



Case Study



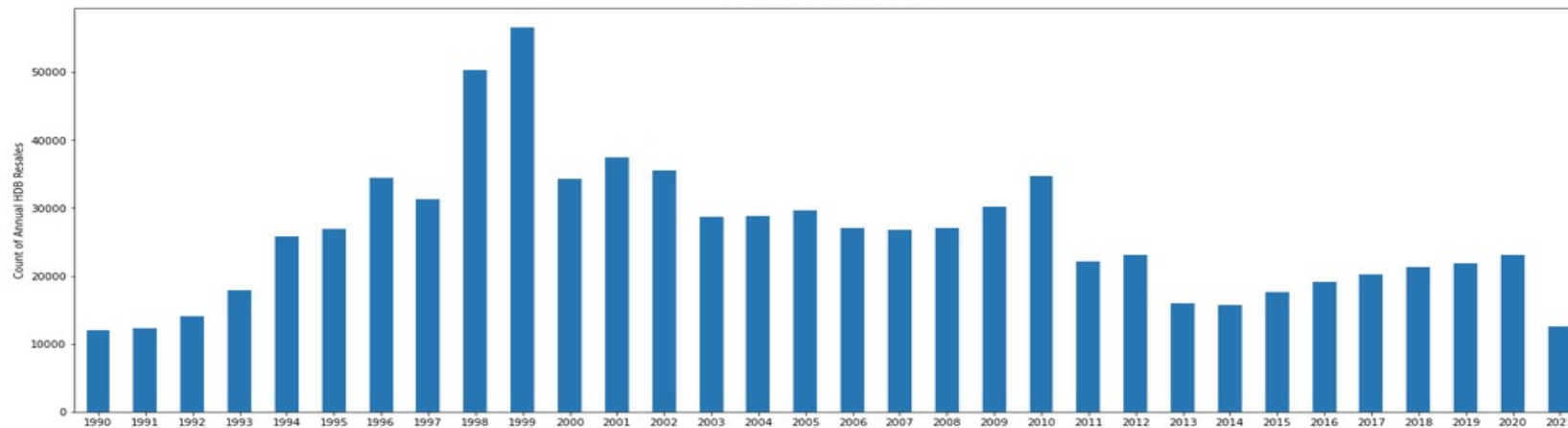
**HOUSING &
DEVELOPMENT
BOARD**

HDB Resale Price Estimator

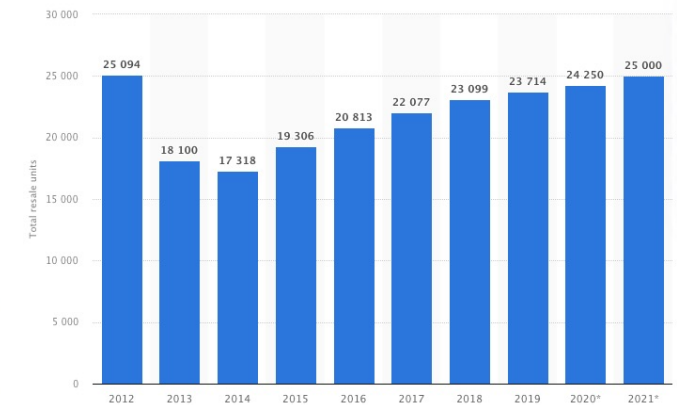
BACKGROUND INFORMATION – RESALE VOLUME

- 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)



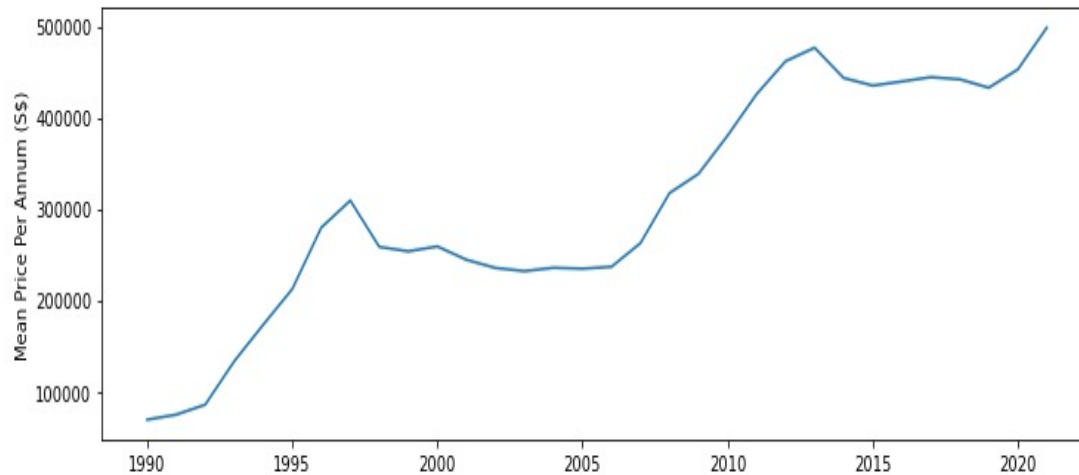
HDB Resale Estimates



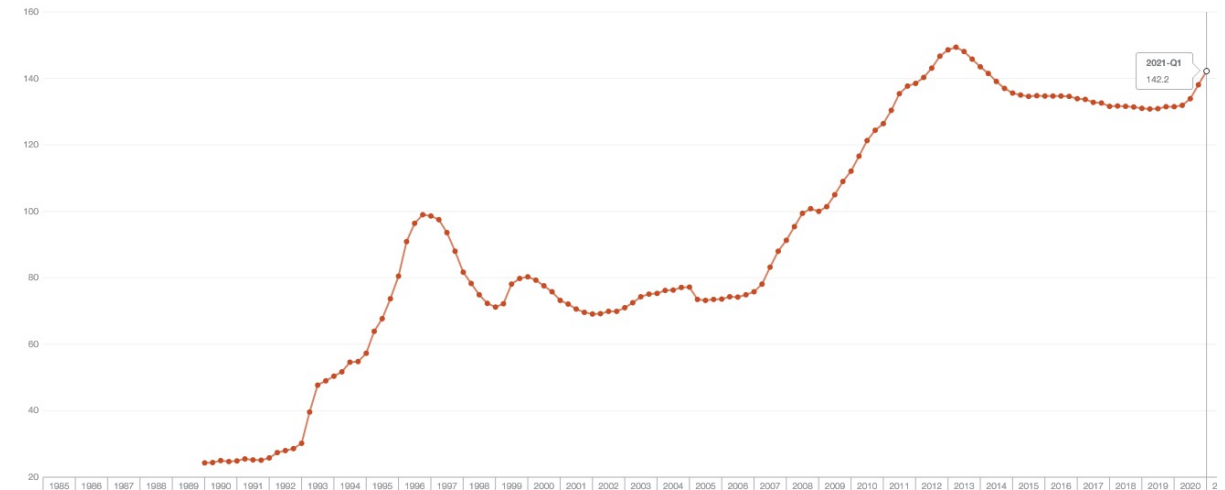
BACKGROUND INFORMATION – RESALE PRICE

- 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



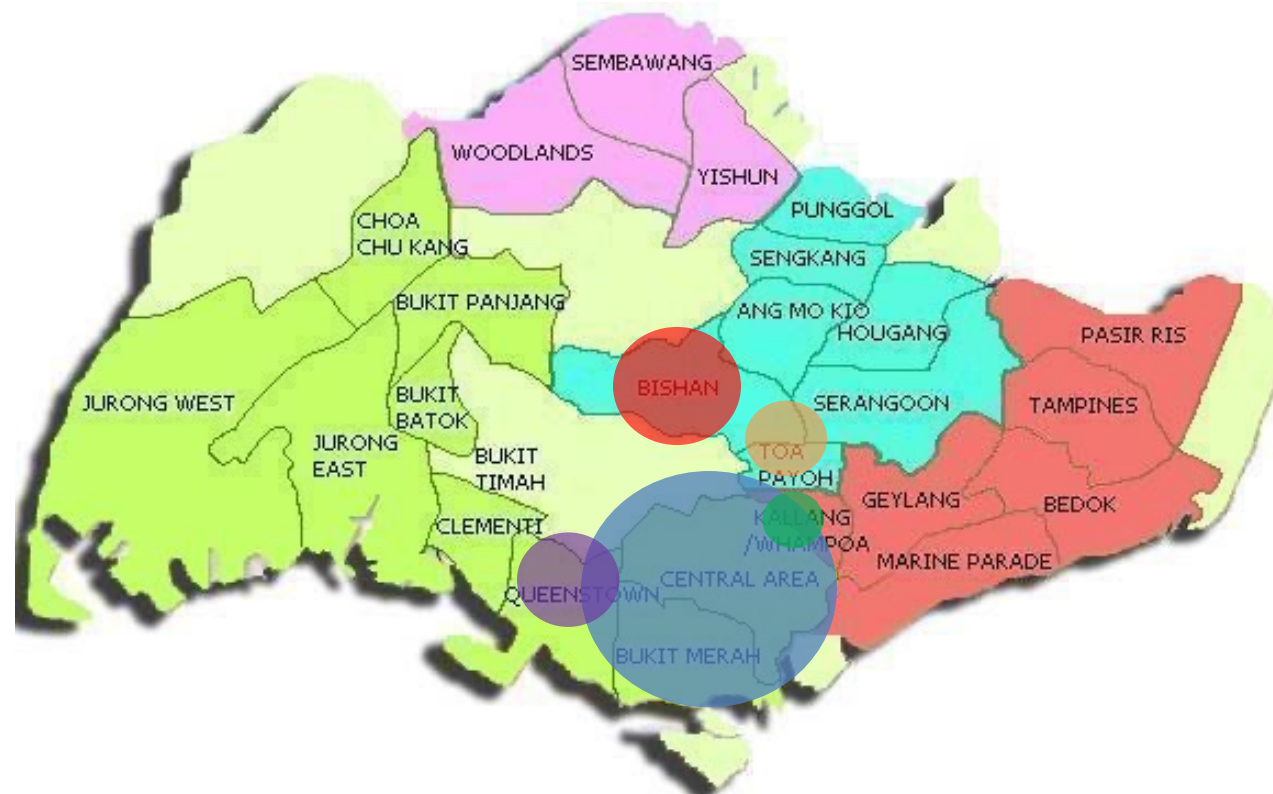
HDB Resale Price Index



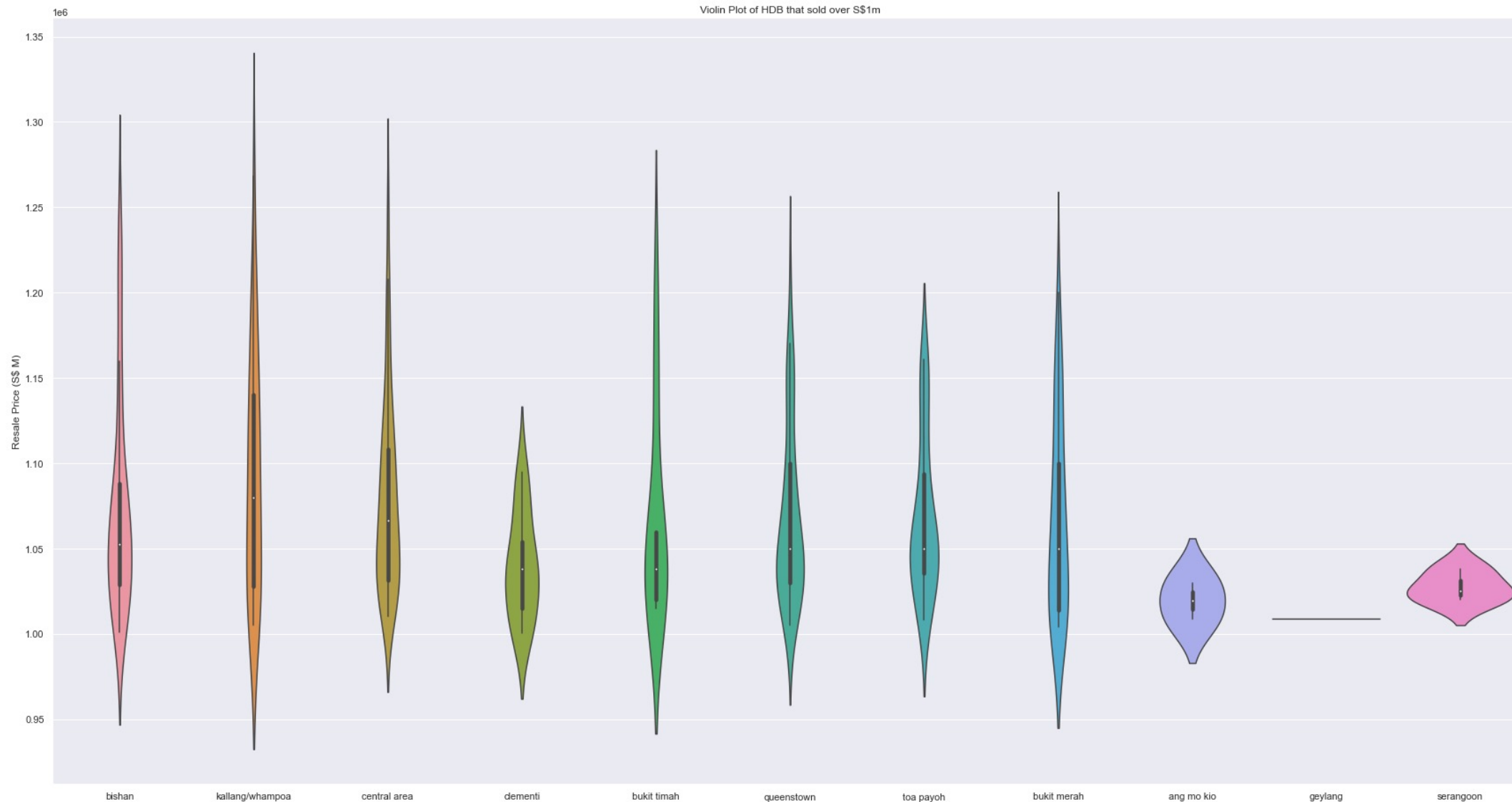
BACKGROUND INFO – MOST EXPENSIVE NEIGHBOURHOODS

TOWNS WITH THE MOST SALES OVER \$1M

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

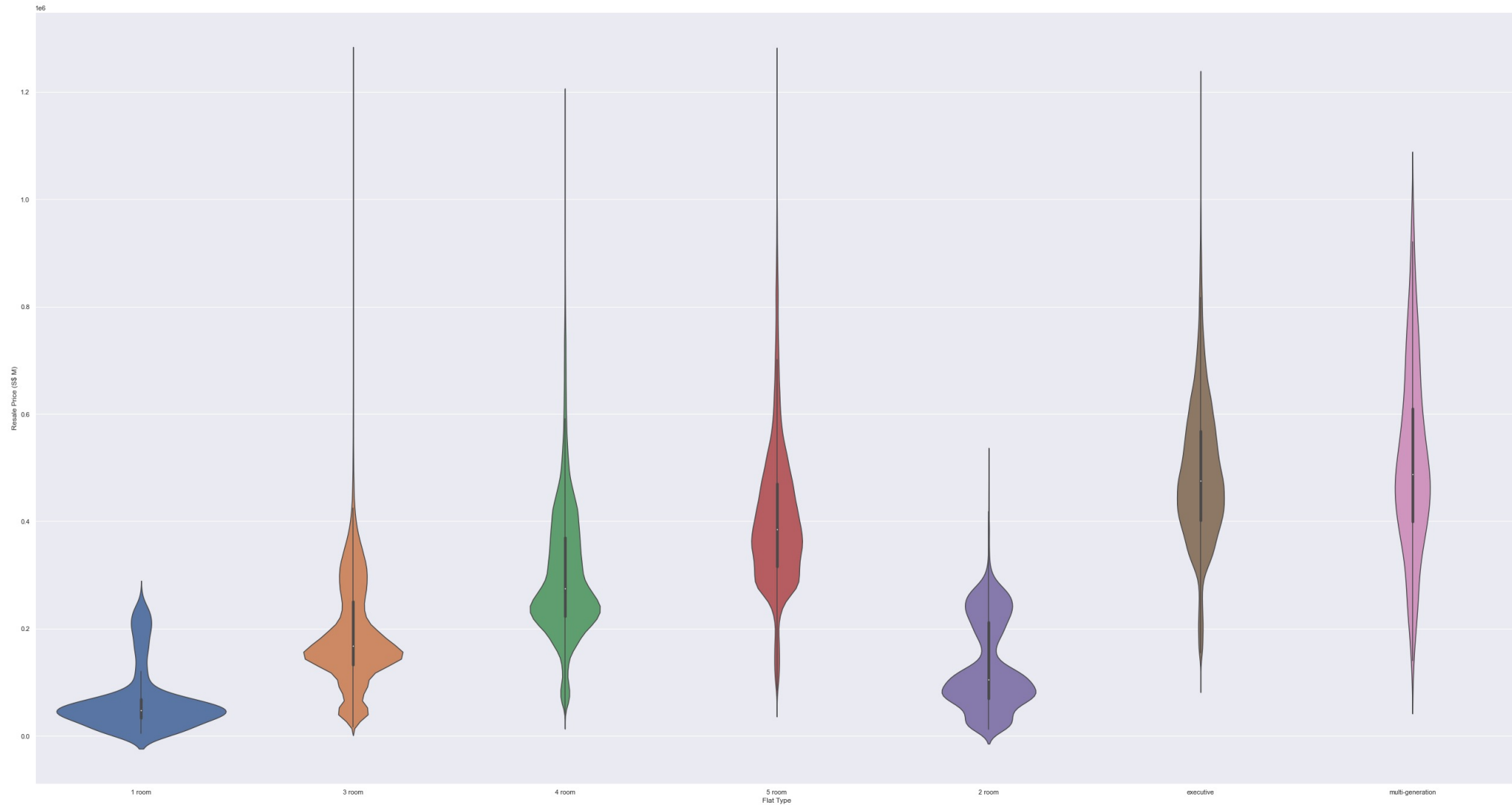


BACKGROUND INFO – MOST EXPENSIVE NEIGHBOURHOODS



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

BACKGROUND INFO – MOST EXPENSIVE FLAT TYPE



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.



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.

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

```
1 for idx,df in enumerate(dflist):  
2     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	8000
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

WHAT ARE THE KEY FACTORS THAT IMPACT RESALE PRICE, AND HOW MUCH SHOULD A STAKEHOLDER BUY/SELL THEIR ASSET FOR

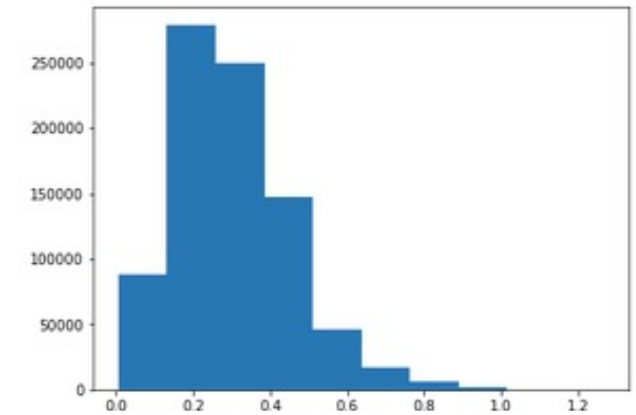
THE DATASET – VALUE COUNTS

TOWN V COUNTS ARE =====	FLAT TYPE V COUNTS ARE =====		FLAT MODEL V COUNTS ARE =====	STORY RANGE V COUNTS ARE =====	
tampines	73779	4 room	314371	model a	232339 04 to 06 210886
yishun	63621	3 room	270536	improved	218654 07 to 09 190110
bedok	61390	5 room	174358	new generation	176515 01 to 03 170456
jurong west	60808	executive	63729	simplified	53893 10 to 12 161366
woodlands	58882	2 room	9896	standard	39888 13 to 15 53591
ang mo kio	47951	1 room	1106	premium apartment	37151 16 to 18 20203
hougang	46038	multi-generation	512	apartment	32535 19 to 21 9726
bukit batok	40429	Name: flat_type, dtype: int64		maisonette	27314 22 to 24 6302
choa chu kang	33964			model a2	9131 25 to 27 2696
bukit merah	30752			dbss	2106 01 to 05 2689
pasir ris	30610			model a-maisonette	1919 06 to 10 2469
toa payoh	28537			adjoined flat	1125 11 to 15 1254
queenstown	26177			terrace	657 28 to 30 1142
geylang	25810			multi generation	512 31 to 33 334
clementi	25737			type s1	311 34 to 36 329
sengkang	25504			type s2	153 37 to 39 320
bukit panjang	24750			improved-maisonette	114 16 to 20 265
kallang/whampoa	24196			premium maisonette	83 40 to 42 157
jurong east	22979			premium apartment loft	65 21 to 25 92
serangoon	21307			2-room	43 26 to 30 39
bishan	19823			Name: flat_model, dtype: int64	
punggol	13990				43 to 45 32
sembawang	11223				46 to 48 30
marine parade	7420				49 to 51 11
central area	6451				36 to 40 7
bukit timah	2318				31 to 35 2
lim chu kang	62			Name: storey_range, dtype: int64	
Name: town, dtype: int64					

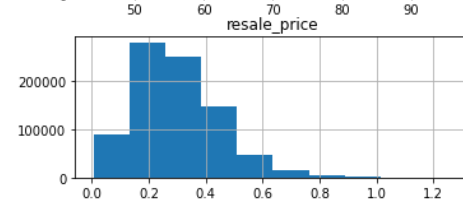
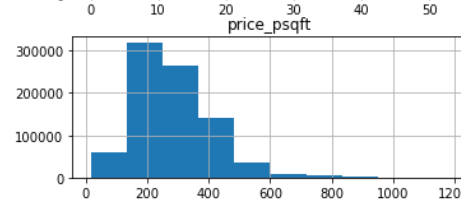
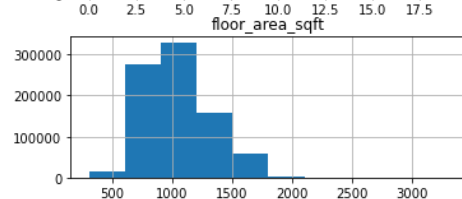
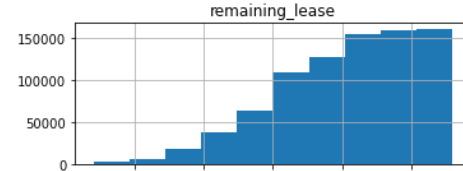
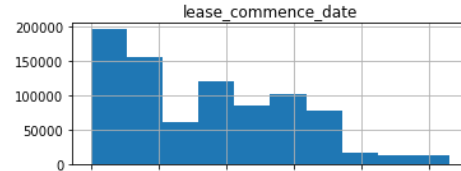
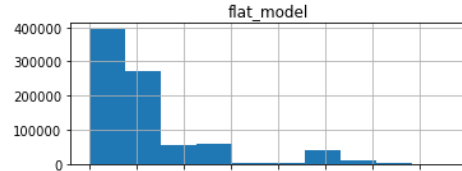
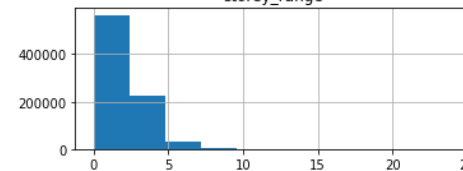
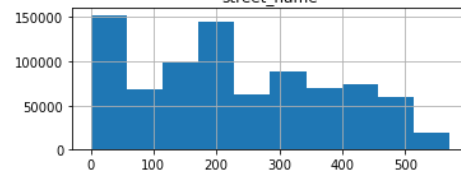
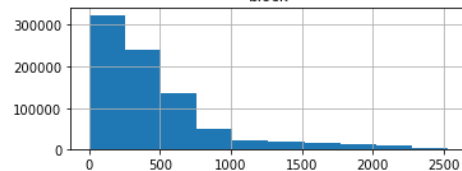
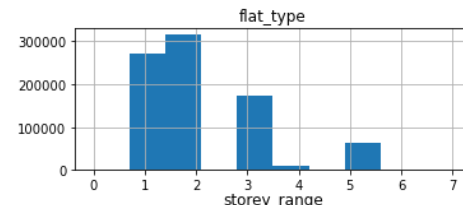
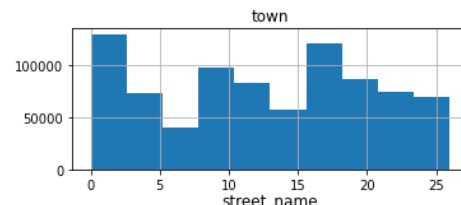
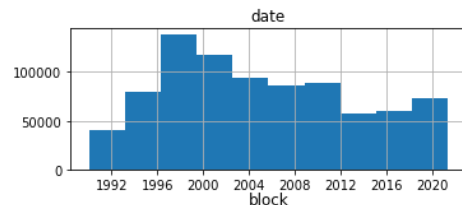
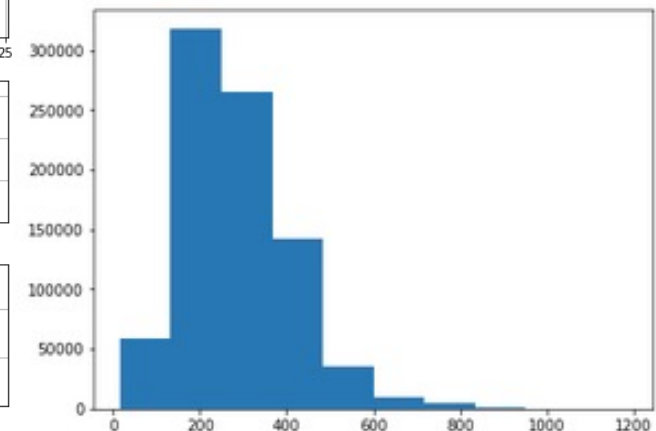
EDA FINDINGS – MEAN/MEDIAN PRICES

	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

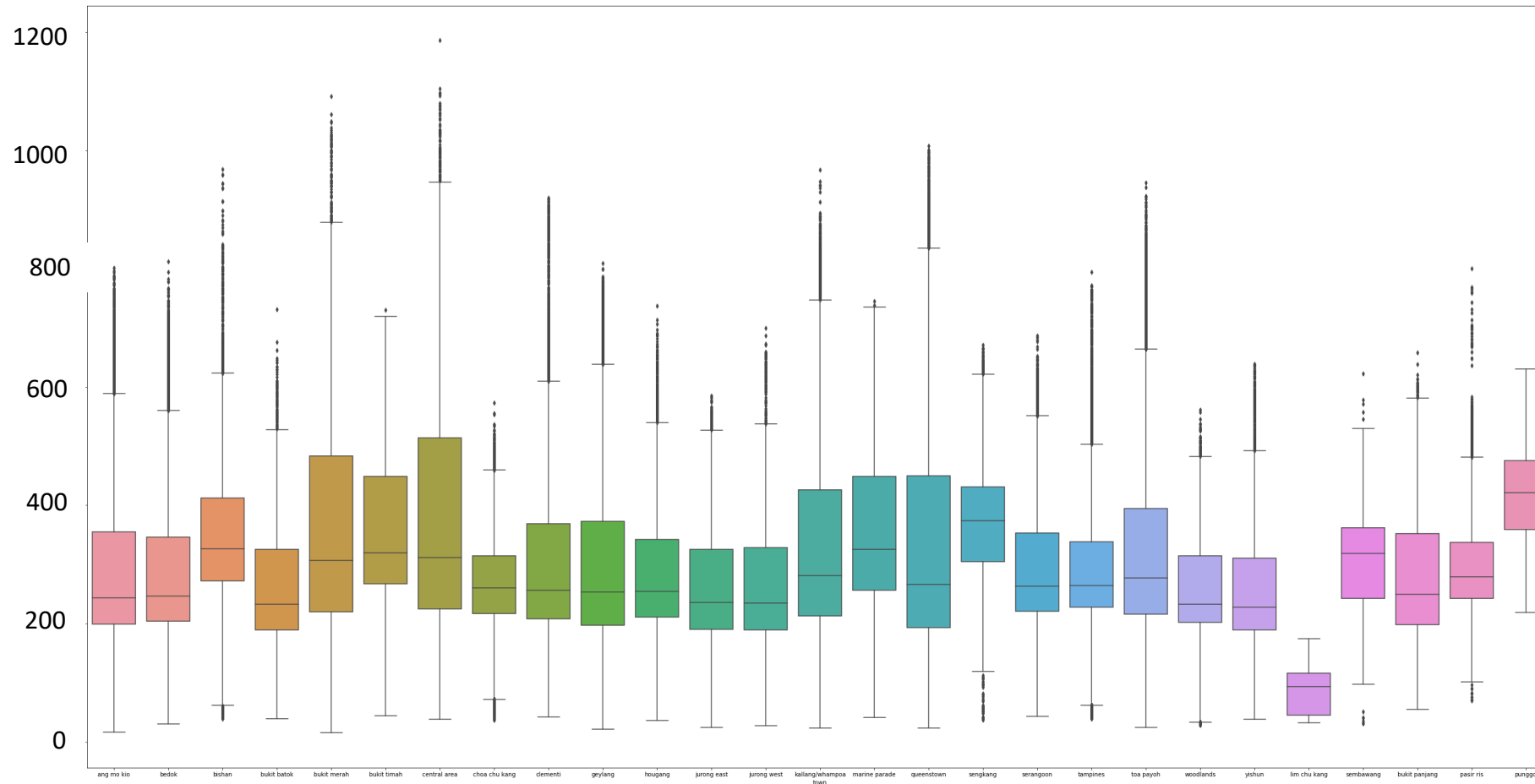


Mean Resale Price PSF



EDA FINDINGS – BOXPLOT OF TOWN PRICE TO PSF

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



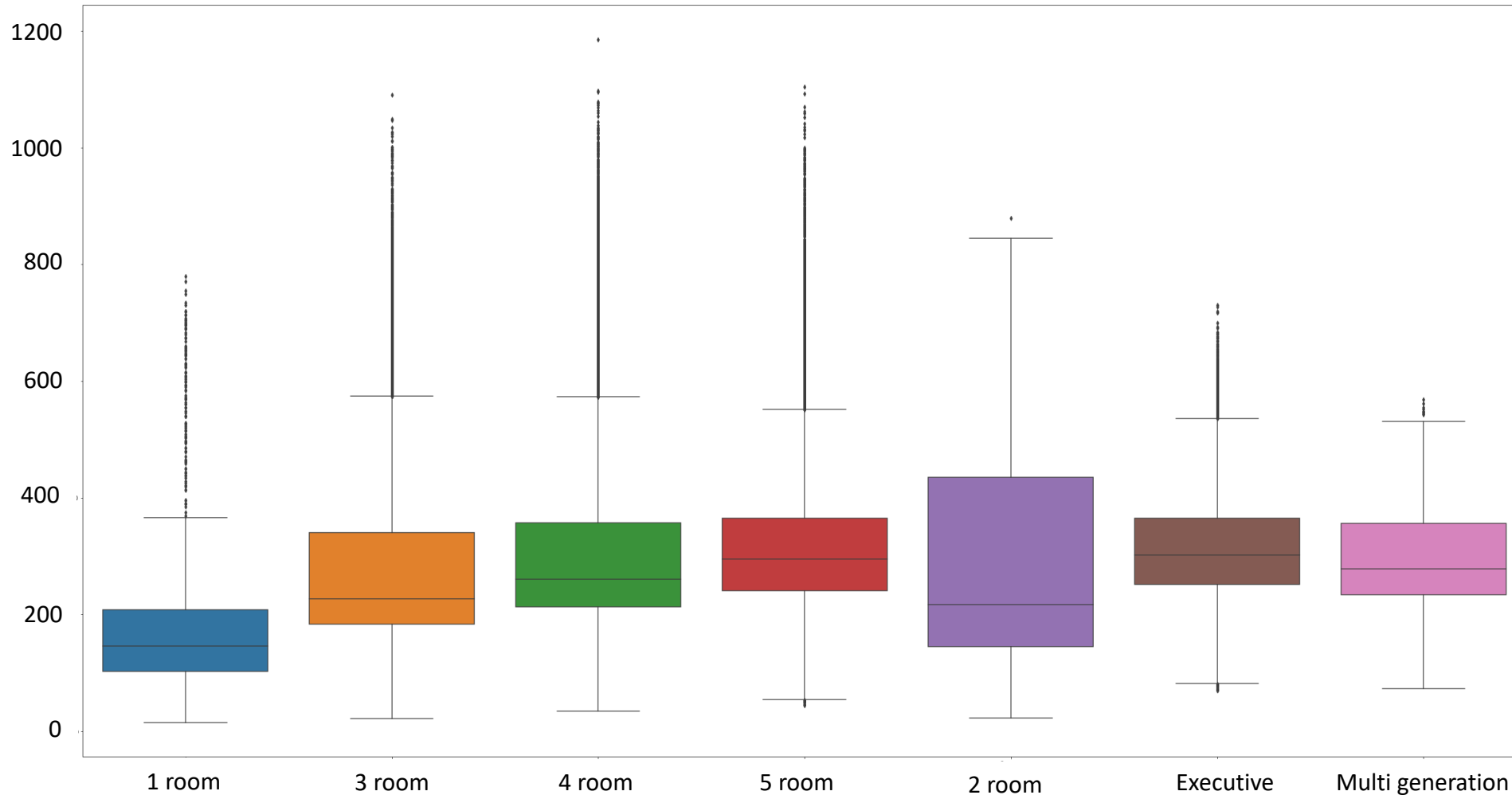
EDA FINDINGS

DESCRIPTIVE STATS OF TOWN PRICE TO PSF

	count	mean	std	min	25%	50%	75%	max
town								
ang mo kio	47951.0	271.415372	127.511440	16.782501	199.468485	243.633718	355.217850	801.416089
bedok	61390.0	270.671335	115.428313	29.861720	203.763628	246.817270	346.367104	811.567572
bishan	19823.0	348.661720	121.607569	39.620452	272.046908	326.612566	412.534136	967.740937
bukit batok	40429.0	253.636062	98.860716	38.709637	189.677224	232.257825	325.160955	730.837955
bukit merah	30752.0	353.830524	186.087902	14.984376	219.851858	306.580329	483.389548	1090.862558
bukit panjang	24750.0	278.546858	97.600924	54.723761	198.568026	249.533200	351.704706	658.489217
bukit timah	2318.0	362.174454	144.263489	44.542597	267.122770	319.577834	448.925741	729.953164
central area	6451.0	377.780620	223.771083	38.442674	224.963556	311.889079	513.699659	1185.651721
choa chu kang	33964.0	268.745968	70.071227	37.663431	217.256966	260.433364	314.101058	573.232078
clementi	25737.0	289.846793	141.481931	42.514992	207.992082	256.523568	368.663214	919.353890
geylang	25810.0	285.368103	139.132853	21.419333	196.612649	253.824623	373.197243	809.156293
hougang	46038.0	273.831598	97.028910	36.216474	210.580428	254.031996	342.322418	737.325691
jurong east	22979.0	255.481674	105.602114	24.846186	190.570523	235.575794	325.160955	584.515526
jurong west	60808.0	255.496878	98.961793	27.381975	188.987587	234.346821	328.591784	699.383113
kallang/whampoa	24196.0	318.653306	157.367279	23.448980	212.551100	280.721758	426.261420	967.145404
lim chu kang	62.0	87.666483	41.612729	31.852502	45.191861	93.026851	116.626608	174.613112
marine parade	7420.0	347.211463	148.491276	41.449089	256.664579	325.160955	448.702256	744.813127
pasir ris	30610.0	292.519707	68.144850	69.995509	242.219484	278.709390	337.715432	800.396196
punggol	13990.0	416.171462	77.907281	218.985949	358.910573	420.967307	475.771192	830.414096
queenstown	26177.0	327.387160	196.675305	22.889177	192.952654	266.591590	450.222860	1007.383336
sembawang	11223.0	310.634155	69.958056	30.919323	242.236309	318.718767	361.833243	622.450970
sengkang	25504.0	369.299550	91.863551	37.387845	304.515815	374.146241	431.483426	670.967049
serangoon	21307.0	289.978932	111.280795	43.548342	220.417230	263.225535	352.853234	686.612194
tampines	73779.0	281.602178	96.925947	38.709637	227.612668	263.967736	337.993345	794.100563
toa payoh	28537.0	309.210445	151.828908	24.265743	215.966347	277.322776	394.838302	945.330094
woodlands	58882.0	254.351011	77.670609	28.109152	201.732511	232.972464	314.231175	561.578621
yishun	63621.0	249.343347	92.116404	38.165610	189.688779	227.663714	310.868165	638.709018

EDA FINDINGS – BOXPLOT OF FLAT TYPE TO PSF

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



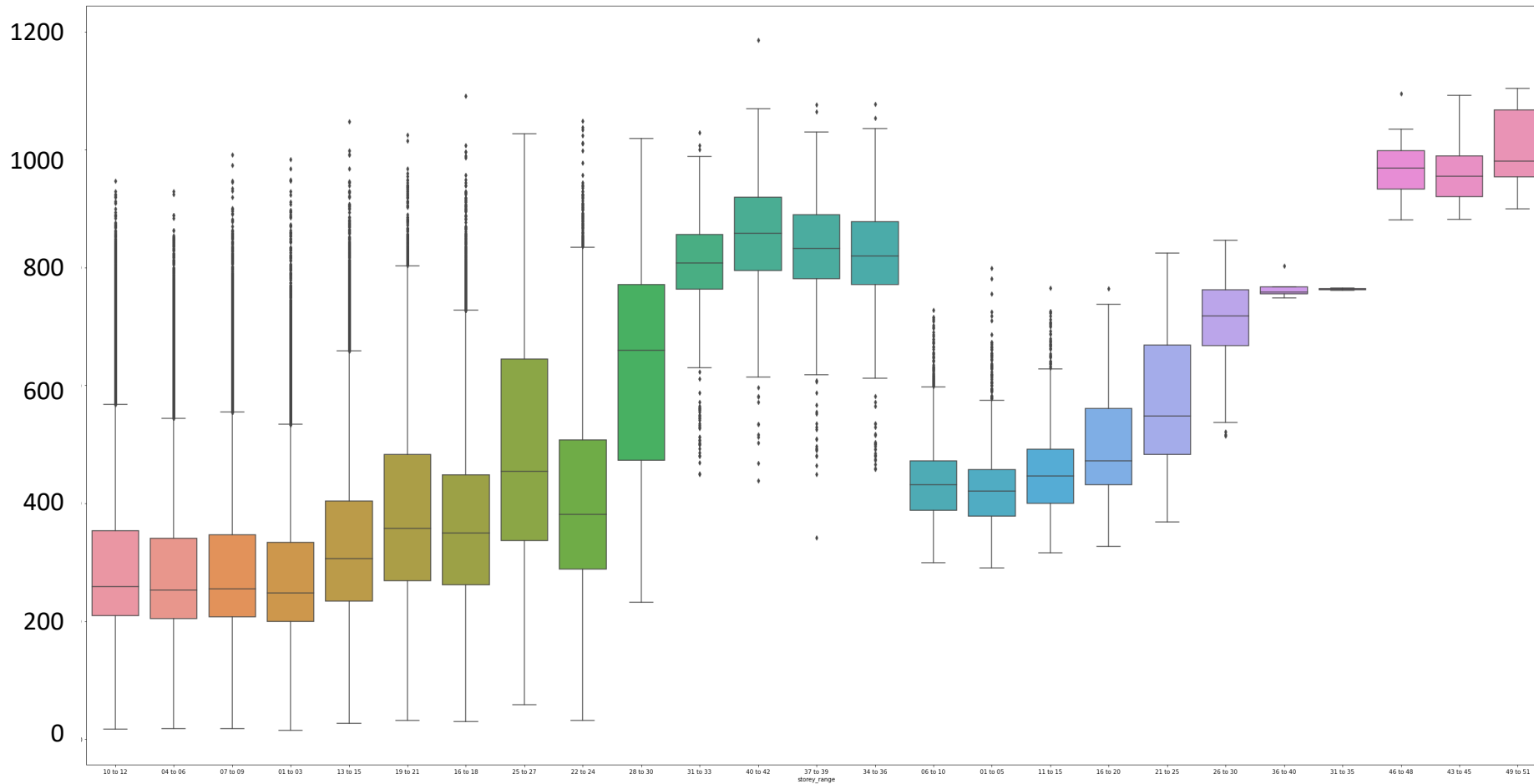
EDA FINDINGS

DESCRIPTIVE STATS OF FLAT TYPE TO PSF

	count	mean	std	min	25%	50%	75%	max
flat_type								
1 room	1106.0	200.129753	175.308593	14.984376	101.968677	146.304345	208.231153	779.187541
2 room	9896.0	271.477473	169.317312	22.296751	144.928883	216.773970	434.865714	879.614741
3 room	270536.0	258.432115	128.123976	21.419333	183.726339	226.530919	339.884271	1090.862558
4 room	314371.0	291.936770	121.529407	34.310815	212.722120	260.128764	356.832476	1185.651721
5 room	174358.0	312.360602	109.463608	44.239586	240.319667	294.537330	364.855928	1104.220058
executive	63729.0	310.131710	84.963760	69.995509	251.612643	301.935172	365.418978	729.953164
multi-generation	512.0	294.039436	98.785008	73.180678	233.555355	277.874381	355.995311	568.511690

EDA FINDINGS – BOXPLOT OF STOREY RANGE TO PSF

➤ Higher floors command the highest prices psf



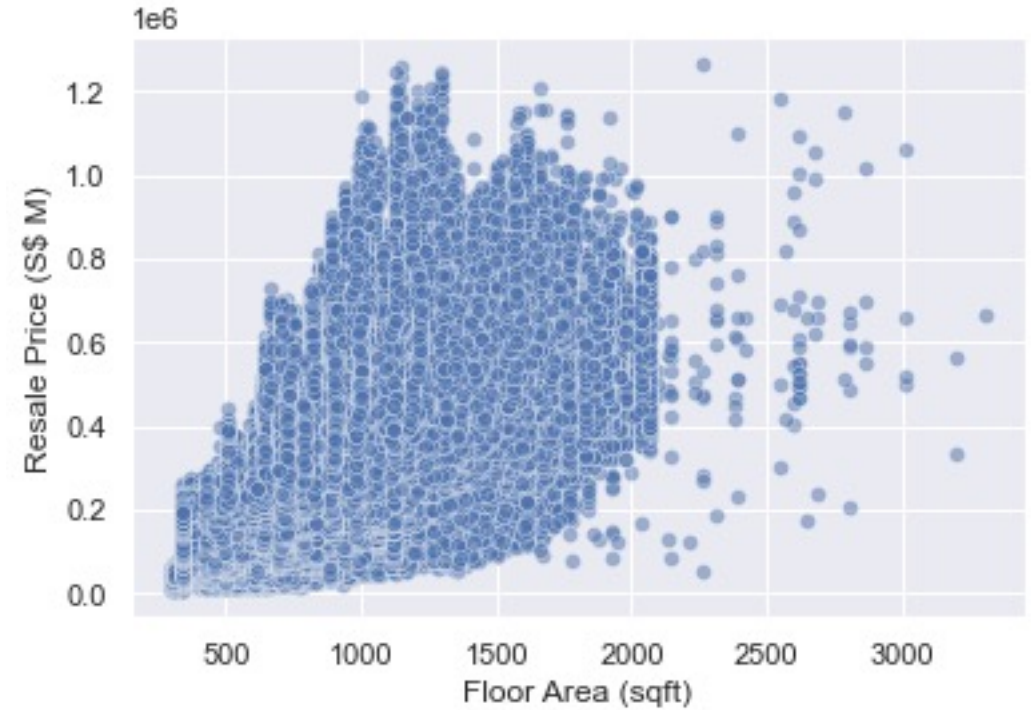
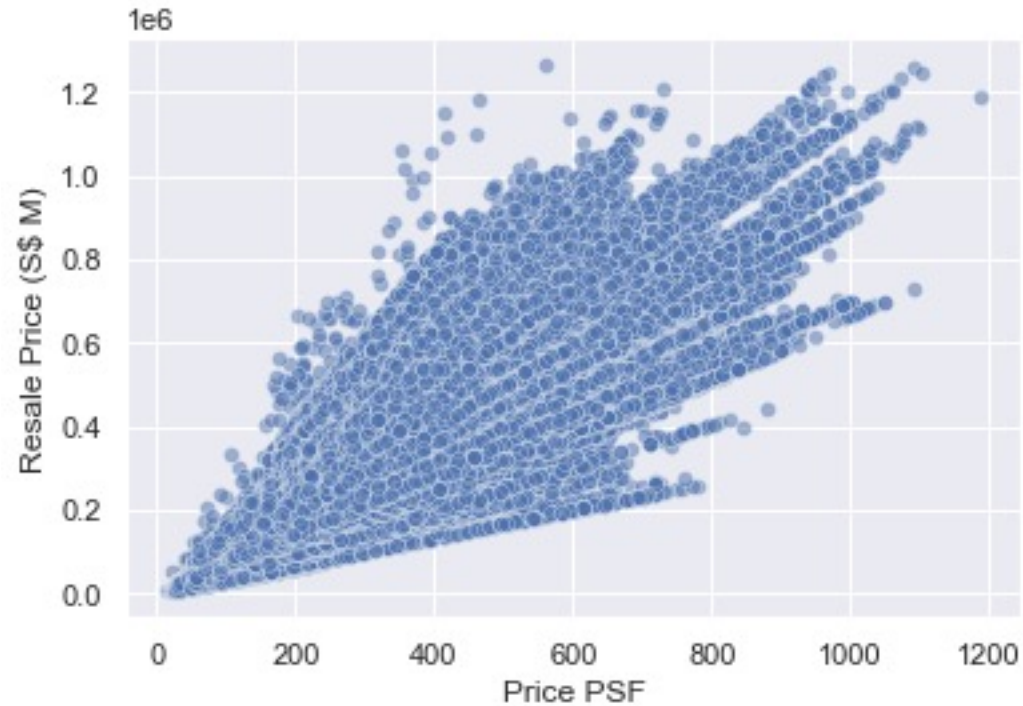
EDA FINDINGS

DESCRIPTIVE STATS OF STOREY RANGE TO PSF

	count	mean	std	min	25%	50%	75%	max
storey_range								
01 to 03	170456.0	267.272150	104.953013	14.984376	200.099049	248.337213	333.606694	983.101894
01 to 05	2689.0	425.816853	64.512489	290.794350	378.612071	420.276064	457.582580	798.966917
04 to 06	210886.0	273.881826	109.117211	17.981251	204.637975	252.580384	340.644810	929.031299
06 to 10	2469.0	438.908650	68.382891	299.179571	388.503998	431.597218	472.257577	727.741184
07 to 09	190110.0	278.201077	113.112535	17.981251	207.483657	255.136954	346.600138	990.966719
10 to 12	161366.0	282.684336	116.155277	17.381876	209.781261	258.386830	353.159158	947.207998
11 to 15	1254.0	455.154516	74.609620	316.029450	400.357679	446.037115	492.028153	765.084599
13 to 15	53591.0	323.899082	131.789938	27.265049	233.953137	306.033840	403.926652	1047.407868
16 to 18	20203.0	366.418211	149.487782	29.543195	262.025164	349.433507	448.198944	1090.862558
16 to 20	265.0	500.647628	96.033212	326.735584	431.999554	471.662044	561.267173	764.283082
19 to 21	9726.0	386.949938	171.805314	32.158776	268.710343	357.905500	482.500729	1025.137985
21 to 25	92.0	573.368301	111.340354	368.615644	483.045230	548.188408	668.466515	824.798519
22 to 24	6302.0	412.015268	180.876597	32.158776	288.835282	380.902833	507.454071	1048.906305
25 to 27	2696.0	482.661001	200.484775	58.738753	336.773846	454.579202	644.536275	1026.429742
26 to 30	39.0	701.829574	87.254901	515.006481	667.152187	717.418614	762.225374	846.617232
28 to 30	1142.0	629.338086	175.906138	232.257825	472.904323	659.895739	771.095978	1018.937554
31 to 33	334.0	792.108691	111.805427	449.368400	762.862101	808.101393	855.646424	1028.880457
31 to 35	2.0	763.210015	2.291597	761.589611	762.399813	763.210015	764.020217	764.830418
34 to 36	329.0	808.349914	121.245272	458.064043	770.898312	819.145877	877.985623	1076.877140
36 to 40	7.0	764.977829	17.908339	748.386324	755.554317	758.708894	766.978953	802.683042
37 to 39	320.0	820.865697	113.752207	341.438854	780.779516	832.138829	889.498052	1075.720452
40 to 42	157.0	843.508617	125.786341	438.709225	794.837889	858.344135	919.353890	1185.651721
43 to 45	32.0	967.186115	54.077443	882.080255	919.790601	954.517236	989.468281	1092.262967
46 to 48	30.0	970.429733	50.256269	880.644252	933.290633	968.895982	997.942961	1095.279005
49 to 51	11.0	1001.470104	70.834398	899.381364	954.233744	980.349218	1067.044687	1104.220058

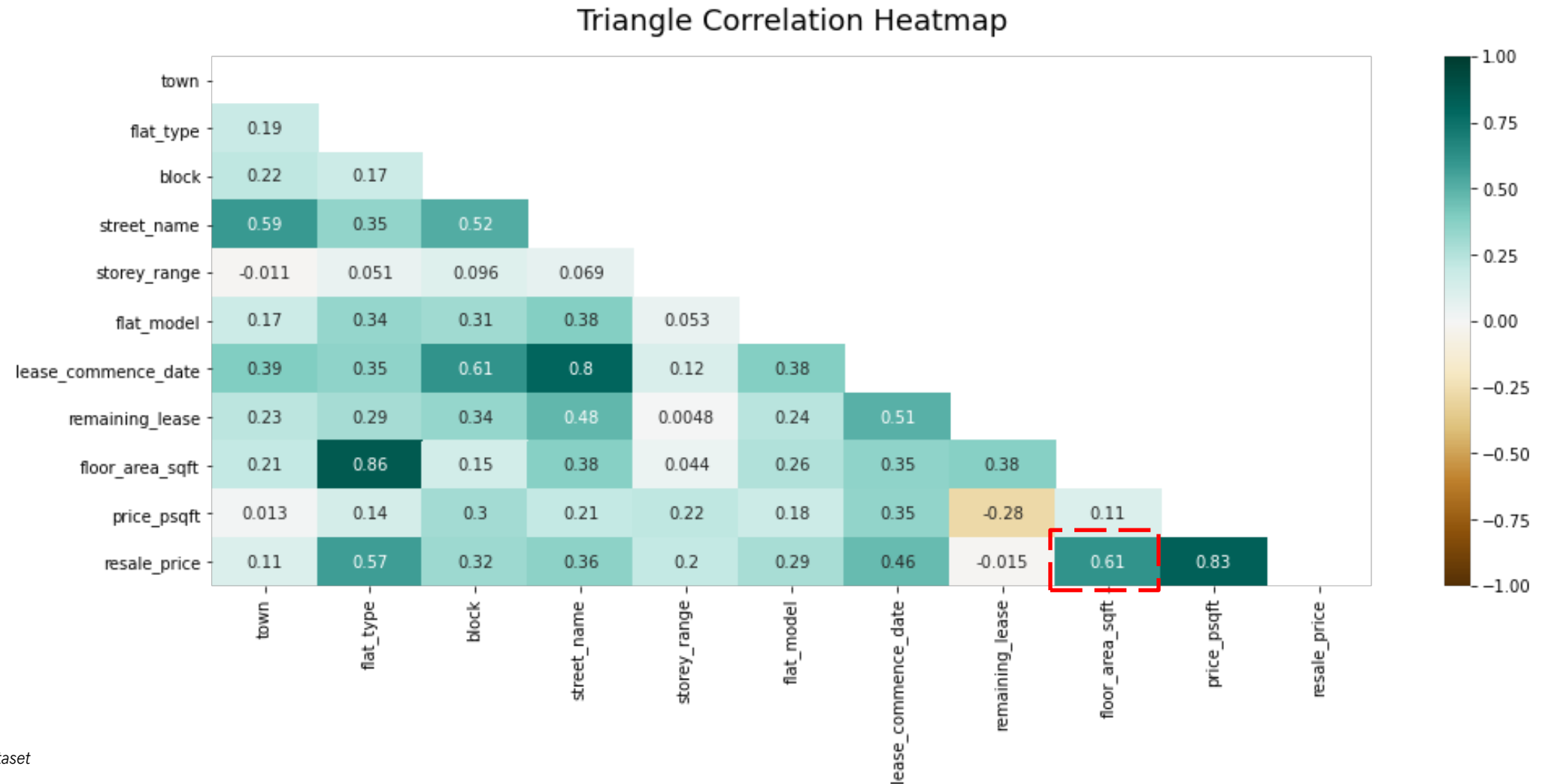
MODEL

LINEAR RELATIONSHIP IDENTIFIED



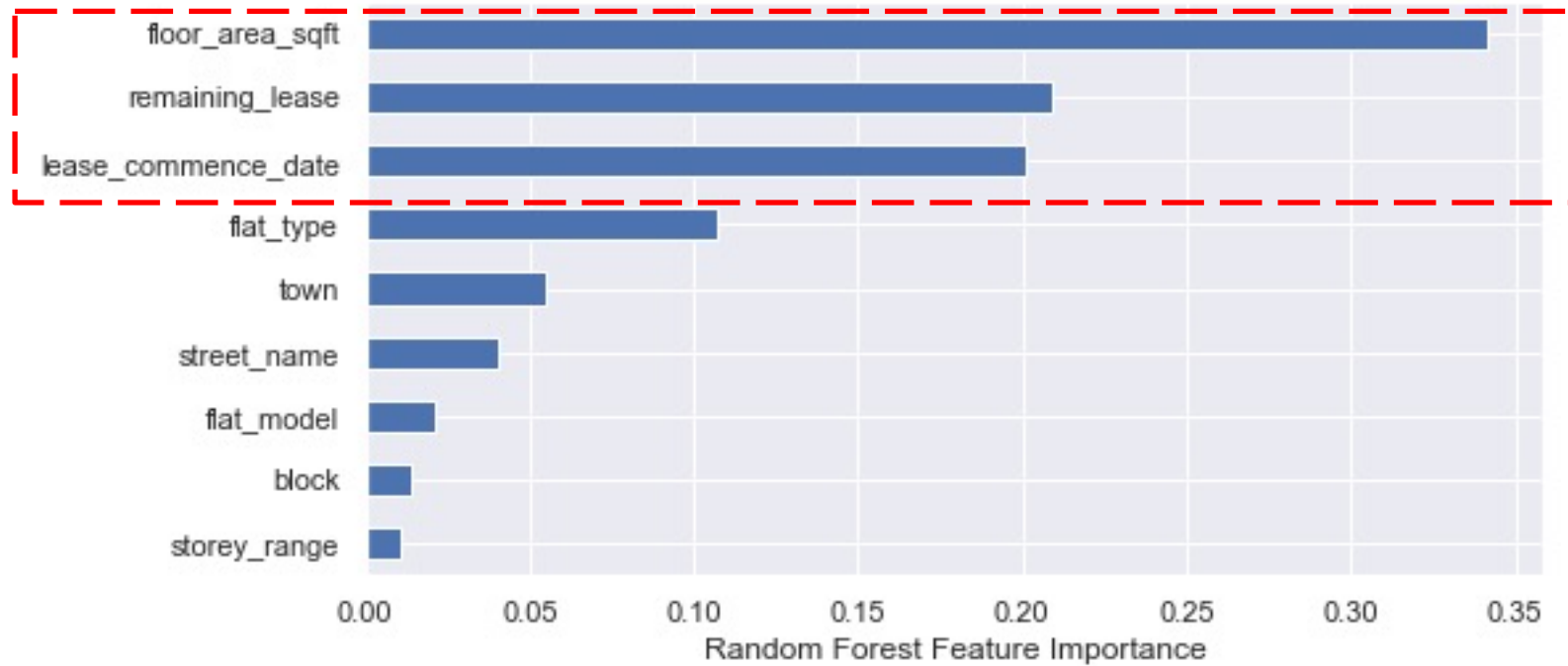
FEATURE SELECTION (SIMPLE LR)

- Floor area by sqft picked as predictor for simple LR



MULTI-FEATURE SELECTION

➤ Top 3 features by RandomForest picked.

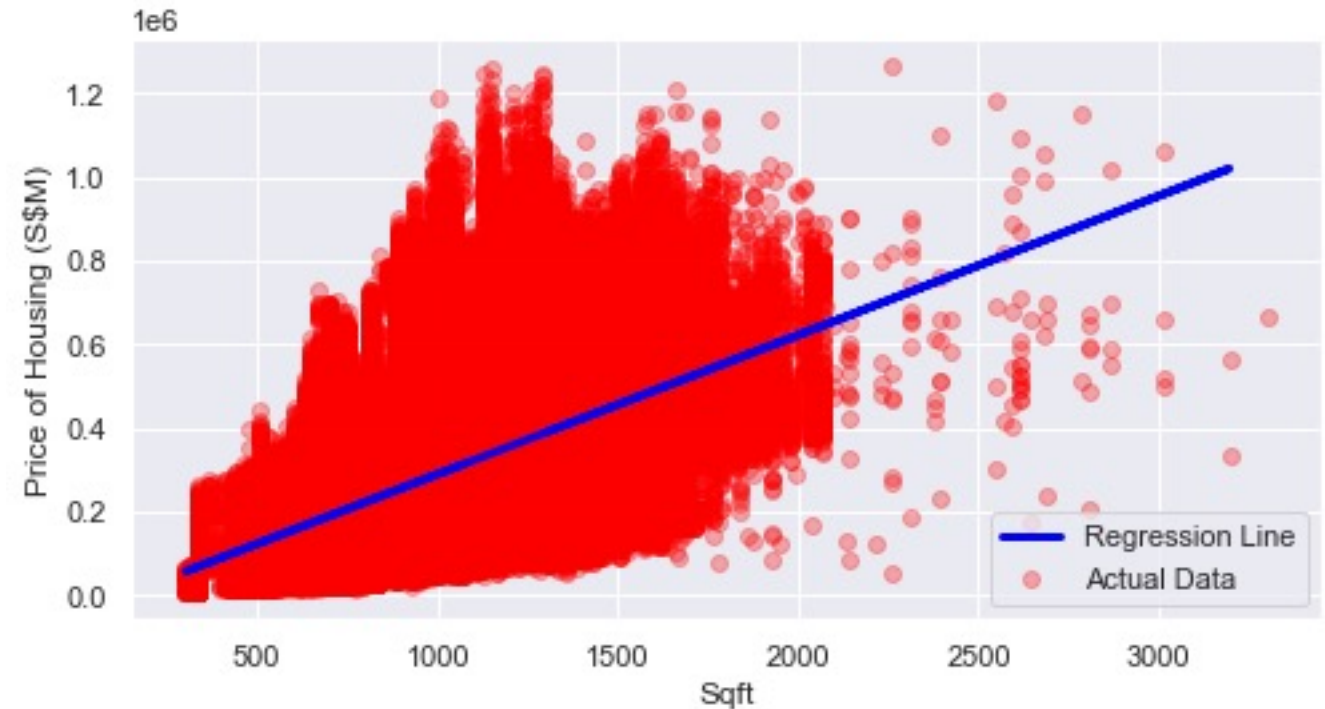


MODEL SELECTION

➤ Selected model can predict price with an **upper/lower bound of \$24,182 (8% of mean price)**

Model Type	Features	R2	RMSE (\$\$)
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)



PREDICTION TESTING

- Price predicted in prediction testing was 14% over actual data

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

town	flat_type	block	street_name	storey_range	flat_model	lease_commence_date	remaining_lease	floor_area_sqft	price_psft	resale_price	
date											
2020-09-01	18	5	944	198	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
```

IMPROVEMENTS

IMPROVEMENTS FOR FUTURE

- Descriptive stats could have been done on only last 1,3,5 year historical results
- More work needs to be done on feature engineering to try and create new features. 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