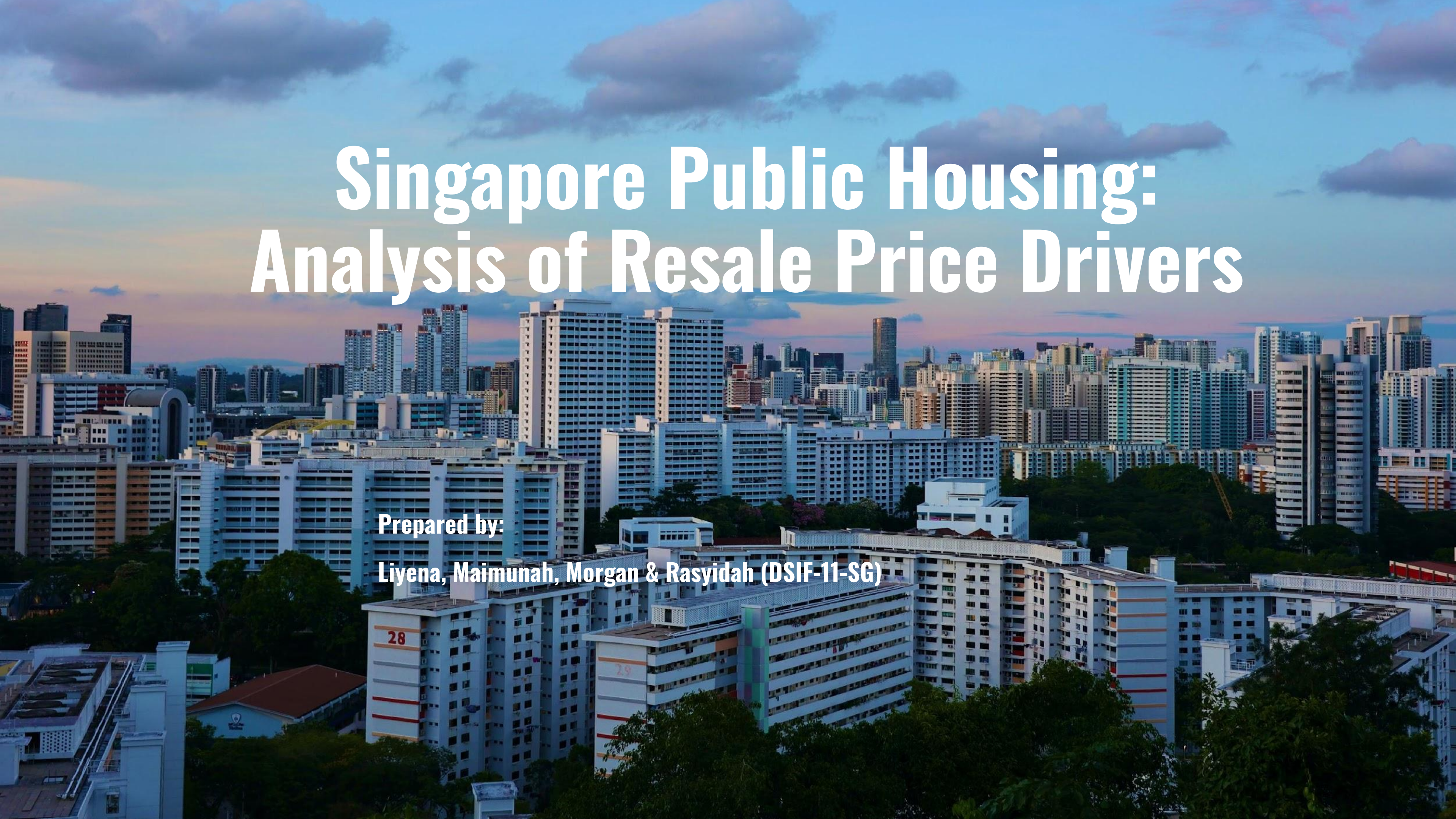


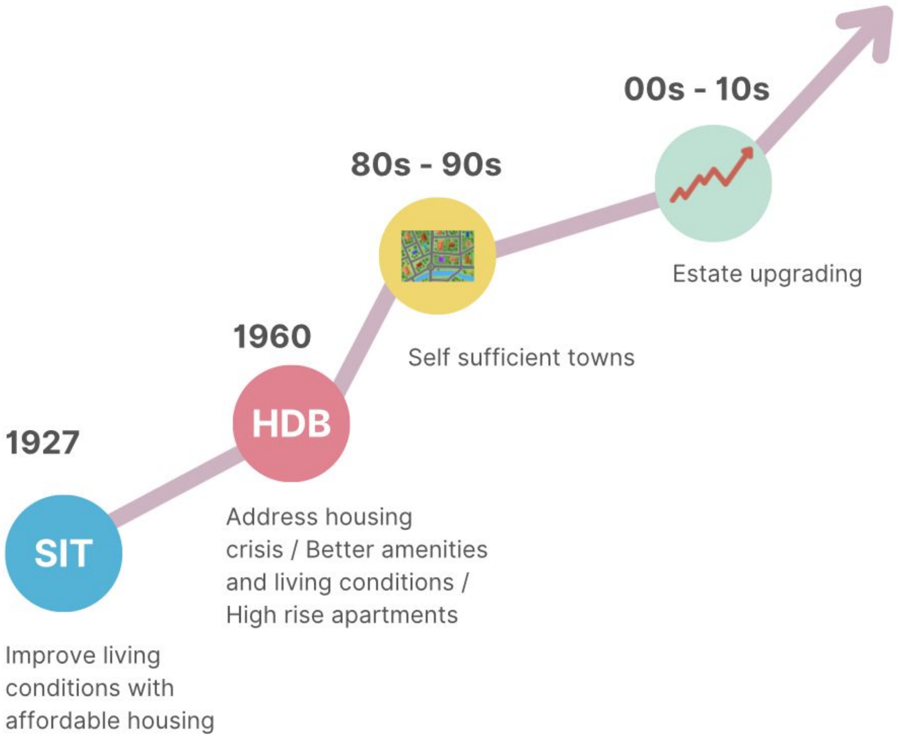
Singapore Public Housing: Analysis of Resale Price Drivers

Prepared by:

Liyena, Maimunah, Morgan & Rasyidah (DSIF-11-SG)

Table of Contents

Background


00s - 10s
80s - 90s
1960
1927
HDB
SIT
Estate upgrading
Self sufficient towns
Address housing
crisis / Better amenities
and living conditions /
High rise apartments
Improve living
conditions with
affordable housing

Problem Statement

Taking a purposive interpretation to the objectives of the HDB - it appears that there is an element of public housing in Singapore that runs counter to such objectives. Obtaining flats from the HDB is similar to a lottery, in that the ability to make a purchase of subsidised public housing is won by what is effectively a game of chance. The winners of this HDB-sponsored lottery are able to resell their flats for a significant profit once they have occupied their flats for a certain amount of time.

Objectives

1. We should not consider bringing down the prices of resale flats in "expensive areas". This damages GDP.

2. We should instead bring up the prices of flats in "less expensive areas" so that Singaporeans who do not win the "HDB Lottery" obtain similar benefits from the growth of the nation.

3. State resources are limited, and we should analyse the cheapest changes which can be made in order to spur value grown in "less expensive areas".

4. In order of preference, we should prioritise recommendations in the following order: (a) Incentives to private industry to spend money, (b) Infrastructure development, since this will need to be done anyway, (c) in the worst case, direct state intervention in the construction of amenities, as these will likely cost the state the most.

Scope

1. Use location and price data of resale flats, containing information about the local area such as the nearest MRT station to the flat being transacted.

2. Consider the impact of the local area on the resale price, specifically, the coefficient of any one feature on the regression coefficient.

Data

1. Price data of resale flats.

2. Features associated with each sale, such as nearest MRT Station, nearest mall, nearest hawker, etc.

Methods and Tools

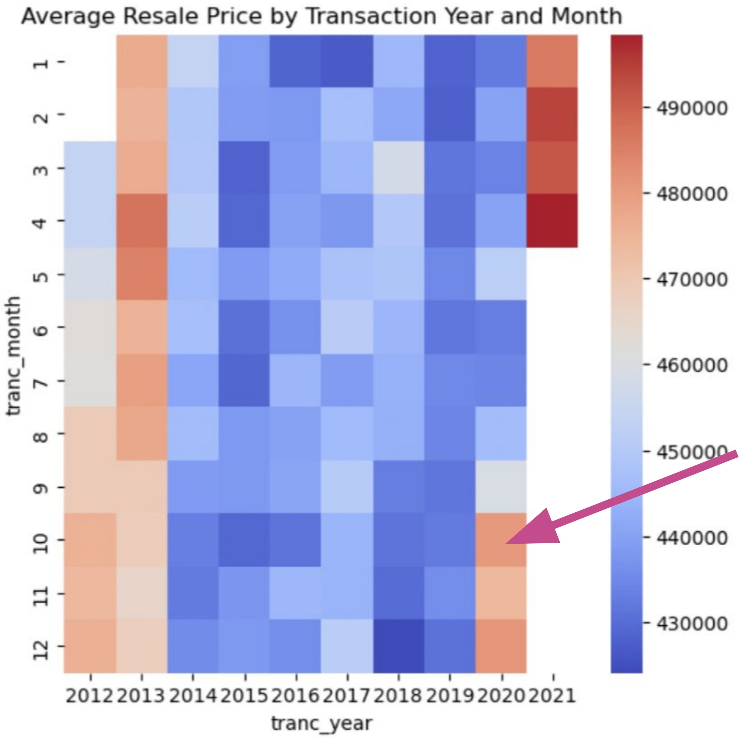
1. Regression analysis.

2. Thereafter, the application of Lasso and Ridge in case these improve performance of a standard regression model.

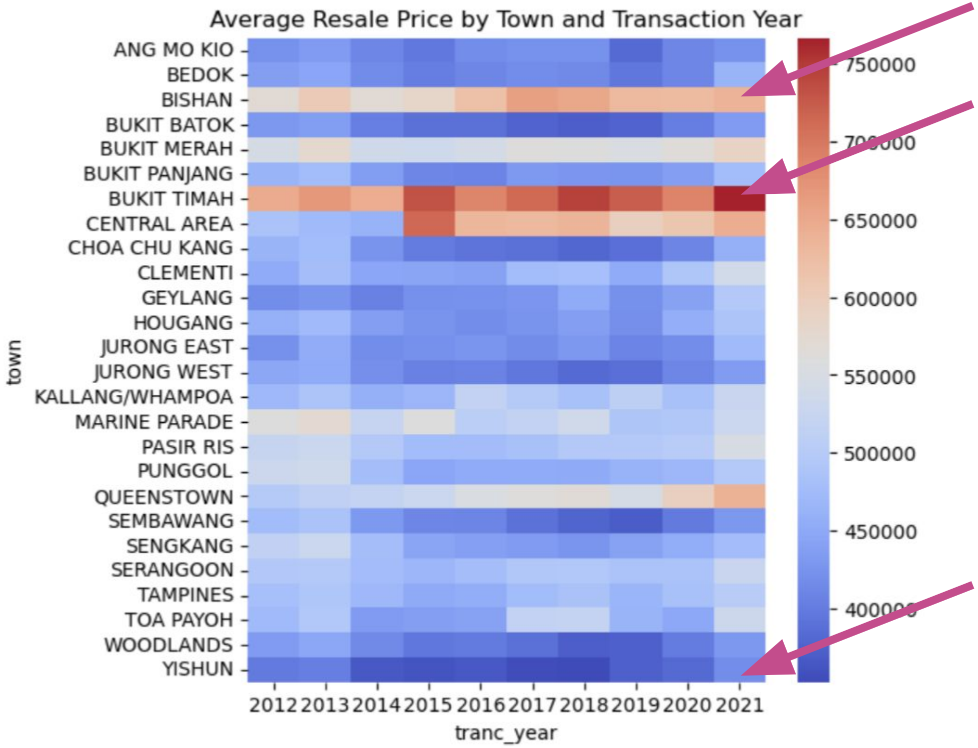
Success Metrics

1. Recommend the most efficient (in terms of deployment of state resources) way to increase the value of HDB flats in "less expensive areas".

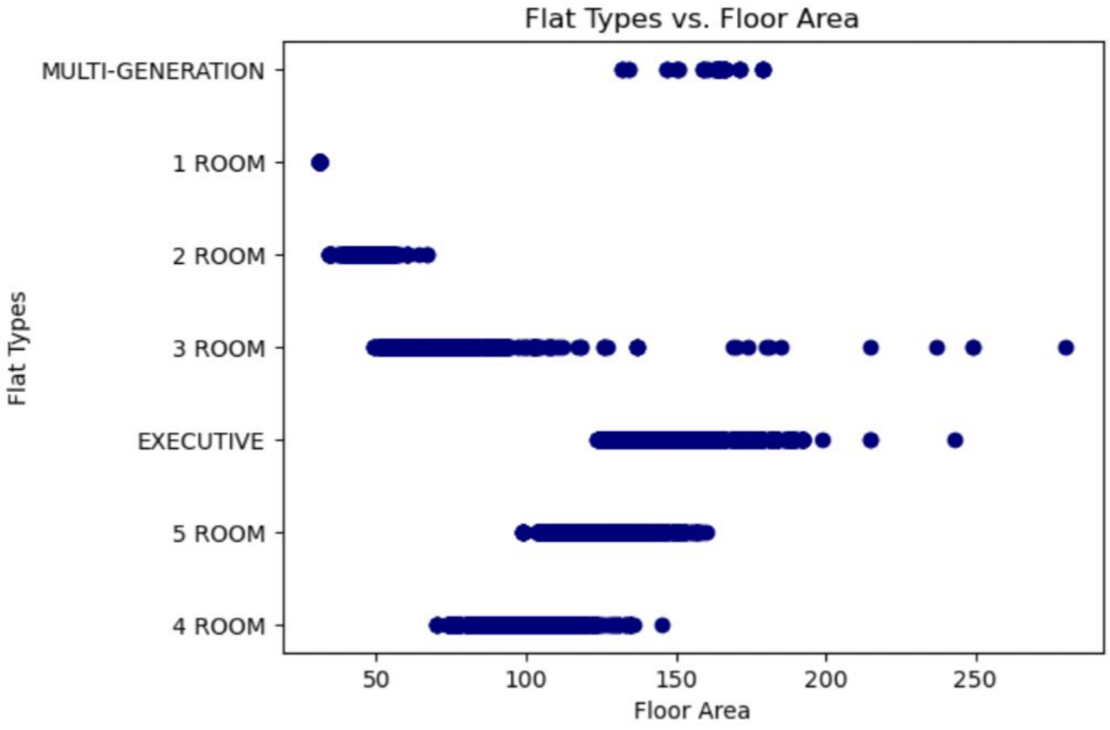
EDA - Stable average prices



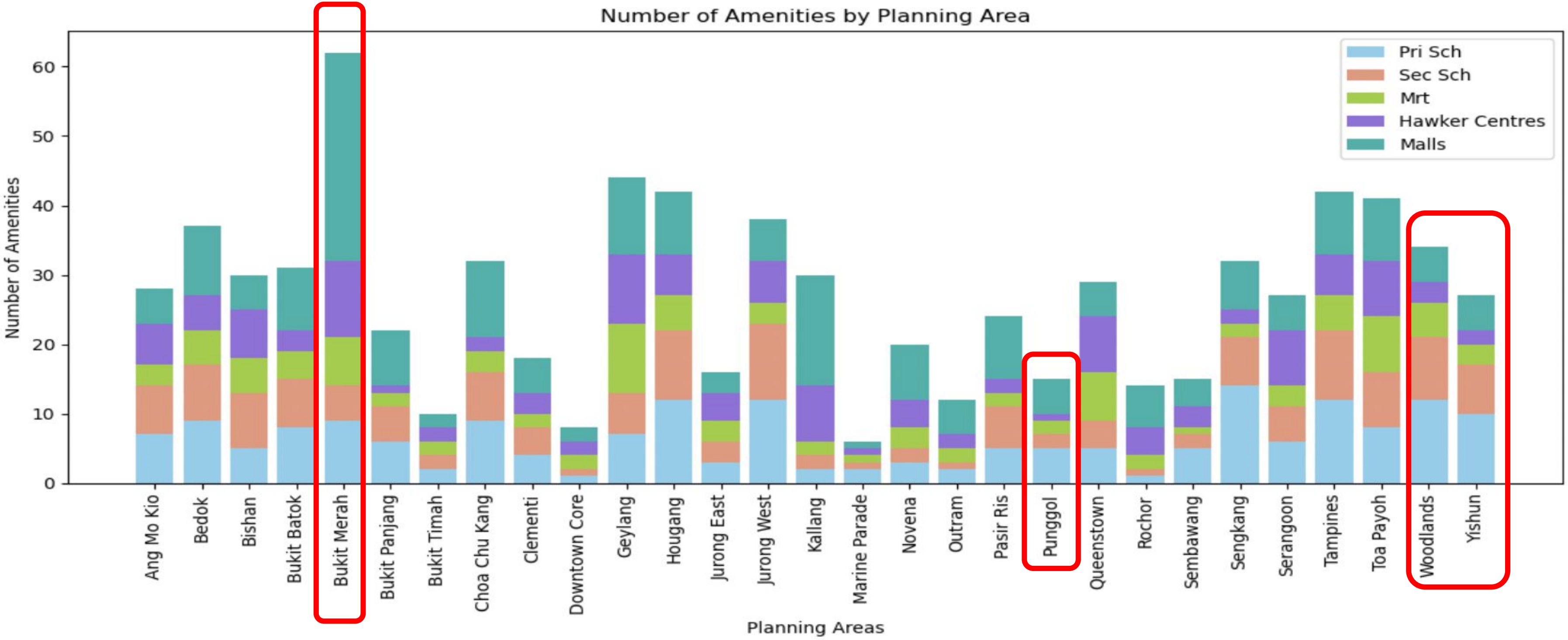
EDA - ‘Hot’ towns & Yishun



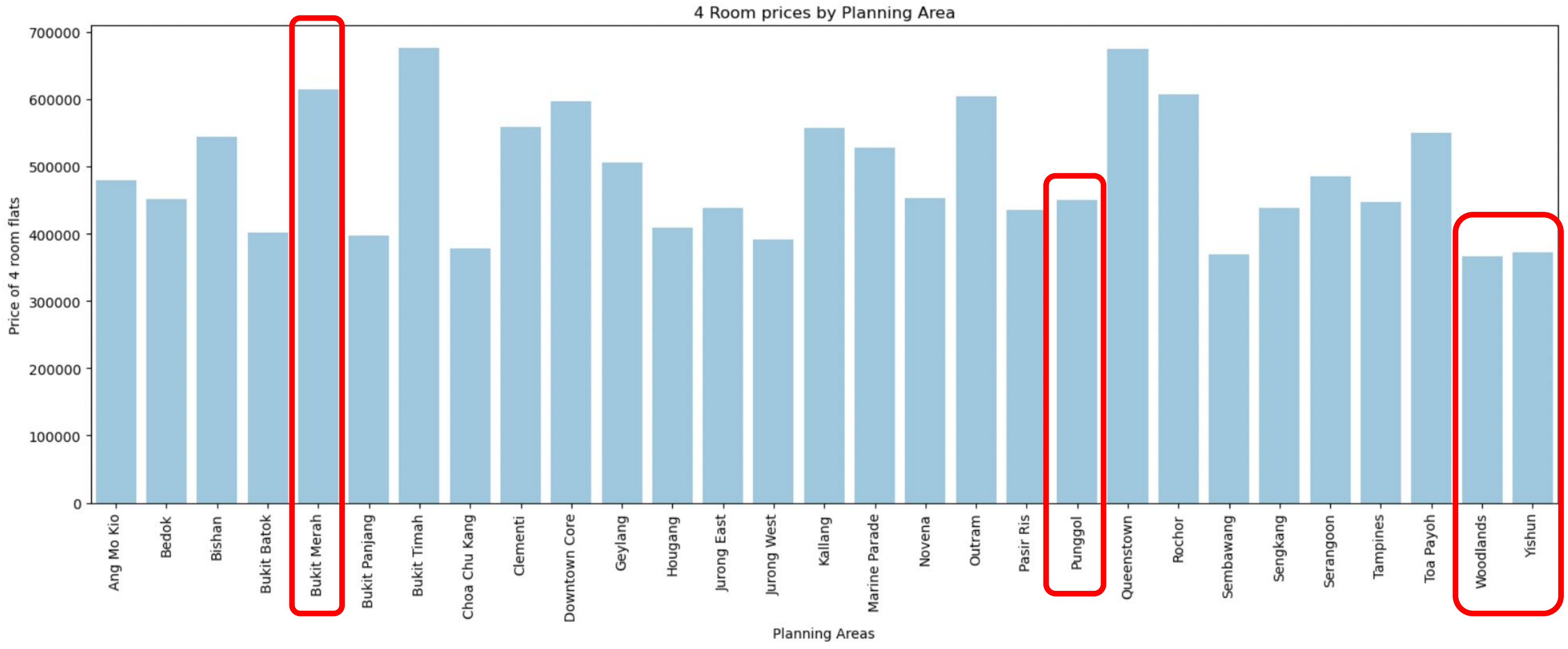
EDA - How big are the flats?



EDA - Which town has the most amenities?

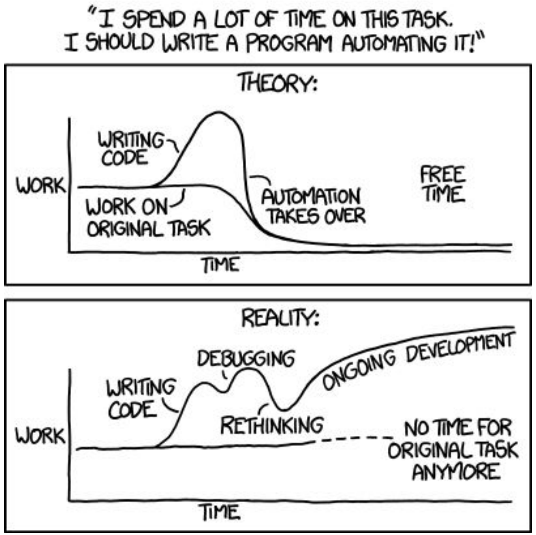


EDA - Yishun 4 room prices is one of the lowest



Challenges - Feature Selection

There were 76 features in the data set. Even after paring away those with no clear correlation with the resale price, there were still 19 features. It was unclear whether we could remove more features. The solution is automation.

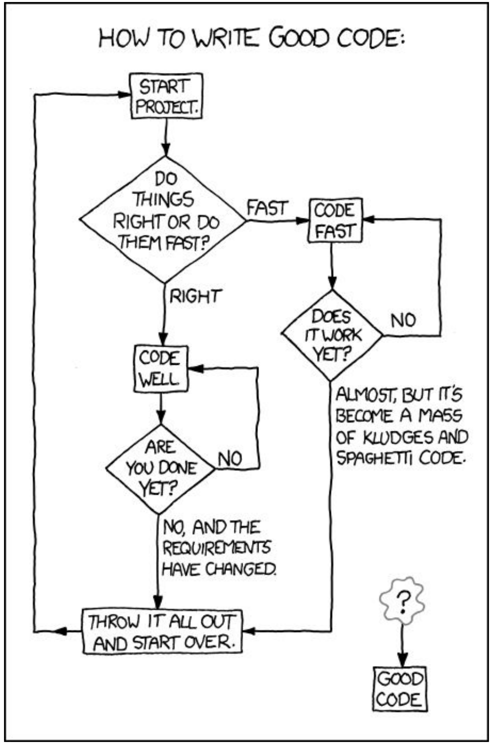


Challenges - Feature Selection

A feature selection program was created. OOP was employed for modularity and maintainability*.

It looked through every single possible combination of between 14 and 19 features (over 60,000) and showed that there were diminishing returns above 14 features.

* For more information, please refer to the technical report.

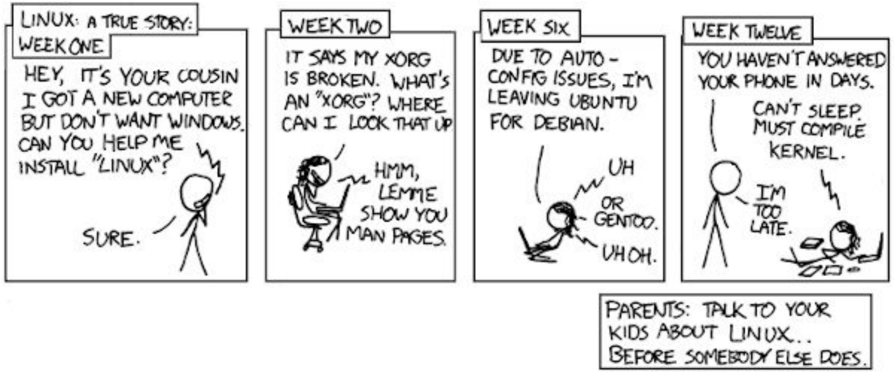


Challenges - Feature Selection

Each run of the program took 30 - 40 hours. Attempts were made to parallelise execution using joblib / multiprocessing and replacing pandas / sk-learn with cudf / cuml (Nvidia RAPIDS).

One Linux install and much tinkering later, the parallelised operation worked but it was slower, likely due to the need to sequentially save results. There was therefore no choice but to wait for complete execution of the ordinary program.

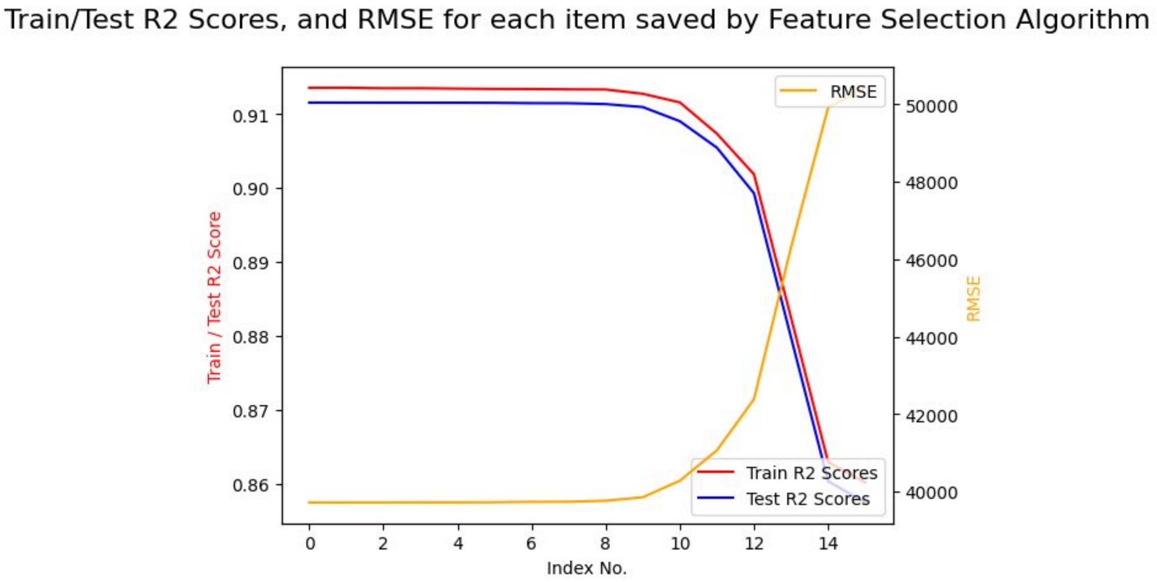
*Source code of experimental alternatives available on request.



Challenges - Feature Selection

Index No. and No. of Features:

0. No. of Features: 19
1. No. of Features: 18
2. No. of Features: 17
3. No. of Features: 17
4. No. of Features: 16
5. No. of Features: 15
6. No. of Features: 15
7. No. of Features: 14
8. No. of Features: 14
9. No. of Features: 14
10. No. of Features: 14
11. No. of Features: 14
12. No. of Features: 14
13. No. of Features: 14
14. No. of Features: 14
15. No. of Features: 14



Challenges - Feature Selection

Index 0 - 7 give almost identical scores, with index 7 having the fewest number of features. We ran the program again to cull more features and found that accuracy only dipped below 91% once we cut down to 11 features.

Challenges - Feature Selection

Turns out we only need 12 features! They are:

Year and month of transaction, flat type, floor area, flat model, approximate flat storey, age of flat, height of building, planning area, distance to nearest mall, distance to nearest MRT station, name of nearest primary school, and name of nearest secondary school.

What about the other features?

Findings - Categorical Variables

The following categorical variables cannot be considered:

1. Planning Area: We cannot teleport flats between planning areas. Therefore, this is a non-starter.

2. Primary / Secondary School Name: A school's name is built over time and having an individual school being reputed for being "better than others" also runs counter to another government objective, that "every school is a good school". Therefore, we should not consider working with this variable.

3. Year and month of transaction: If we were able to manipulate time, we would be solving far more than our problem statement. Therefore, we will not be working with this variable.

4. Flat type: This is likely highly correlated with floor area.

Findings - Categorical Variables

Further research / collaboration with HDB is required. It is not clear what sort of features each flat model has.

At a high level however, we should try to ensure that "less expensive areas" have more of the following flat types being built (as this is the only remaining variable which we can work with):

1. DBSS;
2. Maisonette;
3. Terrace; and
4. Type S1.

These may cost more to build. That is however, a state budgetary matter on whether it is worth allocating additional resources to these areas to promote the egalitarian goals of the HDB.

Findings - Continuous Variables

The following continuous variables cannot be considered:

1. Floor Area: Larger flats will naturally cost more. This is not a variable that can be manipulated.

2. Flat storey: This cannot be manipulated as we cannot have the first 10 storeys of a block unoccupied.

3. Age: We cannot make old flats newer.

4. Market Hawker / Commercial: This variables is a binary variable that states whether or not there is a market/hawker centre or commercial units nearby. While it may have a negative impact on flat value, it would be unwise to separate HDB flats with these establishments, as it would mean residents with mobility issues such as the elderly may experience difficulty buying essentials.

Findings - Continuous Variables

The following variables are what we can work with:

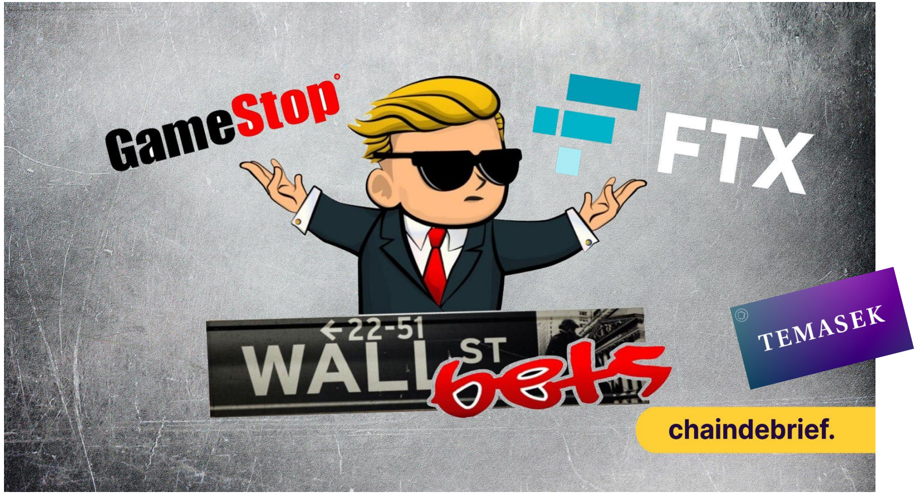
1. Max floor level;
2. Mall Nearest Distance; and
3. MRT nearest distance.

Recommendations

1. Max floor level: Build taller apartment blocks to boost housing density and cut construction costs.

Recommendations

2. Nearest mall: The closer a mall is, the more valuable the property.



Recommendations

3. MRT Nearest Distance: A nearby MRT station boosts flat value

Q&A

Ask Us Questions

Please refer to standard answer chart to the right and check if your question is similar to those before asking.