



Singapore Real Estate Analysis

A Course Application Report
submitted as part of the course
Data Analytics
CSE3025
School Of Computer Science and Engineering
VIT Chennai

FALL 2021-2022

Course Faculty : Dr. B Radhika Selvamani

Submitted By
N.Uday (19BCE1317)
B.Karthik(19BCE1446)
Pushkar Nath Jha (19BCE1478)
S.Suriyakrishnan (19BCE1050)

Abstract

In most nations, land and buildings account for a considerable amount of national wealth, but in Singapore, a tiny city-state with limited territory, real estate plays a critical role in the overall economy. In Singapore, international property and development firms play a vital role. They're a big element of the local stock market, and they've seen a lot of growth in recent years, bringing in a lot of money for governments and a lot of wealth for people. Because of this significant role, the report reclassifies Singapore as a property state, rather than a rent-seeking economy, as the term rent is commonly employed in discussions of Asian capitalism. The prominence of real estate, an immovable asset, is remarkable not just in a global period defined in part by capital mobility, but also since it is publicly held in Singapore. The research examines Singapore's land distribution systems in an attempt to explain the oddity.

1 Introduction

In Singapore, housing is a very valuable asset, not least since just over 90% of all households own their homes. In Singapore, home equity accounts for 40% of household sector net worth, significantly higher than the equivalent share of 14 percent in the United States. Residential construction also accounts for a major portion of overall economic activity. In 2015, expenditures on 'Residential Buildings' accounted for 28 percent of total gross fixed capital creation in Singapore, or little over 7% of GDP.

Governments all across the globe are interested in their housing markets for a variety of societal reasons. Housing is the place where individuals organise their personal and professional life in terms of space. Affordability circumstances can be a major worry for policymakers, hence distributional concerns are frequently significant. Governments, like Singapore's, are compelled to interfere in their housing markets as a result of these circumstances. The government is heavily active in the real estate industry in general, selling public land for private (and public) construction. Following the global financial crisis, the government decided to respond to a huge increase of 40% or more in aggregate home prices. Affordability circumstances worsened by almost a third between 2007 and 2010, as seen by growing price-to-income ratios for individuals purchasing modest-sized Housing and Development Board apartments.

Singapore is a city-state with a finite amount of territory. With a resident population of 5.784 million people living on a total land area of 719.1 square kilometres, it has a high population density of slightly over 8,000 people per square kilometre. Singapore is likewise highly developed, with a gross domestic product (GDP) per capita of Int \$87,855, placing it in the top four economies in the world (International Monetary Fund, 2016).

2 Related Works

Real estate investments have become more popular in last few decades. The existing system involves calculation of house prices without the necessary prediction about future market trends and price increase. The future prices will be predicted by analyzing previous market trend and price ranges and also upcoming developments future prices. Linear Regression is used for prediction since there is no involvement of brokers it reduces the risk involved during monetary transactions. The proposed system is a solution where the clients can view the future value of the real estate and invest accordingly without approaching an agent.[1]

As the COVID-19 pandemic came unexpectedly, many real estate experts claimed that the property values would fall like the 2007 crash. However, this study raises the question of what attributes of an apartment are most likely to influence a price revision during the pandemic. This study used a web-scraping algorithm

and collected a total of 18,992 property listings in the city of Vilnius during the first wave of the COVID-19 pandemic. The retrieved SHAP values conclude that the time-on-the-market variable was by far the most dominant and consistent variable for price revision forecasting. Additionally, the time-on-the-market variable exhibited an inverse U-shaped behaviour.[2]

In the building industry, analytics has altered the fundamental pattern of data processing and forecasting. Delights, the lobbying firm of a newly established real estate buyer, is aiming to penetrate the Melbourne property market. In order to generate insights on various aspects of this booming market, senior management is keen to capitalize on large volumes of historical real estate data. In this article, the descriptive and inferential analytics approach is used to extract insights into the real estate industry.[3]

The real estate market is one of the most competitive in terms of pricing and same tends to vary significantly based on numerous factors; forecasting property price is an important module in decision making for both the buyers and investors in supporting budget allocation, finding property finding stratagems and determining suitable policies. The paper presents various important features to use while predicting housing prices with good accuracy. The regression models can be used, using various features to have lower Residual Sum of Squares error. While using features in a regression model some feature engineering is required for better prediction. Often a set of features multiple regressions or polynomial regression (applying a various set of powers in the features) is used for making better model fit. For these models are expected to be susceptible towards over fitting ridge regression is used to reduce it. So, it directs to the best application of regression models in addition to other techniques to optimize the result.[4]

The concept of big data though relatively new has brought a lot of solutions to modern day challenges. Many authors, particularly in developed countries, have adopted the concept in tackling the numerous challenges unfolding in the real estate profession. However, most of the findings from these authors are on individual bases and as such, there is a need to reach a general consensus on the relevance of big data to the real estate profession. The review shows the impact of big data to include digitization of records, information on user preferences, sensor information on the urban environment and sensor information on movement. The paper concludes that the relevance of big data to the real estate profession cannot be over-emphasised.[5]

3 Proposed Work



3.1 Dataset Description

1. **month**: The month and year in which the flat was sold.
2. **town**: The town the flat is situated.
3. **flat_type**: The category of different flat types.
4. **block**: The block in the town the flat is situated.
5. **street_name**: The street where the flat is situated.
6. **storey_range**: The range of the storey a flat has.
7. **floor_area_sqm**: The dimension for flat types.
8. **flat_model**: The category for the type of flat model.
9. **lease_commence_date**: The year when the flat was left for resales.
10. **resale_price**: The price for which the flat is available.

3.2 Data Exploration

In this phase useful information about the data has been extracted from the dataset. We dealt with Singapore real-estate flat resale data, which helps us analysing about various flats, consisting of various factors like month, town, flat type, block, street name, storey range, floor area, flat model, lease commence date and resale price. These are the variables included in the data set which will be useful for analysing the flat resale trends in Singapore.

3.3 Data Cleaning

Data cleaning is an important task for further processing of data. Any data may consist of inconsistency and also consist missing values in such case handling such missing data is important. It is divided into 2 phases:

1. Numerical
2. Categorical

For the numerical data, missing values have been imputed using mean imputation, factors like floor area and resale price has quiet a few missing values which have been replaces by the mean of those columns. For the categorical data the categories have been converted to numeric factors and they have been replaced by median imputation, factors like flat type(char) had missing values imputed by median of its numeric factors.

The Flat models and street names consisted inconsistent naming, which has been replaced by standard flat models.

3.4 Data Visualization

It is important that apart from analyzing the data statistically it is equally important that that we analyze the data visually. We know that some times numerical data cannot convey the message to the user who is trying to analyze the data. He/she has no idea how big the numerical values after statistical analysis will look like.

That's where the data visualization comes. With the helps of different models and graphs it is easier to analyze and convey the message to the other party. Where it is useful? It is useful in presentations of all companies to learn their companies market. It helps the company boss and sales manager the condition of the company. While we are visualizing the data in our project we could clearly see the results from those graphs.

3.4.1 Why we are using Data Visualization?

1. We are using it to make the data easier in understand and remember.
2. It visualize relationships and patterns quickly.
3. It helped us by allowing us to ask a better question among ourselves and make better decisions.
4. It helped us improve insights. From the Data Visualization we learned how the sales of the flats been for more than a decade and which flats and which town are more expensive or more affordable.

3.5 Problem Analysis

Problem analysis. This is the most important part of the project as we have to do this in a team. This can done only when we as a team forms questions and find answers.

In this project we discussed some questions among the team and tried to analyze the problem successfully. We mentioned some of the questions here:

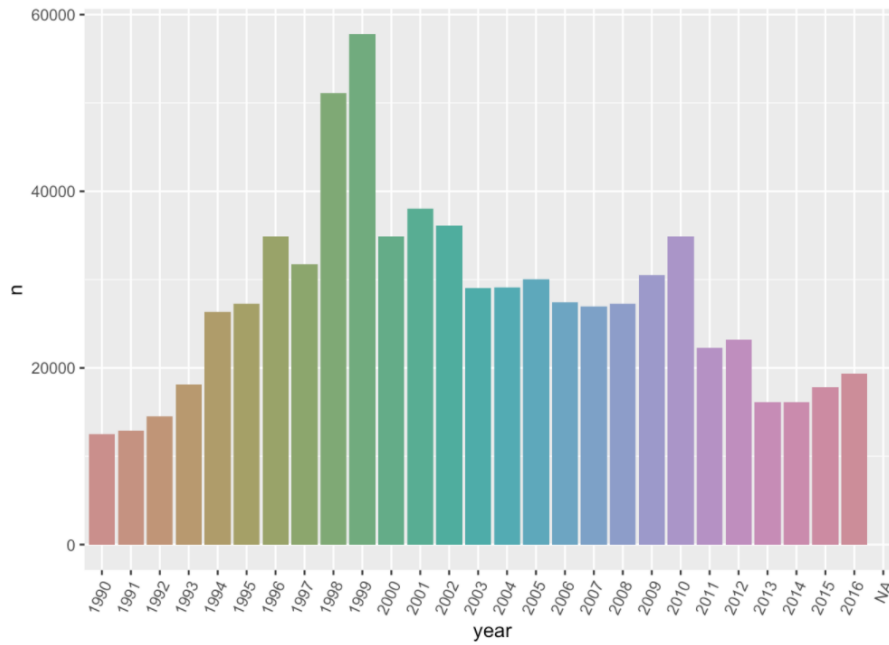
1. Which year between 1990 and 2016 showed a rise in the selling of flats? or the maximum number of resale apartments available in a given year?

2. From 1990 to 2016, which town has the most expensive selling price?
3. Categorize which towns are more expensive and which towns are more affordable for property purchasers in the dataset.

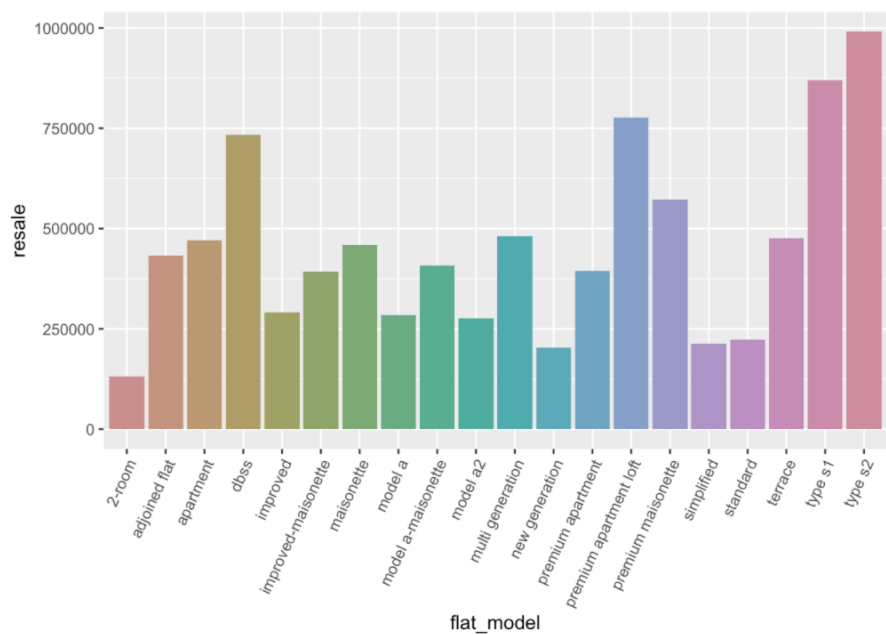
With the help of these questions we are trying to analyze the data statistically and visually to understand the dataset, i.e. the resales of the flats in Singapore's towns and how expensive or affordable can a flat be in these towns.

4 Results & Discussion

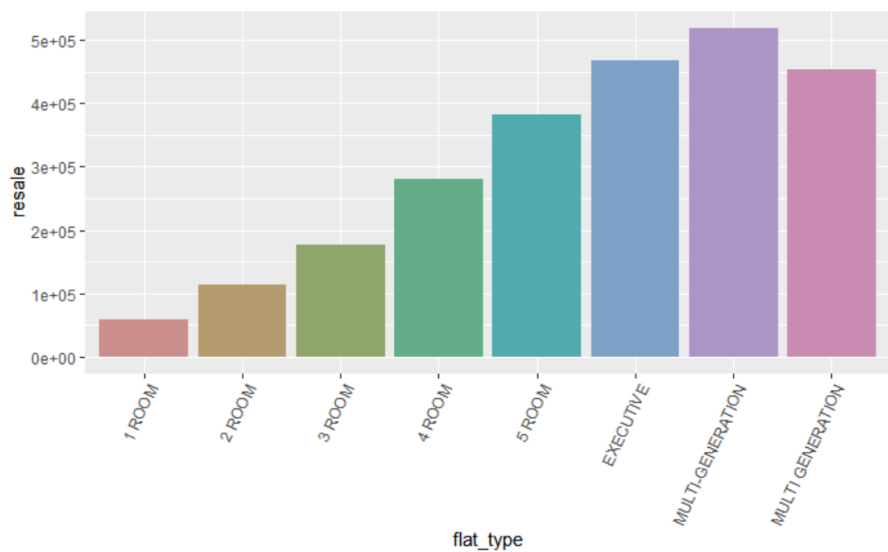
The data we obtained from the detailed analysis/interpretation of the dataset should be given from the following figures:



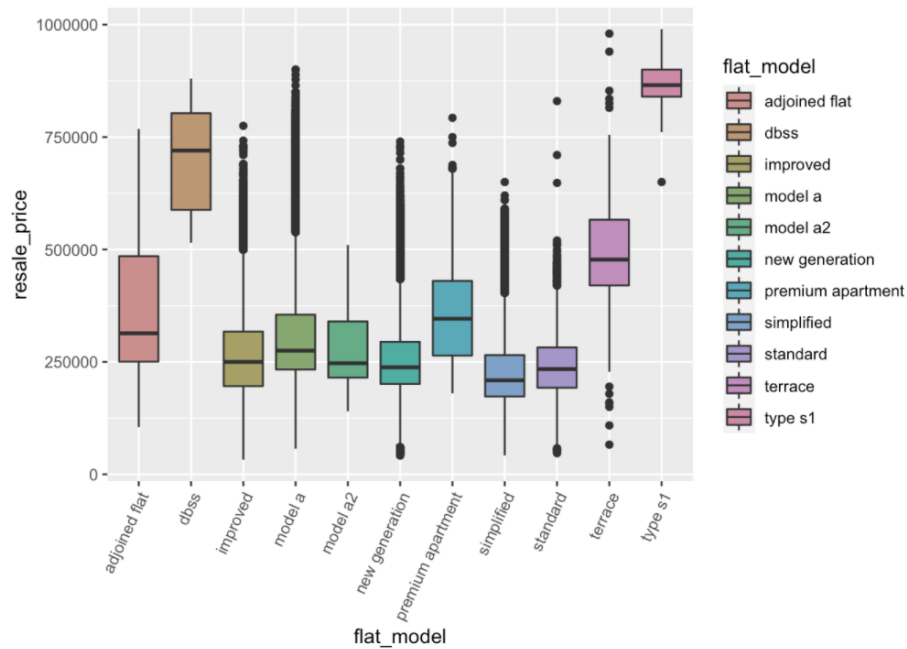
The above figure is a bar chart which shows the number of sales of the flats in the respective years. From the analysis we can conclude that 1999 has seen the maximum number of sales of the flats.



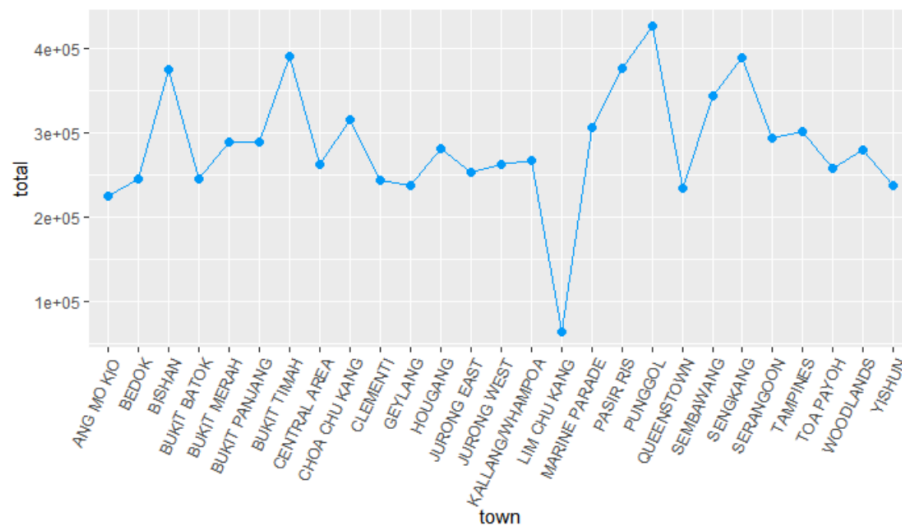
The above graph shows the average resales price for different type of flat models.



The above graph shows the average resales price for different type of flat types.

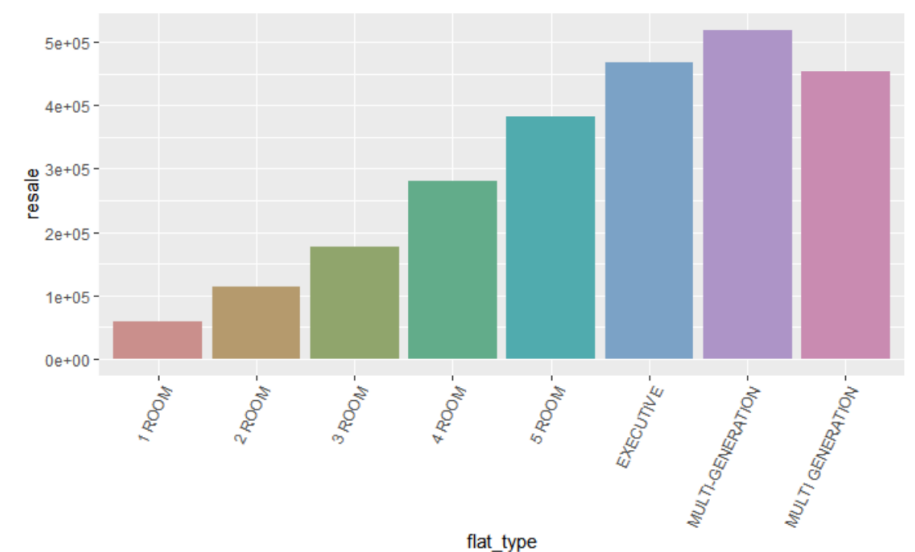


The above bar chart shows the resale price for the different types of flat model for the 4 room flat type.

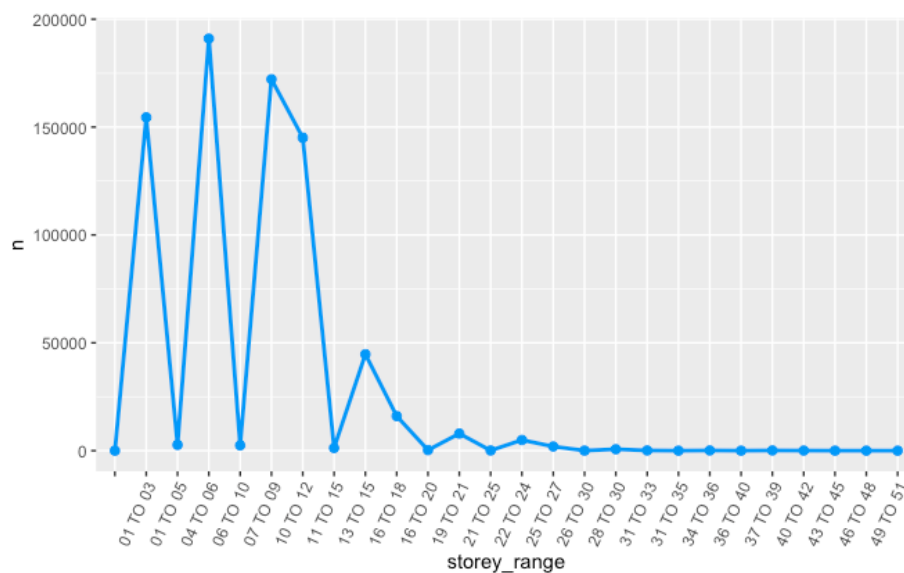


The above line chart shows the average resale price for each town and shows which town has the highest average resale price. The analysis shows that the most expensive resale price is observed in town PUNGGOL with avg. price of 427489.64.

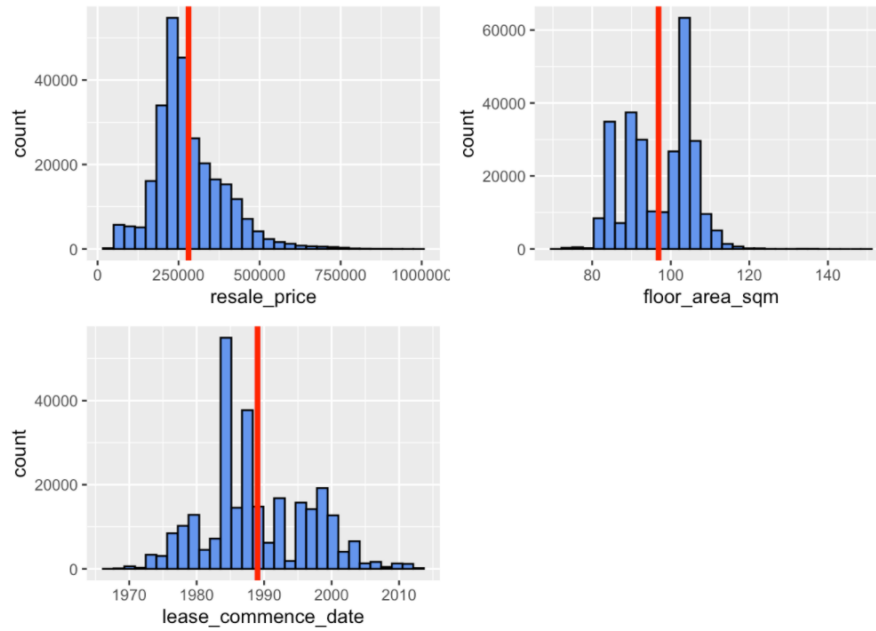
This graph shows which type of flat and in which range the flat type is most bought.



The below graph shows the frequency of the storey range in all types of flat.

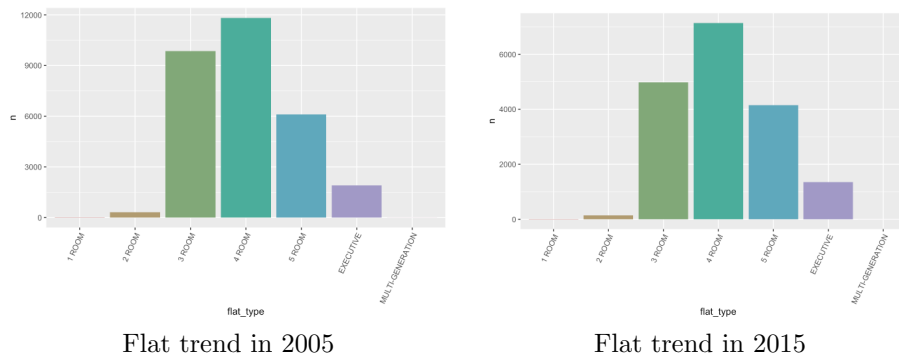


This graph shows the most bought flats in a four bedroom flat.

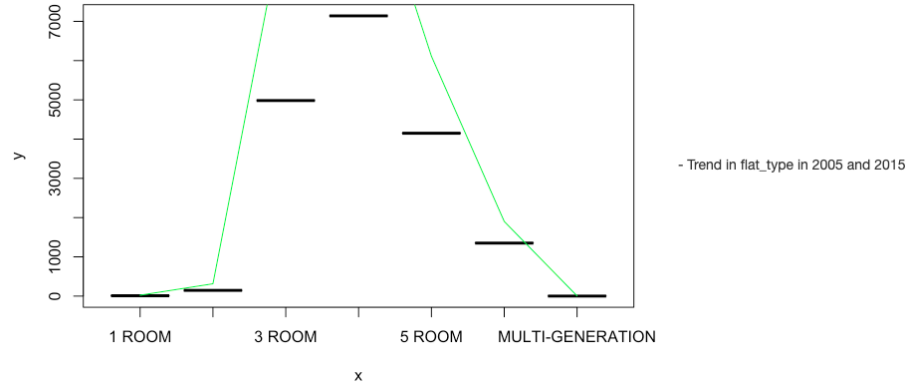


- Most people tend to buy a four bedroom flat with 90-110 floor area
- Most people buy a four bedroom flat with cost between 2-3.5lakhs

The Below graph shows the flat trend for year 2005 and 2015 respectively.

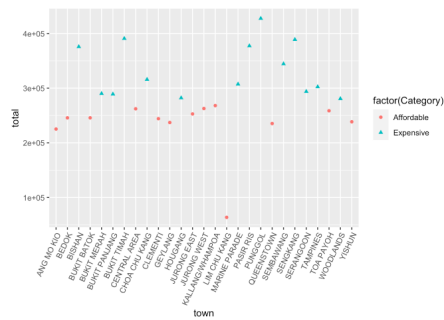


This graph shows the flat trend comparison between year 2005 and 2015.

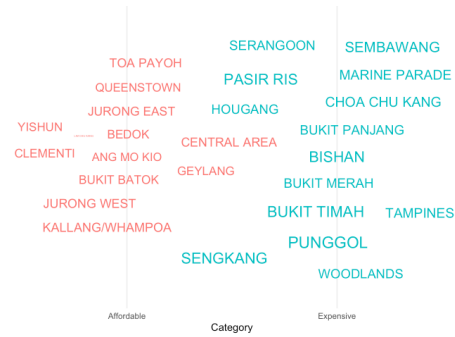


Flat trend comparison between 2005 and 2015

The below graphs shows the comparison between affordable and expensive towns by taking the average of the resale price.



affordable vs expensive towns



affordable vs expensive towns

5 Conclusion & Future Work

Data Analysis plays a crucial role in today's world in improving the businesses. The analysis helps the business leaders and sales executive by giving them the summary of large datasets in just few sentences or lines.

We have done Data Analysis using R to understand the dataset we have been given. With the knowledge we gained and learned on how to use R we used the given dataset to find the answers to the questions which rise. The knowledge from R in doing analysis helps in solving many problems, questions that are stated in the scenario. The plots help in understanding the growth, profits, sales per year for the resales of the flats in Singapore. Various plots give various results that give solutions to most of the questions raised in Business leaders and

sales executives.

From our project we learned about which towns are most expensive and most affordable for property purchases and which type of room is more bought when there was resales of flats are in Singapore. We also get the information regarding the resales of each type of flats in each year. We also learned which town has the highest resales of the flats, and shows us the data by visualizing them into different plots and gives us the summary of such a large dataset which has years worth of data mentioned.

Hence we conclude our data analysis here getting some useful conclusions.

In future work what we wish to complete or accomplish is, the analysis and visualization of the graphs be available to the customers or sales person who wish to buy or sell the flat remotely through an online system. With this system the customers and those who are trying to sell the flats can have the knowledge of the current trend in customers who are trying to buy the flats and also get to know which town is having affordable or expensive flats for sales.

Dataset: here

References

- [1] Geetali Banerji and Kanak Saxena. Analysis of data mining techniques on real estate. *International Journal of Soft Computing and Engineering (IJSCE) ISSN*, pages 2231–2307, 2012.
- [2] Andrius Grybauskas, Vaida Pilinkienė, and Alina Stundžienė. Predictive analytics using big data for the real estate market during the covid-19 pandemic. *Journal of big data*, 8(1):1–20, 2021.
- [3] Mario Bodenbender, Björn-Martin Kurzrock, and Philipp Maximilian Müller. Broad application of artificial intelligence for document classification, information extraction and predictive analytics in real estate. *Journal of general management*, 44(3):170–179, 2019.
- [4] Eulália Santos, Fernando Tavares, Vasco Tavares, and Vanessa Ratten. Comparative analysis of the importance of determining factors in the choice and sale of apartments. *Sustainability*, 13(16):8731, 2021.
- [5] Adedamola O Oluwunmi, BA Role, OM Akinwale, OP Oladayo, and TO Afolabi. Big data and real estate: A review of literature. In *Journal of Physics: Conference Series*, volume 1378, page 032015. IOP Publishing, 2019.