



**NUS**  
National University  
of Singapore

## BT2101 Project

Group 12

Members:

Haris Bin Dzulkifli

(A0235244L e0727244@u.nus.edu)

Jonathan Lim Yu Shun

(A0234165J e0726165@u.nus.edu)

Tan Hui Rong

(A0216246J e0538347@u.nus.edu)

Wong Jing En

(A0229831B e0684546@u.nus.edu)

Yang Yee

(A0218050U e0544086@u.nus.edu)

# Contents Page

<b>Contents Page</b>	<b>1</b>
<b>1 Background Information</b>	<b>2</b>
1.1 HDB flats	2
1.2 Resale HDB flats	3
<b>2 Motivation</b>	<b>3</b>
2.1 Rationale of Project	3
2.2 Target Audience	3
2.2.1 Investors	4
2.2.2 Families	4
<b>3 Project Question</b>	<b>4</b>
<b>4 Hypotheses</b>	<b>4</b>
4.1 Resale price increases as it moves towards the central regions	4
4.2 Resale price decreases as the walking time to the nearest transportation increases	5
4.3 Resale price increases non-linearly as the floor area increases	5
4.4 Relevance of Hypotheses to Target Audiences	5
4.4.1 Investors	5
4.4.2 Families	5
<b>5 Data</b>	<b>5</b>
5.1 Basic Data Statistics	5
5.1.1 Correlation	7
5.1.1.1 Correlation Matrix	7
5.1.1.2 Issues with Multicollinearity	7
5.1.2 Feature Engineering	8
5.1.2.1 Time Taken to Nearest MRT/LRT	8
5.1.2.2 Region of HDB Resale Flat	8
5.2 Charts on Basic Data	9
5.2.1 Breakdown of the HDB Resale Prices	9
5.2.2 Distribution of Region that Resale HDB is located	9
<b>6 Regression Models</b>	<b>10</b>
6.1 Reference Models	10
6.1.1 Ref. Regression 1: Resale Price against Independent Variables	10
6.1.2 Ref. Regression 2: Resale Price per Sq Metre against Independent Variables	10
6.2 Resale price increases as it moves towards the central regions	11
6.2.1 Regions	11
6.2.2 Regression Model and Interpretation	11

6.2.2.1 Regression of Price against RCR and CCR	11
6.2.2.2 Regression of Price per Square Metre against RCR and CCR	12
6.3 Resale price decreases as the walking time to the nearest transportation increases	13
6.3.1 Relationship between Resale Prices and Time Taken to Nearest MRT/LRT	13
6.3.2 Regression Model and Interpretation	13
6.3.2.1 Regression of Price against Time Taken to MRT/LRT	13
6.3.2.2 Regression of Price per Square Metre against Time Taken to MRT/LRT	13
6.3.2.3 Regression of Price against Time Taken to MRT/LRT for Different Regions	14
6.4 Resale price increases non-linearly as the floor area increases	15
6.4.1 Relationship between Resale Prices and Floor Area	15
6.4.2 Regression Model and Interpretation	15
6.4.2.1 Regression of Resale Price against Floor Area	15
6.4.2.2 Regression of Ln of Resale Price against Floor Area	16
<b>7 Discussion</b>	<b>17</b>
7.1 Endogeneity	17
7.2 Best Fit Model	18
7.3 Validity	18
7.4 Insights to Our Target Audience	18
<b>8 Conclusion</b>	<b>19</b>
<b>9 References</b>	<b>19</b>

---

# 1 Background Information

## 1.1 HDB flats

HDB (Housing Development Board) flats in Singapore are subsidised public housing meant for Singaporeans. These housing units are sold to Singaporeans under certain requirements, for a 99-year leasehold. In 2021, 78.7%<sup>1</sup> Singaporeans lived in a flat built by HDB. Buying a HDB flat is one of the most talked about subjects in Singapore among young adults, and it is also possibly one of the largest financial commitments for them. HDB flats can be bought through two different ways, buying a new HDB flat, known as a Build-To-Order (BTO), or buying a resale HDB flat.

---

<sup>1</sup> <https://www.singstat.gov.sg/find-data/search-by-theme/households/households/latest-data>

## 1.2 Resale HDB flats

Pricing of HDB flat resale prices are negotiated between sellers and buyers, and not done through HDB. In recent times, there has been an inconsistent supply of new HDB flats, exacerbated by the COVID-19 pandemic, resulting in multiple construction delays. This, together with the fact that home-owners do not need to wait 4 to 5 years for a resale flat, has resulted in an increasing demand for HDB resale flats. In the month of November 2021 alone, a whopping 29 resale flats were sold for over SGD \$1 million. Resale flat prices in Singapore rose 12.5% for the whole of 2021<sup>2</sup>, showing an ever-increasing trend of HDB resale prices.

BTO flats tend to be heavily subsidised by the Government to help new home-owners get their first house. HDB Resale Flat prices on the other hand are not controlled by the government, but rather by supply and demand forces in the market. Hence, we utilise Resale Flats over BTO Flats because we can omit the influence of the government on the housing prices for Resale Flats.

## 2 Motivation

### 2.1 Rationale of Project

With constantly rising HDB prices, our group is concerned with certain factors that influence resale flat prices. We intend to utilise publicly-available data from data.gov.sg to estimate the relationship between our chosen variables and the prices of resale flats. We aim to use this data to accurately determine the future prices, empowering new home buyers to make an informed decision when looking for HDB resale flats.

### 2.2 Target Audience

Our group aims to help potential buyers use our findings to make more informed decisions when looking for housing in Singapore. We have narrowed down and split our potential buyers into 2 main categories – Investors and Families.

With huge prices incurred by both groups of target audiences, potential buyers wish to gain the most value and purchase the house that best suits their preferences. To make the most informed decision, it is paramount for them to understand more about the factors that affect the prices of the resale flats.

---

<sup>2</sup>

<https://www.straitstimes.com/singapore/housing/hdb-resale-prices-surged-125-in-2021-the-biggest-rise-since-2010-flash-data>

### 2.2.1 Investors

Investors look towards the housing market as an open door for opportunities to increase their wealth. They go into the HDB resale market with the intention of selling it for a profit, and also make earnings from renting out the rooms.

### 2.2.2 Families

Meanwhile, families look for their dream homes that they wish to spend their lives in. Some families are unable to wait for the BTO flats due to personal reasons, and opt for resale houses as an alternative to expedite their home-attaining process. This is highly relevant in the current day because of the extremely long wait time for BTO flats due to the pandemic.

## 3 Project Question

What factors affect the prices of HDB resale flats in Singapore?

## 4 Hypotheses

### 4.1 Resale price increases as it moves towards the central regions

We investigate how the HDB resale prices vary according to the region they are located in. We have segregated the prices by their location data that fall primarily into 3 regions. These are Core Central Region (CCR), Rest of Central Region (RCR), and Outside Central Region (OCR).

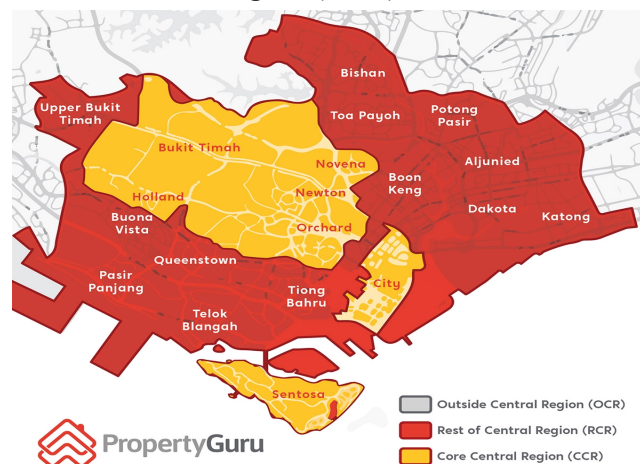


Fig. 1<sup>3</sup>

<sup>3</sup> <https://www.propertyguru.com.sg/property-guides/ccr-ocr-rcr-region-singapore-ura-map-21045>

## 4.2 Resale price decreases as the walking time to the nearest transportation increases

We want to investigate how one additional minute of walking time changes the HDB resale price. Specifically, we will be looking into the time taken to walk to the closest Mass Rapid Transit (MRT) or the closest Light Rail Transit (LRT).

## 4.3 Resale price increases non-linearly as the floor area increases

The relationship between flat price may be modelled differently. We will investigate if the relationship is linear or nonlinear. We want to test non-linear models, such as polynomial models and logarithmic models.

## 4.4 Relevance of Hypotheses to Target Audiences

The hypotheses can be contextualised differently by different stakeholders, who are our two target audiences. As mentioned in Section 2.2, different target audiences have different interests. The four hypotheses aim to serve the different audiences and their interests adequately.

### 4.4.1 Investors

Results from hypothesis 1 (Section 6.2) offer investors insights into good property locations. Different regions attract investors of different portfolios.

Results from hypothesis 2 (Section 6.3) provide investors insights into whether a flat is advantageous in terms of its distance to the nearest public transportation. This might be an important decision for an investor when they purchase it to rent out.

### 4.4.2 Families

Results from hypothesis 3 (Section 6.4) allow families of different sizes to make a better informed decision when purchasing housing. They can compare the size and pricing and make the most appropriate decision when selecting their flat. They can hence weigh their perceived benefits against the cost of the premium and decide whether they are willing to pay for it.

# 5 Data

## 5.1 Basic Data Statistics

<b>Variable</b>	<b>Variable Description</b>	<b>Mean</b>	<b>Std</b>	<b>Min</b>	<b>Max</b>
Resale_Price (Dependent)	Resale Price of the HDB (\$)	463,411	160,313	140,000	1,388,889
ln(Resale_Price) (Dependent)	Natural Logarithm of HDB Resale Price (\$)	12.99	0.331	11.849	14.144
floor_area_sqm	Area of the flat (sqm)	97.806	24.112	31	249
CCR	Binary Variable (HDB in Core Central Region =1, not in = 0)	0.011	0.105	0	1
RCR	Binary Variable (HDB in Rest of Central Region =1, not in = 0)	0.180	0.384	0	1
OCR	Binary Variable (HDB in Outside Central Region =1, not in = 0)	0.809	0.393	0	1
Mean Story Range	Mean of each Storey Range that consists of 3 floors (\$)	8.727	5.911	2	50
remaining_lease	Number of years remaining in the lease (years)	74.409	13.420	43	97
Time_taken_ to_MRT	Time taken to walk to the nearest MRT/LRT (min)	8.901	5.284	1.000	44.000

**Fig. 2**

### 5.1.1 Correlation

#### 5.1.1.1 Correlation Matrix

	Resale_Price	Floor_area_sqm	CCR	RCR	OCR	Mean Story Range	Remaining_lease	Time_taken_to_MRT
Resale_Price	1							
Floor_area_sqm	0.620	1						
CCR	0.134	-0.037	1					
RCR	0.237	0.181	-0.049	1				
OCR	-0.267	0.187	-0.218	<b>-0.963</b>	1			
Mean Story Range	0.370	-0.008	0.100	0.146	-0.169	1		
Remaining_lease	0.335	0.151	-0.044	-0.242	0.248	0.279	1	
Time_taken_to_MRT	-0.118	0.039	-0.083	-0.048	0.069	-0.122	-0.209	1

Fig. 3

#### 5.1.1.2 Issues with Multicollinearity

Multicollinearity is a situation whereby when two or more predictors are highly linearly related, that is when  $|r| \geq 0.7$ , where  $r$  is the correlation coefficient between the 2 variables. Therefore, it is imperative to check for any correlation coefficient that is too high and exclude one of the two variables in our regression model.

From the correlation matrix that we obtained, we see that there is only one pair of predictors with very high correlation coefficient value ( $r = -0.963$ ) that is between OCR (HDB in the Outside Central Region) and RCR (HDB at the rest of the Core Central Region). Hence, we have decided to exclude the OCR variable in our following regression models in order to prevent the issue of multicollinearity. This addresses the dummy variable trap.



## 5.1.2 Feature Engineering

### 5.1.2.1 Time Taken to Nearest MRT/LRT

Another column was created by gathering the coordinates of the MRTs, LRTs and HDB Resale Flats using OneMapAPI.<sup>4</sup> Next, the distance in metres was calculated and walking speed was determined to be 80 metres per minute, consistent with the standard<sup>5</sup> used in real estate listings.

nearest_distance_to_mrt	nearest_mrt	time_to_mrt
1003.555821	Ang Mo Kio	13
1267.331770	Ang Mo Kio	16
1070.949265	Yio Chu Kang	14

**Fig. 4: New column generated through Feature Engineering**

### 5.1.2.2 Region of HDB Resale Flat

We also added in three binary variables: CCR, RCR and OCR. They denote whether a HDB resale flat is within the region or not.

---

<sup>4</sup> <https://www.onemap.gov.sg/>

<sup>5</sup> <https://wikipedia.org/how-long-does-it-take-to-walk-100-meters-2/>

## 5.2 Charts on Basic Data

### 5.2.1 Breakdown of the HDB Resale Prices

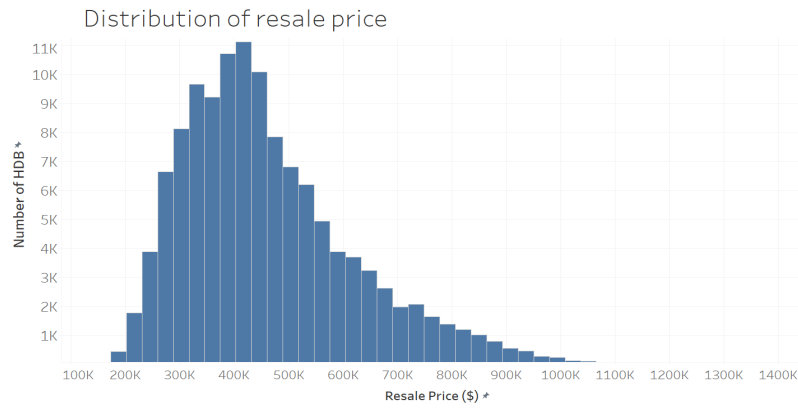


Fig. 5

### 5.2.2 Distribution of Region that Resale HDB is located

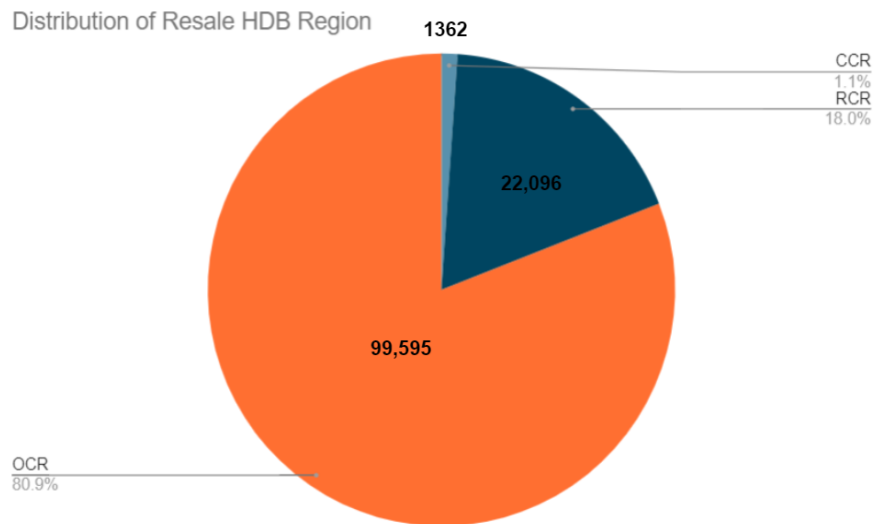


Fig. 6

## 6 Regression Models

### 6.1 Reference Models

#### 6.1.1 Ref. Regression 1: Resale Price against Independent Variables

as Fig. 6

$$\text{Resale\_price} \sim \text{CCR} + \text{RCR} + \text{floor\_area\_sqm} + \text{remaining\_lease} \\ + \text{storey\_range} + \text{time\_to\_mrt}$$

#### 6.1.2 Ref. Regression 2: Resale Price per Sq Metre against Independent Variables

as Fig. 7

$$\text{price\_per\_sqm} \sim \text{CCR} + \text{RCR} + \text{remaining\_lease} \\ + \text{storey\_range} + \text{time\_to\_mrt}$$

<b>Dep. Variable:</b>	resale_price	<b>R-squared:</b>	0.708
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.707
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	4.602e+04
<b>Date:</b>	Mon, 25 Apr 2022	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	00:04:03	<b>Log-Likelihood:</b>	-1.4588e+06
<b>No. Observations:</b>	114162	<b>AIC:</b>	2.918e+06
<b>Df Residuals:</b>	114155	<b>BIC:</b>	2.918e+06
<b>Df Model:</b>	6		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
<b>Intercept</b>	-2.836e+05	1930.797	-146.884	0.000	-2.87e+05	-2.8e+05
<b>floor_area_sqm</b>	4352.6598	10.790	403.383	0.000	4331.511	4373.809
<b>remaining_lease</b>	3207.4852	21.519	149.053	0.000	3165.308	3249.662
<b>storey_range</b>	5941.6767	46.543	127.660	0.000	5850.453	6032.900
<b>time_to_mrt</b>	-819.0765	49.755	-16.462	0.000	-916.595	-721.558
<b>CCR</b>	2.543e+05	2469.997	102.945	0.000	2.49e+05	2.59e+05
<b>RCR</b>	1.66e+05	713.960	232.514	0.000	1.65e+05	1.67e+05

<b>Omnibus:</b>	8365.347	<b>Durbin-Watson:</b>	0.566
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	10984.150
<b>Skew:</b>	0.656	<b>Prob(JB):</b>	0.00
<b>Kurtosis:</b>	3.766	<b>Cond. No.</b>	1.24e+03

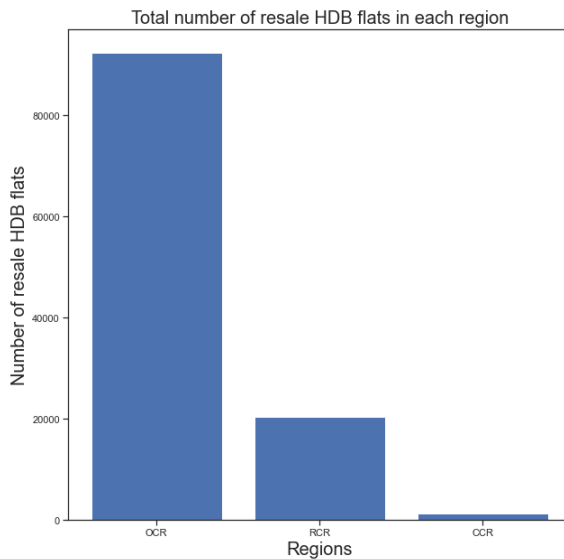
Dep. Variable:	price_per_sqm	R-squared:	0.555			
Model:	OLS	Adj. R-squared:	0.555			
Method:	Least Squares	F-statistic:	2.851e+04			
Date:	Mon, 25 Apr 2022	Prob (F-statistic):	0.00			
Time:	00:28:52	Log-Likelihood:	-9.3216e+05			
No. Observations:	114162	AIC:	1.864e+06			
Df Residuals:	114156	BIC:	1.864e+06			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1274.0280	17.156	74.261	0.000	1240.402	1307.654
remaining_lease	35.5209	0.212	167.621	0.000	35.106	35.936
storey_range	61.2320	0.462	132.603	0.000	60.327	62.137
time_to_mrt	-8.8274	0.493	-17.908	0.000	-9.794	-7.861
CCR	2749.5118	24.495	112.248	0.000	2701.502	2797.522
RCR	1740.5638	7.016	248.090	0.000	1726.813	1754.315
Omnibus:	8790.613	Durbin-Watson:	0.591			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	11927.440			
Skew:	0.664	Prob(JB):	0.00			
Kurtosis:	3.863	Cond. No.	751.			

Fig. 6 and 7

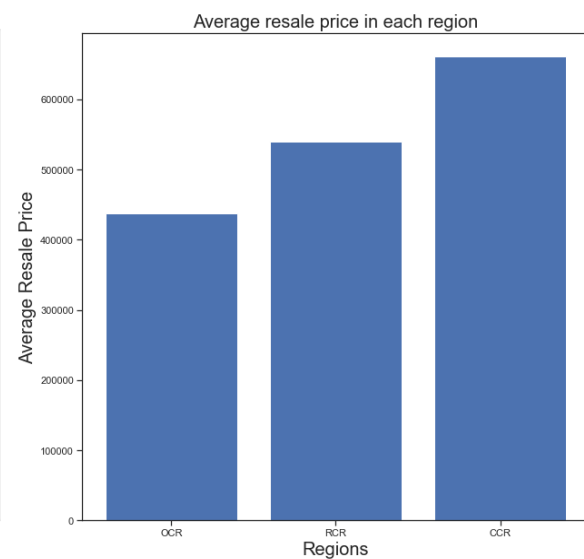
## 6.2 Resale price increases as it moves towards the central regions

### 6.2.1 Regions

HDB flats in Singapore are broken down into 3 regions, CCR, RCR and OCR, as explained earlier in Section 4.1. To investigate this hypothesis, we split the data based on the location of the HDB flat and the corresponding region to create the 3 binary variables.



**Fig. 8** Number of Flats by Regions



**Fig. 9** Average Resale Price by Regions

Figure 8 shows the number of resale HDB flats across the 3 different regions. This illustrates that most HDB flats are located in OCR, followed by RCR and lastly CCR in Singapore. This number can be significant in determining the price due to the relationship of the supply and demand of resale HDB flats.

To visualise an estimated pricing for each region, we calculated the average resale prices of HDB flats in each region, as shown in figure 9.

### 6.2.2 Regression Model and Interpretation

We ran a multiple regression of resale price against RCR and CCR and omitted OCR, as we use OCR as the reference variable while RCR and CCR are the included dummy variables. The remaining variables are kept as constants.

#### 6.2.2.1 Regression of Price against RCR and CCR

Region	Coefficient of Independent Variable	p-value
RCR	166,005.33	~0.000
CCR	254,272.63	~0.000

**Fig. 10**

Based on figure 10, we can interpret that an HDB flat located in the RCR region has a higher price on average by \$166,005.33 than a flat in the OCR region, with all other variables kept constant.

Furthermore, for a HDB flat located in the CCR region, it has a higher price on average by \$254,272.63 than a flat in the OCR region, with all other variables kept constant.

The p-values of the coefficients are all  $\sim 0.000$ , which means that at a 95% confidence level, the coefficients are statistically significant as it is less than 0.05.

With this, we can observe the trend of how moving towards the central region of Singapore, the prices of HDB resale flats generally increase.

#### 6.2.2.2 Regression of Price per Square Metre against RCR and CCR

Furthermore, we wanted to explore if the regions are significant when regressing it on price per square metre, instead of the price itself. This is to get a better indicator of resale price in terms of the floor area, so homebuyers can use it for simple comparisons between houses. To avoid multicollinearity between the dependent variable and the independent variable of floor\_area\_per\_sqm, we omitted it in this regression.

Region	Coefficient of Independent Variable	P-value
RCR	1740.56	$\sim 0.000$
CCR	2749.51	$\sim 0.000$

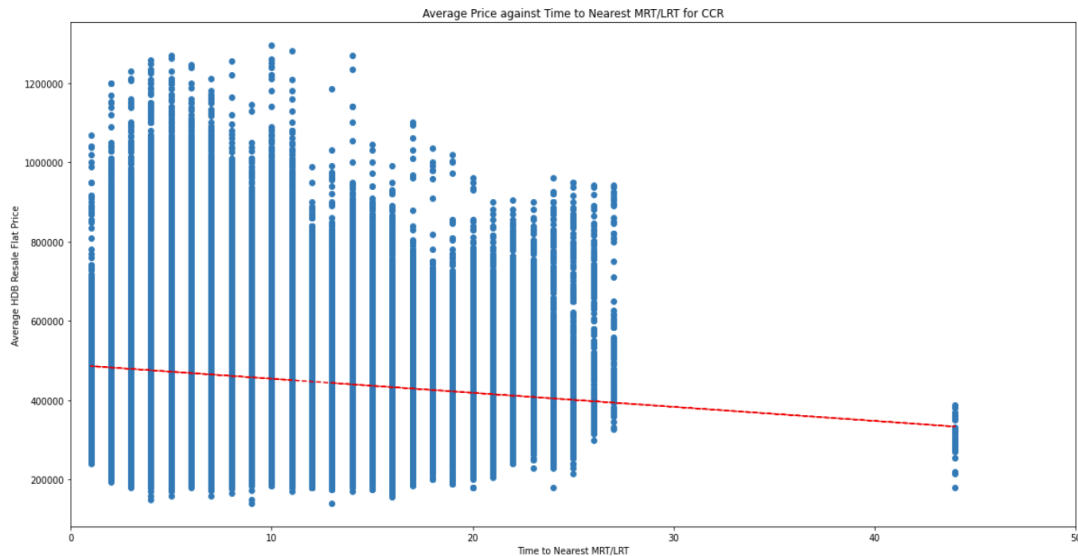
**Fig. 11**

Based on the model, we can interpret that a HDB flat located in the RCR region has a higher price on average by \$1740.56 per square metre than a flat in the OCR region, with all other variables kept constant. Furthermore, the HDB flat located in the CCR region has a higher price on average by \$2749.51 per square metre than a flat in the OCR region, with all other variables kept constant. The p-values of the coefficients are all 0.000, at a 95% confidence level, the coefficients are statistically significant as it is less than 0.05.

As these coefficients are also significant, this can be a better indicator towards homebuyers as it simplifies the prices to per square metre, allowing potential homebuyers to easily compare with other houses, based on per square metre value. This shows that as the HDB flat is located closer to the central regions of Singapore, the price per square metre of the house will increase.

## 6.3 Resale price decreases as the walking time to the nearest transportation increases

### 6.3.1 Relationship between Resale Prices and Time Taken to Nearest MRT/LRT



**Fig. 12:** Scatter Plot of Average HDB Resale Flat Price against Time to Nearest MRT/LRT

As expected, we see a general decrease in average HDB resale flat prices as time to the nearest MRT/LRT increases.

### 6.3.2 Regression Model and Interpretation

#### 6.3.2.1 Regression of Price against Time Taken to MRT/LRT

From our reference regression 1 (Fig. 6), we see that the coefficient associated with “time\_to\_mrt” is  $-819.0765$ . This means that for every one minute increase to reach an MRT/LRT, the flat price decreases by \$819.08 dollars. This is because it is more inconvenient for the flat owner to travel via public transport, hence causing a reduction in price. We see that the p-value associated with this coefficient is  $\sim 0.000$ , which is significant at 1% significance level. This means that the coefficient is statistically significant.

#### 6.3.2.2 Regression of Price per Square Metre against Time Taken to MRT/LRT

As covered in reference model 2 (Fig. 7), we change the dependent variable from price to price\_per\_sqm to account for the different sizes of the HDB flats.

From Figure 7, we see that the coefficient associated with “time\_to\_mrt” is  $-8.8274$ . This means that for every one minute increase to reach an MRT/LRT, the price per square metre decreases by

\$8.8274 dollars. We see that the p-value associated with this coefficient is ~0.000, which is significant at 1% significance level. This means that the coefficient is also statistically significant.

### 6.3.2.3 Regression of Price against Time Taken to MRT/LRT for Different Regions

We go on to determine the relative impact of time taken to MRT/LRT based on the different regions. This is done by filtering the HDBs based on the three regions, then performing the same regression and omitting the independent variables CCR and RCR.

Below are the three regression results:

**Regression Model for each region (CCR, RCR, OCR)**  
*Resale\_price ~ floor\_area\_sqm + remaining\_lease*  
*+ storey\_range + time\_to\_mrt*

```

Regression Model for CCR
Intercept          -433358.164863
floor_area_sqm      6151.334698
remaining_lease     7331.091099
storey_range        7155.705861
time_to_mrt         -13161.385567
dtype: float64

Regression Model for RCR
Intercept          -390251.710998
floor_area_sqm      5567.446028
remaining_lease     5923.732037
storey_range        5657.226023
time_to_mrt         -2641.916984
dtype: float64

Regression Model for OCR
Intercept          -139112.902713
floor_area_sqm      3887.851856
remaining_lease     2055.901497
storey_range        4861.051625
time_to_mrt         -987.029090
dtype: float64

```

**Fig. 13:** Regression Outputs for Different Regions

The important data is summarised in the table below:

Region	Coefficient of Time Taken to MRT/LRT	P-value
CCR	-13161.39	~0.000
RCR	-2641.92	~0.000
OCR	-987.03	~0.000

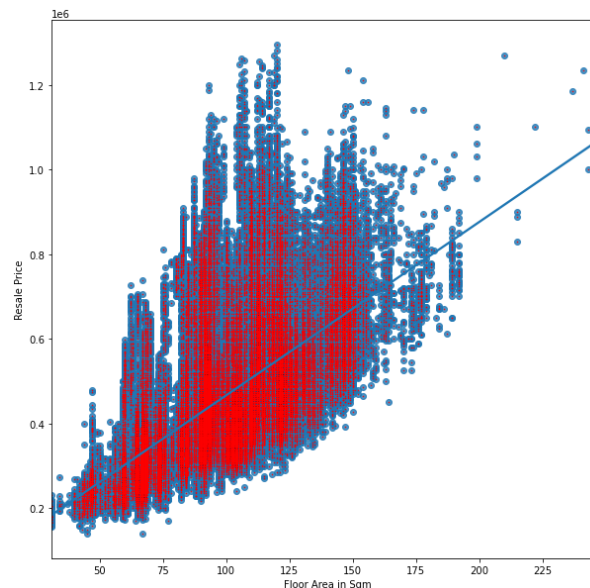
**Fig. 14**

The CCR region primarily consists of the prime areas of Singapore and the Downtown Core, whereby the RCR is considered the intermediate tier consisting of medium-priced tiers of condos and properties, and the OCR being the cheaper-tier region covering the majority of Singapore. Hence, we can generally order the regions from most to least expensive in the following order:  $CCR > RCR > OCR$ .

We see that there is an increasingly negative effect of time taken to the nearest MRT/LRT as the region gets more expensive. For instance, when considering an expensive HDB Resale Flat in the CCR, one additional minute to travel to a MRT/LRT leads to a \$13,161.39 decrease in price, whereby the number is only -987.03 for the OCR.

## 6.4 Resale price increases non-linearly as the floor area increases

### 6.4.1 Relationship between Resale Prices and Floor Area



**Fig. 15**

Floor area of a HDB resale flat is measured in square metres (sqm). Plotting the resale prices of HDB flats against its floor area, we see a general upwards price trend as the HDB unit increases in size measured in square metres. Hence, taking into consideration the other variables in the dataset as control, we came out with the regression model seen in figure 6.

### 6.4.2 Regression Model and Interpretation

#### 6.4.2.1 Regression of Resale Price against Floor Area



From our reference regression 1 (Fig. 6, the coefficient of the interest variable ‘floor\_area\_sqm’ has a p value of ~0.00, which means that at a 95% confidence level, the coefficients are statistically significant as it is less than 0.05. The coefficient is approximately 4352.7, suggesting a positive association between the floor size of a flat and its resale price. In other words, an increase in 1 square metre is associated with an increase in \$4352.7 on its resale price.

With reference to figure 15, we see that the relationship between floor area and resale prices are not exactly linear so even though a linear model is appropriate, we also decided to explore the relationship between floor area of a flat and the natural log of its resale prices.

#### 6.4.2.2 Regression of Ln of Resale Price against Floor Area

$$\ln price \sim CCR + RCR + floor\_area\_sqm + remaining\_lease + storey\_range + time\_to\_mrt$$

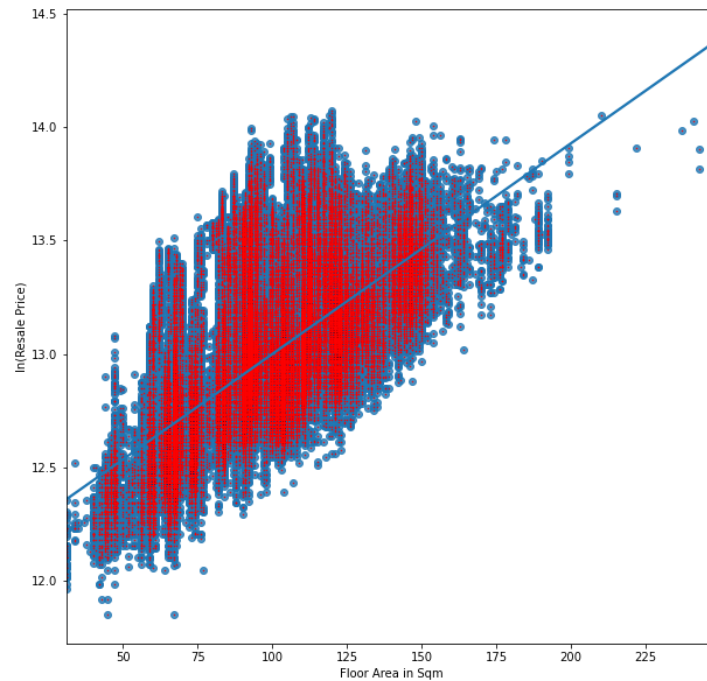
OLS Regression Results						
Dep. Variable:	Inprice		R-squared:	0.743		
Model:	OLS		Adj. R-squared:	0.743		
Method:	Least Squares		F-statistic:	5.499e+04		
Date:	Wed, 20 Apr 2022		Prob (F-statistic):	0.00		
Time:	16:54:46		Log-Likelihood:	41901.		
No. Observations:	114162		AIC:	-8.379e+04		
Df Residuals:	114155		BIC:	-8.372e+04		
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	11.3407	0.004	3004.731	0.000	11.333	11.348
storey_range	0.0108	9.1e-05	118.197	0.000	0.011	0.011
floor_area_sqm	0.0096	2.11e-05	457.314	0.000	0.010	0.010
remaining_lease	0.0075	4.21e-05	179.256	0.000	0.007	0.008
time_to_mrt	-0.0023	9.73e-05	-23.449	0.000	-0.002	-0.002
CCR	0.4572	0.005	94.682	0.000	0.448	0.467
RCR	0.3063	0.001	219.439	0.000	0.304	0.309
Omnibus:	1268.226	Durbin-Watson:	0.518			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1310.655			
Skew:	0.261	Prob(JB):	2.48e-285			
Kurtosis:	3.059	Cond. No.	1.24e+03			

Fig. 16

Using the natural log of resale prices as a dependent variable is appropriate as well in reflecting the association between the 2 variables. From figure 16, we observe that holding other control variables constant, a 1 unit increase in square metre gives a 0.96% increase in resale prices of HDB flats. The coefficient of the ‘floor\_area\_sqm’ also has a very small, negligible P value (~0.00) that shows it is statistically significant at 95%. We observe a positive association

between the natural log of resale prices and the floor area, which corroborates with the previous linear model.

In summary, both models are appropriate in reflecting the association between both variables, but can provide different insights depending on the nature of the resale prices variable. Each model has its strengths and serves the intended different target audiences, based on the type of interpretations required.



**Fig. 17**

Figure 17 is a simple plot of the natural log of resale prices against the flat's floor area measured in square metres. We observe a more uniform graph, with the regression line running through the graph in a more fitting manner.

## 7 Discussion

### 7.1 Endogeneity

When analysing our models, it is important to note that there exists some endogeneity issues such as omitted variables bias that have affected the internal validity of our regression model. Omitted variables bias arises from the limitation of our data set, since it does not come with important data such as the proximity to the nearest shopping mall or school which are correlated with both the response variable - HDB resale price and other explanatory variables in the model such as the region that the HDB falls in. And to address the omitted variables bias, we add in

more control variables such as time taken to nearest MRT through feature engineering as described in section 5.

## 7.2 Best Fit Model

A model which we consider to be over arching most of our findings would be the regression based on the natural logarithm of resale prices as dependent variable. We believe it provides a more linear representation in the prices and it helps our target audiences better relate to how the relationship of the resale prices and each independent variable changes. A curved and wide dataset might not be visually digestible nor understandable for the general audience. Hence it could benefit our target audience by providing a simple statistically significant model for them to base their decisions on. This would allow them to retrieve the important analyses at one glance. The relationship between resale price and square metre is an important indicator for homebuyers and investors as space is an important factor whenever considering purchasing houses. It might be possible to provide a natural logarithm of price per square metre to further enhance the understanding of the resale prices and how it changes with respect to the regions or even lease period.

## 7.3 Validity

In terms of validity, we interpret our findings using two types of validity. We believe our model has high internal validity because we are confident that there is a cause-and-effect relationship established in our study. We are confident in our model's external validity as well because we believe our model's findings can be generalised to outside our population, for instance, when predicting the prices of Private Condominiums or Landed Properties in Singapore. This is because the variables we selected are generally applicable for all prospective home buyers, regardless of whether they are looking for HDB Resale Flats specifically or not.

## 7.4 Insights to Our Target Audience

We summarise our findings and present them to our target audiences.

For the investors, we saw that the regions do heavily influence resale flat prices, as does the distance to the nearest public transportation. Investors can and should focus more heavily on these factors when considering the purchase of a HDB resale flat because they can have a great impact on the current price, as well as on the appreciation of the property's value over time. Investors can also look at anomalies, such as resale flats that are potentially undervalued from our model. This could point towards a possible appreciation of the flat value once future buyers realise the true value of the resale flat. Such buying opportunities help our investors to profit off the purchasing and selling of properties in the Singapore property market.

For the family looking towards the flats for accommodation, we have concluded that size, together with storey level, play a heavy impact on resale prices as well. Our model shows that for every additional increase in storey level, the flats costs an additional \$5,941. When considering the pros and cons of a unit, families can take this into consideration to have a better understanding about where and how to channel their funds. For instance, they could opt for a lower, and cheaper, floor and spend the additional money saved towards renovating their house. Such trade-offs empower families to make informed decisions.

## 8 Conclusion

In conclusion, our group has successfully identified some variables that contribute to the increase in HDB Resale Flat prices. Key variables include the region, remaining lease, and time taken to the nearest MRT/LRT. There are certain HDB flats that do not follow the general trend, for instance, they are very close to the MRT, yet they are cheaper than other houses in their proximity. These are houses that both investors and families can seek out to enjoy the benefits of living closer to public transportation while forking out less money. We can utilise our results in a variety of ways. As a potential home buyer, I will need to consider whether I am willing to fork out \$819 for enjoying an additional minute of proximity to the MRT/LRT station. As a resale flat seller on the market, I can price my flat accordingly, knowing the value that my property holds due to its proximity to public transportation.

## 9 References

1. *Households - latest data*. Base. (n.d.). Retrieved April 27, 2022, from <https://www.singstat.gov.sg/find-data/search-by-theme/households/households/latest-data>
2. Ng, M. (2022, January 3). *HDB resale prices surged 12.5% in 2021, biggest rise since 2010: Flash Data*. The Straits Times. Retrieved April 27, 2022, from <https://www.straitstimes.com/singapore/housing/hdb-resale-prices-surged-125-in-2021-the-biggest-rise-since-2010-flash-data>
3. *Singapore district map: Defining the CCR, RCR and OCR by ...* (n.d.). Retrieved April 27, 2022, from <https://www.propertyguru.com.sg/property-guides/ccr-ocr-rcr-region-singapore-ura-map-21045>
4. OneMap. (n.d.). Retrieved April 27, 2022, from <https://www.onemap.gov.sg/>
5. Editors, W. (2021, September 2). *How long does it take to walk 100 meters?* Wikipedia Encyclopedia. Retrieved April 27, 2022, from <https://wikipedikia.org/how-long-does-it-take-to-walk-100-meters-2>