BT2101 Final Report

Name	Matriculation Number	Email
Angel Koh Xue Qi	A0223001J	e0564042@u.nus.edu
Nguyen Hong Phuc Nguyen	A0243332R	e0817834@u.nus.edu
Eve Wang Zhi Ru	A0221476H	e0556668@u.nus.edu
Guan Xinyi	A0221713U	e0559245@u.nus.edu
Zhao Xi	A0223275J	e0564316@u.nus.edu

Table of Contents

Introduction 3
- Background 3
- Problem Statement 4
- Assumptions 5

Dataset 6
- Description of Dataset 6
  - Description of variables 6
  - Investigating the variables 8

Hypotheses 10
- Proximity from Central Area 10
- Number of Hospitals 11
- Number of MRT Stations 13
- Storey of the HDB flat 14

Conclusion 16
- Decision 16
- Evaluation 16

References 18

Introduction

1. Background

HDB resale prices have been a long-standing discussion in Singapore, as it is an affordable alternative to Build-To-Order (BTO) houses. Generally, the geographics of Singapore can be broadly categorised into three different regions — Core Central Region (CCR), Rest of Central Region (RCR) and Outside Central Region (OCR), depending on the distance of the districts from the city area.

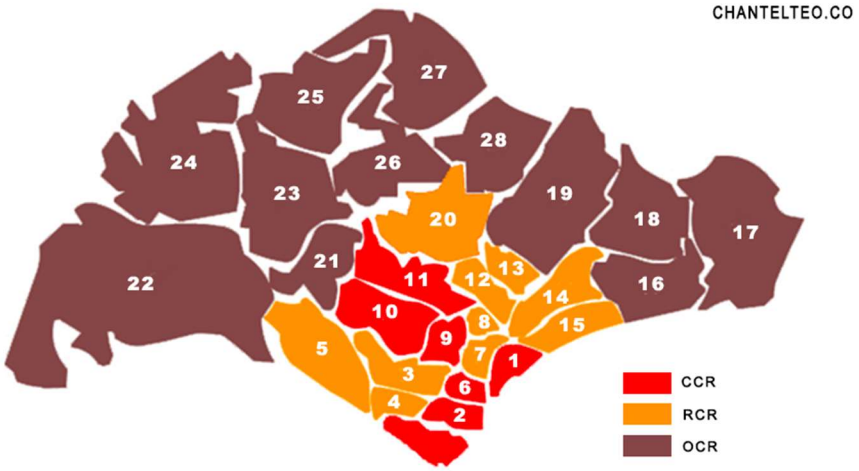

Figure 1: Regions in Singapore (Teo, 2020)

The central region has always been the most coveted area to live in, due to its prime location and its reputation as the most luxurious area in Singapore. Therefore, the prices of property are the highest in this region. In addition, this region is the most densely populated area in Singapore, putting a heavy strain on infrastructure such as public transport and roads especially during peak hours (My Exclusive Condo, 2020). As such, the government has been putting in conscientious efforts into decentralising the demand for housing across all regions instead of having it concentrated in the CCR and RCR (Cheong, 2018; Housing & Development Board, 2022; Lin, 2021).

Currently, the strain on central areas is expected to rise with the increasing demand for HDB flats as there are many corporate offices in the vicinity. Therefore, decentralising is important to relieve the strain on this area and spread out the demand to other regions. In addition, having offices and housing concentrated in the central region can bring about diseconomies of scale. Thus, decentralising can prevent this and bring about economic benefits to the country. Decentralising by having more malls, offices etc. in non-central regions also improves accessibility and affordability to residents, thereby enhancing quality of life and standard of living. To add on, with decentralisation, proximity to various amenities and office buildings is increased, commute time is significantly reduced and the number of vehicles will decrease, thus helping to meet governmental air quality targets and reducing the need for wide roads and flyovers, allowing cost savings for the government. (Condo, 2020)

The authorities have implemented multi-pronged policies to promote decentralisation of the central area. One of the policies is Land Transport Authority's (LTA) Land Transport Master Plan 2040 (LTMP 2040), which aims to reduce the commute time to downtown city to 45-minute during most peak hours by connecting more places by train (Land Transport Authority, 2020). New Mass Rapid Transit (MRT) lines such as Thomson-East Coast Line and Cross Island Line are also under construction to improve connectivity in non-central areas.

Another strategy to decentralise reliance on the Central Business District (CBD) area is to develop other office spaces in non-central areas such as Paya Lebar, Tampines and Jurong, which belong to the outskirts of the city (My Exclusive Condo, 2020). They have also provided subsidies for commercial buildings in these areas to incentivise more companies to take up these rental spaces.

Figure 2: Previous and Current urban planning (PropertyLimBrothers, 2022)

2. Problem Statement

We are a group of policymakers who have been working on the decentralisation of housing demand in Singapore. **As such, we are trying to determine whether the effect of proximity on HDB resale prices across the regions is significant, and the possible factors influencing this effect.** In this report, we have decided to interpret the resale prices as the demand for the properties in Singapore, and we would like to conclude whether the policies implemented, like increasing the number of mrt stations, have caused this paradigm shift. In addition, we will also be checking whether new policies such as building more hospitals or higher level HDBs will help with the decentralisation.

3. Assumptions

- Results are constant across different time periods.
- Only 4 room flat type data are used in this analysis as it is the most common.

Dataset

1. Description of Dataset

The dataset we are using is “Singapore Resale Flat Prices” found on the Singapore government Singstats website (https://data.gov.sg/dataset/resale-flat-prices). This dataset includes over 120,000 records, each including a HDB flat being resold and its other relevant information from January 2017 onwards to April 2022.

Description of variables

1. month: The month of the resale transaction in YYYY-MM format.
2. town: The town where the flat was located.
3. flat_type: The type of HDB flat (7 distinct types). → We filtered for 4-room flats only since it is the most popular flat type
4. block: The block number.
5. street_name: The street on which the block of flats was located.
6. storey_range: The floor interval on which the flat was located. → We took the average of the storey range
7. floor_area_sqm: The area of the flat, in square metres.
8. flat_model: The model of the flat (20 distinct models).
9. lease_commence_date: The start year of the lease.
10. remaining_lease: The remaining time on the lease of the flat, when the sale was transacted, in years.
11. Remain_lease_m: The remaining time on the lease of the flat, when the same was transacted, in months.
12. resale_price: The resale price of the flat.
13. log_resale_price: We applied log transformation to the resale price as it does not follow normal distribution and has a large magnitude
13. n_hospitals: the number of hospitals in its town
14. n_mrt: the number of mrt stations in its town
15. Proximity: the distance between its town to the central area, with 0 for core central, 1 for rest of the central area, 2 to 4 for near central to far central.
16. Central: binary, =1 if town is in core central or rest of central area, = 0 for the non-central regions

Figure 3: Classification of proximity

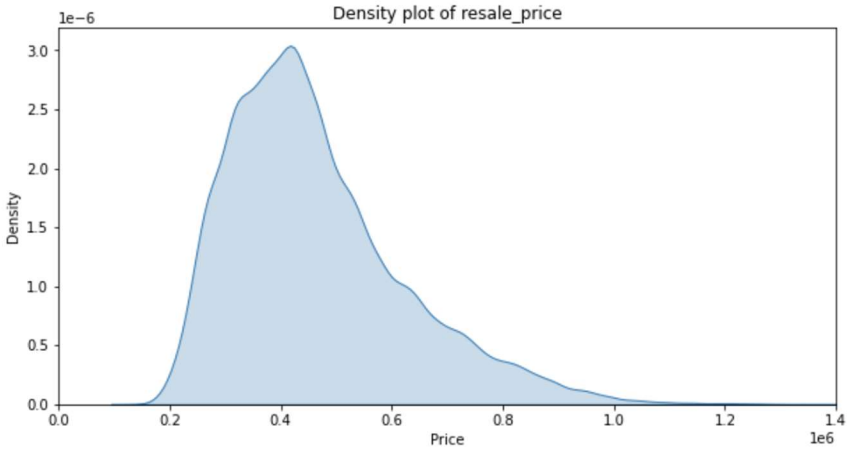

Figure 4: Density plot of Y variable (resale price)

As seen, the resale price is highly right-skewed. Observation in the skewed data will have a disproportionate effect on the parameter estimates. Skewness will also violate the assumption of normally distributed error at each X, hence the confidence interval calculated from the model might be too narrow or too wide. Hence, we applied log transformation to resale_price.

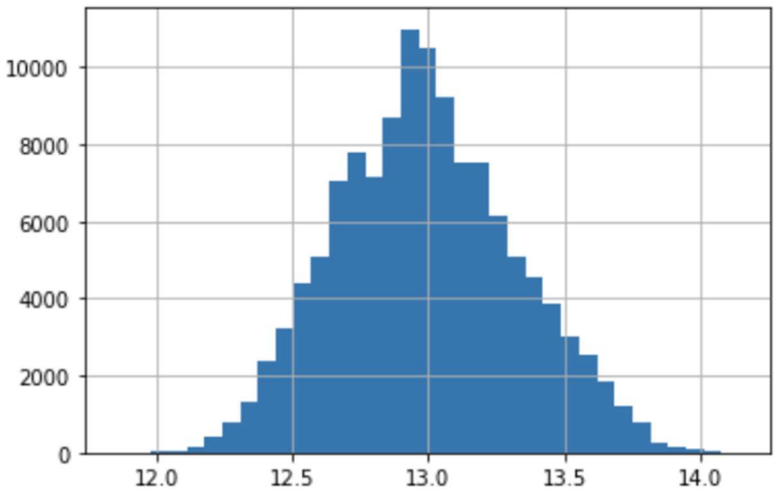

Figure 5: Histogram of log(resale_price), symmetrical as compared to resale_price

As shown, it has become much more symmetric, having its skewness reduced to only 0.165 as compared to 1.017 before the log transformation. The distribution is now normal.

Investigating the variables

	log_resale_price	floor_area_sqm	proximity	average_level	remain_lease_m	n_hospitals	n_mrt
count	51638.000000	51638.000000	51638.000000	51638.000000	51638.000000	51638.000000	51638.000000
mean	13.003457	95.129426	2.697703	6.625160	942.348813	1.288799	3.140575
std	0.248454	7.135004	1.073108	3.343173	152.202563	1.243035	2.860462
min	12.292250	70.000000	0.000000	2.000000	534.000000	0.000000	0.000000
25%	12.829334	91.000000	2.000000	5.000000	803.000000	0.000000	1.000000
50%	12.971540	93.000000	3.000000	6.000000	945.000000	1.000000	3.000000
75%	13.128345	102.000000	4.000000	8.000000	1109.000000	2.000000	4.000000
max	14.014361	145.000000	4.000000	27.000000	1171.000000	7.000000	23.000000

Figure 6: Descriptive statistics for all continuous variables

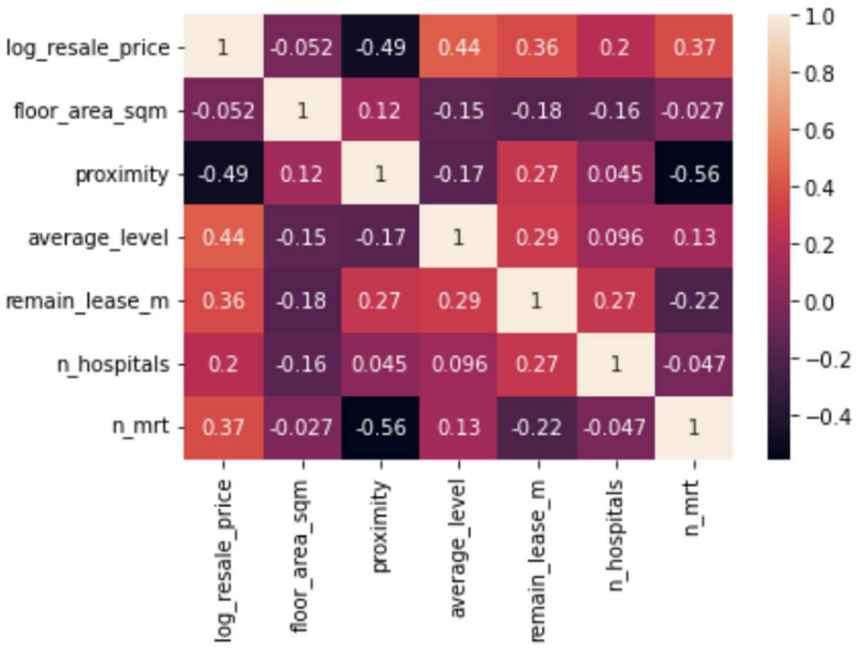

Figure 7: Correlation matrix visualised

As shown in the correlation matrix, all of the correlation between pairs of independent variables does not exceed 0.6, suggesting that there is no multicollinearity issue and fulfils the assumption of ordinary regression models.

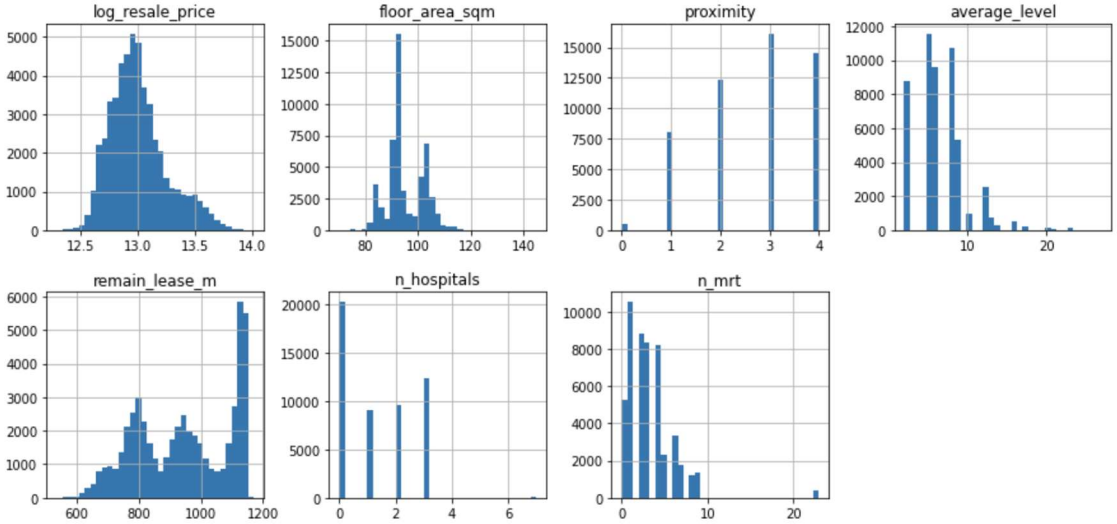

Figure 8: Histogram of respective variables

Although floor_area_sqm and remain_lease_m are skewed and do not show a normal distribution, we do not have to log transform them as these variables are only included as control in the model.

Hypotheses

1. Proximity from Central Area

Statement: Proximity to central Singapore has an effect on resale price of 4-room HDB from 2017 onwards.

$$log(resale\_price) = \beta_0 + \beta_1 * proximity + controls + \varepsilon$$
$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

We have decided to include floor area as our control variable, as it correlates with both the independent variable (IV), proximity and dependent variable (DV), resale price, which would result in omitted variable bias if not included in the regression model. This is owing to the fact that the larger the HDB housing unit, the higher its resale price. Moreover, the closer the proximity to the central region of Singapore, housing units for the same flat type tend to be smaller as there is limited land space nearer to the city area (99.co, 2022). Hence, it is an endogenous variable and we should include it as a control variable.

Remaining lease is also included as a control variable as it correlates with both the IV and DV. The shorter the remaining lease of a HDB, the lower its resale price due to depreciation over time (lifefinance, 2020). Remaining lease also correlates with the proximity, since HDBs nearer to central region are likely older and more mature, hence the remaining lease will tend to be lower than that farther from the central area (Adis, 2020).

Thus, running the multiple regression model at 5% significance level, we obtain the following estimated regression model:

	np.log(resale_price)
Intercept	11.5698***
	(0.0032)
proximity	-0.1459***
	(0.0005)
floor_area_sqm	0.0099***
	(0.0000)
remain_lease_m	0.0009***
	(0.0000)
R-squared	0.7259
R-squared Adj.	0.7259

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

$$log(resale\_price) = 11.5698\ (0.0032) - 0.1459\ (0.0005) * proximity$$
$$+0.0099\ (0.0000) * floor\_area\_sqm + 0.0009(0.0000) * remain\_lease\_m$$

Interpretation: Based on the estimated regression model, distance from the central area is significant as p-value < 0.05. Its coefficient has a value of -0.1459, which means that an increase in distance by 1 unit is associated with a -14.59% change in resale price.

The adjusted R-squared value is 0.7259, which means that around 72.59% of variation in resale prices can be explained by the model. Since the adjusted R-squared value greatly exceeds 30%, the model has high explainability power.

Contextual meaning:
Thus, as distance from the central area increases, there will be a 14.59% fall in HDB resale prices. This is likely due to the fact that central areas tend to be the most prime estates to live in, and its reputation of being high-end and luxurious likely gave rise to the high demand for housing in these areas.

2. Number of Hospitals

Statement: The effect of the number of hospitals on the resale price of 4-room HDB depends on whether the flat is located in the central area or not.
(i.e: the relationship between the number of hospitals around a flat and its resale price is different in central regions and non-central regions.)

$$log(resale\ price) = \beta_0 + \beta_1\, central + \beta_2\, n\_hospitals + \beta_3\ central * n\_hospitals + controls + \varepsilon$$

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0$$

Average storey level, floor area, and remaining lease are kept as the control variables.

	w/o interaction	with interaction
Intercept	11.8286***	11.8370***
	(0.0118)	(0.0118)
R-squared	0.5673	0.5680
R-squared Adj.	0.5673	0.5680
average_level	0.0178***	0.0177***
	(0.0002)	(0.0002)
central	0.3784***	0.3624***
	(0.0020)	(0.0027)
central:n_hospitals		0.0133***
		(0.0014)
floor_area_sqm	0.0046***	0.0045***
	(0.0001)	(0.0001)
n_hospitals	0.0257***	0.0228***
	(0.0006)	(0.0007)
remain_lease_m	0.0006***	0.0006***
	(0.0000)	(0.0000)

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

$$log(resale\ price) = 11.8370\ (0.0118) + 0.3624\ (0.0027) * central$$
$$+0.0228\ (0.0007) * n\_hospitals + 0.0133\ (0.0014) *\ central * n\_hospitals$$
$$+0.0045\ (0.0001) * floor\_area\_sqm + 0.0006\ (0.0000) * remain\_lease\_m$$

Interpretation: Compare the 2 models, as the coefficient for the interaction term is significant, and adjusted R^2 value has also increased slightly from 0.5673 to 0.5680, the effect of the number of hospitals on HDB resale price indeed depends on its region.

For a central flat, increasing the number of hospitals by 1 will lead to a $100 \times (0.0228 + 0.0133)\% = 3.61\%$ increase in resale price.
For a non-central flat with all the other variables kept the same, increasing the number of hospitals by 1 will lead to a $100 \times 0.0228\% = 2.28\%$ increase in resale price only.

Contextual meaning:
Especially in the era of pandemics, the access to medical resources and services is of utmost importance for families especially those with children and elderly. Hence, families would be willing to pay higher prices for more hospitals in the surroundings.

Increase of 1 hospital in the central region increases the resale price by a larger percentage. This could possibly be because the hospitals that choose to locate in the central areas are the largest such as Singapore General Hospital and Tan Tock Seng Hospital. These hospitals have larger capacity and more specialised doctors, thereby increasing the nearby flat price by a greater percentage.

3. Number of MRT Stations

Statement: The effect of the number of MRT stations on resale prices of 4-room HDB depends on the region that the flat is in.

$$log(resale\_price) = \beta_0 + \beta_1\, central + \beta_2\, n\_mrt + \beta_3\ central * n\_mrt + controls\ +\ \varepsilon$$

$$H_0: \beta_3\ = 0$$

$$H_1: \beta_3\ \neq 0$$

	w/o interaction	with interaction
Intercept	11.8024***	11.7781***
	(0.0119)	(0.0120)
R-squared	0.5656	0.5668
R-squared Adj.	0.5656	0.5668
average_level	0.0172***	0.0174***
	(0.0002)	(0.0002)
central	0.3214***	0.3563***
	(0.0024)	(0.0038)
central:n_mrt		-0.0080***
		(0.0007)
floor_area_sqm	0.0040***	0.0040***
	(0.0001)	(0.0001)
n_mrt	0.0127***	0.0174***
	(0.0003)	(0.0005)
remain_lease_m	0.0006***	0.0007***
	(0.0000)	(0.0000)

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

Interpretation: Comparing the 2 models, as the coefficient for the interaction term is significant, and adjusted R^2 value has also increased slightly from 0.5656 to 0.5668, the effect of the number of mrt stations nearby on HDB resale price indeed depends on its region.

For a central flat, increasing the number of mrt stations in its neighbourhood by 1 will lead to a $100 \times (0.0174 - 0.0080)\% = 0.94\%$ increase in resale price.
Meanwhile, for a non-central flat with all the other variables kept the same, increasing the number of hospitals by 1 will lead to a $100 \times 0.0174\% = 1.74\%$ increase in resale price.

Contextual meaning:
As MRT is one of the main public transport in Singapore, the number of MRT stations in the neighbourhood would largely represent the connectivity and accessibility of the housing unit. Greater accessibility leads to greater convenience and thus contributes to higher resale price.

Increase of 1 MRT station in the **non-central** region increases the resale price by a larger percentage. This could possibly be because of the immaturity of the public transport network in the non-central areas, thus building a new MRT station has a larger potential benefit in non-central areas than in central areas that have already been highly connected.

4. Storey of the HDB flat

Statement: The effect of the HDB flat's storey on resale prices depends on the region that the flat is in.

$$log(resale\_price) = \beta_0 + \beta_1\, central + \beta_2\, average\_level + \beta_3\ central * average\_level + controls\ +\ \varepsilon$$

$$H_0: \beta_3\ = 0$$

$$H_1: \beta_3\ \neq 0$$

We have decided to keep the floor area and remaining lease as control variables. Floor area is correlated with both resale price and storey level as flats with larger floor areas tend to have a higher resale price. In addition, flats on lower floors, especially the ground floors, tend to have larger floor areas due to the presence of Private Enclosed Spaces (PES) like an outdoor yard. Remaining lease is also controlled as resale prices tend to be higher with a longer remaining lease, and newer flats with longer remaining lease generally have higher storey levels due to land scarcity and changes in design of the HDB.

	w/o interaction	with interaction
Intercept	11.2740***	11.3059***
	(0.0034)	(0.0036)
R-squared	0.7254	0.7280
R-squared Adj.	0.7254	0.7280
average_level	0.0184***	0.0147***
	(0.0002)	(0.0002)
central	0.3207***	0.2409***
	(0.0013)	(0.0027)
central:average_level		0.0112***
		(0.0003)
floor_area_sqm	0.0096***	0.0096***
	(0.0000)	(0.0000)
remain_lease_m	0.0007***	0.0007***
	(0.0000)	(0.0000)

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

$$log(resale\_price) = 11.3059\ (0.0036)\ + 0.2409\ (0.0027) * central + 0.0147\ (0.0002) * average\_level\ + 0.0112\ (0.0003) * central * average\_level + 0.0096\ (0.0000) * floor\_area\_sqm + \ 0.0007(0.0000)\ remain\_lease\_m$$

Interpretation: Comparing the two models, the R2 value increased from 0.7254 to 0.7280, showing that the second model that includes the interaction term is a better fit, as more of the observed variations can be explained by the model's inputs. The standard error of central, average_level, floor_area_sqm and remain_lease_m remained roughly the same between the two models, showing that we did not introduce any new omitted variables by including the interaction term.

The coefficient of the interaction term central*average_level is statistically significant at the 1% level of significance as it has a p-value of less than 0.01. Thus, we reject the null hypothesis and conclude that the storey that the HDB flat is on affects resale prices depending on the region that the flat is in.

Increasing the floor of a HDB flat in a central area by 1 will lead to a $100 * (0.0147 + 0.0112)\% = 2.59\%$ increase in resale prices.
Meanwhile, keeping the storey level the same, a HDB flat in a non-central area will only experience a $100 * (0.0147)\% = 1.47\%$ increase in resale prices.

Contextual meaning:
The resale price of a HDB flat on higher levels in central areas is higher probably due to better views, as central areas tend to have more modern buildings and decorations. On the other hand, the view in non central areas may not be as ideal as the HDBs might be clustered together, and the surrounding area may consist of older infrastructure. Central areas also tend to be more noisy and lively, which may affect those living at lower levels, so it is more attractive to purchase a flat on a higher level. This is also the case in noncentral areas, as higher levels are more sought after.

Insights:
The demand for HDB flats on higher floors in non central areas may increase if we improve the views from the flats. This can be done by putting up more festive decorations in these areas, and ensuring that these areas are kept clean. Since non-central areas tend to be quieter, we can use this as a selling point for HDB flats there to increase the demand for them. We can also build higher flats in non central areas.

Conclusion

1. Decision

For decentralising strategies that are currently in place, we can conclude that increasing the number of MRT stations in non-central areas is effective. Increasing the number of MRT stations in non-central areas is positively correlated with an increase in resale price, which indicates an increase in demand for HDBs in non-central areas. However, we were unable to evaluate the effectiveness of increasing office spaces in non-central areas due to the lack of information.

A new decentralising strategy that should be implemented is the construction of more public hospitals in non-central areas. Increasing the number of public hospitals in non-central areas is positively correlated with an increase in resale price, which indicates an increase in demand for HDBs in non-central areas. To enhance this strategy, public hospitals should not be constructed in central areas. Increasing the number of public hospitals in central areas is positively correlated with an increase in resale price greater than that in non-central areas, which means that the demand increase would be higher if hospitals were constructed in central areas than in non-central areas. Thus, to prevent an increase in demand for HDBs in central areas, public hospitals should not be constructed in these areas.

Another decentralising strategy that can be implemented is the construction of higher HDBs in non-central areas. An increase in storey level in non-central areas is positively correlated with an increase in resale price. This shows that there is higher demand for HDBs on higher levels in non-central areas, thus higher HDBs can be constructed to encourage staying in non-central areas. In addition, we can improve the view from flats in non-central areas as people still find it more desirable to stay in higher level HDBs in central areas. This can be seen from the greater positive correlation between storey level and resale price in central areas as compared to non-central areas. The view can be improved by ensuring that surrounding areas are clean, by building more modern infrastructure or by putting up more festive decorations.

Lastly, marketing campaigns can also be introduced to publicise the benefits of purchasing housing in non-central areas, including less traffic and more affordable shops in the area, to aid with decentralisation efforts.

2. Evaluation

a. Multiple Regression Model

We have decided to use multiple regression models for our hypothesis instead of simple regression models due to the significant presence of omitted variable bias. For example, in hypothesis 1, we included omitted variables such as floor area and remaining lease in the model as these are factors in the error term that correlate with both the independent and dependent variable, proximity and resale price respectively. Using a multiple regression model will give a less biased result, and increase model explainability as well.

b. Omitted Variable Bias

Although we included various factors to control for omitted variable bias, it may still be present. This is due to the absence of omitted variables, specifically office spaces, which is positively correlated with resale prices and proximity to central areas as control in our model. Resale prices are higher at places near many office spaces due to the convenience of getting to work. There are also more office spaces nearer to central areas, especially the CBD area.(My Exclusive Condo, 2020) However, we were unable to include office spaces in our model due to the lack of information and data from online resources.

c. Assumptions may not hold

In this analysis, we are assuming that the results are not changing over time. However, this assumption may not hold due to external effects such as inflation. When analysing data that changes over time, a fixed effects model can be used to control for time invariant omitted variables and fix time to yield accurate coefficients and results. In addition, a 2SLS model can also be used to further strengthen the accuracy of the test by controlling for time variant variables.

In addition, our analysis is based on 4 room flats as it is the most common flat type in Singapore. The results obtained may not be generalisable to other flat types.

References

99.co. (2022, March 29). *Does prime location of a property still matter for buyers?* 99.co. Retrieved May 1, 2022, from https://www.99.co/singapore/insider/which-location-location-location/

Adis, K. (2020, January 22). *Living in a mature HDB estate? your property may be seeing a decline in value*. khaliladis.com. Retrieved May 1, 2022, from https://www.khaliladis.com/blog/living-in-a-mature-hdb-estate-your-property-may-be-seeing-a-decline-in-value

Cheong, H. K. (2018, April 25). *Shaping the future of Singapore's heartlands*. TODAY. Retrieved May 1, 2022, from https://www.todayonline.com/commentary/shaping-future-singapores-heartlands

Condo, M. E. (2020, May 23). *Shining chance to decentralise CBD*. New Launch Condo. https://www.myexclusivecondo.com/buyer-seller-must-know/shining-chance-to-decentralise-cbd/

Housing & Development Board. (2022). *Right in the heart of Town*. HDB. Retrieved May 1, 2022, from https://www.hdb.gov.sg/cs/infoweb/about-us/news-and-publications/publications/dwellings/Right%20in%20the%20Heart%20of%20Town

Land Transport Authority. (2020, June 6). *Land Transport master plan 2040*. LTA. Retrieved May 1, 2022, from https://www.lta.gov.sg/content/ltagov/en/who_we_are/our_work/land_transport_master_plan_2040.html

lifefinance. (2020, January 22). *What's the value of my leasehold condomium? revisiting Bala's table*. The Life Finance Blog. Retrieved May 1, 2022, from https://lifefinance.com.sg/whats-the-value-of-my-leasehold-condominium-revisiting-balas-table/?shared=email&msg=fail

Lin, C. (2021, October 28). *New prime area housing model could fuel demand for nearby units without the same restrictions, say analysts*. CNA. Retrieved May 1, 2022, from https://www.channelnewsasia.com/singapore/prime-area-housing-model-demand-nearby-units-subsidies-2272651

My Exclusive Condo. (2020, May 24). *Shining chance to decentralise CBD*. New Launch Condo. Retrieved May 1, 2022, from https://www.myexclusivecondo.com/buyer-seller-must-know/shining-chance-to-decentralise-cbd/

PropertyLimBrothers, E. (2022, March 6). *RCR vs CCR - CBD decentralisation happening faster than anticipated, may affect property prices*. PLB. Retrieved May 1, 2022, from

https://www.propertylimbrothers.com/insights-posts/rcr-vs-ccr-cbd-decentralisation-happening-faster-than-anticipated-may-affect-property-prices

Teo, C. (2020, May 17). *How to be a happy property buyer part 1: Planning areas*. Chantel Teo OrangeTee & Tie. Retrieved May 1, 2022, from https://www.chantelteo.co/post/how-to-be-a-happy-property-buyer-part-1-planning-areas