*Young Zhan Heng*

*Matriculation Number: A0239197N*

*Dse1101 Project*

Factors affecting HDB Resale Prices

Table of Contents

[Introduction 2](#_Toc87561087)

[Data Validation 2](#_Toc87561088)

[Data Source and Cleaning 2](#_Toc87561089)

[General Data Observations 2](#_Toc87561090)

[Findings 3](#_Toc87561091)

[Flat properties 3](#_Toc87561092)

[Amenities 3](#_Toc87561093)

[Geographic factors 4](#_Toc87561094)

[Making a new model 5](#_Toc87561095)

[Testing the model 5](#_Toc87561096)

[Limitations of model 5](#_Toc87561097)

[Analysis 6](#_Toc87561098)

[Flat Properties 6](#_Toc87561099)

[Amenities 6](#_Toc87561100)

[Geographical Factors 6](#_Toc87561101)

[Table of Figures 7](#_Toc87561102)

[References 8](#_Toc87561103)

# Introduction

As we are living in a Covid-19 pandemic, jobs are being lost and the economy seems to be taking a downturn, Singapore is not spared. Between 2019 and 2020, our Year-on-Year GDP fell by shrank by a record by 5.4% (World Bank, 2020). However, between this period of economic recession, the index of resale flats still hit an all-time high at 134.4 (HDB, 2021). In this report, I would be discussing the possible factors that might have affected the resale prices of HDB in 2021 to peaked.

# Data Validation

## Data Source and Cleaning

The data contains 6000 observations with 230 variables. By reading through the variables, I was able to split the data down into 3 main components, namely: Flat Properties, Amenities and Geography, all of which would be further discussed in the report. The goal of my report is to determine the factors that will affects the price of HDB resale flats. I would be using a model that returns a continuous Y variable, in this report; thus, I would be using the linear regression model and the decision trees model. By running through all the variables into a multiple linear regression model, the adjusted R-Squared value was 0.9227, showing that the linear regression model is very suitable for this set of data.

## General Data Observations

Since the number of observations is high, the central limit theorem applies, forming a normal curve as shown in *Figure 1*. The normal curve is highly right skewed with the mode being within the $400,000 - $500,000 range. The mean HDB price is at $497,486 and the median is $468,000, there is a standard deviation in price of $163,165.

Chart, histogram

Description automatically generated

Figure : Distribution of HDB Resale Prices in 2021

After running a Tree Plot to determine the factors that will affect the price of the resale flat the most, these are my findings. From *Figure 2*, I was able to identify that the most important flat properties that determines the price of the house is the Floor Area, Flat Type, and the Maximum Floor level. Other factors that play an important role includes the distance to the CBD, which will be discussed in geographic section.



Figure : Decision Tree

# Findings

## Flat properties

As shown in *Figure 3*, there is an evident relationship between the size of the house and the resale price, an increase in floor area results in higher resale price. I can link this statistic to the number of rooms in Singapore, as the size of the house increases by the increase in the number of rooms. Thus, it is not surprising to find out that the price of the house increases for each extra room. In addition, different flat types such as DBSS which may have the same Floor Area, might also affect the price of resale units and my new model will need to take note of such factors

Chart, scatter chart

Description automatically generated

Figure : Relationship between Resale Price and Floor Area

As stated above, another factor that played an important role in determining the price of the house is the maximum floors. By forming a simple linear regression between the maximum floor level and the resale price, I was able to form a linear regression model as seen in *Figure 4*

![A picture containing text, device

Description automatically generated]()

Figure : Relationship between the Maximum Floor Level and Resale Price

## Amenities

In this section, I would be studying the effect of the distance of amenities to the price of the houses. Amenities includes shopping malls, markets, stations, and schools.

Table

Description automatically generated

Figure : How different Amenities affect the price of Resale HDB

From *Figure 5*, I identified that the predictors with the low p-values (<0.001) would have played a significant role in the price of the houses. Among all these factors, most of the amenities have a negative coefficients to the predictor, meaning that the closer the house is to the amenity, the higher the price of the house due to the lower penalty.

Since the distance to the nearest station played a significant role to the pricing of the house, I would be studying which station line would be the most “expensive” line.

Table

Description automatically generated

Figure : How different MRT Lines will affect the Resale Price

By creating a multiple linear model which takes in a binary of “0” or “1”, with “0” meaning that the nearest MRT line is not the station and “1” being the nearest MRT line. There can be multiple “1” in a model if the nearest station is an interchange, e.g Bishan which serves as an interchange for the CCL and NSL. From this model, I was able to identify that house along the CCL and EWL are the most expensive.

## Geographic factors

One key component that might have affected the price of HDB resale price to peak is the distance to the CBD from the resale unit. Once again, by forming a simple linear regression model, I was able to identify a relationship between the distance of the HDB to the CBD area and the resale price.

![A picture containing text, sky, screenshot

Description automatically generated]()

Figure : Relationship Between the Distance to CBD and Resale Price

# Making a new model

With over 230 variables, it would be difficult to determine the price of the house should I use include every variable. Thus, using the decision tree in *Figure 2* and the findings above, I created a linear regression model which would provide us the highest accuracy using the least variables. The linear regression is as such:

*new\_model = lm(resale\_price ~ Dist\_CBD + max\_floor\_lvl + flat\_type\_dbss + floor\_area\_sqm + Remaining\_lease + mature + Dist\_nearest\_station + NSL + EWL + NEL + CCL + DTL)*

Using only 12 variables out of 230 variables provided, the new model achieved an adjusted R Squared value of 0.8541 which was quite comparable to the original model which had an adjusted R Squared of 0.9227

## Testing the model

With this simplified model, I would be able to have a good estimate of the price of a house. For this example, I would be using Block 154 Toa Payoh Lorong 2 as an estimator.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Distance to CBD | Max Floor Level | Flat Type DBSS | Floor Area (sqm) | Remaining Lease | Mature | Distance to MRT | MRT Line |
| 5.3km | 40 | No | 101 | 77 | Yes | 0.1km | NSL |

Based on my model, the estimated price of the unit should be $711,985. In October 2021, this same unit sold for $766,618 (PropertyGuru, 2021). There is a 7.67% difference in the price of my model as compared to the original price. Despite the evident decrease in the number of variables as compared to the original dataset, the model was able to give a decent estimate of the resale unit price. One reason for the price different may be due to the property price rising and I was extrapolating the data to October.

## Limitations of model

One evident limitation of the model is that there were too many variables that were not covered. Factors that were not covered includes the distance to other amenities and what some sociologists describe as a “Town Premium”, where the location of the town would determine the price of the unit (Lam, 2021). In addition, my model did not account for which floor the unit was sold at. On average, every additional level result in a $5000 increase in price but my model was not able to account for that as the data provided for the floor range was a dummy variable, making it hard to determine the exact increase in price per floor. (PropertyGuru, 2021)

# Analysis

## Flat Properties

Singapore’s public housing is a unique system as it runs on a 99-year lease system. When the house reaches its 99-year lease, the house will be returned to the government with no payments made to the owner. Essentially, the house will be worth $0 after 99 years (Parliament Singapore, 2014). This means that the price of the house should be depreciating at 1% of the original price. My findings above were above to prove this as there was a proportionate relationship between the remaining lease and the price of the house. However, the data did not provide the initial price the house was purchased for and thus it would be hard to predict if the house is depreciating at the expected rate. I was able to identify a relationship between the price of HDB resale flats and the size of the resale flats.

An interesting observation was that there was a correlation with the district being mature and the price of the flat. Flat that are within the mature estates tend to be more expensive. As mentioned before, a limitation of my model was that it did not account for the “town premium”, but by including a maturity factor, I was able to account for the “town premium” to a certain extent as mature estates tend to go for a higher price that non-mature estates (Fatti, 2018)

## Amenities

The amenity that played a significant role was the distance to the MRT station, amongst which, houses along the CCL had the highest resale value. One possible reason may be because houses that are near the CCL stations are also nearer to the CBD. In addition, houses with CCL as the nearest station had the highest correlation to the distance to the CBD.

## Geographical Factors

The key component I measured is the distance from the CBD to the resale flat. From Figure 2, it was evident that the distance played an important role in the price of the resale flat, this may be due to the high office density within the CBD region, where 12.9% of Singapore’s workforce work at (Department of Statistics, 2021) .With the CBD region having more office buildings being built, as well as the rising cost to own a car, I can foresee the distance to the CBD to be a more significant factor in the future.

# Table of Figures

[Figure 1: Distribution of HDB Resale Prices in 2021 2](#_Toc87551947)

[Figure 2: Decision Tree 3](#_Toc87551948)

[Figure 3: Relationship between Resale Price and Floor Area 3](#_Toc87551949)

[Figure 4: Relationship between the Maximum Floor Level and Resale Price 3](#_Toc87551950)

[Figure 5: How different Amenities affect the price of Resale HDB 4](#_Toc87551951)

[Figure 6: How different MRT Lines will affect the Resale Price 4](#_Toc87551952)

[Figure 7: Relationship Between the Distance to CBD and Resale Price 5](#_Toc87551953)

# References

Department of Statistics. (2021). *Singapore census of population 2020, Statistical Release 2: Households, geographic distribution, transport and difficulty in basic activities*. Singapore Census of Population 2020, Statistical Release 2: Households, Geographic Distribution, Transport and Difficulty in Basic Activities. Retrieved November 10, 2021, from https://www.singstat.gov.sg/publications/reference/cop2020/cop2020-sr2/census20\_stat\_release2.

Fatti, M. (2018, October 19). *99.Co's guides: Mature estate vs non-mature estate*. 99.co. Retrieved November 11, 2021, from https://www.99.co/singapore/insider/mature-estate-vs-non-mature-estate/.

HDB. (n.d.). *Resale statistics*. Resale Price Index (RPI) RPI for 1st Quarter 1990 to 3rd Quarter 2021. Retrieved November 10, 2021, from https://www.hdb.gov.sg/residential/buying-a-flat/resale/getting-started/resale-statistics.

Lam, F. (2021, June 4). *Price premium of HDB resale flats in mature estates may shrink further: Report*. The Straits Times. Retrieved November 10, 2021, from https://www.straitstimes.com/business/property/price-premium-of-hdb-resale-flats-in-mature-estates-may-shrink-further-report.

Parliament Singapore. (n.d.). Official Reports. Retrieved November 10, 2021, from https://www.parliament.gov.sg/.

PropertyGuru. (2021). *Blk 154 Lorong 2 Toa Payoh, (S)310154*. Recent Transactions In 154 Lorong 2 Toa Payoh. Retrieved November 10, 2021, from <https://www.propertyguru.com.sg/singapore-property-listing/hdb/toa-payoh/lorong-2-toa-payoh_107362/175>.

PropertyGuru. (2021, March 30). *High floor vs low floor units: Which one is better?* PropertyGuru. Retrieved November 11, 2021, from https://www.propertyguru.com.sg/property-guides/high-floor-vs-low-floor-unit-which-is-better-45449.

The World Bank. (n.d.). *GDP growth (annual %) - singapore*. GDP growth (annual %) - Singapore. Retrieved November 10, 2021, from https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=SG.