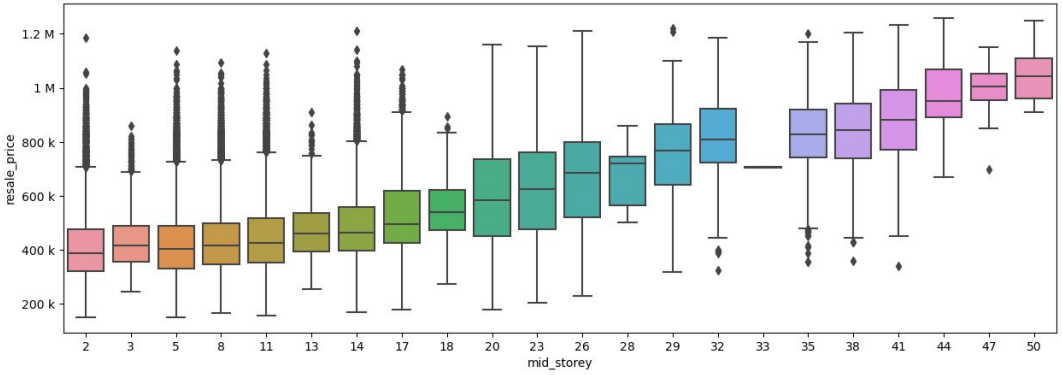
- **Age at Transaction**
  - Tranc_Year - Lease_comm_date
  - Negatively correlated with price



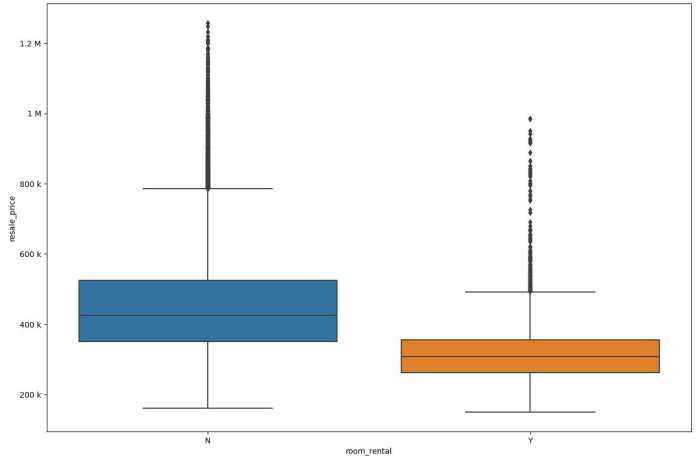
jaks

HDB Characteristics

- **Mid_storey**
  - Positively correlated with price



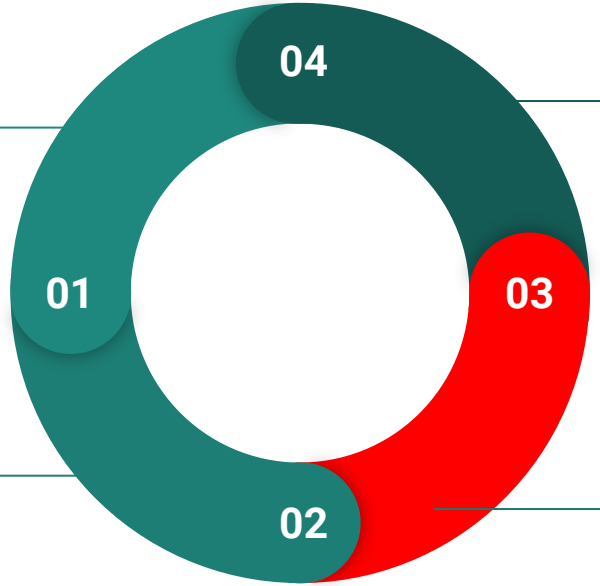
- **HDB Public rentals**
  - For citizens who are unable to afford other housing options
  - Perceived undesirable traits of tenants



jaks

Location

HDB Characteristic



Education

Transport

Transport Features

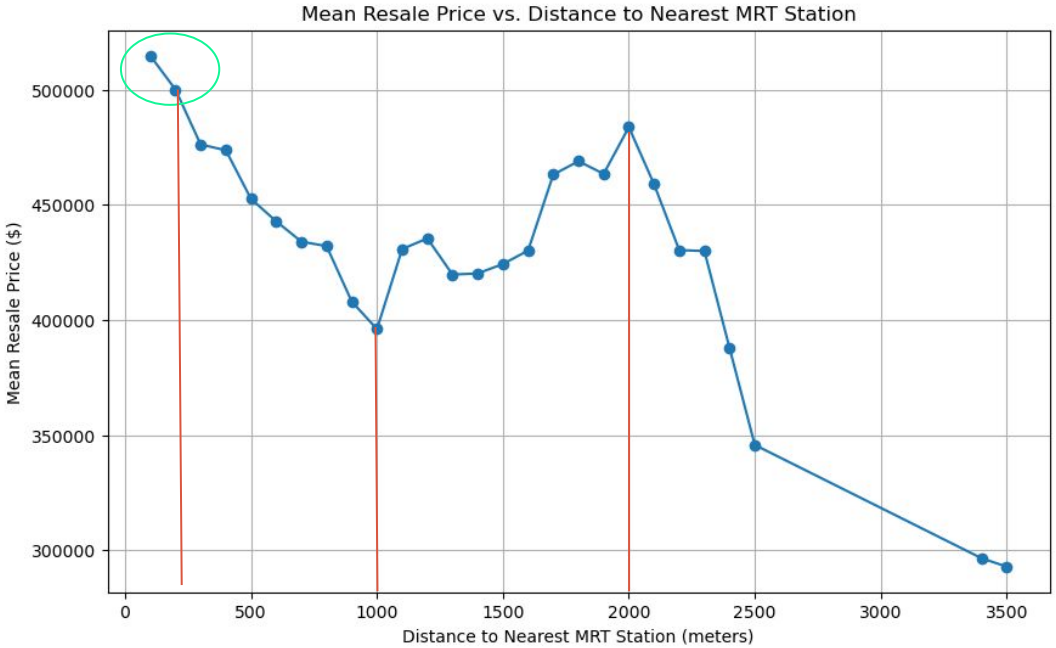
Correlations with resale_price:

mrt_nearest_distance	-0.128222
bus_interchange	0.002547
mrt_interchange	0.066413
mrt_latitude	-0.214939
mrt_longitude	0.106506
bus_stop_nearest_distance	0.031940
bus_stop_latitude	-0.215760
bus_stop_longitude	0.106394

- mrt_latitude & bus_stop_latitude has the strongest correlation (negative) to the target variable amongst all the transport related features
  - However, this was not selected as a feature of choice as it was not logical
- For potential homeowners - proximity to MRT will be a priority concern
  - Selected mrt_nearest_distance
    - Logical point of consideration for potential homeowners
    - Negative correlation suggests that being nearer to an MRT station reduces resale price

jaks

Transport Features



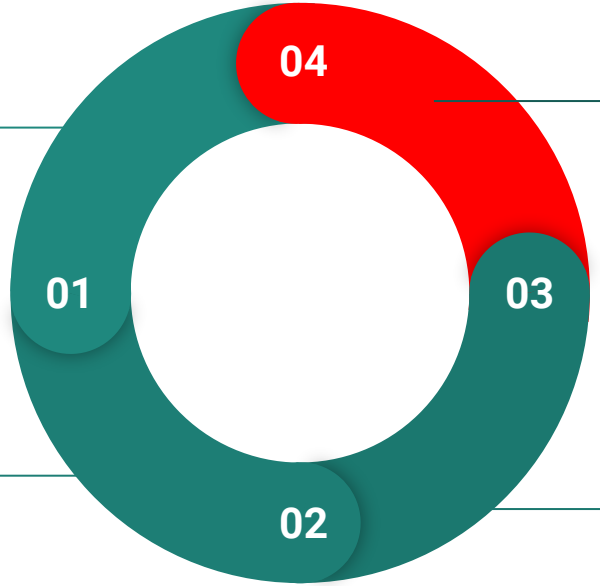
Mrt_nearest_distance:

- Mean resale prices highest when the distance to MRT station is extremely low
  - Begins to drop as the distance increases
  - Starts rising again after 1000m, but drops sharply after 2000m (too inaccessible)

jaks

Location

HDB Characteristic



Education

Transport

Education

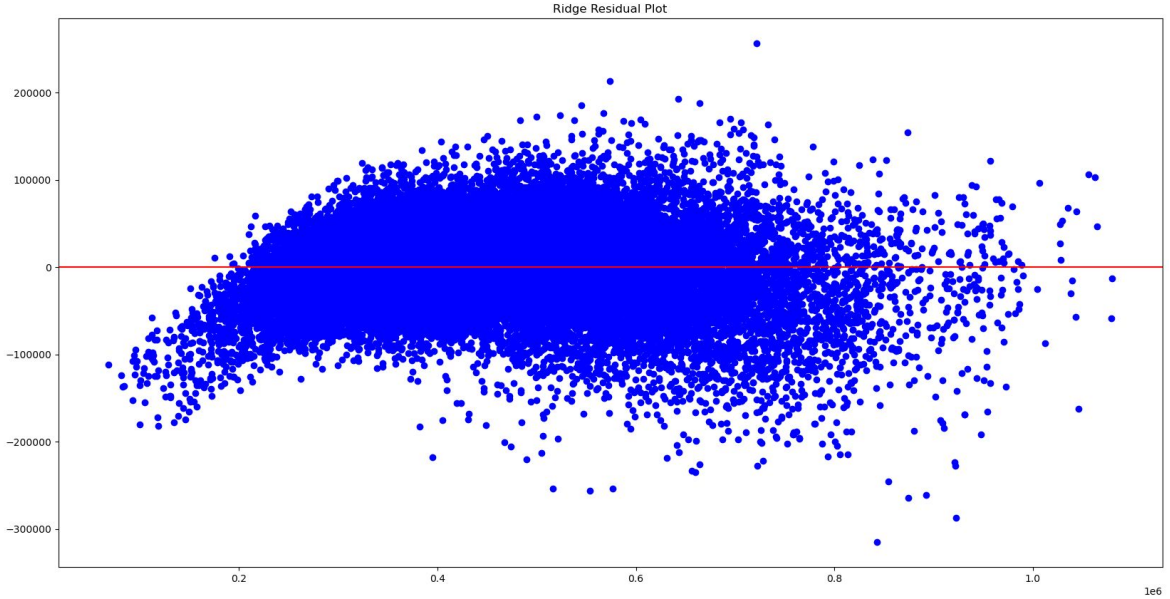
- Proximity to schools are not directly correlated to resales HDB resales prices
- As young adults / families mainly go for BTO for subsidies.
- However, proximity to top schools are still a consideration when it comes to purchasing on resale unit.
- Resales units are mainly for personal capital gain.

jaks

Linear Regression (L2) Results

```
Ridge Search Alpha 13.257
Ridge Train Score: 0.902
Ridge Test Score: 0.902
Ridge RMSE: 44,753
```

	Predictions	Actual
0	301,580	305,000
1	459,355	475,000
2	350,215	310,000
3	466,040	438,000
4	376,184	366,000



jaks

Recommendations

- Include Polyclinic & Health related Amenities
- Account for Change in Population growth and influx of new residents
- Factor in economic conditions and loan interest rates as data points

jaks

Summary

- Based on Category selection and perspectives, we were able to identify certain key features, which can influence resale values for a HDB flat
- We identified a high scoring (R2 = 0.902) Ridge Linear Regression model, and used it to predict resale prices with relatively low differences from actual values.
- Finally, to improve the model accuracy further, and with an eye on the future, we recommended additional data points such as Health amenities, loan interest rates etc.

jaks

Thank you

jaks