

2018 QUANTITATIVE STRATEGY CASE

By: Traci Lim

Datasets Used

- HDB Resale Flat Transactions from year 1990 to 2018
- HDB Addresses with Geographical Coordinates
- HDB Addresses with Geographical Coordinates
- School Addresses and MRT data with Geographical Coordinates
- Certificate of Entitlement (COE) Bidding Results

A Peek

	block	flat_model	flat_type	floor_area_sqm	lease_commence_date	month	remaining_lease	resale_price	storey_range	street_name	town
0	309	IMPROVED	1 ROOM	31.0	1977	1990-01	NaN	9000.0	10 TO 12	ANG MO KIO AVE 1	ANG MO KIO

	ADDRESS	BLK_NO	BUILDING	LATITUDE	LONGITUDE	LONGTITUDE	POSTAL	ROAD_NAME	SEARCHVAL	X	Y
0	216 ANG MO KIO AVENUE 1 ANG MO KIO 22 SINGAPOR...	216	ANG MO KIO 22	1.36619678831055	103.841505011903	103.841505011903	560216	ANG MO KIO AVENUE 1	ANG MO KIO 22	28911.0522399386	38692.616790574

	ADDRESS	BLK_NO	BUILDING	LATITUDE	LONGITUDE	LONGTITUDE	POSTAL	ROAD_NAME	SEARCHVAL	X	Y
0	EUNOS MRT STATION EXIT C SINGAPORE		EUNOS MRT STATION EXIT C	1.31960866729867	103.903178612002	103.903178612002	NIL	NIL	EUNOS MRT STATION EXIT C	35774.7365172636	33541.2481965665

	ADDRESS	BLK_NO	BUILDING	LATITUDE	LONGITUDE	LONGTITUDE	POSTAL	ROAD_NAME	SEARCHVAL	X	Y
0	37 HILLCREST ROAD NATIONAL JUNIOR COLLEGE SING...	37	NATIONAL JUNIOR COLLEGE	1.33029291072176	103.804110905425	103.804110905425	288913	HILLCREST ROAD	NATIONAL JUNIOR COLLEGE	24749.4869699031	34722.567163026

Q1: Data Viz

Question 1A:

Show an overview of the number of property transactions, median price across the years.

Provide both a view at the national level, as well as by HDB towns.

Your dashboard should also provide functionality to filter based on Flat Type (e.g. look at only 5 Room flats).

[Link](#) to Tableau Dashboard

Trends on HDB Resale Prices

Yearly HDB Median Resale Price by Towns

	year								
Town	2011	2012	2013	2014	2015	2016	2017	2018	
ANG MO KIO	350,000	385,000	390,000	370,000	352,000	358,000	358,000	355,000	
BEDOK	370,000	408,000	418,000	387,500	368,944	375,000	385,000	380,000	
BISHAN	530,000	542,000	580,000	536,000	560,000	610,000	650,000	655,000	
BUKIT BATOK	365,000	393,000	393,000	373,500	368,000	360,000	345,000	371,000	
BUKIT MERAH	484,000	538,000	599,000	519,000	518,888	552,500	580,000	585,000	
BUKIT PANJANG	415,000	448,000	455,000	420,000	390,000	380,000	418,000	410,000	
BUKIT TIMAH	545,000	603,500	654,444	690,000	794,000	700,000	698,000	770,000	
CENTRAL AREA	425,000	460,000	471,000	445,000	818,000	522,500	524,000	618,000	
CHOA CHU KANG	416,000	450,000	460,000	407,500	380,000	375,000	370,000	360,000	
CLEMENTI	384,500	408,000	432,000	402,000	390,000	378,000	420,000	425,000	
GEYLANG	344,500	371,000	373,000	355,000	374,000	370,000	370,000	442,500	
HOUGANG	412,000	443,000	450,000	420,000	400,000	388,000	402,000	408,000	
JURONG EAST	393,000	380,000	420,000	380,000	390,000	395,000	388,000	415,000	
JURONG WEST	407,000	445,000	453,000	418,000	400,000	402,000	389,000	389,000	
KALLANG/WHAMPOA	400,000	423,000	448,000	410,000	430,000	485,000	463,750	444,000	
MARINE PARADE	452,000	506,500	507,500	475,000	510,000	475,888	468,000	472,800	
PASIR RIS	470,000	503,000	509,000	477,000	459,444	452,000	465,000	480,000	
PUNGGOL	495,000	528,000	535,000	478,000	440,000	446,000	440,000	436,000	
QUEENSTOWN	378,000	419,000	433,944	438,000	450,000	520,000	565,000	613,000	
SEMBAWANG	425,000	462,000	477,000	420,000	400,000	400,000	380,000	368,000	
SENGKANG	470,750	508,944	528,888	470,000	440,000	435,000	426,000	418,444	
SERANGOON	423,000	480,000	473,000	438,500	450,000	450,000	461,500	486,500	
TAMPINES	434,000	470,000	488,000	455,000	430,500	435,000	450,500	480,000	
TOA PAYOH	353,000	396,000	406,000	368,000	365,500	355,000	445,000	510,000	

Consistently high prices for flats in Bishan and Bukit Timah.

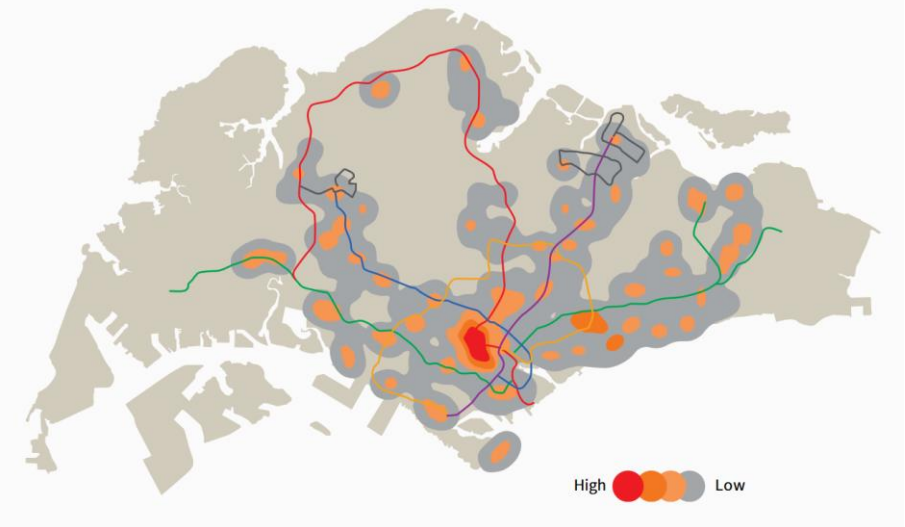
Approx. 2x Jump!

Why 2x Jump?

Town	≡ ↕	year						
		2012	2013	2014	2015	2016	2017	2018
ANG MO KIO		385,000	390,000	370,000	352,000	358,000	358,000	355,000
BEDOK		408,000	418,000	387,500	368,944	375,000	385,000	380,000
BISHAN		542,000	580,000	536,000	560,000	610,000	650,000	655,000
BUKIT BATOK		393,000	393,000	373,500	368,000	360,000	345,000	371,000
BUKIT MERAH		538,000	599,000	519,000	518,888	552,500	580,000	585,000
BUKIT PANJANG		448,000	455,000	420,000	390,000	380,000	418,000	410,000
BUKIT TIMAH		603,500	654,444	690,000	794,000	700,000	698,000	770,000
CENTRAL AREA		460,000	471,000	445,000	818,000	522,500	524,000	618,000
CHOA CHU KANG		450,000	460,000	407,500	380,000	375,000	370,000	360,000
CLEMENTI		408,000	432,000	402,000	390,000	378,000	420,000	425,000

- Report from JLL shows out of 1610 purchases made by foreigners in 2016/2017, 60% were located in Central Area
- But foreigners can't buy HDBs
- Could be due to a **trickle-down effect**

Figure 2 Heat Map of Singapore Residential Properties Purchased by Foreigners in 2016/ 2017 (Jan-Apr)

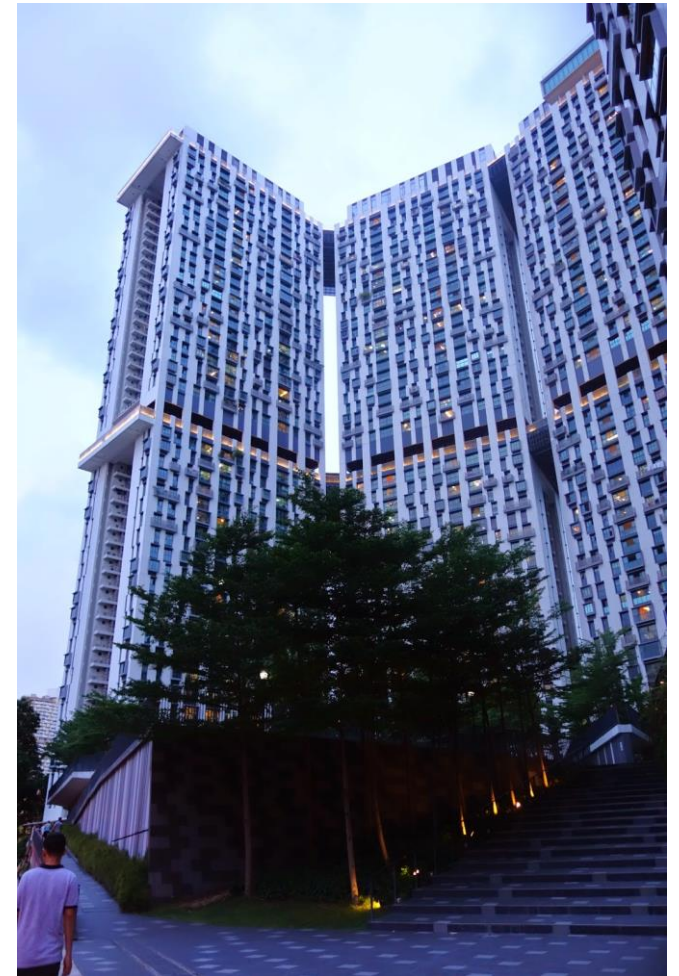


<http://www.ap.jll.com/asia-pacific/en-gb/Research/SG-hotspots.pdf?587a0557-2140-49bc-b82c-46b7d96b1df2>

Why 2x Jump?

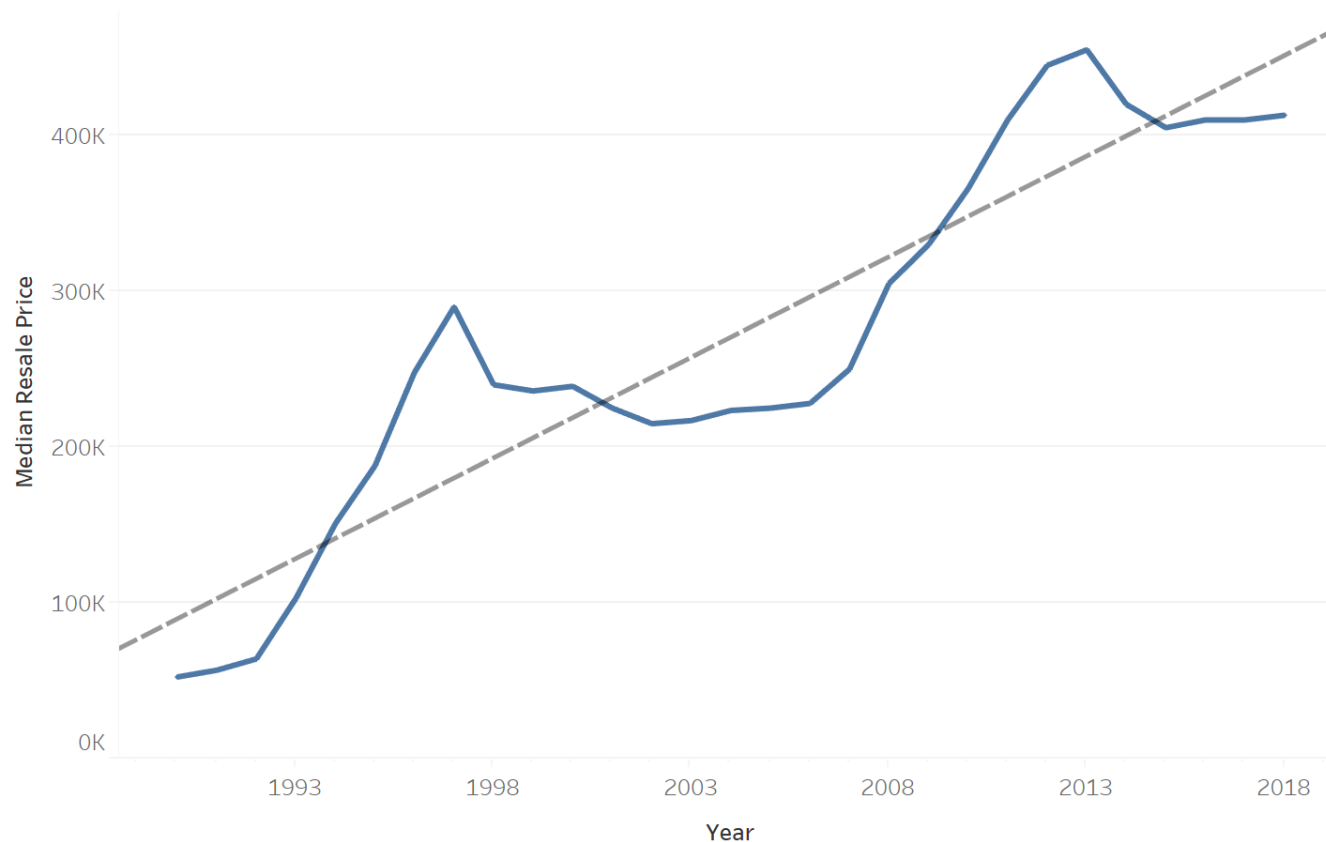
Town	↕	year						
		2012	2013	2014	2015	2016	2017	2018
ANG MO KIO		385,000	390,000	370,000	352,000	358,000	358,000	355,000
BEDOK		408,000	418,000	387,500	368,944	375,000	385,000	380,000
BISHAN		542,000	580,000	536,000	560,000	610,000	650,000	655,000
BUKIT BATOK		393,000	393,000	373,500	368,000	360,000	345,000	371,000
BUKIT MERAH		538,000	599,000	519,000	518,888	552,500	580,000	585,000
BUKIT PANJANG		448,000	455,000	420,000	390,000	380,000	418,000	410,000
BUKIT TIMAH		603,500	654,444	690,000	794,000	700,000	698,000	770,000
CENTRAL AREA		460,000	471,000	445,000	818,000	522,500	524,000	618,000
CHOA CHU KANG		450,000	460,000	407,500	380,000	375,000	370,000	360,000
CLEMENTI		408,000	432,000	402,000	390,000	378,000	420,000	425,000

- Pinnacle@duxtun is the culprit.
- Out of **756** transactions in Central Area for year 2015, **472** are Pinnacle transactions priced above 800,000.
- If Pinnacle is car, it would be the Rolls-Royce of HDBs.



Up and Up and Up

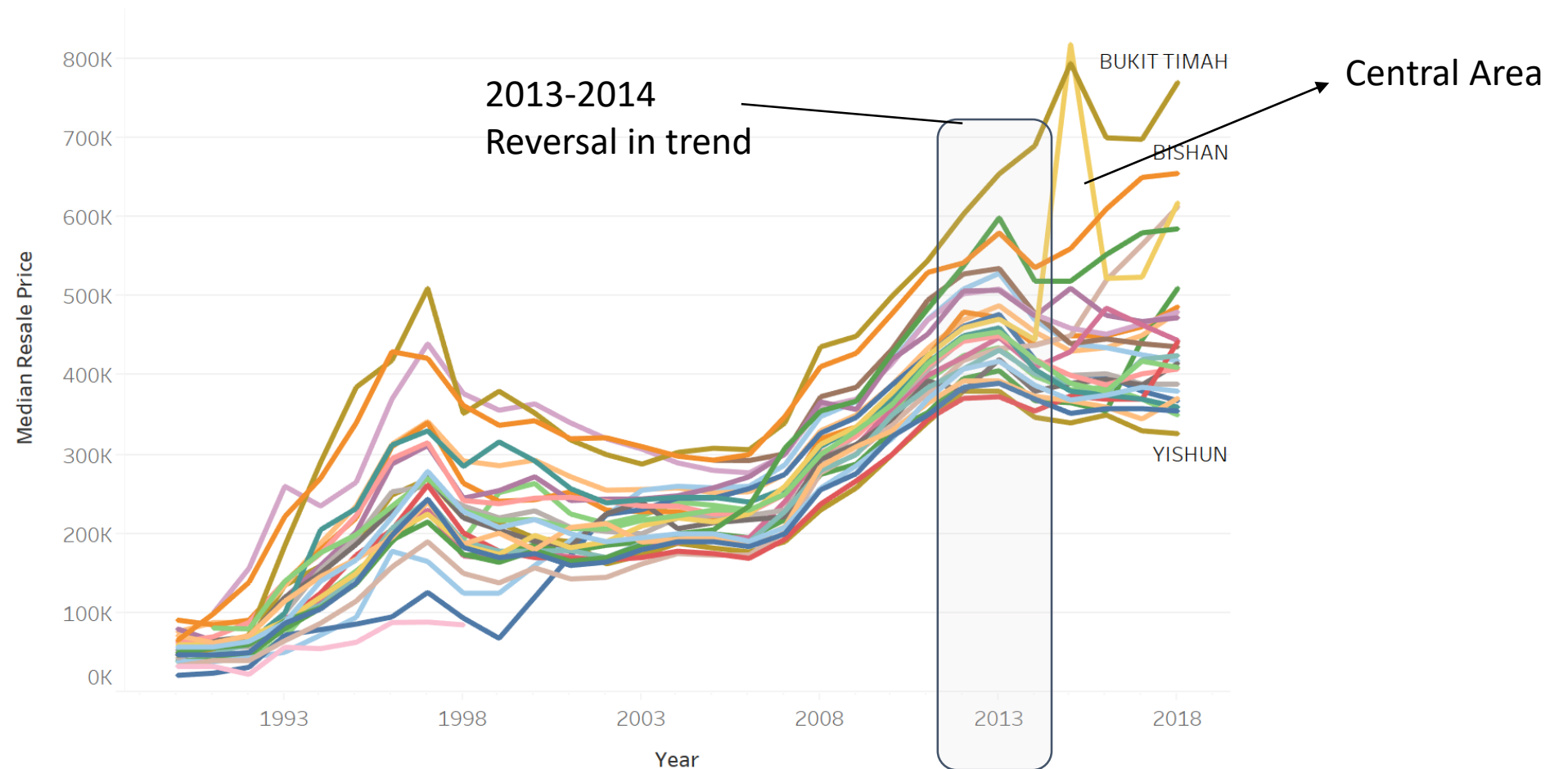
Median Resale Price by Year



Although prices are not adjusted by inflation, median resale prices have **increased by a factor of 4** since 1990.

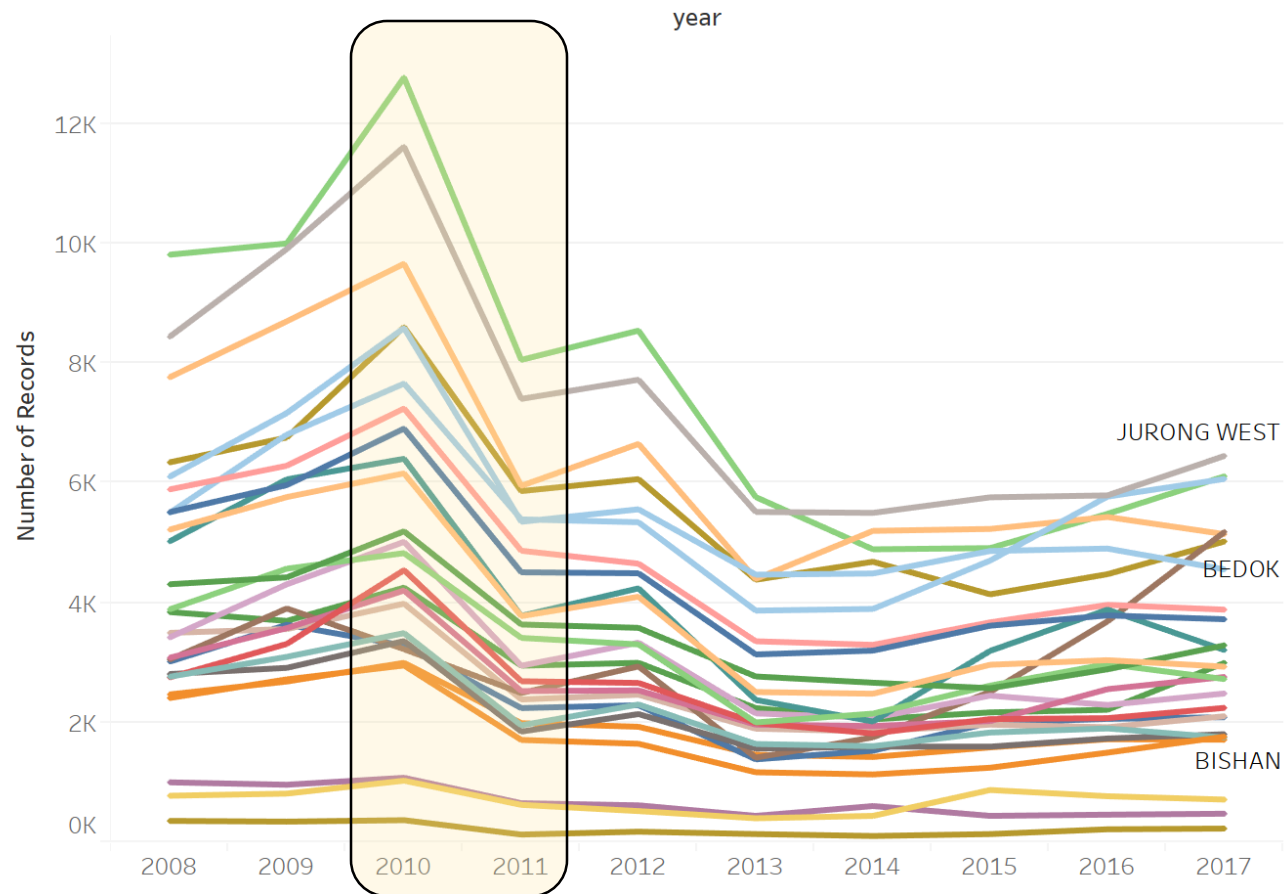
Yearly Trend

Yearly trend on Median Resale Price across Towns



First wave of cooling measures amidst an overheated market

Number of Transactions by Year across Towns



In January 2011, four **Singapore property cooling measures** were introduced.

Reversal in trend from 2013 to 2014

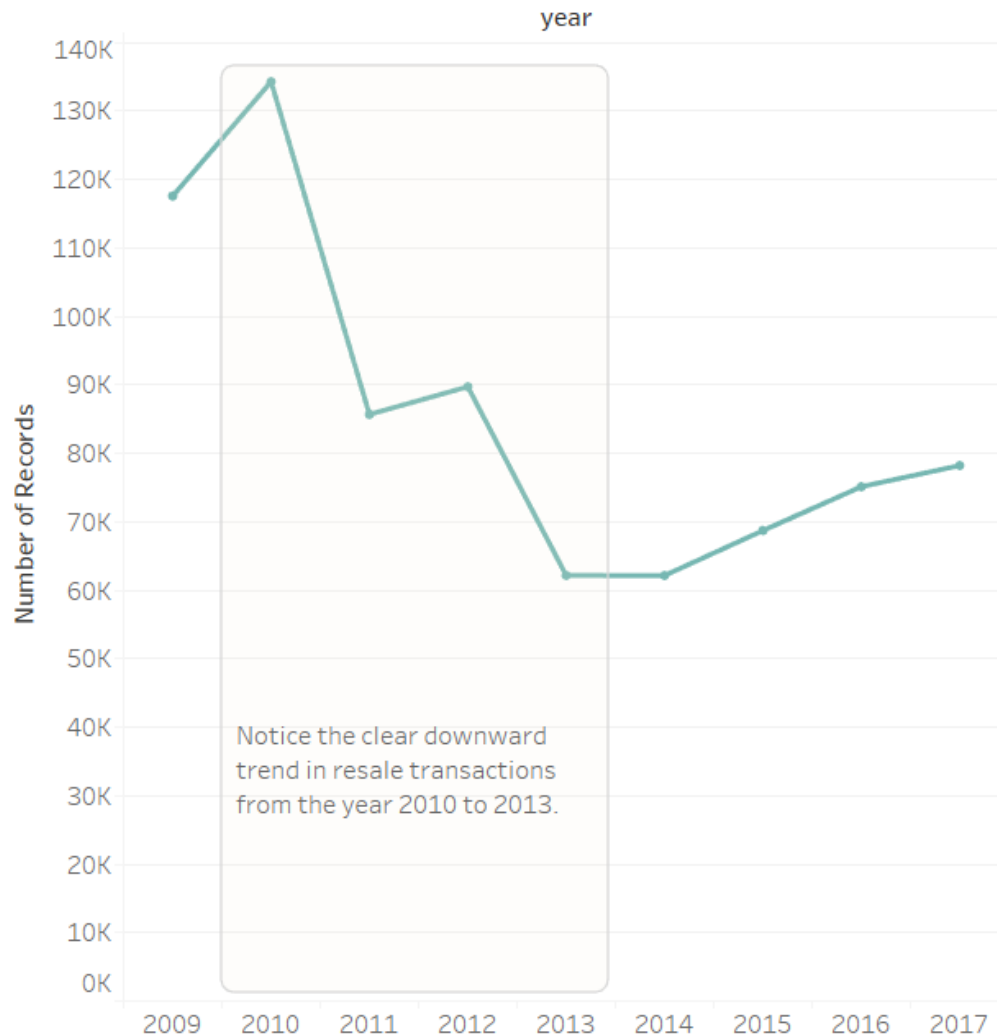


<http://www.mas.gov.sg/News-and-Publications/Media-Releases/2013/Additional-Measures-To-Ensure-A-Stable-And-Sustainable-Property-Market.aspx>

More cooling measures?



Yearly HDB Resale Transactions



10 cooling measures were implemented from Sep 2009 to Dec 2013.

- Seller's Stamp Duty (SSD)
- Lowered loan-to-value (LTV) ratio limits
- Additional Buyers' Stamp Duty (ABSD)
- Mortgage service ratio (MSR)
- Some other stuff...

Q1: Data Viz

Question 1B:**

Some buyers would want to get the largest flat possible within a given budget. Create a dashboard to allow potential buyers to input their budget, and then suggest towns where such flats exist based on historical transactions.

Question 1C*:**

There are buyers who would like to optimise given the proximity of a flat to important locations in the neighbourhood, such as the nearest MRT station.

Create a dashboard that allows buyers to: i) input their budget; and ii) optimise flat selection given distance to important locations around the neighbourhood.

Question 1B, 1C

HDB Resale Flat Suggestion Tool

In the right panel, please input the following preferences to optimize the suggestions given to you.

- 1) Estimated budget range.
- 2) Distance to the nearest MRT station.
- 3) Distance to the nearest school.

Estimated Budget (in \$)

304934



416250

Distance To Nearest MRT Station's Exit (in kilometres)

0.0200



0.4520

Distance To Nearest School (in kilometres)

0.0000



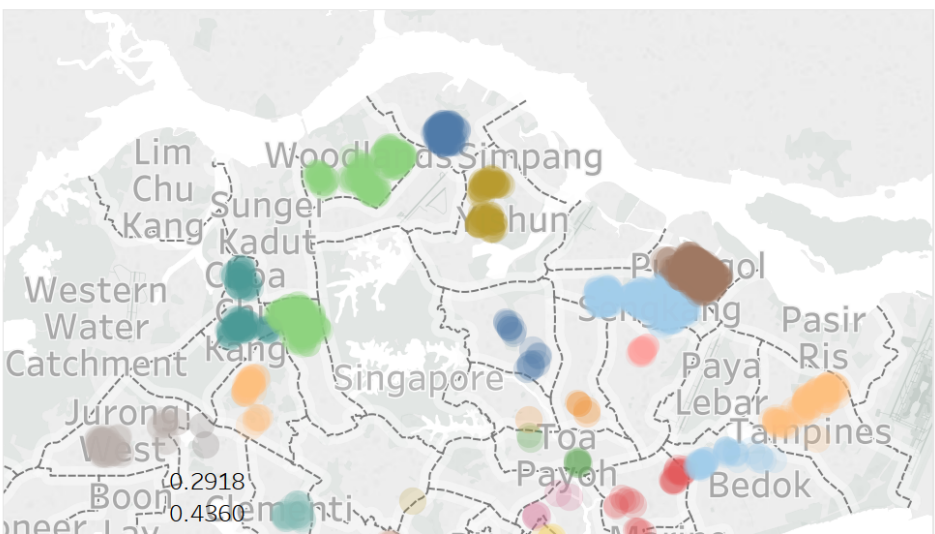
0.3040

Flat Types Available to you:

- ☒ (All)
- ☐ 2 ROOM
- ☐ 3 ROOM
- ☐ 4 ROOM
- ☐ 5 ROOM

Suggested towns and flat types based on historical resale transactions

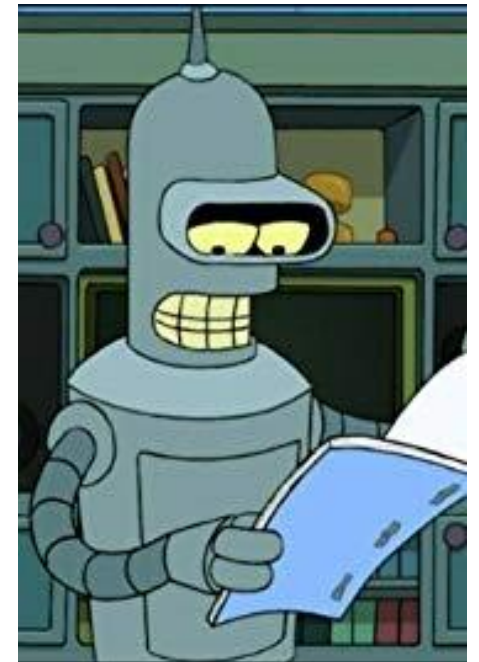
Town	Flat Type			
	2 ROOM	3 ROOM	4 ROOM	5 ROOM
ANG MO KIO		350,000	402,500	
BEDOK		315,000	390,000	
BISHAN		374,000		
BUKIT BATOK		308,000	354,400	391,500
BUKIT MERAH		345,000		
BUKIT PANJANG		329,000	350,000	400,000
BUKIT TIMAH		385,000		
CENTRAL AREA		376,000		



Q2: Data Modeling

HDB wants to know if a resale flat transaction fits market expectations. Your task is to create statistical models to answer the following questions.

Question 2A*: Predict a resale flat transaction price in 2014. Use the following characteristics: flat type, flat age and town. Propose and implement a minimum of three models, select the best model, and explain the reasons for your choice.



Predicting Resale Price with 3 Features

```
In [526]: train.iloc[:, :-1].head()
```

Out[526]:

	flat_type	town	flat_age	resale_price
1519039	3 ROOM	ANG MO KIO	42	375000.0
1519040	4 ROOM	ANG MO KIO	41	520000.0
1519041	5 ROOM	ANG MO KIO	41	560000.0
1519042	3 ROOM	ANG MO KIO	37	350000.0
1519043	4 ROOM	ANG MO KIO	36	480000.0

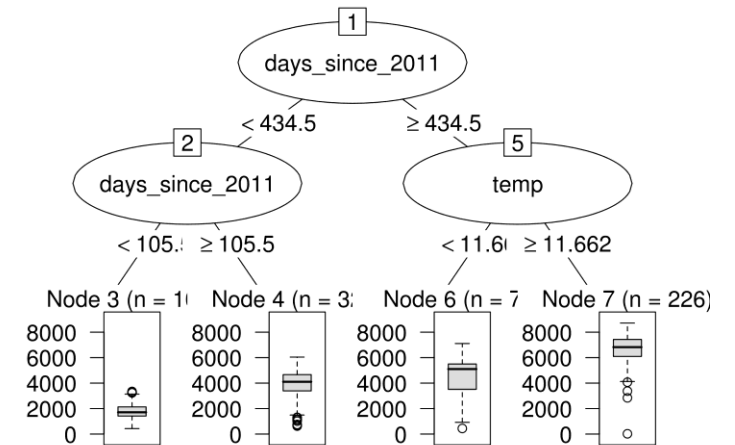
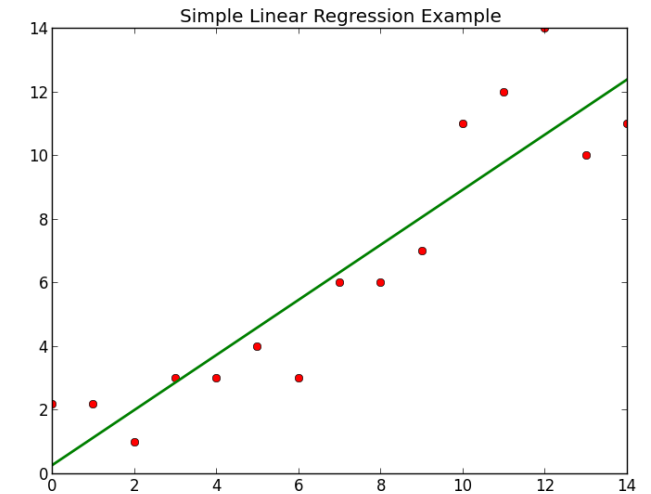
```
In [525]: test.iloc[:, :-2].head()
```

Out[525]:

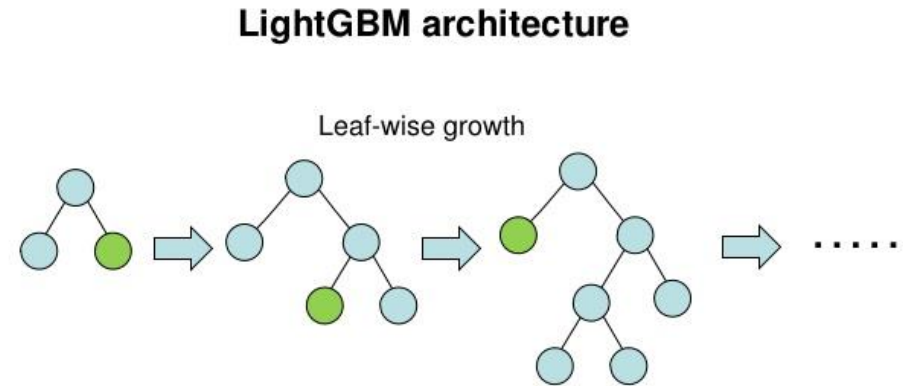
	flat_type	town	flat_age
2928291	2 ROOM	ANG MO KIO	38
2928292	3 ROOM	ANG MO KIO	38
2928293	3 ROOM	ANG MO KIO	39
2928294	3 ROOM	ANG MO KIO	38
2928295	3 ROOM	ANG MO KIO	38

Models used

- Linear Regression with Regularization (Elastic Net)
 - Simple
 - Easy to understand
- Decision Trees (Regression)
 - Simple
 - Good when there are lots of categorical variables
 - Can capture non-linear relationships
- LightGBM (Microsoft's Gradient Boosting Algorithm)
 - Fast
 - Cutting Edge
 - Mega-improved version of decision trees



What is LightGBM?



#StrataHadoop

Dr. Mathew Salvaris @msalvaris

Dr. Miguel Fierro @miguelgferro

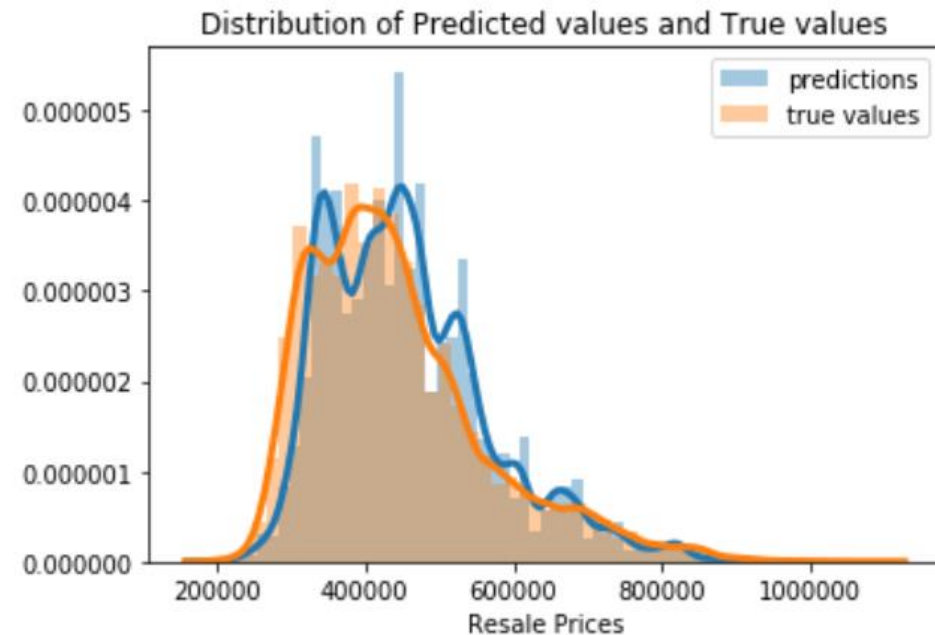
Strata Hadoop
WORLD

- LightGBM (Microsoft's Gradient Boosting Algorithm)
 - Fast (took 3 seconds)
 - Cutting Edge
 - Mega-improved version of decision trees

Every prediction is \$33,000 off on average

```
In [637]: mean_absolute_error(y_test, model_lgb.predict(X_test))|  
Out[637]: 32821.64796678499
```

- This is one of the cases where Tree models > Linear models.
- MAE tells us how big of an error we can expect from the predicted value on average.



Q2: Data Modeling

Question 2B:** A flat was sold in Nov 2017 with the following characteristics:

Flat type: 4 ROOM

Town: Yishun

Flat Model: New Generation

Storey Range: 10 to 12

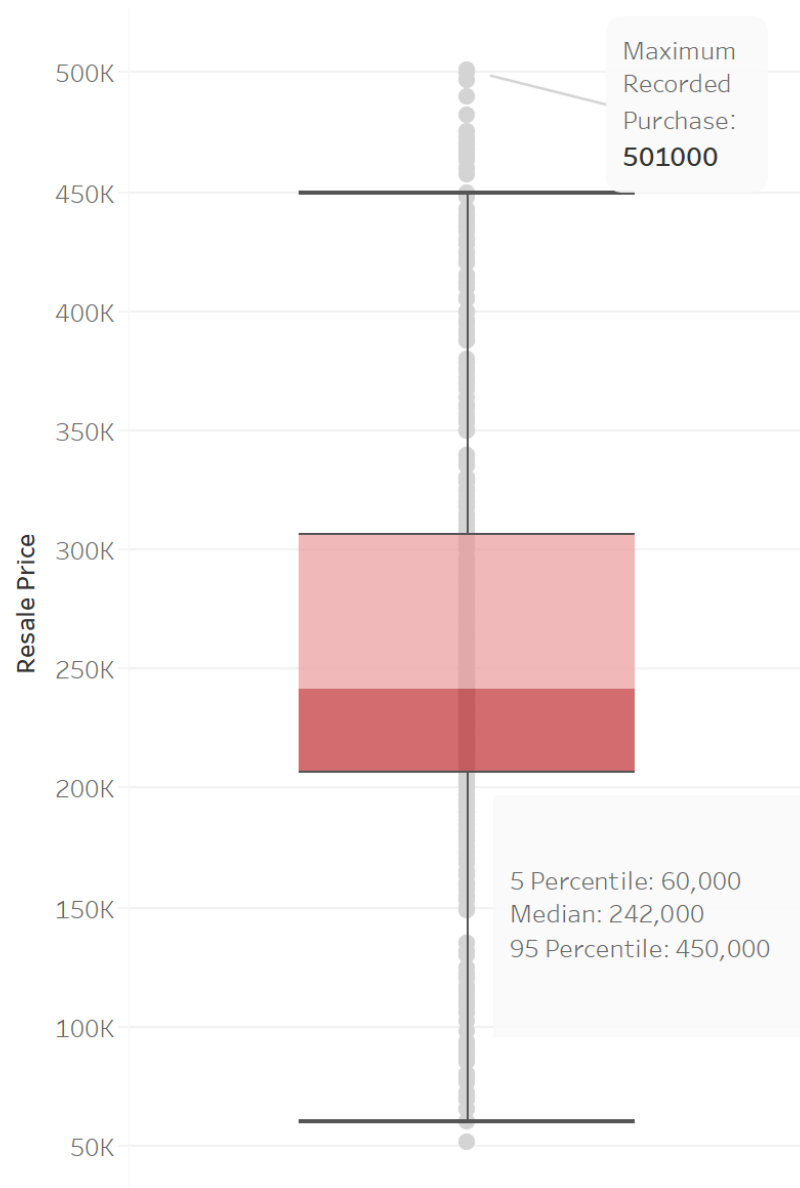
Floor Area (sqm): 91

Lease Commence Date: 1984

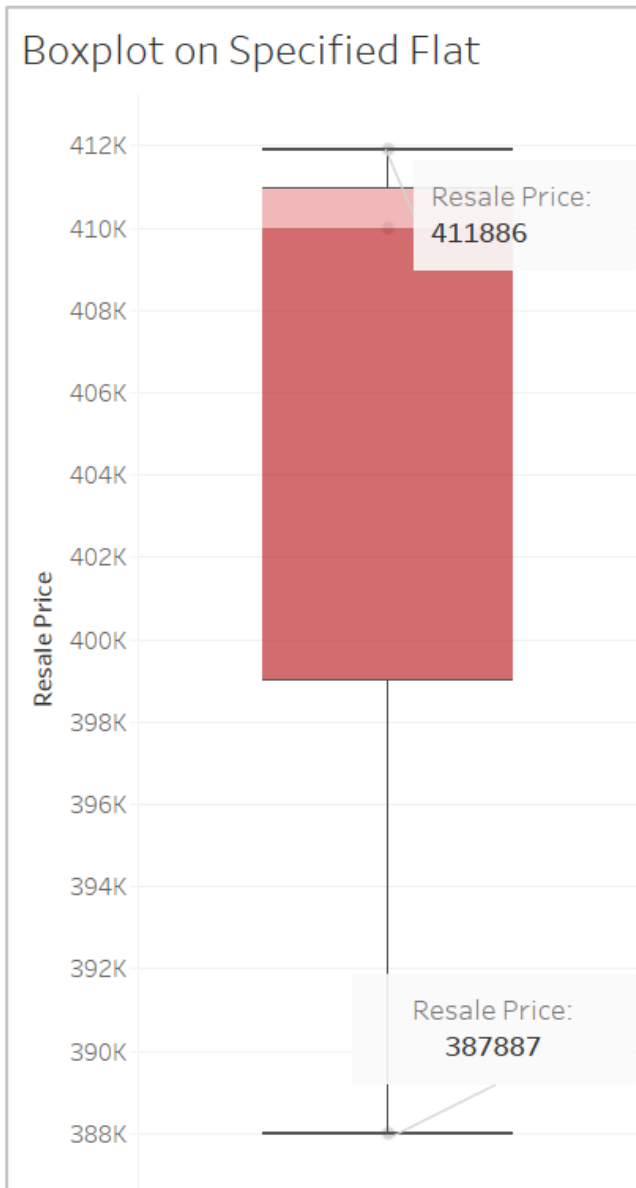
Resale Price: 550,800

Was this a reasonable price for the transaction? How confident are you in your assessment?

Boxplot on Specified Flat



- Highly overpriced transaction
- The word 'reasonable' is highly subjective, since the transaction consists of a willing buyer and seller
- Must be good fengshui



- Highly overpriced transaction
- The word 'reasonable' is highly subjective, since the transaction consists of a willing buyer and seller
- Must be good fengshui



Predicting Flat Types

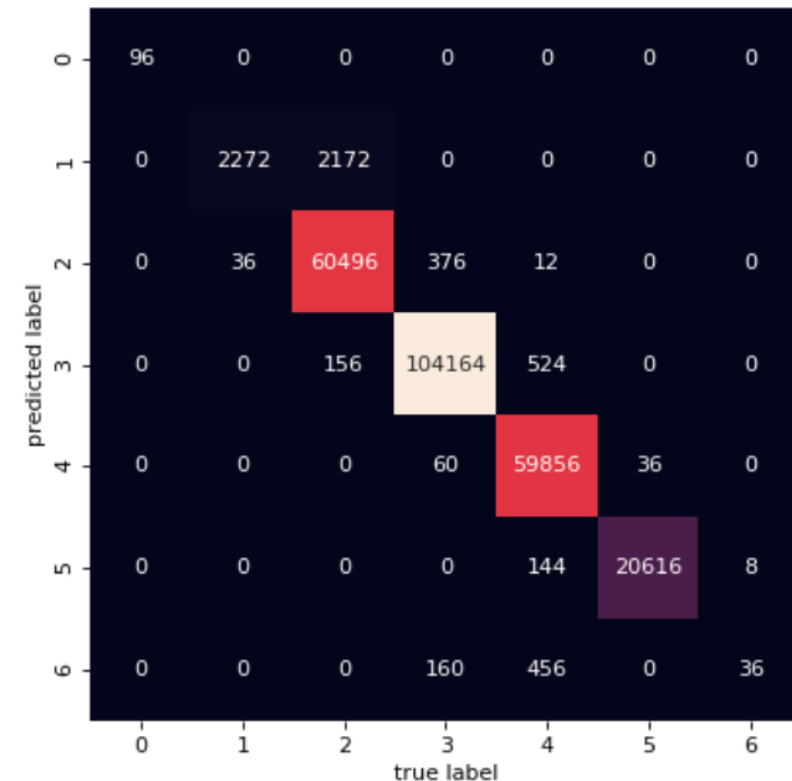
- **Question 2C***:** Someone mistakenly deleted the column containing data on Flat Type in the database from 2015 onwards. While backups exist, these data are critical to HDB's daily operations, and time would be needed to restore these data from the backup. Senior management would like you to create a model to predict flat type given a transaction's other characteristics. Explain the reasons for choosing this model.

Model correctly predicts flat type 98% of the time

	precision	recall	f1-score	support
1RM	1.00	1.00	1.00	96
2RM	0.51	0.98	0.67	2308
3RM	0.99	0.96	0.98	62824
4RM	0.99	0.99	0.99	104760
5RM	1.00	0.98	0.99	60992
6EX	0.99	1.00	1.00	20652
7MG	0.06	0.82	0.10	44
avg / total	0.99	0.98	0.99	251676

Accuracy Score is: 0.983550278930053

Confusion Matrix



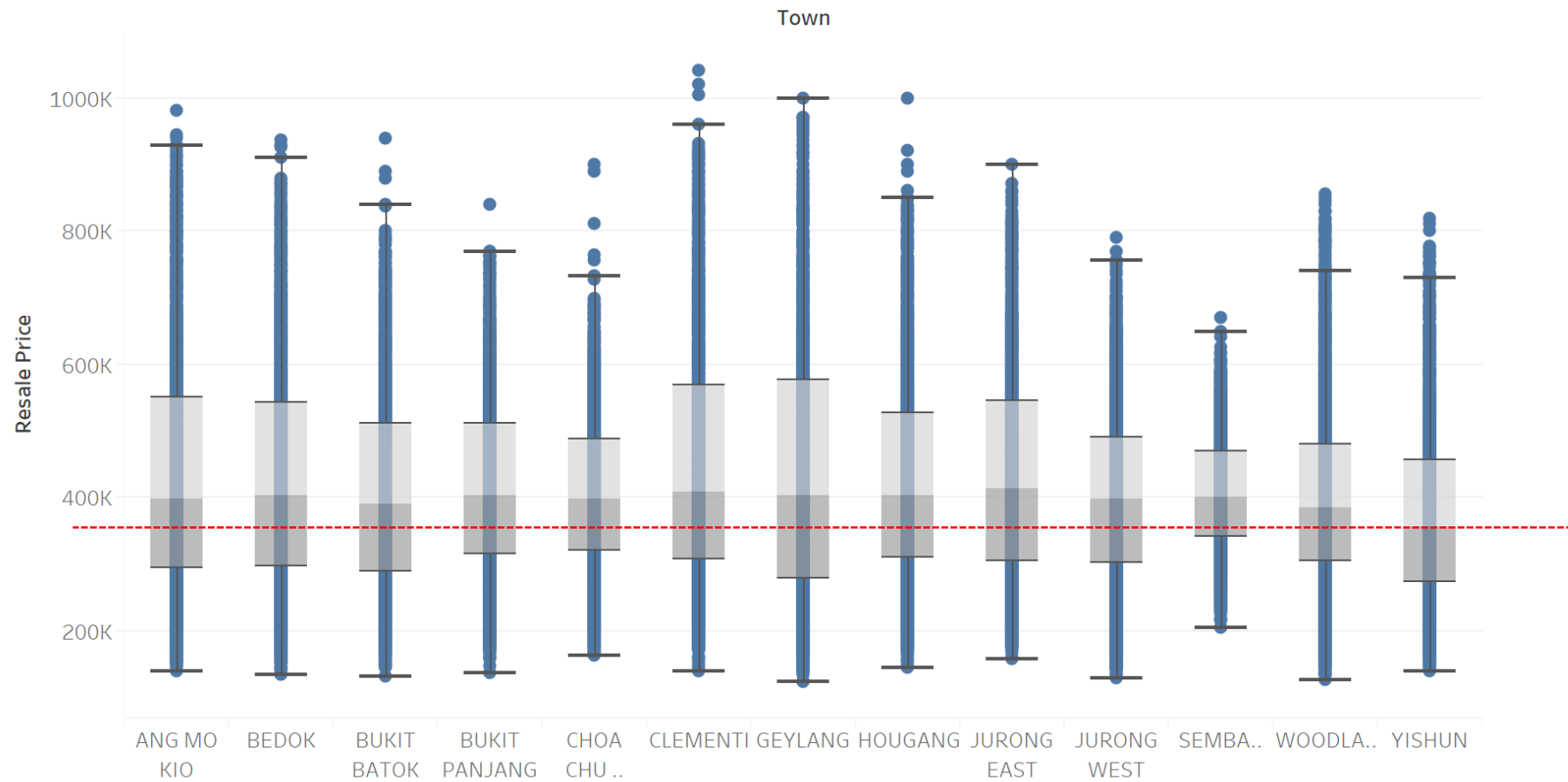
Q3: Policy Analysis

Question 3A*:

Yishun has received a negative reputation as “Crazy Town”, and property prices might have been impacted. Are Yishun flats the cheapest in the country?

Can't Beat Yishun Cheap Prices

Boxplots of Resale Prices from 2008 to 2018



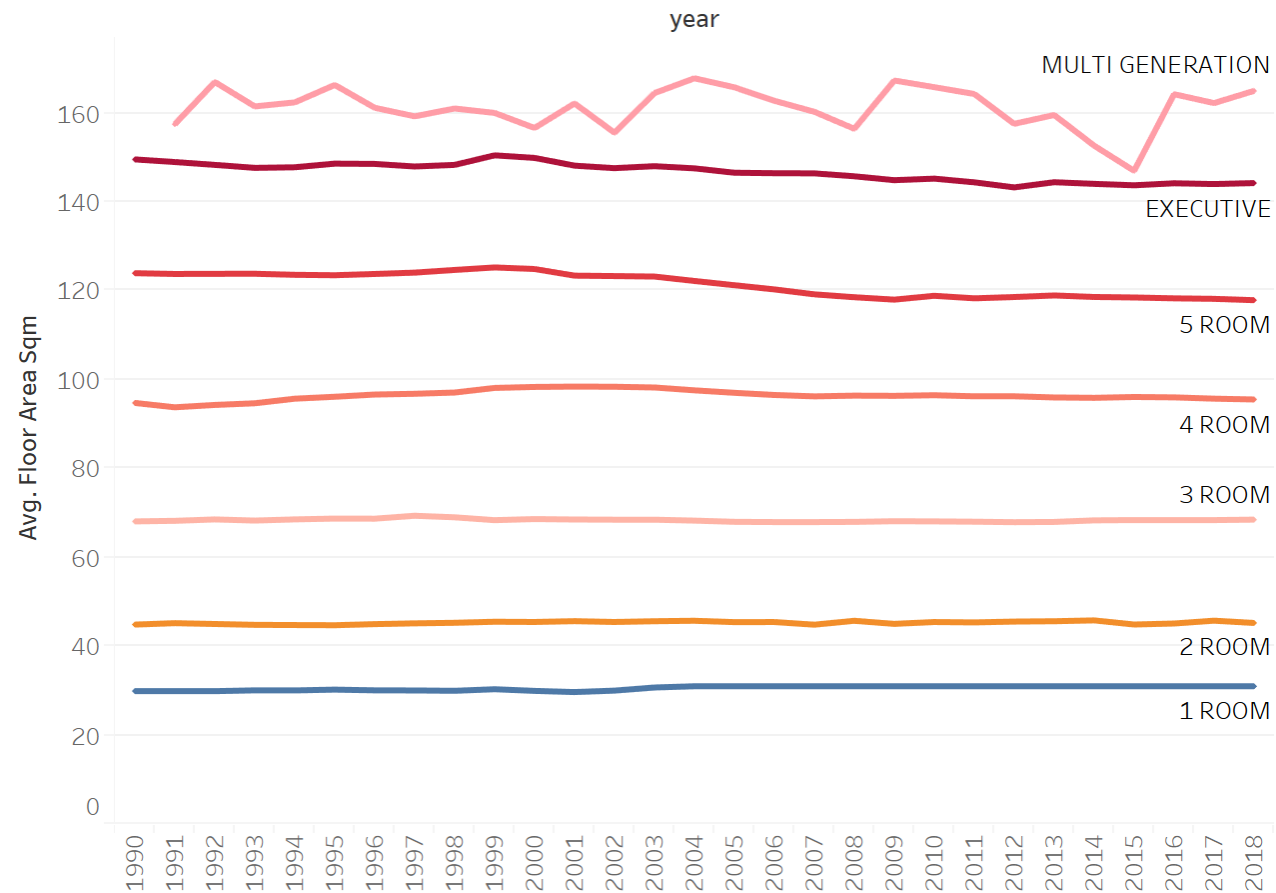
Q3: Policy Analysis

Question 3B*:

Some members of public have been saying that flat sizes have gotten smaller over the years. Is there any truth in this statement?

Nearly Parallel lines

Average Floor Area among HDB Resale Flats



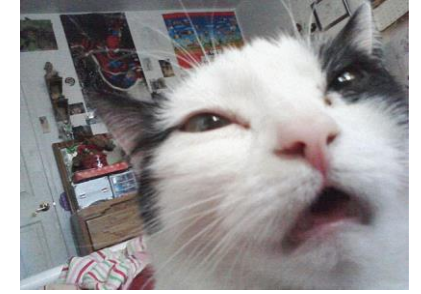
Near-constant trend

Q3: Policy Analysis

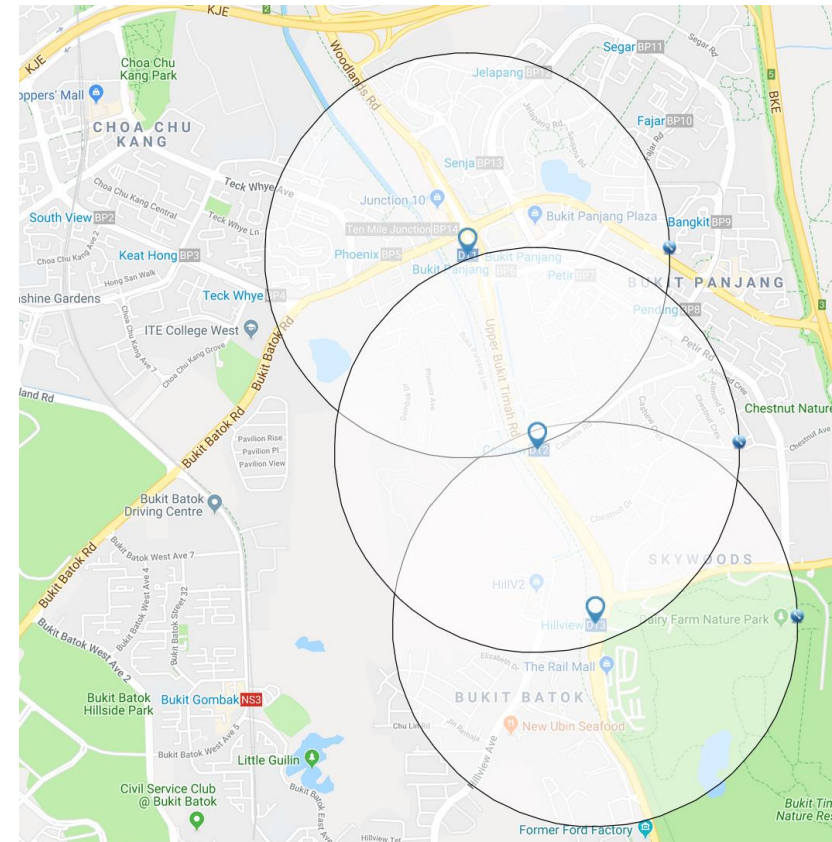
Question 3C:**

The Downtown Line Stage 2 connects the Bukit Panjang heartland to the city. Have prices increased for resale flats in the towns served by this Line? You might want to use a difference-in-differences model for this task.

Difference-in-Differences Model



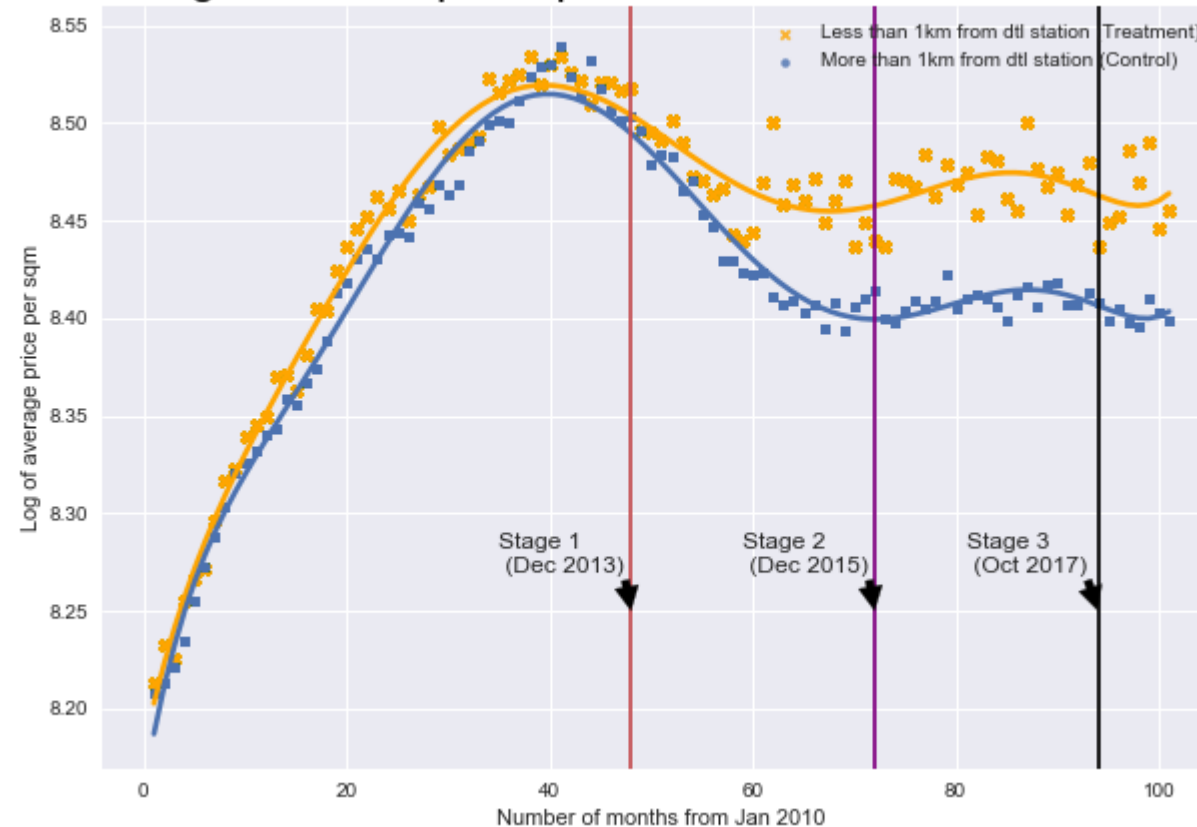
- Used to **estimate causal effects of the opening of Singapore's Downtown Line**
- Two groups: Control and Treatment
- Control group: Flats *outside* 1km radius of any DTL station
- Treatment group: Flats *within* 1km radius of any DTL station



Did Opening of DTL affect Average Price per Sqm of HDB Resale Flats?



Temporal Changes of Price per Square Metre in Treatment and Control Groups



Estimating the DiD estimator



```
didreg_1 = lm(log_avg_price_per_sqm ~ treated + time + did, data = df)
summary(didreg_1)
```

Call:

```
lm(formula = log_avg_price_per_sqm ~ treated + time + did, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.187719	-0.022983	-0.007261	0.052923	0.149679

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.3901195	0.0001393	60214.24	<2e-16 ***
treated	0.0105795	0.0003263	32.43	<2e-16 ***
time	0.0278074	0.0002050	135.63	<2e-16 ***
did	0.0388501	0.0004766	81.52	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07642 on 685640 degrees of freedom

Multiple R-squared: 0.07719, Adjusted R-squared: 0.07718

F-statistic: 1.912e+04 on 3 and 685640 DF, p-value: < 2.2e-16

- Helped to increase average price per sqm of flats near a DTL station by 3.9%.
- But the average price per sqm of flats was experiencing a 2.8% growth after the opening of DTL, regardless of its proximity to a DTL station.

Q3: Policy Analysis

Question 3D*:**

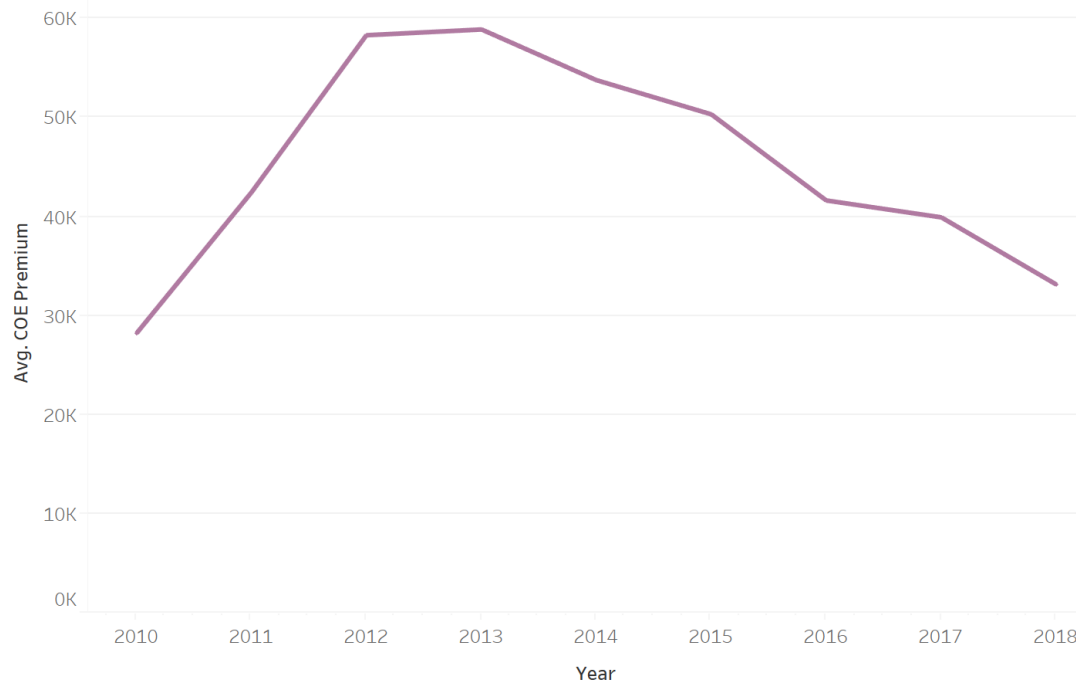
There have been comments online that people are buying flats in towns further from the city so that the cost savings can be used for a car.

Are resale prices in HDB estates in areas further away from the city (i.e. Sengkang and Punggol) impacted by Certificate of Entitlement (COE) prices for cars?

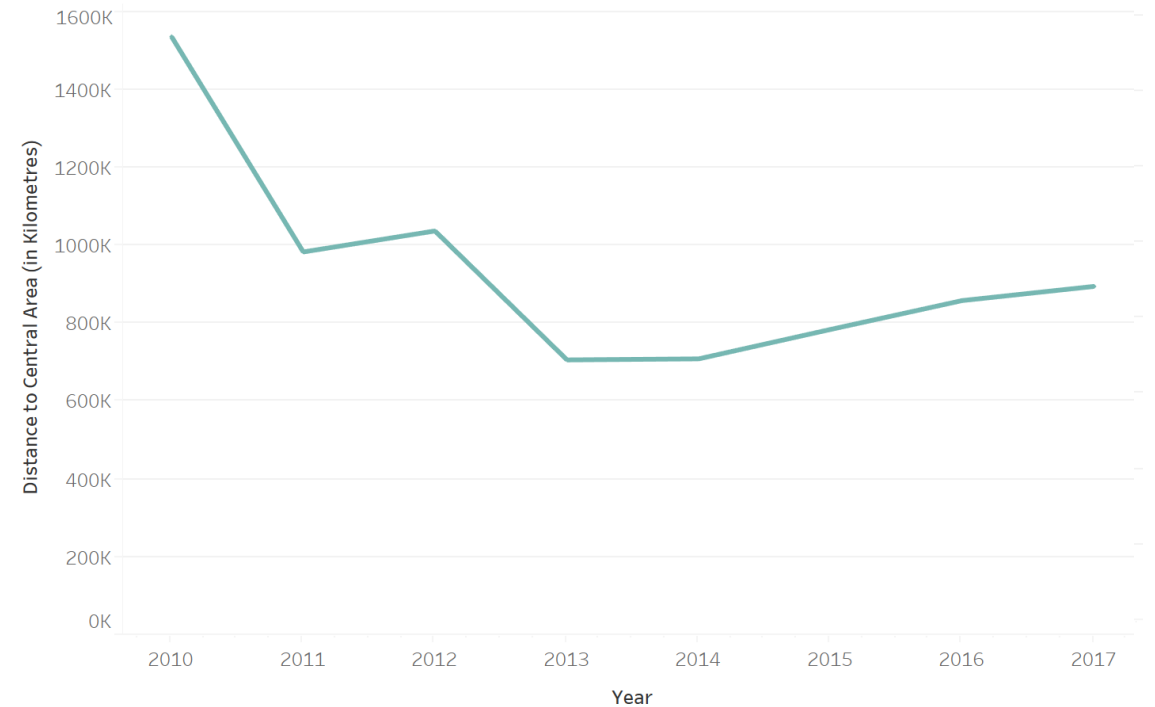
About COE Prices

- **COE prices** are determined by the interaction of **demand** and supply.

Average COE Premium



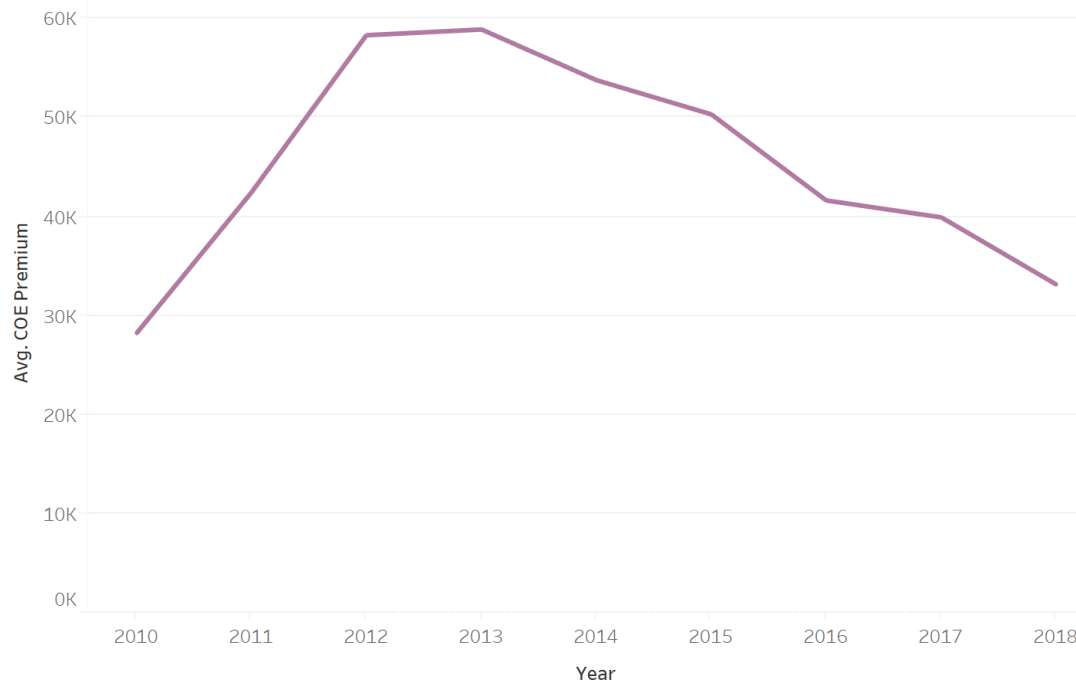
Sum of Resale Flats' Distance to Central Area



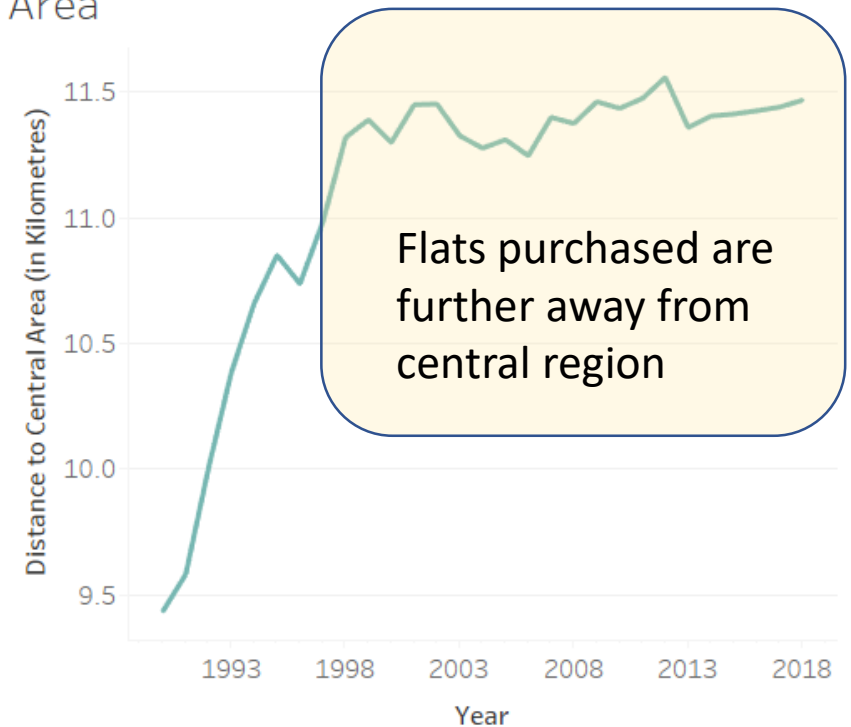
Flats bought are further away to central area



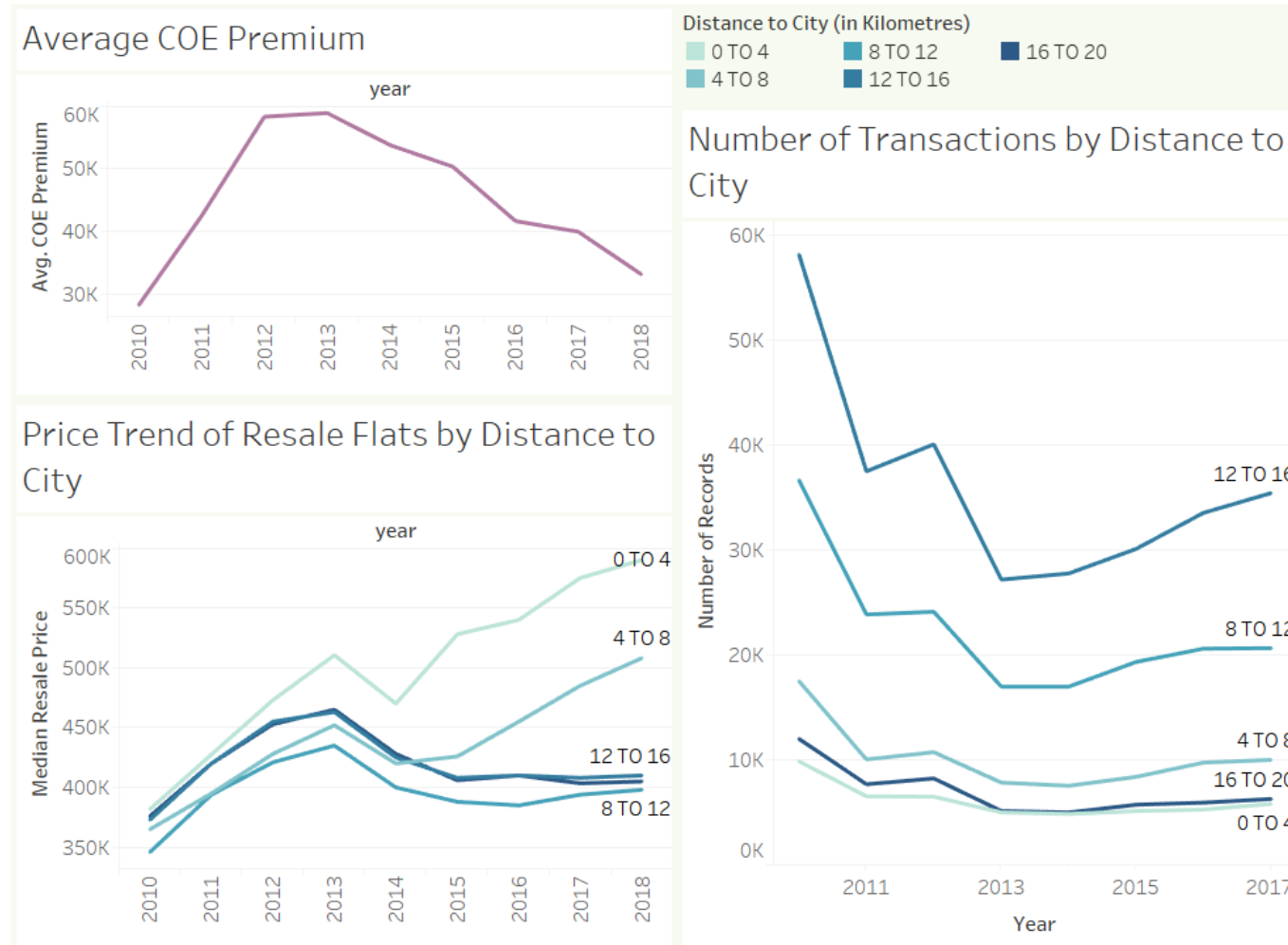
Average COE Premium



Average Resale Flats' Distance to Central Area



Q3D: Looking at Resale Prices and COE



Sengkang and Punggol fall in the 12 to 16 km mark.

Price trend of flats that are 8 to 20 km away from the city actually follows a similar pattern with COE prices.

End.