REALVIS: Making Smart Real Estate Decisions using Visual Analytics

Mak Han Ren, Sia Heng Guang, Tan Jit Kai
ISSS608 Visual Analytics and Application 2021/2022 T2, MITB SMU

Project Motivation & Objectives

Real estate is an industry where information is critical to make the right decisions, especially when looking to purchase expensive homes for stay or investment and much of the information while publicly available, is not easily filtered by the average home buyer.

Currently, property listing websites like 99.co and Propnex allow buyers to do specific searches for listed properties in the market, however any analytics regarding past transactions or further information is limited to their paid service applications designed and built for real estate agents.

This means potential buyers are not able to easily study the price trends for themselves without consulting an agent directly who may not be willing to share certain information with them. Our project aims to provide a data visualization tool so that users can see basic analytics regarding past real estate trends in Singapore to make more informed decisions with regards to property transactions.

Dashboard Use Cases

Potential Home Seekers

With over a hundred thousand past property transactions records, buyers will be able to look at the trends of the property market over the previous years and isolate certain key factors such as location, type of property etc. to know what are the trends and past prices for similar properties as they search for their dream home.

Investment Seekers

Users will be able to see the trends in turnover profits for past properties with similar characteristics to properties they are planning to purchase.
Certain factors which could affect the flipping prices can be identified to help them decide what needs to be considered in their decisions.

Data

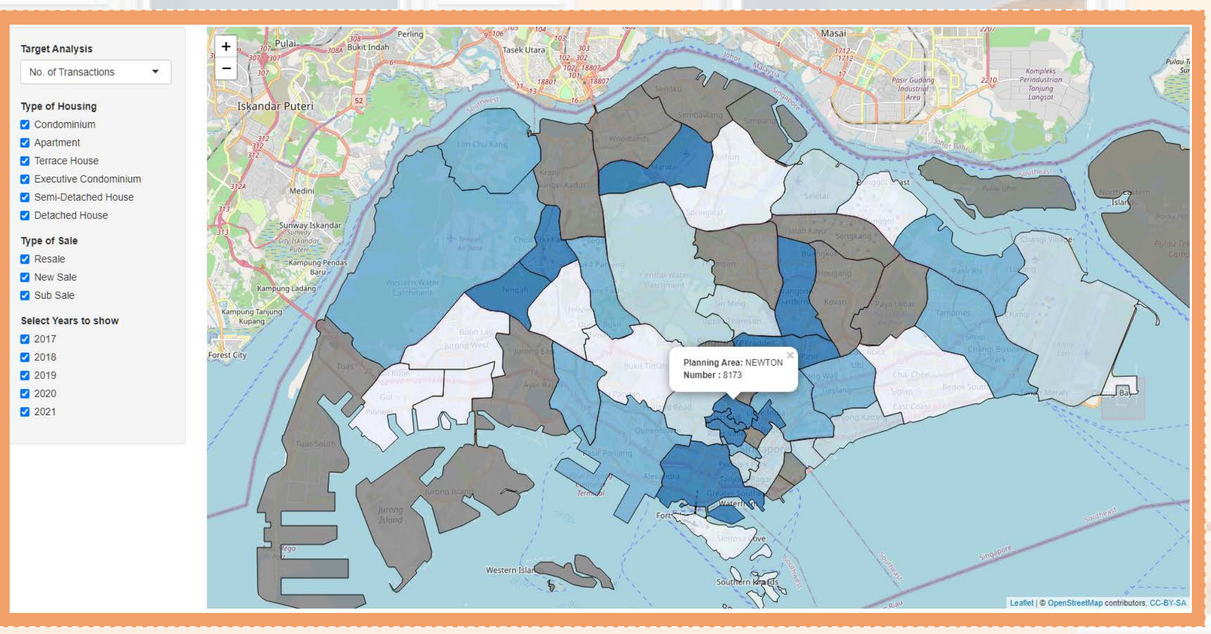
Data set used is obtained from the Realis database which has comprehensive transactional information regarding the Singapore private property market recorded by the Urban Redevelopment Authority.

We will be using the transaction data for Private Residential properties for a period of 5 years from January 2017 to December 2021.

Data Preparation

The raw Realis data sets comprised of several csv files for each year depending on the number of records stored, each csv file usually contained between 3 to 4 thousand records. To make a single csv file for easy access, we imported the csv files into a single data frame using R before making the following edits. First we filtered the `Number of Units` to exclude all transactions with more than 1 unit because we wanted to focus on residential transactions for home-buyers instead of commercial purchases. Next we change the format of the `Sale Date` column to `%d %b %Y and lastly we split the `Tenure` column into 2 separate columns for date of tenure start and duration of tenure. Finally, we save this data frame as the base data frame csv for our visualizations. Further visualization-specific data changes were made off the base data frame by us individually.

Making Smart Real Estate Decisions using Visual Analytics

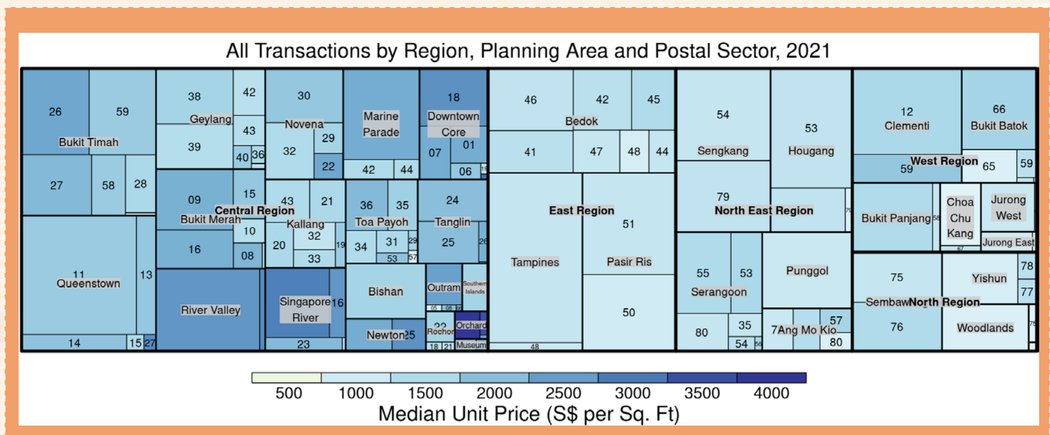


Choropleth Map

A map module was to show the prices and distribution of transactions in Singapore during the selected time periods.

The leaflet library was used to overlay the map as it is highly customisable and produces the best visuals. To make use of Shiny's capabilities, the input for the left panel are all reactive and the chart will update based on user input.

Users can adjust the filters to view the statistics of the number of transactions, average price (S$) and median price (S$) based on the various types of housing, the types of sale and the years to show.



Tree Map

Our Treemap module was used to show the prices and distribution of transactions in Singapore during the selected time periods. Users can adjust the filters to view transactions that are more relevant to them.

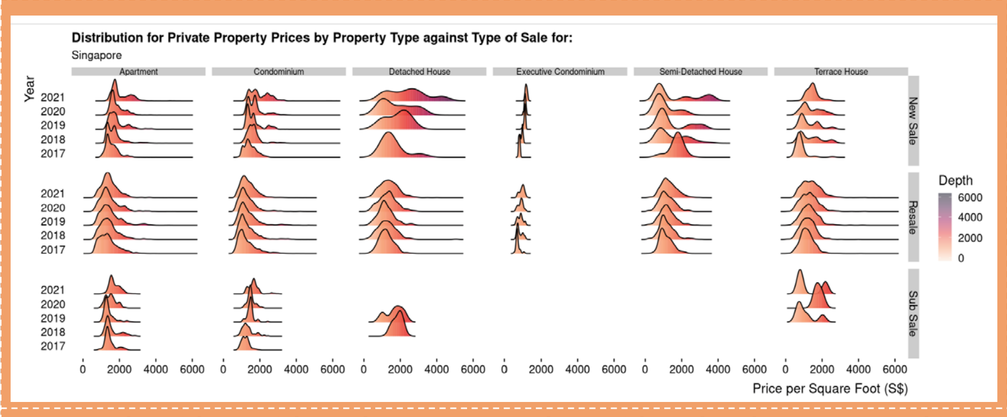
To create the treemap, we used the treemap package to create a static treemap based on the indexes: `Planning Region`, `Planning Area`, `Postal Sector` in order. The size of the treemap was based on the total units sold, which was the sum for `Number of Units` sold grouped by the index. Finally, the tree map color is based on `Median Unit Price (S$ per square foot) which was the median of `Unit Price (S PSF)`.
Based on our observations, the Central region had the most transactions in general with the highest Median Unit Price areas being Newton, River Valley and Orchard. Those 3 areas were usually the darkest rectangle in the tree map when they appeared.

Confirmatory Data Analysis

From our secondary research, we have identified key issues that our dashboard users could be facing. We summarised them into 3 hypothesis that we will be testing:

1. Are Landed Properties more expensive than Condominiums?
2. Are Prices of Free-hold properties different from Lease-Hold properties?
3. Is there a correlation between Floor and Prices?

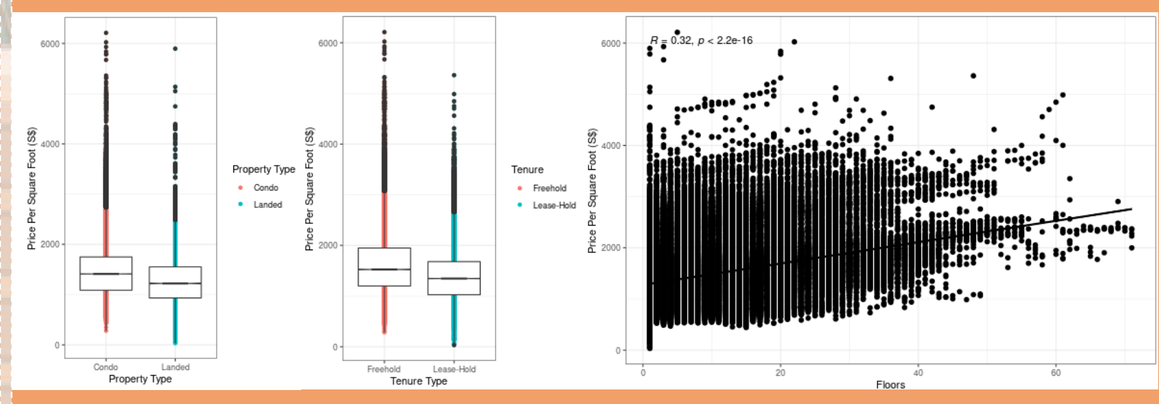
We used ggplot to plot out a boxplot diagram containing the 'Price Per Square Foot' value of the property types. We notice a difference in the mean values or both "Freehold vs Leasehold" and "Landed vs Condo".
We then performed a t-test using both Wilcoxon and two sample means at 0.95 significance level. Given that p-value is less than 0.05, we reject the null hypothesis that there are no differences in prices. We then used a scatterplot to identify a weak positive correlation between the number of floors and prices.



Ridge Plot

The ridge plot was used to show the prices and distribution of transactions in Singapore during the selected time periods.

We used ggridges with ggplot to create a base ridge plot that we turned into a facet grid based on the Property Type and Type of Sale. This was so that users could observe and compare different filtered ridge plots more easily. Additionally, users can decide how many years do they want to compare across to look for changes in the trend over time.



Condo vs Landed

Freehold vs Leasehold

Floor (Level) vs Prices

Future Directions

We can extend the number of years of data that we are using to review trends during huge economic crisis such as 1997 Asian Financial Crisis and 2007/08 Global Financial Crisis. As for the scope of the project, we can add in a property price predictor using ARIMA. Based on the user's input on variables such as preferred Property Type, Unit Price per Square foot and Planning Area, we can recommend the user a certain property in the market for them to look at. We can also expand our project's geographical scope where we perform similar property price analysis on countries such as Hong Kong and Taiwan.