HOMEINTEL: A Visual Analytics Tool for Exploring and Analyzing Singapore Housing Resale Prices using R Studio

Xuan Ni Rachel CHUA, Nurul Khairina Binte ABDUL KADIR, Qian Ling ONG

Abstract — There is a vast amount of data and information available online on Singapore's housing resale price industry. It is wise for current and future homeowners to be aware of trends in the market if they are planning to buy a new home or sell an existing one. The Housing Development Board offers a service where users can check the transacted prices for resale flats within the past years. However, users have to select a specific flat type and HDB town to perform a search, and this does not allow users to make comparisons across towns and flat types. The datasets on resale price history from Data.gov.sg are in CSV format and do not offer much insights. To address the lack of a one-stop and in-depth analysis platform on housing prices, we created Home Intel to allow users to explore the resale price trends and key factors that affect resale prices. We adopted various visualization techniques to provide users with useful insights, and the key features include a choropleth map, scatter plots, and heatmap. As such, users can make a more informed decision when buying a house.

Index Terms — Singapore, HDB Resale Prices, Housing, Property,

1 Introduction

Singapore still ranks as the second priciest housing market in the world in 2019, just behind Hong Kong [1]. This is despite cooling measures that the government has come up with to curb rising property prices. Based on the PropertyGuru's Consumer Sentiment Survey[1], 2 in 3 millennials own their first property between 28-34 years old[2]. At this age, the high property prices might be a financial burden for these young working adults as they seek to purchase their first house. In this paper, we will be looking at the Singapore resale Housing Development Board (HDB) flats in detail to create visualizations that can best help potential buyers and investors who are looking to buy a new house.

This paper consists of 10 sections where we will be walking through the current resale housing industry in Singapore and the thought process that the creators had when developing these visualizations. The final aim is to ensure that users would be well informed of the choices they can make. Section 1 will introduce the housing industry in Singapore, Section 2 will include the motivations and objectives of the research conducted. Next, in Section 3, we will be going through related works that had been done and the key takeaways from these visualizations. Section 4 will highlight the visualization approach that the team has taken.

In Section 5, we will proceed to discuss the Data Collection, Data Exploration and Preprocessing work that we have done. In Section 6, we will discuss the design considerations of the application. In Section 7, we have a case study to demonstrate the functional use of the visualization tool. Next, in Section 8, we will be going through the key findings and insights. In Section 9, we have included the future works that could potentially be done in the future. Finally, in Section 10 and 11, we have the conclusion and the acknowledgment, respectively.

[1] Conducted bi-annually since 2009. It aims to measure property sentiments and expectations around the property market to help consumers, property agents and developers gain a better perspective of the local property market.

[2] Source: IBTimes

2 Motivation & Objectives

2.1 Motivation

82% of Singaporeans live in a HDB flat. Most Singaporeans only start dreaming of buying their own flats when they are getting married [2]. After all, owning a flat in Singapore is a huge financial commitment, and many are worried about the financial burden that comes together with owning their dream house. Thus, many potential buyers take time to look up the internet for information only to find out that these data do not offer great insights to them.

In addition, housing is also considered a form of investment for many other potential buyers. Many also buy houses with the intention of selling them one day to upgrade into a better house. They would like to see how each flat with specific traits like region, floor area, stories, and various other factors may contribute to the price a few years down the road.

Hence, this makes it worthwhile to create visualization dashboards to help potential buyers see data that best represent the current resale flat industry. We hope that through these visualizations, potential buyers can analyse and compare the costs of owning a flat across different towns, prioritize what is important to them, before deciding on their dream home.

2.2 Objectives

The objectives that we aim to create with this visualization includes:

i. Allow users to gain overall insights on resale price trends over the last 5 years (2015-2019)

ii. Understand the key factors that affect resale prices (such as average resale price by region, storey, flat types etc.)

iii. Gain insight on the town area which has the highest investment potential[3] for housing flat.

[3] Investment potential may vary from investors to investors but some general information that investors look out for in the housing market

3 Related Works

Before working on this project, we took references from various websites to learn about the best practices in the hope of creating easy to use and easy to understand dashboards. Here is a summary of our findings.

i. Singapore Property Prices Heatmap [3]

Figure 1: Choropleth map showing the price trends for private properties in Singapore

The map focuses on the price per square foot by the district. Users can compare the performance of the private property market by districts across several quarters. Users can select a date range, and tooltips provide more information on the trend, number of transactions and towns within the district.

The chart is very clear such that at one glance, you can see that certain areas tend to have a higher price per square foot compared to others. We incorporated this design in our dashboard to show the variation of average resale prices across towns.

ii. Historical Average of HDB Resale Prices in Ang Mo Kio [4]

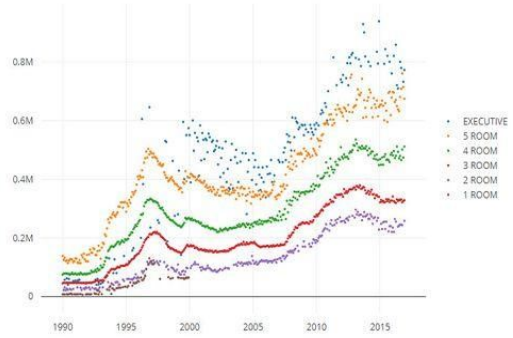

Figure 2: Line chart showing the resale price trends over time

The line chart shows the trend in resale prices by flat type for a specific neighborhood over time. Initially, we wanted to implement this in our dashboard, but we noticed that it is common for 1 room flats to be the cheapest, followed by 2 room, 3 room, 4 room, etc. We decided to tweak this idea and do a line chart that indicates the resale price trends over time by the town. The user can select multiple towns using the Town filter, and it will be displayed in the same chart, which will provide the user with more insights as compared to the chart above.

includes comparable sales figure, gross rent multiplier and cash on cash return etc.

iii. Heatmap of Median Resale Price [5]

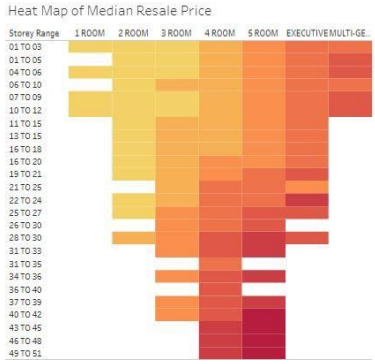

Figure 3: Heatmap showing the median resale price

The heatmap shows the relationship between flat type and storey range. It is easy to interpret from above that higher. We can improve this visualization by creating heat maps based on Estate Type (mature and non-mature estates) for easier comparison.

Overall, these visualizations were useful in guiding us during the brainstorming and storyboarding process. Each chart has its strengths and limitations that we seek to address. Lastly, we aim to enhance these visualizations by using filters and tooltips to convey relevant and useful information.

4 Visualization Approach

Three main steps were taken to develop the visualization: 1) Data exploration and preprocessing, 2) Planning the visualization tool, and 3) Implementing the visualization.

5 Data Exploration & Preprocessing

5.1 Data Collection

Data.gov.sg provides a comprehensive collection of datasets on resale flat prices from 1990 to 2019 based on the registration and approval date [6]. For the purpose of this project, we have decided to analyze the data from 2015 to 2019 since it serves as a meaningful range to identify relevant trends and the features available in each dataset is enough for analysis to be done. The initial data set consist of 11 features – Month, Town, Flat Type, Block, Street Name, Storey Range, Floor Area per Square Meter, Flat Model, Lease Commencement Date, Remaining Lease and Resale Price.

Stackedhomes.com provide detailed information on the region and estate type of each town [7]. This information is not included in the dataset above. As such, we created a dataset manually using Excel using information from the website as a reference. The data set called region.csv, consist of 3 features – Town, Region and Estate Type.

5.2 Data Exploration and Preprocessing

During the initial phase of our project, we sketched out the various variations of graphs that we would like to implement in the project on a whiteboard. We also used the feedback given from Prof Kam Tin Seong and revised our storyboards. We used Tableau to explore the dataset and create the dashboards we are planning to implement in R. Overall, the steps taken were an iterative process. The dataset is clean, free of errors, and there were no missing data. Thus, no problems were identified in the original dataset.

5.3 Data Cleaning And Feature Creation

Before creating the visualisation model, we used R to prepare our data. Firstly, we used R to merge two Excel source files together to obtain a final dataset file with 5 years' worth of resale price transactions (2015 to 2019). After merging the files, we used R to generate the extra columns needed such as age and month name for current and future analysis. Lastly, the region.csv file is merged with the revised dataset using Town as the join condition. Once all the changes were made, we exported the dataset as an Excel CSV file.

6 Design Considerations

Before we started on developing our applications, we researched the various visualization libraries we can use to present the data. We have decided to use libraries such as Plotly due to ease and flexibility in creating visualizations, Leaflet, and Tmap to create the map and Heatmaply to create heatmaps. Key considerations taken when planning the design of the dashboards include the clarity of the visualization, interactivity using filters and tooltips as well as aesthetic considerations.

For our visual application, we have also decided to go with a storytelling approach. In the initial iterations, we were unsure of how to organize the charts and separate them into dashboards. After much consideration, we reorganized the charts in the dashboards to ensure that they complement each other well. We decided to adopt a sequential approach where each dashboard tells its own story. The first dashboard tells the overview of resale prices. The second dashboard is a resale price analysis, where we take a step further and go deeper into the analysis. Lastly, the third dashboard shows the key factors affecting resale price trends, which can be especially useful for investors and buyers.

The following descriptions show some of the design considerations in the visualization tool:

A. Choropleth Map

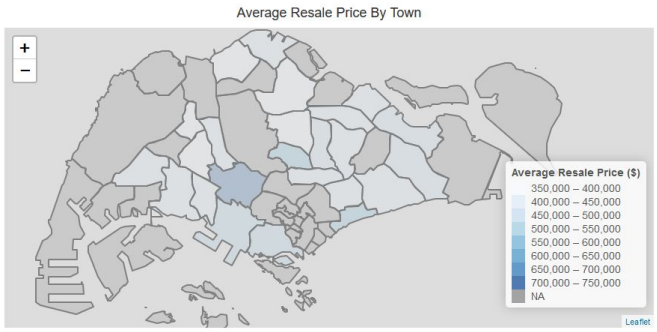

Figure 4: Choropleth map showing the average resale price by town

The choropleth visualization enables us to display geographical locations of each town area by the average resale price. With an overview of the resale price marked by the intensity of colour, users can zoom in to particular town areas and explore detailed trends.

One of the limitations we encounter in the visualization process is plotting the boundaries for town areas. Singapore Planning Area and Subzone are different from the town boundaries which Housing Development Board uses. Singapore has a total of 23 new towns, but 55 planning areas. However, there wasn't any shapefile available on the new town area, we then decided to work with the planning zone for the time being.

B. Heatmap

Figure 5: HeatMap showing the comparison against flat type and storey

Heatmap is a graphical representation of data that quickly communicates its data to the viewer, using easy-to-understand color gradations. With the heatmap visualization, the user will be able to quickly visualise the volume of the location/events within the dataset and focus on the areas that matter most. Therefore, we decided to make use of the heatmap to show the median resale price against the flat type and the storey range for better visualisation.

C. Bar and Line Charts

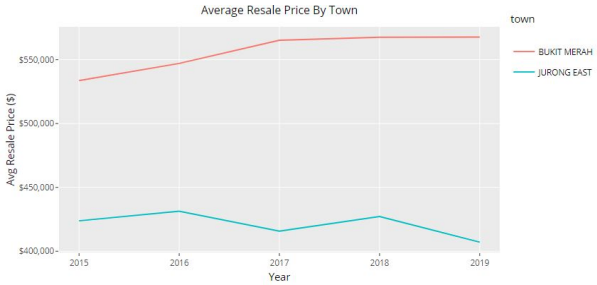

Figure 6: Line Chart showing the comparison

The bar and line charts are some of the most common type of data visualization tool. They are simple yet informative. A bar chart is good for comparison whereas a line chart is designed to reveal a trend or changes over time. Therefore, we will be creating a bar chart to show the top 10 towns and line charts to show the yearly trend by town.

7 Case Study

To demonstrate how our application can be used, we will be focusing on one town, Pasir Ris as it is a mature estate and it is a popular residential community with plenty of family-friendly amenities. We will set the Year filter to 2018 for demonstration purposes.

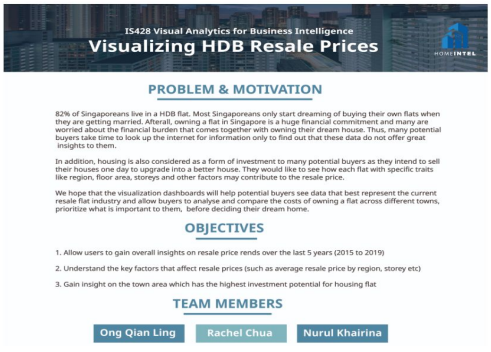

Figure 7 : Home Page

This is our homepage where users can learn more about the problem, motivation and the objectives of the project.

7.1 Dashboard 1: Overview of resale prices

To begin our analysis, we will start off with the first dashboard which aims to provide the user with a rough idea on the variation in resale prices across towns.

Figure 8: Order - Average resale price by town (left), Top ten towns by average resale price (right) for Pasir Ris in 2018

We will set the filters first. The user can click on the Year slider and select 2018. From the **choropleth map**, we can see that Pasir Ris has an average resale price of $498,272, median resale price of $470,000 and the total number of flats sold in this town is 625 in 2018. We can also see how expensive Pasir Ris is compared to the other towns through this map where the average resale price is represented by the color intensity. The darker it is, the higher the amount.

This is also supplemented by the **box plot** which shows the distribution of resale prices of the top 10 towns sorted in descending order by average resale price (left to right). These visualizations can be filtered by year and flat types.

7.2 Dashboard 2: Resale Price Analysis

We will move on to the second dashboard: Resale Price Analysis to go deeper and uncover more insights.

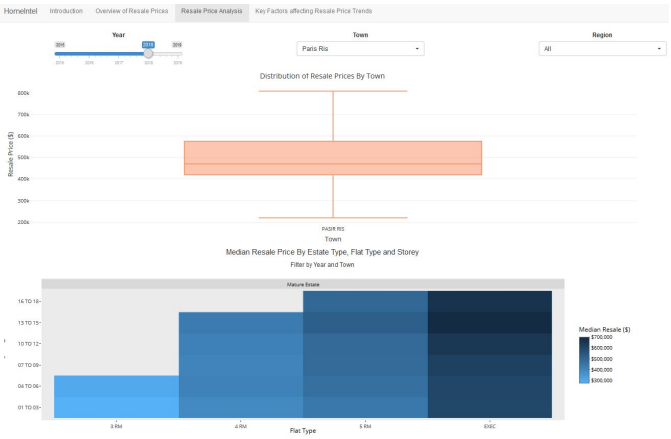

Figure 9: Order - Distribution of resale prices by town (top), Median resale price by estate type, flat type and storey (bottom) for Pasir Ris in 2018

In Figure 9 we set the filters as follows - Year: 2018, Town: Pasir Ris. We would like to focus on Pasir Ris alone first. From the **boxplot**, we can gather the following resale amount details – Minimum: $220,000, Q1: $419,500, Median: $470,000, Q3: $575,000 and Maximum: $808,000.

The boxplot is particularly useful in visualizing skewness[4]. In this case, the resale price is positively (right) skewed as denoted by the median price being closer to the lower/bottom quartile and the mean of $497,272 obtained from the previous dashboard is greater than the median of $470,000. This suggests that the transactions in Pasir Ris constitute higher frequency of high values amounts.

Figure 10: Distribution of Resale Prices By Town in the East region in 2018 (cropped)

In Figure 10, we set the filter as follows – Year: 2018, Region: East. We would like 'zoom out' and look at the big picture by comparing Pasir Ris with other towns in the East region. This can be useful for buyers who don't mind living in other towns if it is within the same region.

From the **boxplot**, we can see that Tampines, Pasir Ris and Bedok have a median resale price of $460,000, $470,000 and $370,00 respectively. The shape of the boxplot for Tampines and Bedok is positively (right) skewed as well. However, outliers can be identified in both towns. Therefore, if the user is seeking for a cheaper option, Bedok can be considered as it has the lowest median and minimum price as compared to the other towns in the East.

Figure 11: Median resale price by estate type, flat type and storey for Pasir Ris in 2018 (cropped)

In Figure 11 we set the filters as follows - Year: 2018, Town: Pasir Ris. Some buyers prefer living on a higher storey but is it worth the price? The color intensity of the **heatmap** is represented by the median resale price. The darker the color, the higher the amount. Some of the cells are empty (in grey) since some towns do not have a specific flat type/storey range available.

[4] Boxplots can be interpreted as follows:
1) Normal / Symmetric Distribution: box plot has equal proportions around the median. 2) Positively / Right Skewed: Median is closer to the lower/bottom quartile and when the mean is greater than the median.
3) Negatively / Left Skewed: Median closer to the upper/top quartile and when the mean is less than the median.
4) Outliers: A value that is higher or lower than the 1.5* InterQuartile Range (IQR = Q3 – Q1).
Source: Listen Data

From the heatmap, we can see that the flats in Pasir Ris has a storey range from 0 to 18 and flat types that are 3 room, 4 room, 5 room and executive. We can also see that executive flat types have the highest median resale prices as compared to the other flat types. In this case, the cheapest option would be a 3 room flat.

7.3 Dashboard 3: Key Factors Affecting Resale Price Trends

Lastly, the third dashboard shows the Key Factors affecting Resale Price Trends.

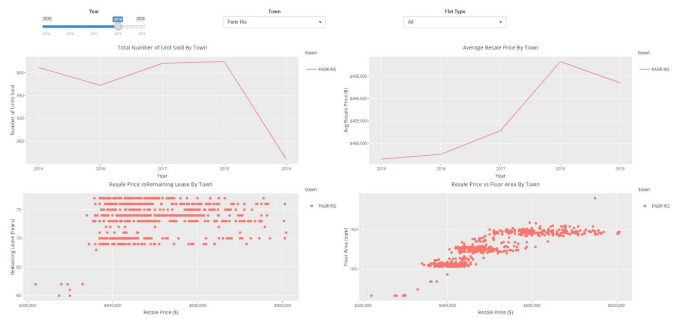

Figure 12: Comparison between resale price and key factors

In Figure 12 we set the filter as follows - Year: 2018, Town: Pasir Ris. On top left left hand corner, you can see the number of flats sold and together with the graph on the top right hand corner that shows the average resale price, we can see the total sales generated. The scatterplots at the bottom of the dashboard shows the relationship between resale price, remaining lease and floor area.

We can see that 628 units were sold and the average resale price has been increasing consistently in 2018. The data for 2019 is incomplete which explains for the dip between 2018 and 2019. In general, we can see that as the remaining lease/floor area increase, the resale price also increases.

8 Key Findings And Insights

The list below features some of the key findings and insights that can be derived from our visualization application:

1. There has been an increase in the number of flats sold in Sengkang, Punggol, Tampines and Sembawang.

As seen in the graph below, apart from 2019 where the data has yet to be completed, there has been an increasing trend in the number of resale flats sold in these towns. In these towns, there was quite a bit of HDB flats under the BTO scheme in the last 10 years. All owners of HDB flats purchased under the BTO scheme are only allowed to sell their unit after the five-year minimum occupation period. Therefore, this might be one reason explaining the increasing trend in the number of flats sold.

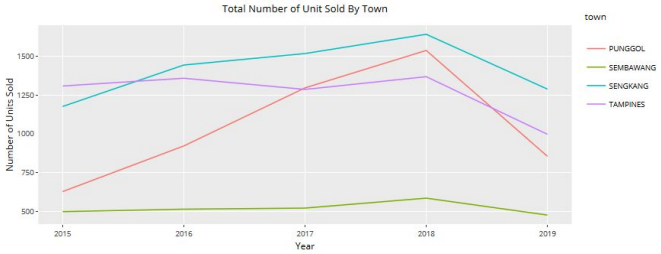

Figure 13: Total Number of Units Sold of 4 towns

2. Resale Price difference between Mature Estate and Non Mature Estate

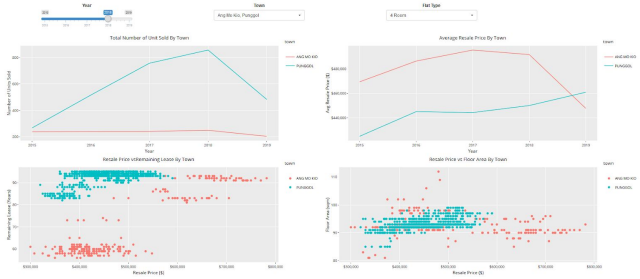

Figure 14: Comparison of resale price and key factors between Ang Mo Kio & Punggol

According to an article by Channel News Asia [8], owners of older flat units are anxious about the short lease years left and the wear and tear caused by the old flats; the two factors that could adversely affect the value of their flats.

However, based on the comparison between mature estate (Ang Mo Kio) and non-mature estate (Punggol), we can clearly see a big difference in average resale price between these two towns. However we notice from the graph floor area vs resale price in 2018 that for 4 room flat, mature estates have a higher price per floor area despite having a lower remaining lease period. This is an interesting finding because despite knowing that there are factors that would hint at a drop in housing prices for mature estates, our graph's findings show otherwise.

3. Limitation of Certain Towns with Specific Room Type

There are certain towns such as Marine Parade and Bukit Timah that do not have any 1 room flats. This would indicate that if Marine Parade and Bukit Timah were to be affected by the nearing lease year-end, the probability that a large number of huge households will be affected. This should be taken note of especially when catering to the flat types to the demand should there be a need to take back the flats. Especially for Marine Parade with limited lease years left.

4. Singapore's Most Expensive HDB Resale Areas

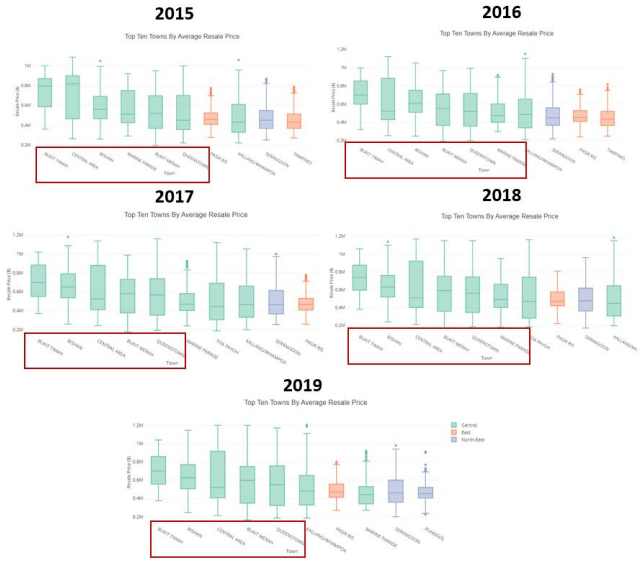

Figure 15: Top 10 Towns By Average Resale Price (2015 to 2019)

The cost of buying a resale flat varies across different towns. From 2015 to 2019, the top 5 towns ranked by average resale price for all flat types have consistently been towns in the Central region: Bukit Timah, Bishan, Central Area, Bukit Merah and Queenstown. These towns are popular and mature and continue to see BTO launches that are significantly oversubscribed (if they are launched within these towns). It is also interesting that Pasir Ris is the only town in the East region that made it to the top 10 towns from 2015 to 2019. These results do not consider the qualitative aspects such as age, remaining lease, floor, internal decoration and condition etc and these factors can introduce a certain amount of bias into the data [9].

Figure 16: Top 10 Towns By Average Resale Price for 3 room flats

We decided to narrow down and filter the results by flat type. We discovered that 2 non-mature estates, Punggol and Sengkang have consistently been in the top 10 towns from 2015 to 2019 for 3-room flats. This shows how demand continues to be strong in these areas. The newer towns such as Punggol and Sengkang attract buyers due to the exciting future developments in these locations [10].

9 Future Work

In this paper, the dataset used was only from one source and we did not include other factors that might have played a part in influencing the choices made by the buyers. Given more time, here are some other future works that could be considered to provide an even better visualizations for potential buyers in Singapore:

i. Distance to Amenities

There are various other amenities that potential buyers/investors would consider before buying a house. These amenities may include distance to educational institution, distance to prestigious primary school[5]the nearest bus stops, nearest MRT, number of MRT lines within 1 kilometres of the house etc. As a result, it can have a huge effect on property prices, and it can be an attractive option for investors.

ii. Other Housing Data (other than resale HDB flats)

In future, when housing data for other types of houses are available such as Build-To-Order flats, landed properties etc, a more sophisticated data visualization can be created with the inclusion of these different housing datasets, making the process of buying a house in Singapore easier for Singaporeans considering different housing options.

iii. Side by Side Comparison (between flat type/towns)

With the proposed idea stated above, we would also like to further enhance the user experience by including a customisable comparison. User will then be able to select two of their preferred housing options either by flat type or town.

With those two additional options, users will be able to analyse the difference, thereby positively influence their decision-making process.

10 Conclusion

In this paper, we have outlined the key issues that potential resale flat buyers in Singapore faces when they are looking for a flat to purchase. The main issue includes the lack of valuable visualizations tools that could allow potential buyers to gain insights of the resale housing industry in Singapore. Using the visualizations we have produced, it serves as a guide for potential buyers to make informed decisions and we hope that potential buyers would find them useful.

11 Acknowledgement

The authors would like to thank Dr KAM Tin Seong, Associate Professor of Information Systems (practice) at the School of Information Systems, Singapore Management University, for his valuable discussions, comments and invaluable help in making this project a success.

12 References

[1] Poh, J. (2019, July 2). *Should You Buy or Rent Your Next Home in Singapore?* Retrieved from Money Smart: https://blog.moneysmart.sg/property/buy-rent-next-home-singapore/

[2] *Singapore Still the Second Priciest Housing Market in the World*. (2019, April 12). Retrieved from Singapore Business Review: https://sbr.com.sg/residential-property/in-focus/singapore-still-second-priciest-housing-market-in-world

[3]*Singapore Property market District Price Trends Heatmap*. Retrieved from https://www.greyloft.com/singapore/property-prices-heatmap

[4] *Home Value Appreciation - A Web Visualization Project*. Retrieved from Get A Flat: https://getaflat.herokuapp.com/charts

[5] Darrick (2019, Sep 12). *HDB Resale Analysis DDO*. Retrieved from Tableau: https://public.tableau.com/profile/darrick8462#!/vizhome/HDBResaleAnalysisDDO/HDBRESALEDASHBOARD

[6] *Resale Flat Prices*, Retrieved from Data.gov.sg: https://data.gov.sg/dataset/resale-flat-prices?resource_id=adbbddd3-30e2-445f-a123-29bee150a6fe

[7] *HDB Estate Information, Retrieved from Stacked Homes*: https://stackedhomes.com/hdb-estates.html

[8]Sing , Dinesh (2018, Sep 23) *Commentary: A nightmare scenario Ang Mo Kio emptying out in 20 years.*Retrieved from Channel News Asia:https://www.channelnewsasia.com/news/commentary/hdb-housing-schemes-vers-seres-hip-ang-mo-kio-10743838

[9] Dayani, Dinesh (2019, March 5) *HDB Price Guide: 5 Most Expensive HDB Estates in 2019*, Retrieved from Dollars and Sense: https://www.businesstimes.com.sg/hub-projects/property-2019-september-issue/hdb-resale-market-sees-strong-demand

[10] Mak, Nicholas (2019, Sep 5) *HDB resale market sees strong demand*, Retrieved from Business Times: https://www.businesstimes.com.sg/hub-projects/property-2019-september-issue/hdb-resale-market-sees-strong-demand

[5] Homeowners who live within 1km from the school will be given priority during the primary school registration exercise. It is thus not uncommon for parents to rent or buy houses nearby their choice of school, especially among top-ranking ones. Source: 99.co