

# DR.HDB: Dissecting Resale HDB Pricing by Analysing HDB Public Housing Data

Team 11: Lu ZhiMao, Lim Pao Ling Rachel, Koh Teck Yun

**Abstract**—In the past, Singaporeans would turn to the primary source giver of Resale housing information, formally known as The Housing Development Board (HDB), to understand the changes surrounding real estate. With the advent of Singstat and other public domain data sources, there are various online visualisations that Singaporeans seek guidance from, to understand the Housing situation in Singapore. To meet the analytical needs of HDB buyers or sellers, our team has contributed the DR.HDB tool which identifies numerous dashboards that provide an overview of resale markets in Singapore. At times, such displays can be confusing for users to understand. In this research paper and project, we would like to demonstrate the potential of DR.HDB by breaking down resale and unit prices to allow users to understand the costs across various granularities from planning regions, towns to floor level. Through the use of treemaps, time-series charts and correlation boxplots dynamically linked to interactive maps, the user would have access to an interactive overview of the general resale price level and volume across the years.

**Index Terms** — Resale Price, Unit Price, Market Trends, Floor Level, Hierarchical Mapping .

## 1 INTRODUCTION

SINGAPORE PROPERTY CHEAT SHEET			
	<b>BTO</b> (BRAND NEW HDB FLAT)	<b>RESALE</b> (OLDER HDB FLATS)	<b>EC</b> (FULL FACILITY SEMI-PRIVATE)
<b>3-ROOM PRICE</b>	\$160K - \$380K	\$350K - \$450K	\$750K - \$950K
<b>PROS</b>	- Brand new & empty - Lowest ever prices - High appreciation	- Some furnishing - Choose better locations - Move in immediately	- Fully furnished - Highest appreciation - Full amenities
<b>CONS</b>	- Wait for 3 to 4 years - Stay minimum 5 years - Rent/Sale restrictions	- Limited appreciation - Limited subsidies - Potential Re-design	- Most expensive - No HDB Loan option - Stay minimum 5 years

Fig. 1. Singapore Property Cheatsheet – Pros & Cons

Due to COVID-19, it is now mandatory for Singaporeans to stay at home for a month. With that in mind, the importance of home became very vital and evident as it is part of our livelihood. Singaporeans generally have proper living conditions and a shelter over their heads compared to other countries such as Hong Kong where housing is typically small which may sometimes consist of a large family. This circuit breaker which started on 7th April 2020 has also made many Singaporeans realise that home is always open for them, ready to shelter when in times of crisis.

With that in mind, the importance of technology also comes into play where it is one of the most importance sources to stay connected and updated. There are often fake news or misleading prices on resale flats which many online property information that claims to be “cheat sheets” that could help Singaporeans decide on their desired choice of homes. In the context of this project, Resale HDBs is the focal point of our project.

- Teck Yun is an undergraduate in the School of Information Systems, Singapore Management University,
- E-Mail: teckyun.koh.2017@sis.smu.edu.sg
- Rachel is an undergraduate in the School of Information Systems, Singapore Management University,
- E-Mail: rachel.lim.2017@sis.smu.edu.sg
- Zhimao is an undergraduate in the School of Information Systems, Singapore Management University,

Choosing a resale HDB has never been easy as there are many factors to consider such as location, HDB type, number of remaining lease years, resale value, etc. On top of that, thousands of Resale HDBs transactions are happening each month, making it almost impossible for an owner to get a view of every transaction. Therefore, most buyers and sellers do have to consult property agents for their services.

Often or not, the services provided by them would not have a satisfactory outcome for the consultees. While not consulting property agents is an option, it is still advisable for the buyer and/or seller to have a better understanding of the resale market before making an informed decision.

In most cases, people tend to rely on multiple platforms to help them in understanding and predicting the market trend. On that note, our team would like to know whether the information provided to them is truly insightful or usable. Thus, WeHouse comes into play to provide greater insights for people to better understand the HDB resale market which can provide better aid in making more informed decisions.

## 2 MOTIVATION AND OBJECTIVES

Our team would like to minimise the number of visualisations a prospective buyer would have to see. The data visualisations provided were generally overloaded with information, hence, we aim to create concise visualisations surrounding resale HDB trends. It is ever so important to relay critical information pertaining to yearly sales trends, average resale prices and volume based on floor level category.

In this project, we aim to deliver a focused and compact visualisation to allow Singaporeans to be well-informed of the average HDBs resale prices around their desired location with the following objectives in mind.

- #1 Comparing and identifying the change in median per unit price trends of resale HDB Town across different months within the selected year and floor level.
- #2 Comparing the prices of the resale flats within individual HDB Town area based on their remaining lease year of the HDB flat.
- #3 Determine which year had the highest or lowest resale price sold and number of transactions
- #4 Identify the most expensive town given the flat type category

### 3 RELATED WORKS

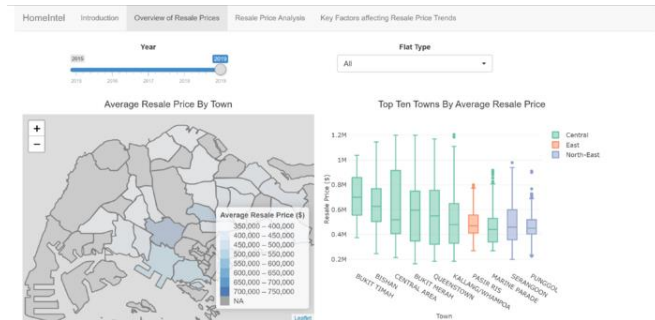


Fig. 2. Overview of Avg. HDB Resale Price by Town

From Fig 2, there were desired benefits from the box plot. Boxplot can depict outlier resale prices within each town listed, indicating which towns have more certainty with flat resale prices. On top of that, the user can understand the general median resale price of each town.

However, improvements could be made with the order in which the box plots are displayed, it could be sorted in a descending order according to the coloured region for a more uniform chart output. The heat map provided an overall view on the distribution of average resale flat prices given the year selected.

Furthermore, improvements could be made about the gradient colour type; a darker varied gradient could have been used to make the shaded regions stand out more on the grey scaled mapping. If not, the whiter shade portions are hard to distinguish upon the grey unshaded tones.

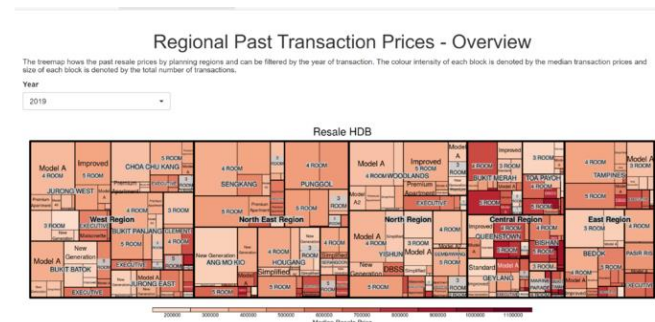


Fig. 3. Treemap of Past HDB Transactional Pricing

From Fig 3, treemap can deliver multitude of benefits in the visualisation standpoint where median resale prices and sales volume transaction can be understood across 5 regions within Singapore given the year selected. The gradient tone depicted the varying median resale prices which can be understood by the legend.

However, improvements could be made through removing the model type and only having the number of room type as it has a greater decision impact on the viewers rather than HDB model types. Also, having more interactivity would give the user more information when it comes to region and pricing such as Unit Price.

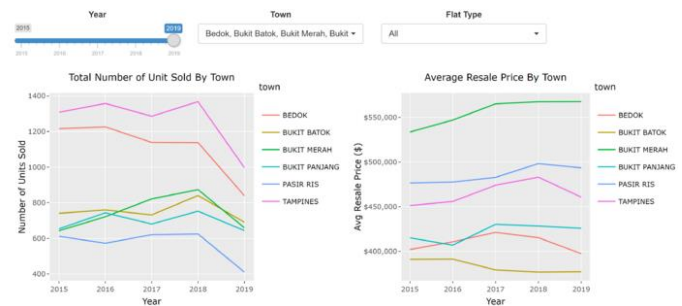


Fig. 4. Factors Affecting Yearly HDB Resale Price Trends

From Fig 4, the gathered benefits of using these line charts is that the overall increasing/decreasing transaction trends and the average price trends can be observed within a span of 5 years.

However, it could be better illustrated in a chart that incorporates both elements which would create a more visually appealing chart with just a single chart to focus on.



Fig.5. Correlational Scatter Plot: Remaining Lease vs. Resale Price

From Fig 5, the correlation charts provided an overview how the factors (remaining lease years, floor area) influence the resale prices over the years and with the colours to signify the various HDB Towns.

However, improvements could be made to the chart by excluding the exact address of the flats and only keeping the comparison specifically between remaining lease years and Resale Price by town area to reduce the cluttered scatter plots.

### 4 VISUALISATION APPROACH

After visiting Data.gov, Singapore's public statistics data site, our team collated the relevant data on resale flat prices and resident households by planning area in line with our previously mentioned objectives, motivations and brainstorming storyboards. Our team has taken a 4-step approach to develop our visualisations:

- [1] Exploratory Data Analysis (EDA)
- [2] sketching and brainstorming storyboards
- [3] data cleaning and pre-processing with tableau prep
- [4] visualisation implementations.

## 4.1 Exploratory Data Analysis (EDA)

There were in total 8 datasets found via data.gov.sg, within those datasets we discovered a diverse range of categories covering various aspects of the Resale HDB situation in Singapore. Amongst them, we decided to use the Resale Flat Prices dataset that classifies each HDB resale instance by year, month, HDB Towns, Street Names, Blocks, Floor Level, Unit size, Resale Price and lastly Remaining Lease Duration.

Our team began exploring the various floor categories and planning areas to gain a better understand on the resale price trends pertaining specifically to these identified areas.

We were able to spot a noticeable trend where "Top" floor levels are generally more desirable and a few areas which are usually prime real estate places as denoted by its high resale pricing.

In order to visualise the capabilities of the data sets, we first experimented the graphical potential by creating simple charts on a tableau dashboard.

## 4.2 Brainstorming & Storyboard Ideation

After exploring and understanding how much we could work with the data set chosen, the team proceeded to research on the various visualisation creations and r packages that could be of assistance in producing the right graphs for our project.

All of us considered the viability of each graph type and how likely the overall dashboard design would make sense to our targeted audience. Moreover, we also planned to include graphical interactivity for the various dashboard views.

In order to display various HDB price trends and number of transactions over a 9-year period, we aim to utilise the following charts and dashboards below or more:

### 4.2.1 Line-bar Chart and Trellis Line Chart

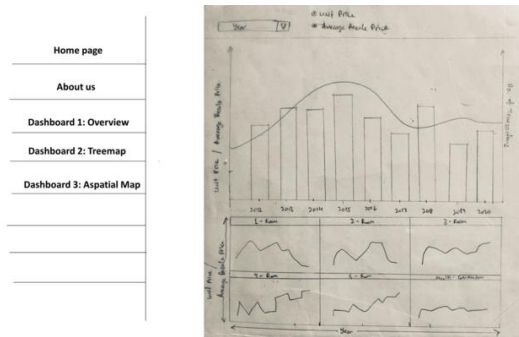


Fig. 6. Storyboard Sketch 1

After a thorough discussion amongst ourselves together with the professor, we decided to revise our overview dashboard to include a trellis plot by Flat Types on top of the existing line-bar chart we had in mind to provide an overview of resale prices over the past few years.

By being simple, the line-bar chart can give a trending overview of number of HDB transactions by the average resale price annually. The inclusion of having a trellis plot is to provide a snapshot view of the various HDB pricings by room types. In addition, a radio button

filter is considered to allow the user to toggle between unit price view or average resale price view.

### 4.2.2 Treemap Plotting

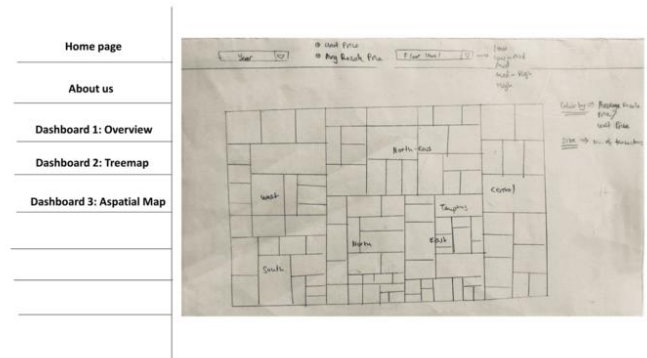


Fig. 7. Storyboard Sketch 2

A treemap chart will be used to display the hierarchical data with the use of nested rectangles. The size of the treemaps are based on sales volume the layers of treemap are in the following sequence: Planning Region → HDB Town → HDB Floor Category

The HDB floor levels were categorised according to Low, Low-Mid, Mid, Mid-High and High. The user will be able to switch between Unit Price and Average Resale Price that affects the colour intensity of the rectangles to allow an interactive view of all pricing per hierarchy.

### 4.2.3 Geo-Facet Grid, Scatterplot with Marginal Boxplot

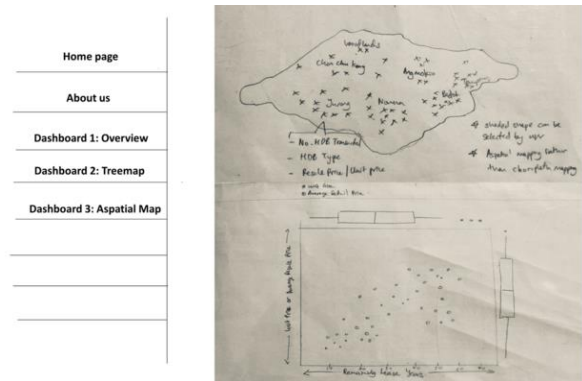


Fig. 8. Storyboard Sketch 3

The geofacet R package allowed us to build a flexible visualization data for different geographical regions of Singapore by providing a ggplot2 faceting function. In order to proceed with geofacet, it requires a set of grids that is aligned with Singapore planning areas.

In order to proceed with the geofacet, it requires a set of grids that maps to Singapore planning areas. Online templates do not provide the matched name of the estates based on the data we have on-hand hence modification has to be done on our side where we will edit the names of the estates that provide a more accurate depiction of the district. Out of these estates, we are only focusing on those that are classified under the "HDB Town" as defined under the Housing & Development Board (HDB).

## 5 DATA VISUALISATION WALKTHROUGH

There are 3 main dashboards with 2 sub dashboards available for viewing through our shiny application:

### 5.1 Overview

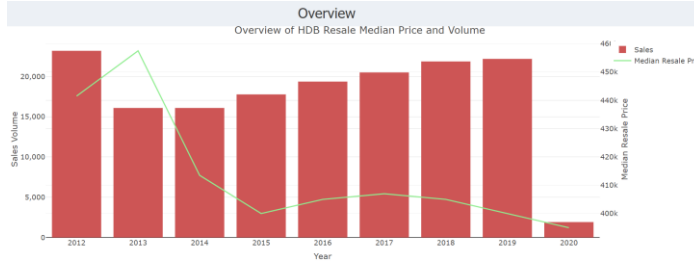


Fig. 9. Overview of Median Resale HDB Price and Sales Volume

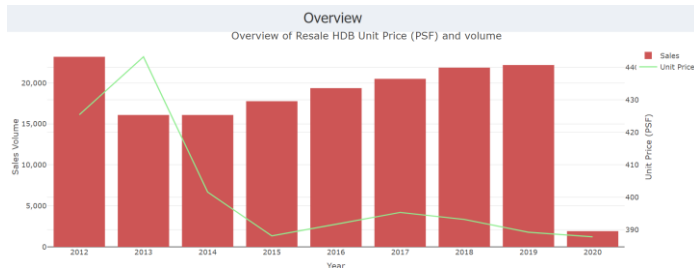


Fig. 10. Overview of Resale HDB Unit Price (PDF) and Sales Volume

The overview page in figure 10 and 11 is designed to provide an overall insight of the resale flats sales volume and Price. From figure 9 and 10, when a user hovers over the chart, the user will be able to see the sales volume and median resale price / Unit Price (PSF) from each year.

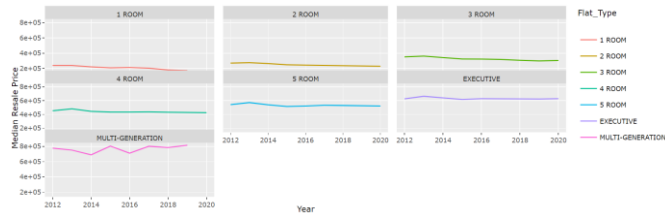


Fig. 11. Trellis Plotting of HDB Flat sold by Room Type

The overview has figure 11 below figure 10/11 where the user will be able to gather insights on which room types contribute to the overall median resale price / Unit Price (PSF). The user will be able to understand the lows and highs of from the different room types.

### 5.2 Segregated Pricings

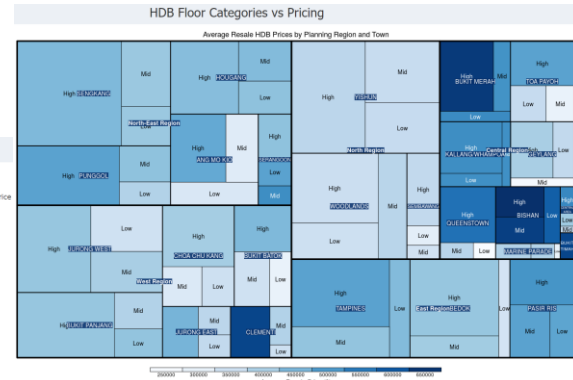


Fig. 12. Average Resale HDB Prices by Planning Region and Town

From figure 12, a treemap illustration is used to depict the hierarchical view by Planning Region, HDB Town and HDB Floor Level category. The user will be able to toggle between Resale Price and Unit price, together with a slider year input to further understand each layer of the treemap and to notice noticeable difference between both visualisations.

### 5.3 HDB Town

#### 5.3.1 Geofacet

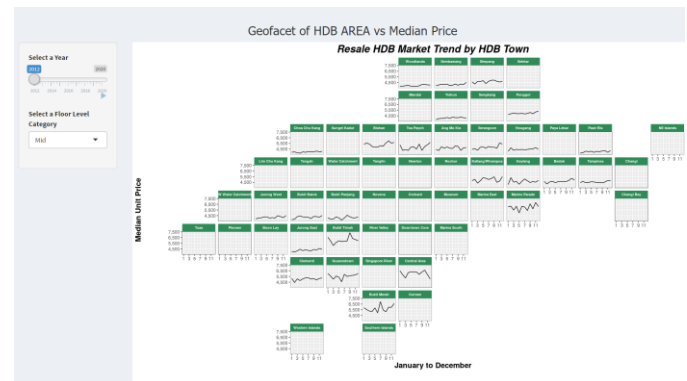


Fig. 13. Yearly Resale HDB Market Trend by HDB Town

From Figure 14, a Geofacet map contains grids that represents each HDB Town in Singapore. From the above visualisation, the user will be able to navigate through each grid, which is the HDB Town to further understand the median per unit price within that area.

The X axis is represented by month and Y axis is represented by the medium per unit price. It forms a line graph showing the transaction of medium per unit price as median pricing gives a more accurate representation of pricing. The map consists of an interactive filter for the user to toggle between past transactions ranging from 2012-2020 and floor level category which consist of low, medium and high.

### 5.3.2 Scatterplot



Fig. 15. Resale HDB Price vs. HDB Remaining Lease Years

This graph shows a combination of a scatter plot and box plot where the x-axis is the remaining lease in years and Y-axis is the resale price.

The data is then represented as data points on the scatter plot. The red box plot summarizes all the data points on the x-axis while the green box plot summarizes all the data points on the Y-axis. When hovering around the box plot elements, the user is able to see 6 features: minimum, lower fence, quarter one, median, quarter three, maximum. Moreover, the user is able to toggle between 3 filters namely, year, HDB town estate and flat type.

## 6 KEY FINDINGS AND OBSERVATIONS

There were several key findings observed from the various visualizations created.

**Objective:** Determine which year had the highest or lowest resale price sold and number of transactions

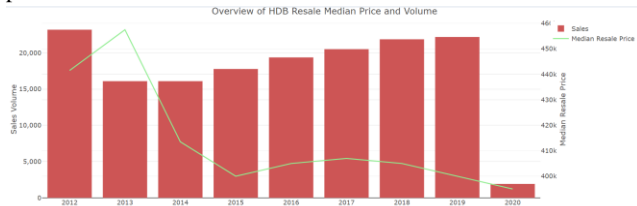


Fig. 16. Overview of Resale Prices vs Volume

In 2012, post economic crisis in 2011, we can see that the housing boom is evident from 2012 to 2013 as it shows higher resale price. However, between 2013 to 2015, it experienced a dip in resale price. Generally, Resale Volume are on a rise, yet the overall resale prices are generally dipping throughout which signify that supply of resale HDBs are enough to prevent spikes in median resale prices.

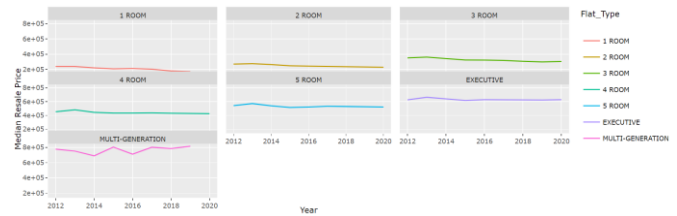


Fig. 17. Trellis Plotting of HDB Flat sold by Room Type

From the trellis plot, Singaporeans can generally decipher that executive and multi-generation experienced high resale price from 2012 to 2020. However, the overall median resale price trend is at a price considerably lower than these two flat types, hence they would be able to make a conclusion on what type of Flat Type to purchase if they wish to purchase within this year based on the year.

**Objective:** Comparing and identifying the change in median per unit price trends of resale HDB Town across different months within the selected year and floor level

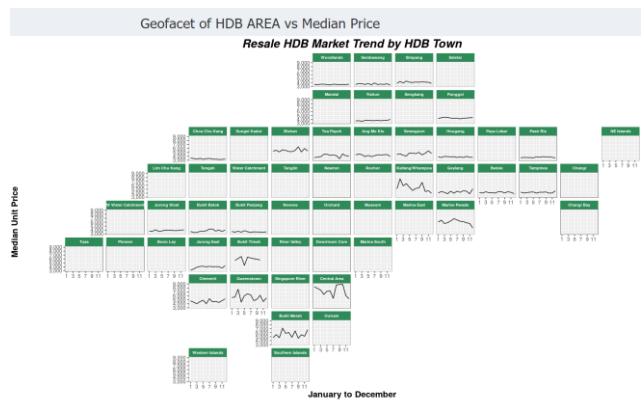


Fig. 18. Geofacet of HDB AREA vs Median Price (Year 2016)

From 2012-2015, unit price in each estate classified under HDB town is seeing fluctuations throughout the months. However, from 2016 onwards, it becomes evident that these fluctuations are only occurring to estates located nearer to town such as Queenstown, Bishan, Whampoa, Kallang and Toa Payoh. Whereas other estates located further away from town such as Punggol, Sembawang, Woodlands, Jurong West are more or less stagnant in their prices. This is especially evident for Central area, in 2019, where floor level is selected as mid. The prices can be as low as \$5000 per unit price to \$9000 in the next month.

**Objective:** Identify the most expensive town given the flat type category



Fig. 19. Resale HDB Price vs. HDB Remaining Lease Years (Yishun, 3 Room)

In the following HDB Town, Yishun, under the standard room flat - 2-room, 3-room, 4-room and 5-room have seen a common trend where the flats tend to sell at a higher price when its number of lease years is closer to 99 years.



Fig. 20. Resale HDB Price vs. HDB Remaining Lease Years (Yishun, Multi-Generation)

However, this trend does not apply to executive flat and Multi-generation flats where the number of years lease left does not affect the resale flat price. In the case of a multi-generation flat with a remaining of 68 lease years, it holds a record of selling at \$868k in Yishun. The resale price of such a multi-generation flat is much higher than those with a longer lease year.

## 7 LIMITATIONS

Although the HDB data used to create the various visualisations had the various street addresses of the Flats and town area, it lacked the information on MRT stations that could be situated near the HDB flats. It could have aided the visualizations to have that decisive edge for targeted audiences to pinpoint their decisions on a desired HDB housing location.

## 8 FUTURE WORK

In this paper, the dataset used was from a single source and we did not include other datasets which may play a part in influencing the property purchasing decisions made by the users. If we have the chance to refine it in the future, here is a list of things our team intends to explore and expand our capabilities to provide more information to the users, enabling them detailed decision making when it comes to purchasing the perfect property.

#1 Rather than providing only general information with regards to HDB Town pricings by Floor Level, future visualizations could provide a deeper insight by the user to have a more customizable filter such as block-district avenue, other amenities such as distance to schools, distance to bus stops, nearest MRT, number of MRT lines within 1 kilometers radius of the house etc. All these factors play a part on the property prices, and it can be an attractive option for investors. All these factors play a part in property pricing and it might also be an attractive trait for investors.

#2 Our Team might consider looking into providing more datasets consisting of essential facilities around the estates. This enhances the user experience as they can view if their chosen estate meets their desired housing needs.

#3 Lastly, we could also look to provide visualizations for other housing types such as landed properties, condominiums, rental flats and Built-To-Order (BTO) flats. Due to policies in Singapore, non-Singaporeans have a lot of criteria for housing eligibility. Hence, including more variety of housing types can cater to a wider target audience.

## 9 CONCLUSION

After the whole experience with the Visual Analytics Final Project, our team found that analyzing vast amounts of data and creating sensible visualizations was quite challenging. Much effort was required for extensive research on Singapore's property market despite the narrowed focus of HDB resale market for our project. Nevertheless, throughout the vigorous process of storyboard ideation and technical development with R programming, our team has learnt many valuable lessons. For the technical aspect, we gained exposure to various R packages such as 'treemap', 'tidyverse', 'shiny', 'shinydashboard', 'dplyr', 'ggplot2', 'ggExtra', 'lattice', 'geofacet', 'plotly'.

It was very crucial to have perfectly "cleaned" data set to begin with as it was the bulk of our R Shiny app codes, therefore the tedious steps in data cleaning should be done meticulously. Given our HDB Resale project topic, we were exposed to the various HDB market trends and it opened our eyes to understanding the considerations buyers or sellers might go through before purchasing a property.

## ACKNOWLEDGMENTS

Our team would like to thank Professor Kam Tin Seong for his never-ending guidance during classes as well as consultation sessions. He provided our team with valuable inputs and enabled us to exercise much critical thinking during our project implementations. This aided in improving our visualisations; ensuring that our team was on the right track.

## REFERENCES

- [1] <https://hdbviz.shinyapps.io/hdbviz/>
- [2] <https://homeintel.shinyapps.io/HomeIntel/>
- [3] <https://plotly-r.com/client-side-linking.html>
- [4] <https://www.musgraveanalytics.com/blog/2018/8/24/how-to-make-ggplot2-charts-interactive-with-plotly>
- [5] <https://towardsdatascience.com/plotly-with-r-shiny-495f19f4aba3>
- [6] <https://hafen.github.io/geofacet/>
- [7] <https://plotly-r.com/>
- [8] <https://rpubs.com/tskam/treemap>