Value of HDB Flats: Expressways + Height

Salman Yusuf
DSAID QS - Data Science Intern

TABLE OF CONTENTS

01 ABOUT THE DATASET

02 ONEMAP API

03 APPROACH

04 ANALYSIS

05 FUTURE IMPROVEMENTS

01

ABOUT THE DATASET

908,041

Rows of HDB Sales Data

13

Columns, including address features, flat type, flat_model, floor_area_sqm, resale_price, etc.

1990 - 2023

Various datasets were extracted from https://beta.data.gov.sg/datasets/189/view to give a combined dataset that provides information on all resale HDB flats from 1990 to 2023

02

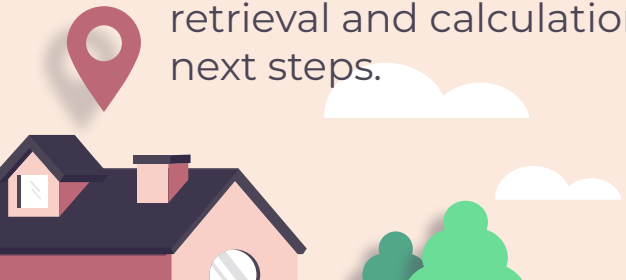
ONEMAP API

Expressways Data

Hypothesis: Distance to nearest expressway affects HDB resale value

To get all the coordinates of the 10 expressways in Singapore, the Onemap Search API was used.

The list of coordinates of each expressway were stored in a python dictionary for efficient retrieval and calculations in the next steps.



TAGGING OF EXPRESSWAY COORDINATES USING FOLIUM

03

APPROACH

THE GOAL

NEAREST EXPRESSWAY

Finding out the distance of the nearest expressway to each HDB flat will allow us to study the relation and degree of impact on the resale value

APPROACH

ADDRESS COORDINATES

Coordinates of each unique address (Block + Street Name) in the dataset was calculated using **Onemap API**. 118 addresses had no coordinates, hence they were removed from the data.

HAVERSINE FORMULA

Using for-loops, the distances were calculated between address coordinates and expressway coordinates by the Haversine Formula.

NEAREST HIGHWAY

The shortest distance between an address coordinate and an expressway was stored as a value in a dictionary, with the key being the address coordinate.

MERGE TO DATASET

Using a map function, the nearest expressway and its distance from the address were added to the dataset based on the address coordinates.

04

ANALYSIS

FEATURE ENGINEERING

DISTANCE RANGE

As proximity to expressway is of interest, the distance was binned into 6 categories. i.e 0-0.1km ... >0.5km

PRICE PER SQM

The resale value was divided by the floor area (sqm) to get a new feature, price per sqm, which is a more accurate and robust response variable.

HEIGHT AND FLAT TYPE

The datasets was grouped by height type and flat type independently to make line plots showing price per sqm against distance range.

Figure 1: Average Price per sqm vs Distance Range (by height type)

The figure shows that there is a general increase in price per sqm as the HDB flat is located further away from the expressways. High rise flats in particular seem to have a sharper increase compared to Low and Mid rise flats. Hence, the assumption that expressway proximity is good for residents of higher floors due to unblocked views may not be true. Further investigation is required.

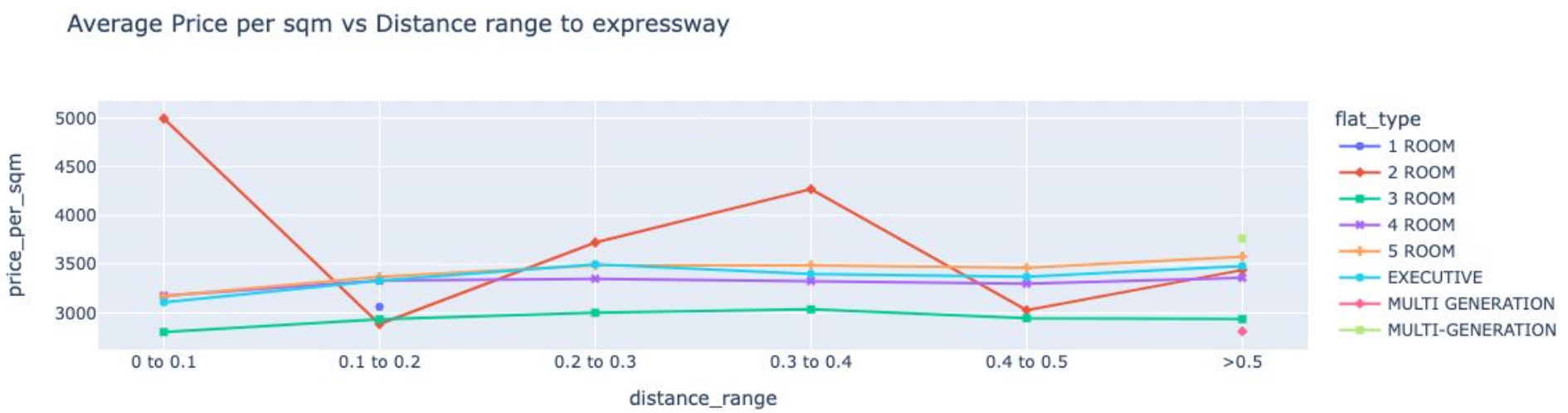
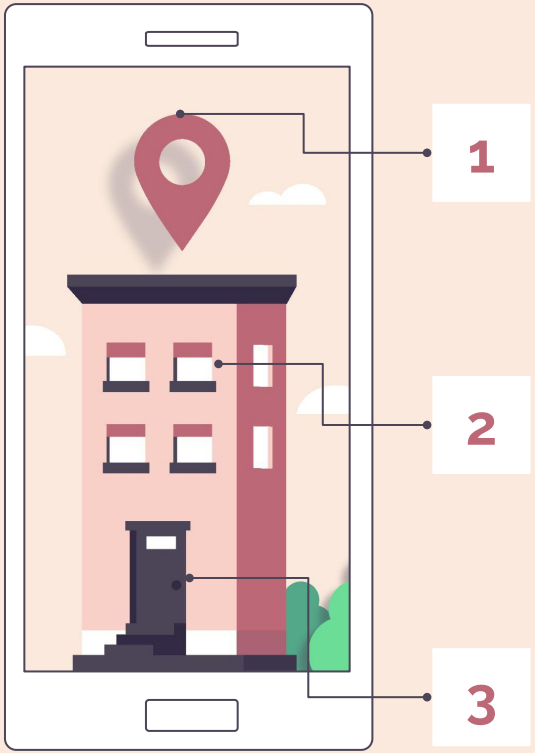


Figure 2: Average Price per sqm vs Distance Range (by flat type)

In the figure above, the 2 room flat shows fluctuations in price per sqm and the distance from the expressway increases. Perhaps, younger couples are likelier to occupy 2 room flats and may value the close proximity to expressways for convenience to work, etc. However, the price per sqm falls drastically before climbing again as the distance increases, maybe due to diminishing marginal returns.

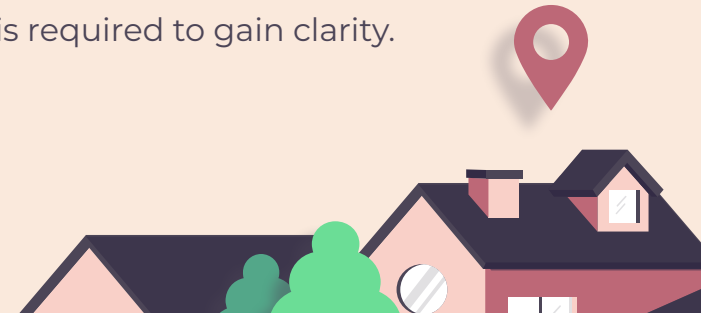


SUMMARY

1. Prices per sqm of HDB flats generally tend to increase as distance from expressways increases. Assumption that HDB resale value suffer due to close proximity to expressways may be true, however it cannot be confirmed if it's due to noise at this stage.

2. The assumption that expressway proximity is good for residents of higher floors due to unblocked views may not be true as prices tend to increase, especially for high rise flats, as distance from expressways increase.

3. More analysis is required to gain clarity.



05

FUTURE IMPROVEMENTS

FUTURE IMPROVEMENTS

FEATURE IMPORTANCE

Carrying out feature importance using statistical tests, correlation matrix or through SHAP plots, could provide better clarity on impact of specific features.

MODELLING

Identify more features or confounders that affect current features, limiting the accuracy of results. Test model accuracies in predicting prices per sqm in specific locations.

THANK YOU!