Case Study

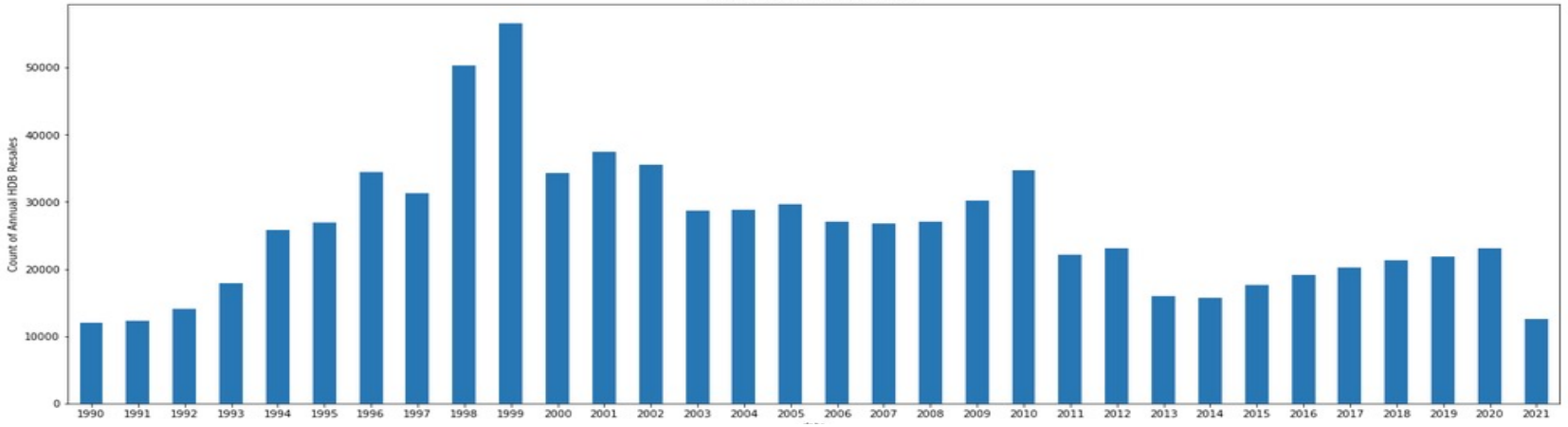
HOUSING & DEVELOPMENT BOARD

HDB Resale Price Estimator

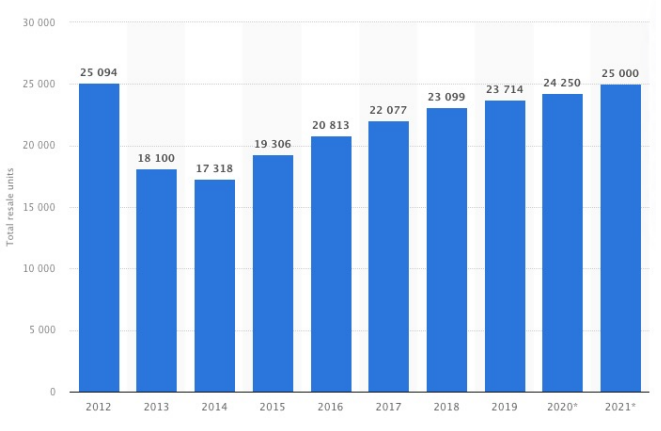
BACKGROUND INFORMATION – RESALE VOLUME

- HDB dataset given shows at least 10k HDB resale transactions per annum, and external sources show in recent years this number goes past 20k.

Annual HDB Resale Transaction Volume (From Dataset)



HDB Resale Estimates

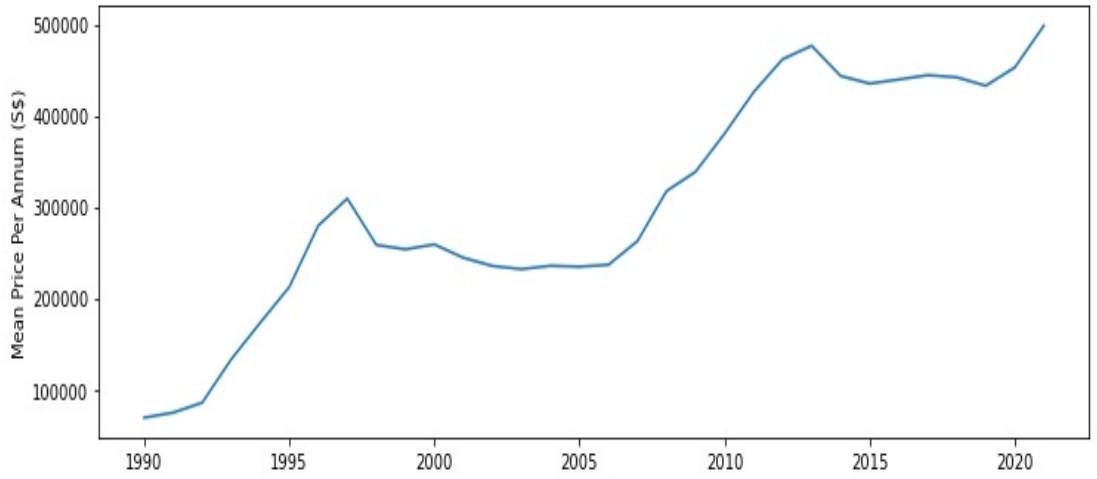


Source: HDB Dataset, Statista

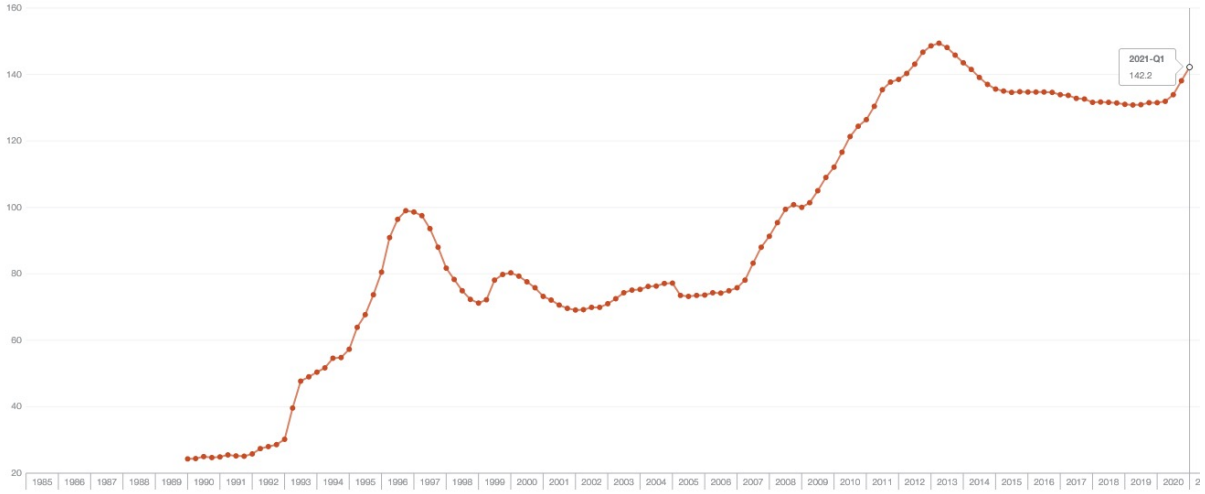
BACKGROUND INFORMATION – RESALE PRICE

- HDB resale prices have remained buoyant since its conception
- One of the reasons for the resilience of the public housing market is it is a basic necessity and the most affordable type of housing in Singapore.
- Demand for public housing is unlikely to evaporate even during economic downturns.

Mean HDB Resale Price Per Annum



HDB Resale Price Index

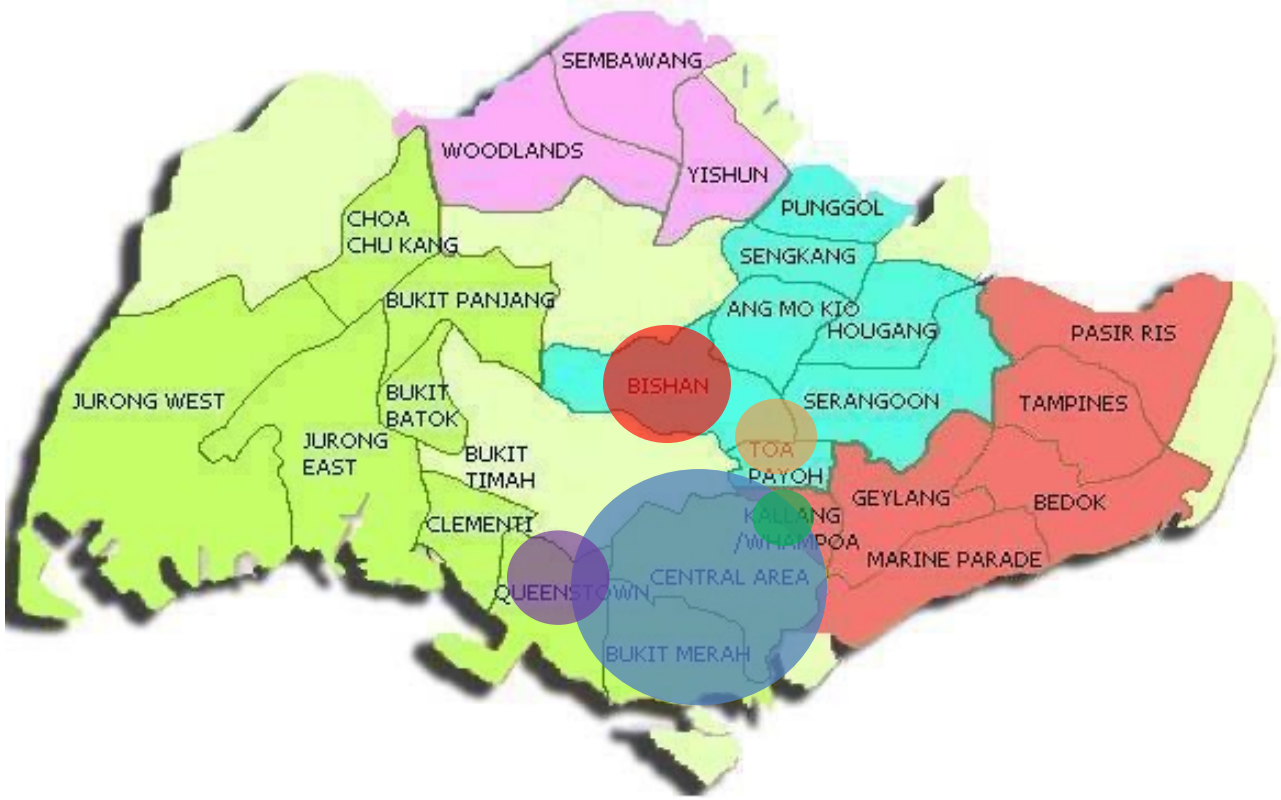


Source: HDB Dataset, Data.gov.sg

BACKGROUND INFO – MOST EXPENSIVE NEIGHBOURHOODS

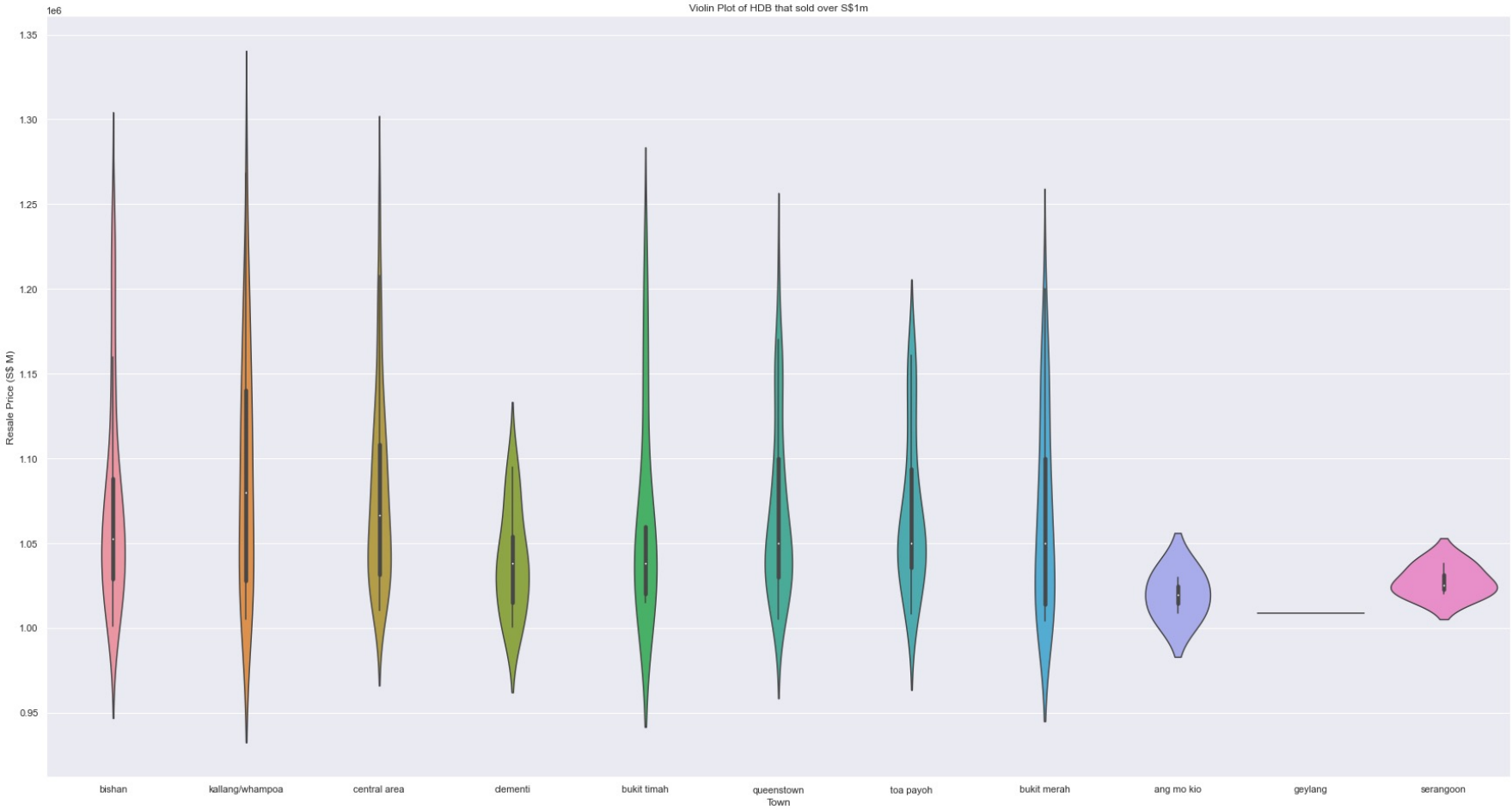
TOWNS WITH THE MOST SALES OVER S$1M

Town Name	Count
Central Area (CBD)	112
Bishan	68
Queenstown	49
Toa Payoh	38
Kallang/Whampoah	29



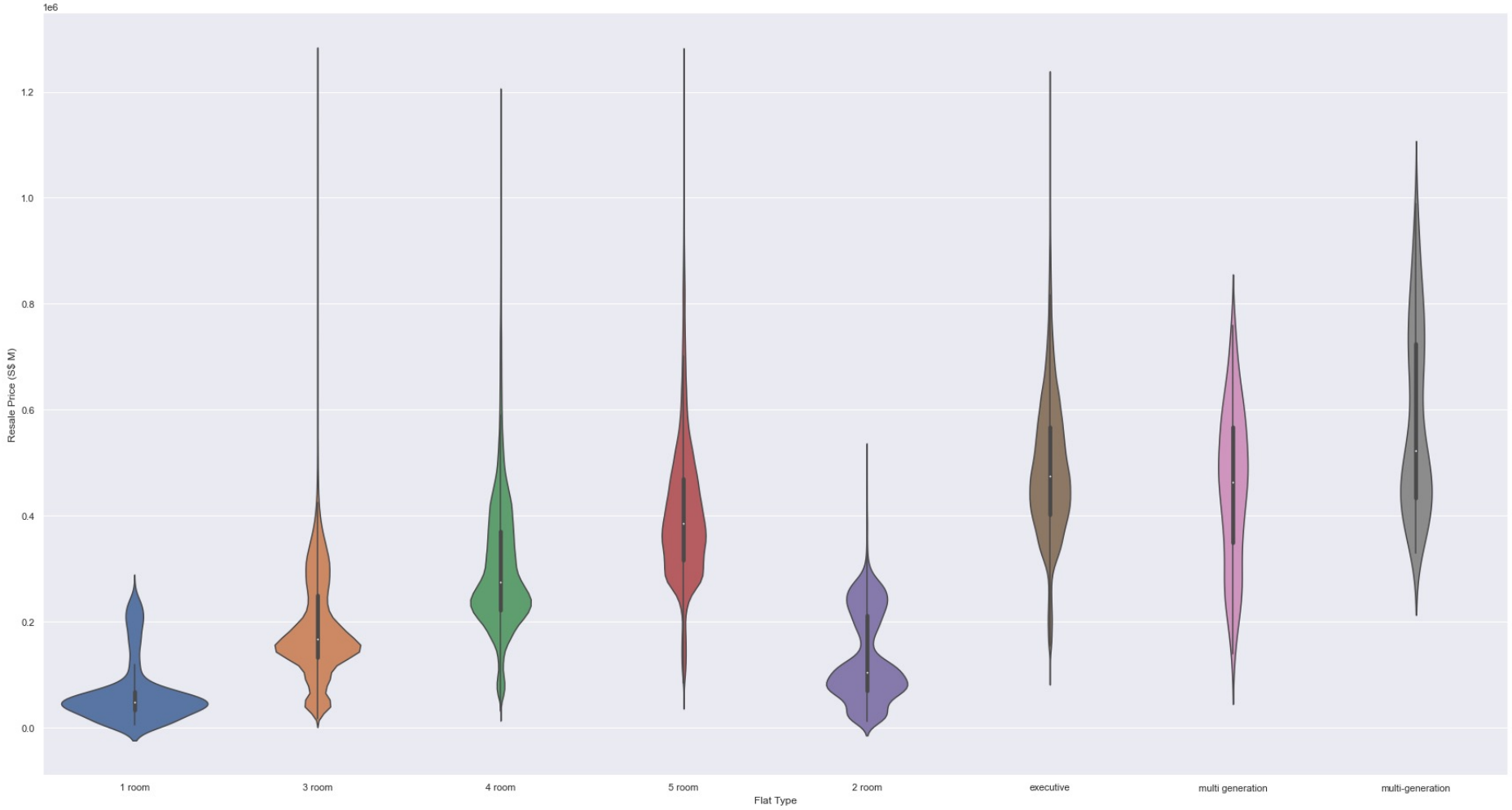
Source: HDB Dataset
Notes: Graph not to scale

BACKGROUND INFO – MOST EXPENSIVE NEIGHBOURHOODS



Source: HDB Dataset – Only HDB resale transactions over S$1m were analyzed

BACKGROUND INFO – MOST EXPENSIVE FLAT TYPE



Source: HDB Dataset

BACKGROUND INFORMATION – GOVT REGULATION

- The ownership and transaction of HDB flats are regulated. Only Singaporeans and permanent residents can own HDB flats. Each resident household can own only one HDB flat, which will be their primary place of residence. Each flat owner must live in his flat for a minimum of five years before he can sell the flat.

- There are other rules and regulations to ensure that HDB flats are treated as a home rather than a trading asset, which contribute to its price stability.

Source: HDB.gov.sg

BUSINESS PROBLEM

At least 80% of Singaporean resident population live in HDB's, of which 90% own their own homes (HDB.gov.sg). So it is very likely that most Singaporeans either are current homeowners or are on the market to purchase a HDB.

But due to government regulation, the public housing market is not as liquid as the private housing market, so most consumers have little knowledge on the market.

Homeowners and buyers often rely on the expertise of real estate agents that specialize in the HDB sector and have the potential to "go in blind" without fully understanding the worth of the asset in question.

We propose a data driven approach for homeowners and potential buyers to better understand the market and use this tool as a point of negotiation to value this asset class.

Source: HDB.gov.sg, The Business Times

The Dataset

- Given 5 datasets with information on HDB resale transaction information from 1990 to 2020 were merged
- Around 25k duplicate entries were removed
- Final dataset is 834,508 entries with 11 feature columns

```
for idx,df in enumerate(dflist):
    display(df.head())
```

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price
0	1990-01	ANG MO KIO	1 ROOM	309	ANG MO KIO AVE 1	10 TO 12	31.0	IMPROVED	1977	9000
1	1990-01	ANG MO KIO	1 ROOM	309	ANG MO KIO AVE 1	04 TO 06	31.0	IMPROVED	1977	6000
2	1990-01	ANG MO KIO	1 ROOM	309	ANG MO KIO AVE 1	10 TO 12	31.0	IMPROVED	1977	8000
3	1990-01	ANG MO KIO	1 ROOM	309	ANG MO KIO AVE 1	07 TO 09	31.0	IMPROVED	1977	6000
4	1990-01	ANG MO KIO	3 ROOM	216	ANG MO KIO AVE 1	04 TO 06	73.0	NEW GENERATION	1976	47200

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price
0	2000-01	ANG MO KIO	3 ROOM	170	ANG MO KIO AVE 4	07 TO 09	69.0	Improved	1986	147000.0
1	2000-01	ANG MO KIO	3 ROOM	174	ANG MO KIO AVE 4	04 TO 06	61.0	Improved	1986	144000.0
2	2000-01	ANG MO KIO	3 ROOM	216	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	159000.0
3	2000-01	ANG MO KIO	3 ROOM	215	ANG MO KIO AVE 1	07 TO 09	73.0	New Generation	1976	167000.0
4	2000-01	ANG MO KIO	3 ROOM	218	ANG MO KIO AVE 1	07 TO 09	67.0	New Generation	1976	163000.0

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price
0	2012-03	ANG MO KIO	2 ROOM	172	ANG MO KIO AVE 4	06 TO 10	45.0	Improved	1986	250000.0
1	2012-03	ANG MO KIO	2 ROOM	510	ANG MO KIO AVE 8	01 TO 05	44.0	Improved	1980	265000.0
2	2012-03	ANG MO KIO	3 ROOM	610	ANG MO KIO AVE 4	06 TO 10	68.0	New Generation	1980	315000.0
3	2012-03	ANG MO KIO	3 ROOM	474	ANG MO KIO AVE 10	01 TO 05	67.0	New Generation	1984	320000.0
4	2012-03	ANG MO KIO	3 ROOM	604	ANG MO KIO AVE 5	06 TO 10	67.0	New Generation	1980	321000.0

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price
0	2015-01	ANG MO KIO	3 ROOM	174	ANG MO KIO AVE 4	07 TO 09	60.0	Improved	1986	255000.0
1	2015-01	ANG MO KIO	3 ROOM	541	ANG MO KIO AVE 10	01 TO 03	68.0	New Generation	1981	275000.0
2	2015-01	ANG MO KIO	3 ROOM	163	ANG MO KIO AVE 4	01 TO 03	69.0	New Generation	1980	285000.0
3	2015-01	ANG MO KIO	3 ROOM	446	ANG MO KIO AVE 10	01 TO 03	68.0	New Generation	1979	290000.0
4	2015-01	ANG MO KIO	3 ROOM	557	ANG MO KIO AVE 10	07 TO 09	68.0	New Generation	1980	290000.0

	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price
0	2017-01	ANG MO KIO	2 ROOM	406	ANG MO KIO AVE 10	10 TO 12	44.0	Improved	1979	232000.0
1	2017-01	ANG MO KIO	3 ROOM	108	ANG MO KIO AVE 4	01 TO 03	67.0	New Generation	1978	250000.0
2	2017-01	ANG MO KIO	3 ROOM	602	ANG MO KIO AVE 5	01 TO 03	67.0	New Generation	1980	262000.0
3	2017-01	ANG MO KIO	3 ROOM	465	ANG MO KIO AVE 10	04 TO 06	68.0	New Generation	1980	265000.0
4	2017-01	ANG MO KIO	3 ROOM	601	ANG MO KIO AVE 5	01 TO 03	67.0	New Generation	1980	265000.0

Source: HDB Dataset

SOLUTION

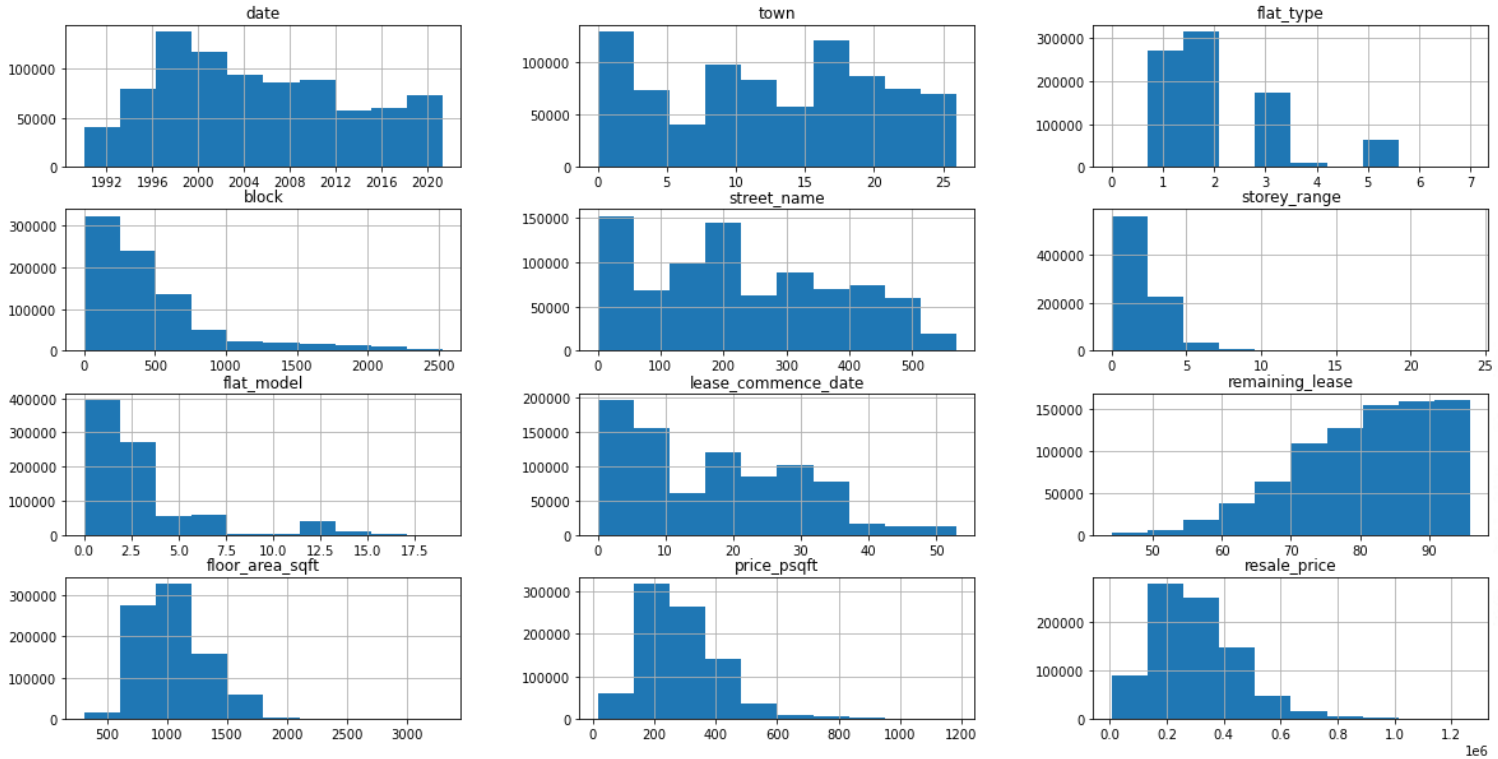
Parties in the market should use the next few slides to have a better understanding of their asset in order to purchase/sell a property at least near the median value of previous transactions

A regression model was also built for homeowners/buyers that proposes a purchase/sale price

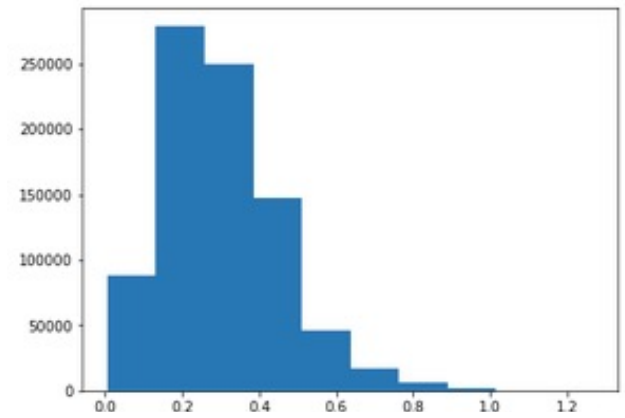
Source: HDB.gov.sg, The Business Times

EDA FINDINGS

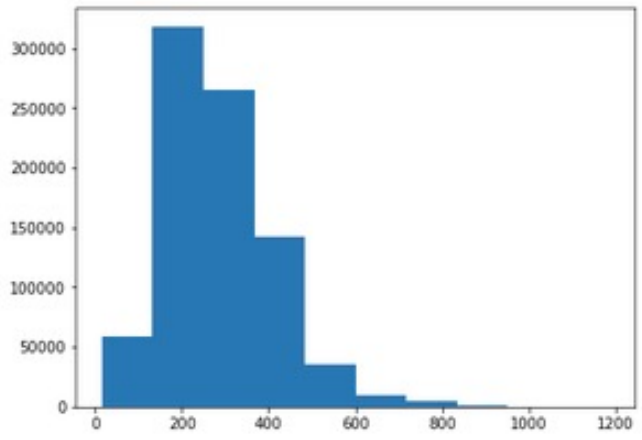
	Mean	Min	Max	Median
Mean Resale Price:	S$ 299,175	S$5,000	S$1.268 M	S$280,000
Mean Resale Price PSF:	S$ 286	S$15	S$1186	S$261



Mean Resale Price



Mean Resale Price PSF

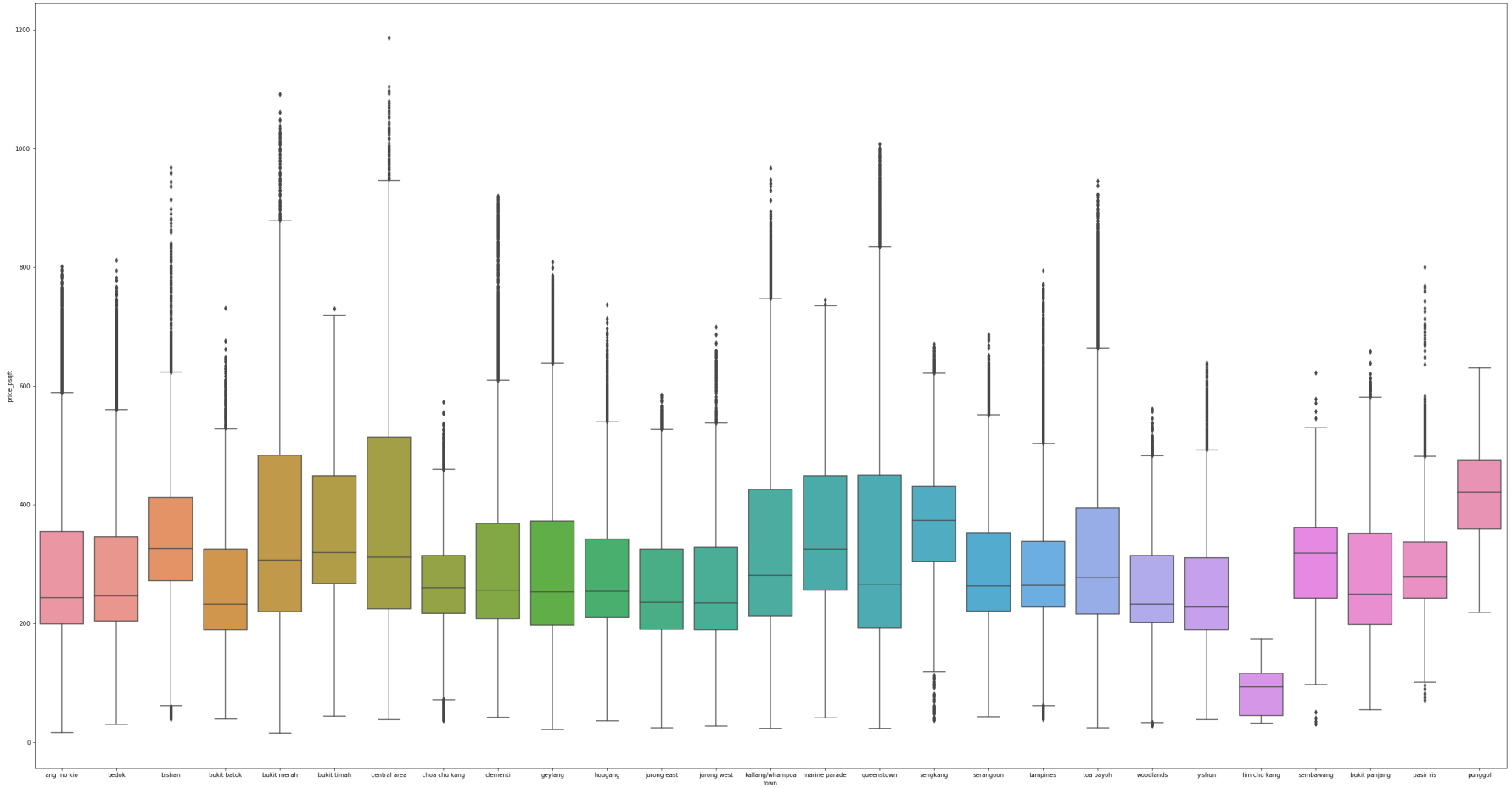


Source: HDB Dataset
Notes: The figures are subject to rounding errors

EDA FINDINGS

- Pungoll has the highest median price per square foot
- Lim Chu Kang has the lowest median price per square foot

Boxplot of Town to Price PSF

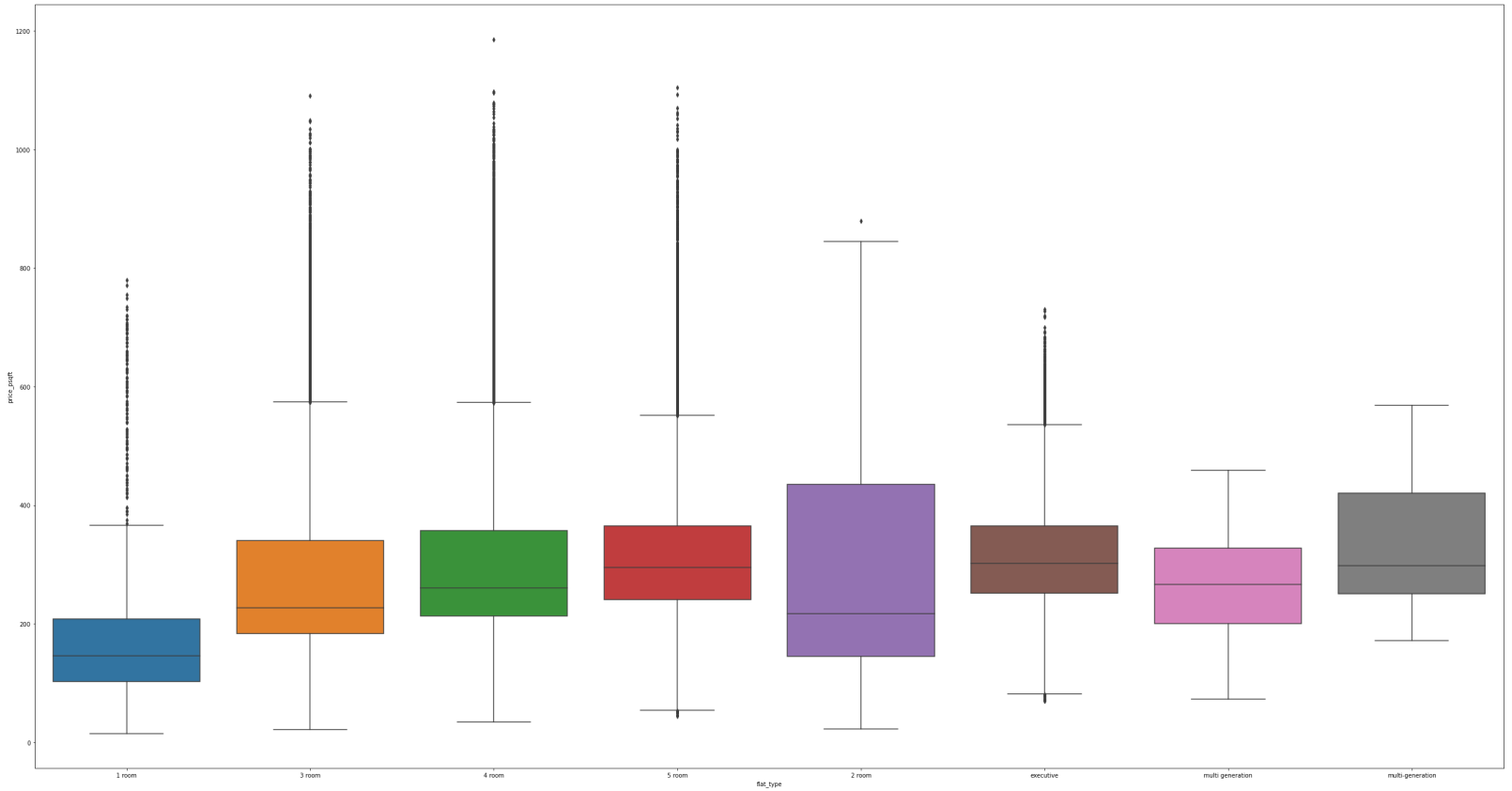


Source: HDB Dataset

EDA FINDINGS

- Multi-generation flats have the highest median price per square foot
- 1-room flats have the lowest median price per square foot

Boxplot of Flat Type to Price PSF

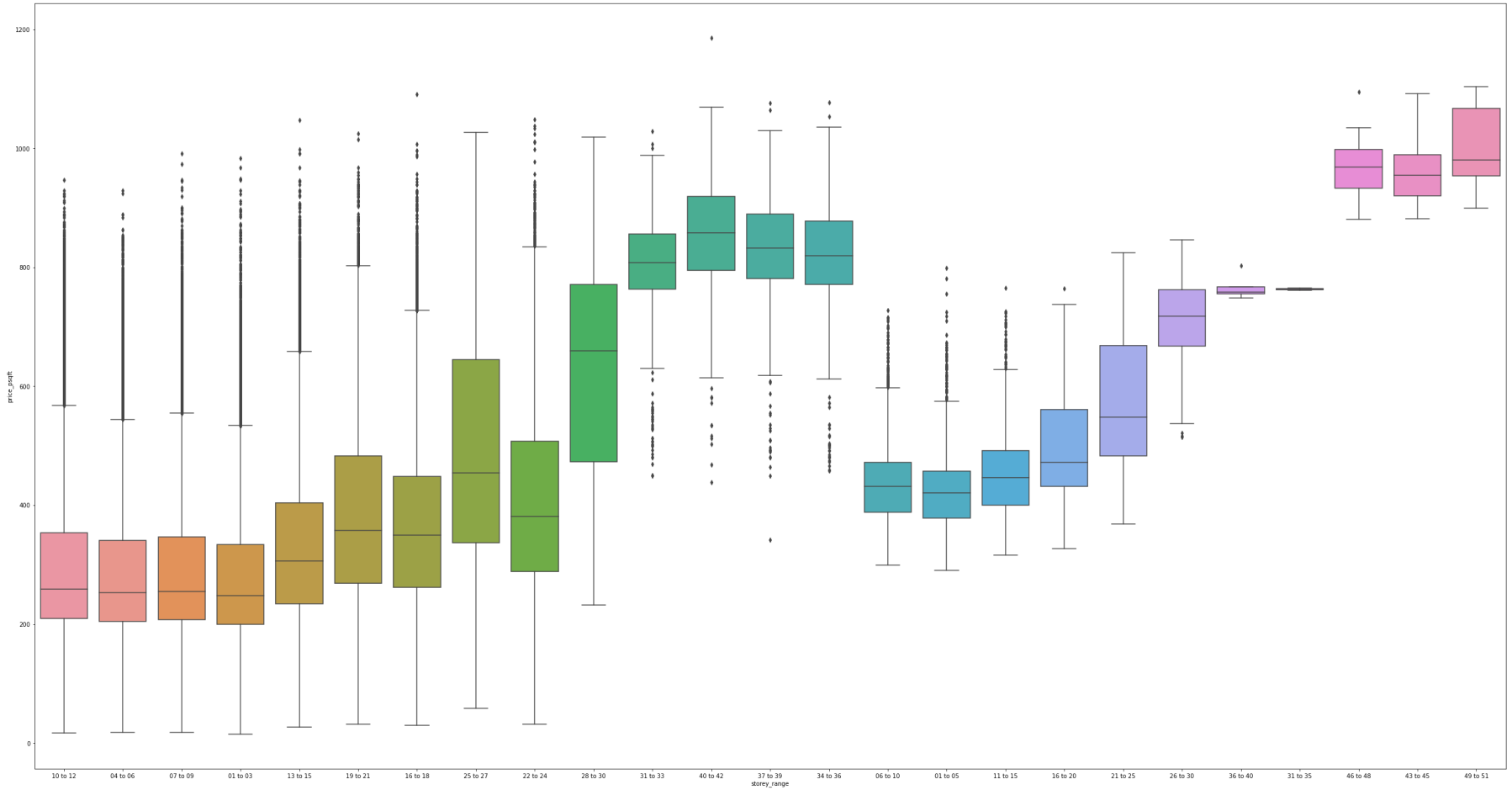


Source: HDB Dataset

EDA FINDINGS

- Higher floors command the highest prices psf

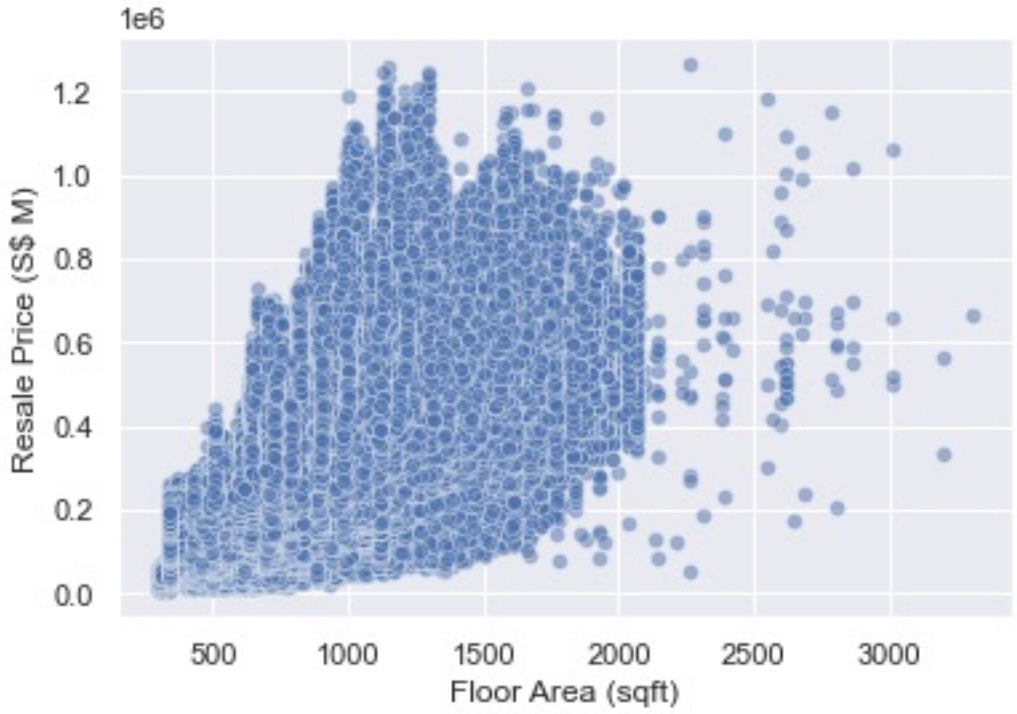
Boxplot of Storey Range to Price PSF



Source: HDB Dataset

MODEL

LINEAR RELATIONSHIP IDENTIFIED

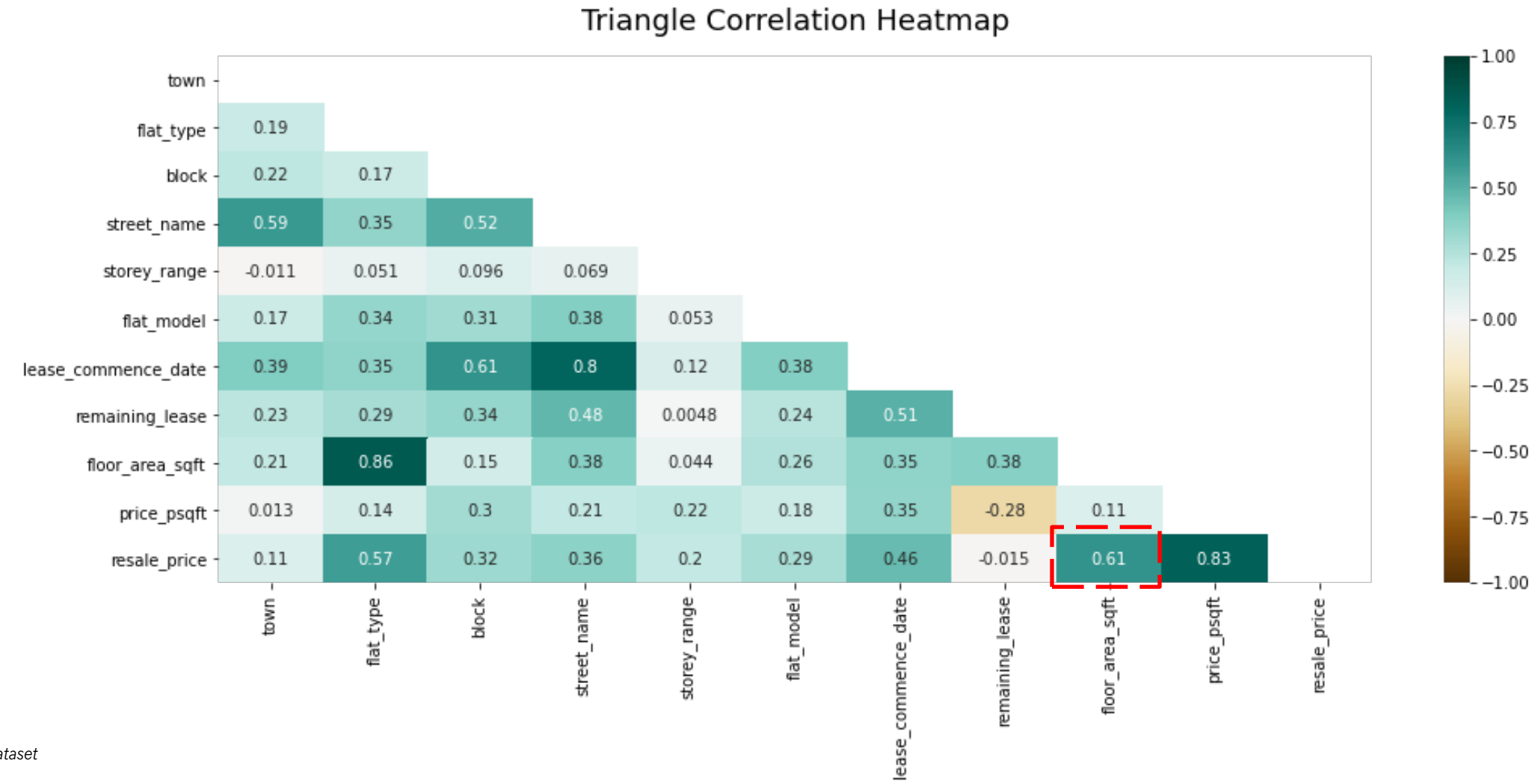


Source: HDB Dataset

FEATURE SELECTION (SIMPLE LR)

Institute of Data

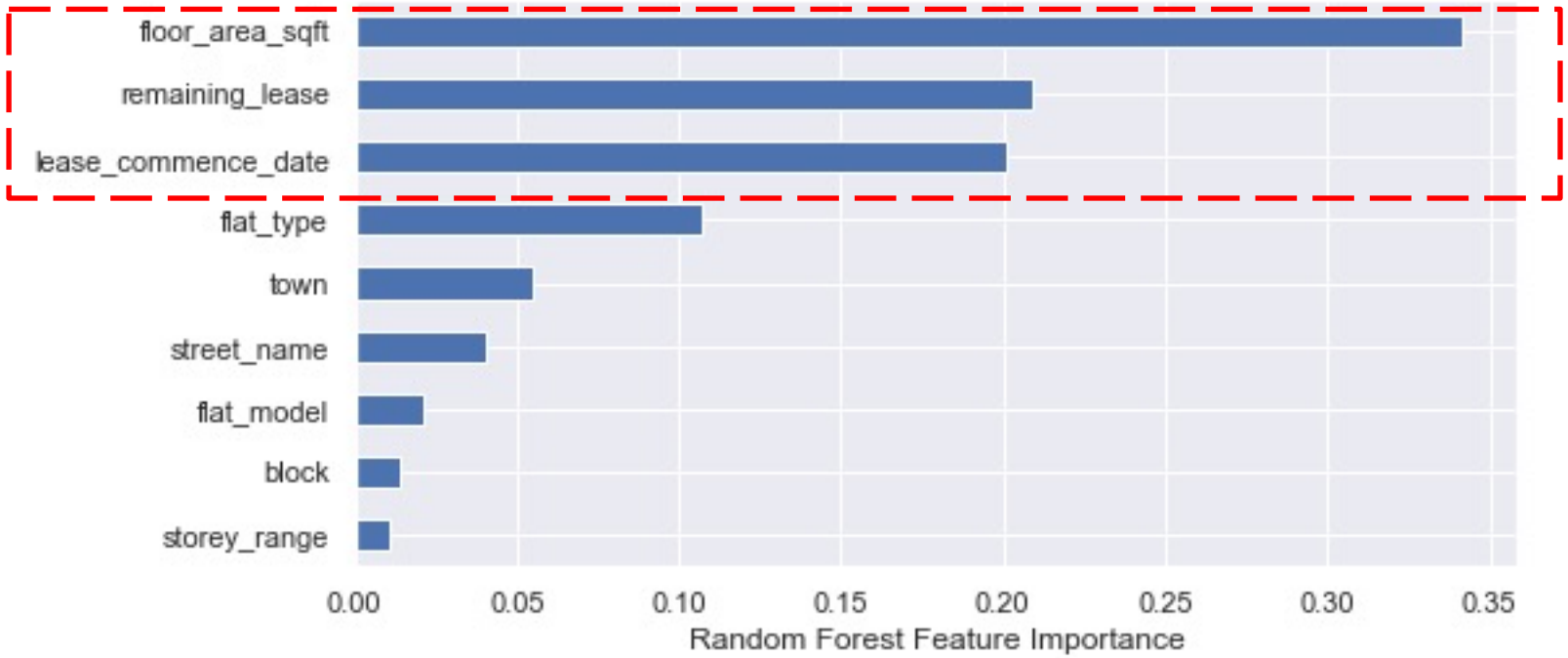
- Floor area by sqft picked as predictor for simple LR



Source: HDB Dataset

MULTI-FEATURE SELECTION

- Top 3 features by RandomForest picked.

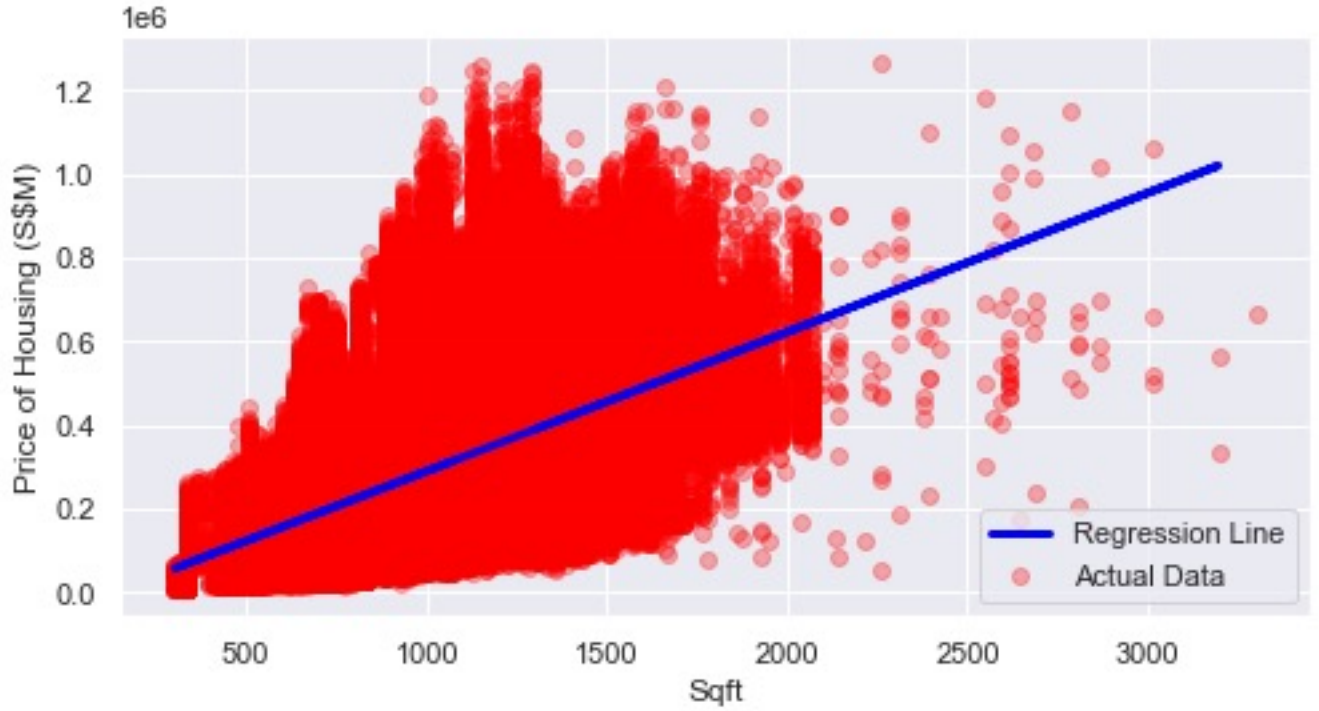


Source: HDB Dataset
Notes: RandomForest was used as it can handle both continuous and categorical data at ease

MODEL SELECTION

Model Type	Features	R2	RMSE (S$)
Simple LR (Not Scaled)	1	0.38	119,710
Simple LR (Scaled)	1	0.38	119,513
Multi LR (Scaled)	9	0.66	88,307
RFR	9	0.97	24,182
RRR	3	0.89	49,880

SIMPLE LINEAR REGRESSION (NOT SCALED)



Source: HDB Dataset

PREDICTION TESTING

Predictor Features	Characteristic
Town	Tampines
Flat Type	Executive
Block	942
Street Name	Tampines Ave 5
Storey Range	10 to 12
Flat Model	Maisonette
Lease Commence Date	1992
Remaining Lease	66
Floor Area Sqft	1,657

date	town	flat_type	block	street_name	storey_range	flat_model	lease_commence_date	remaining_lease	floor_area_sqft	price_psqft	resale_price
2020-09-01	18	5	944	196	0	7	20	66	1657.6406	446.417637	740000.0

```
In [132]: 1 RFR.predict(x_pred) #- predicted price is SGD 850k. 14% higher than actual value of S$740k

[Parallel(n_jobs=8)]: Using backend ThreadingBackend with 8 concurrent workers.
[Parallel(n_jobs=8)]: Done   2 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done   9 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done  16 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done  25 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done  34 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done  45 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done  56 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done  69 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done  82 tasks      | elapsed:    0.0s
[Parallel(n_jobs=8)]: Done  96 out of 100 | elapsed:    0.0s remaining:    0.0s
[Parallel(n_jobs=8)]: Done 100 out of 100 | elapsed:    0.0s finished
Out[132]: array([850383.76])

In [134]: 1 850/740
Out[134]: 1.1486486486486487
```

Source: HDB Dataset

IMPROVEMENTS

IMPROVEMENTS FOR FUTURE

- More work needs to be done on feature engineering. Price psf imputed from target variable too simplistic
- Could include geospatial data as a new feature and identify proximity to CBD, public transport, schools as these could possible increase asset valuations

Thank you