

VINCENT TAN, 29TH JUNE 2020

# SINGAPORE PUBLIC HOUSING (HDB) RESALE PRICE PREDICTION MODEL





---

VINCENT TAN, 29TH JUNE 2020

# Problem Statement

---

- Singapore has always had one of the most expensive housing market in the world
- Housing price are still benchmarked manually by experienced appraiser today
- Hence, it would really be helpful for home buyer if there is a predictive model available to find out any undervalued property listing and maximise their savings.

# Methodology

- The data source for the analysis came from three different means:



## Official Data

- ▶ Core dataset consist of HDB resale flat transaction record from January 2017 to March 2020
- ▶ Resale Price along with basic flat attributes such as:
  - ▶ Floor Level, Area, Lease
  - ▶ Flat Model, Transaction Date



## Web Scraping

- ▶ Amenities and Infrastructure in Singapore obtained through multiple website:
  - ▶ Primary & Secondary Schools
  - ▶ MRT Stations & Bus Stops
  - ▶ Shopping Malls



## API Calls

- ▶ HDB full address were used to send request to OneMap API to acquire its respective coordinates
- ▶ Obtain commuting time from flat unit to employment hub in Singapore
  - ▶ By public transport and also by driving

# Geospatial Studies

- As of 2016, almost 60 percent of Singapore working population are relying on public transport as their primary mode of transport
- Concentrate project study on geospatial features of a flat unit
- Connectivity & Proximity:
  1. Distance to MRT Station or Bus Stops
  2. Proximity to highway
  3. Commuting time to Major Employment Hub
  4. Proximity to schools



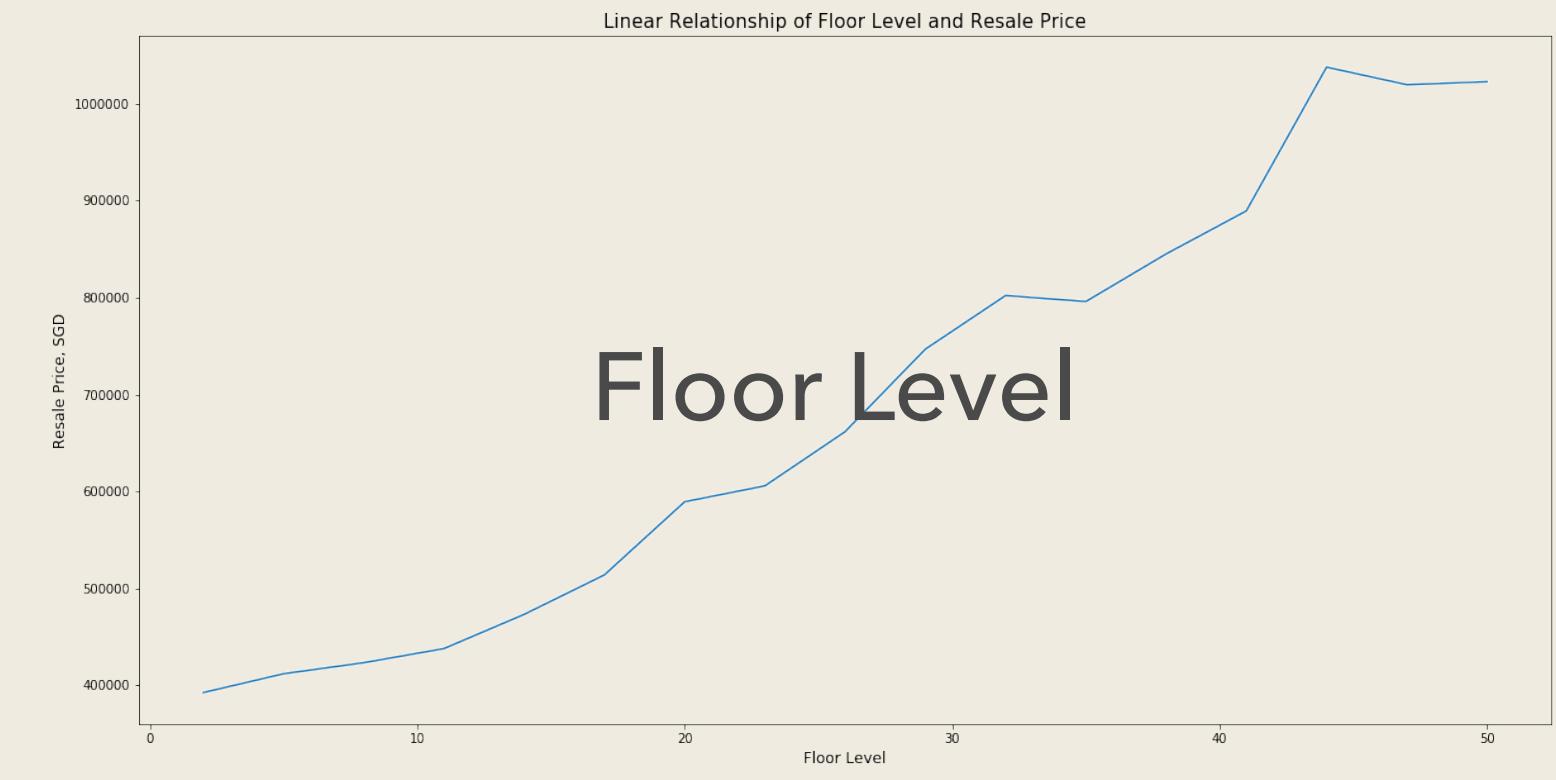
# Key Indicators



Floor Area



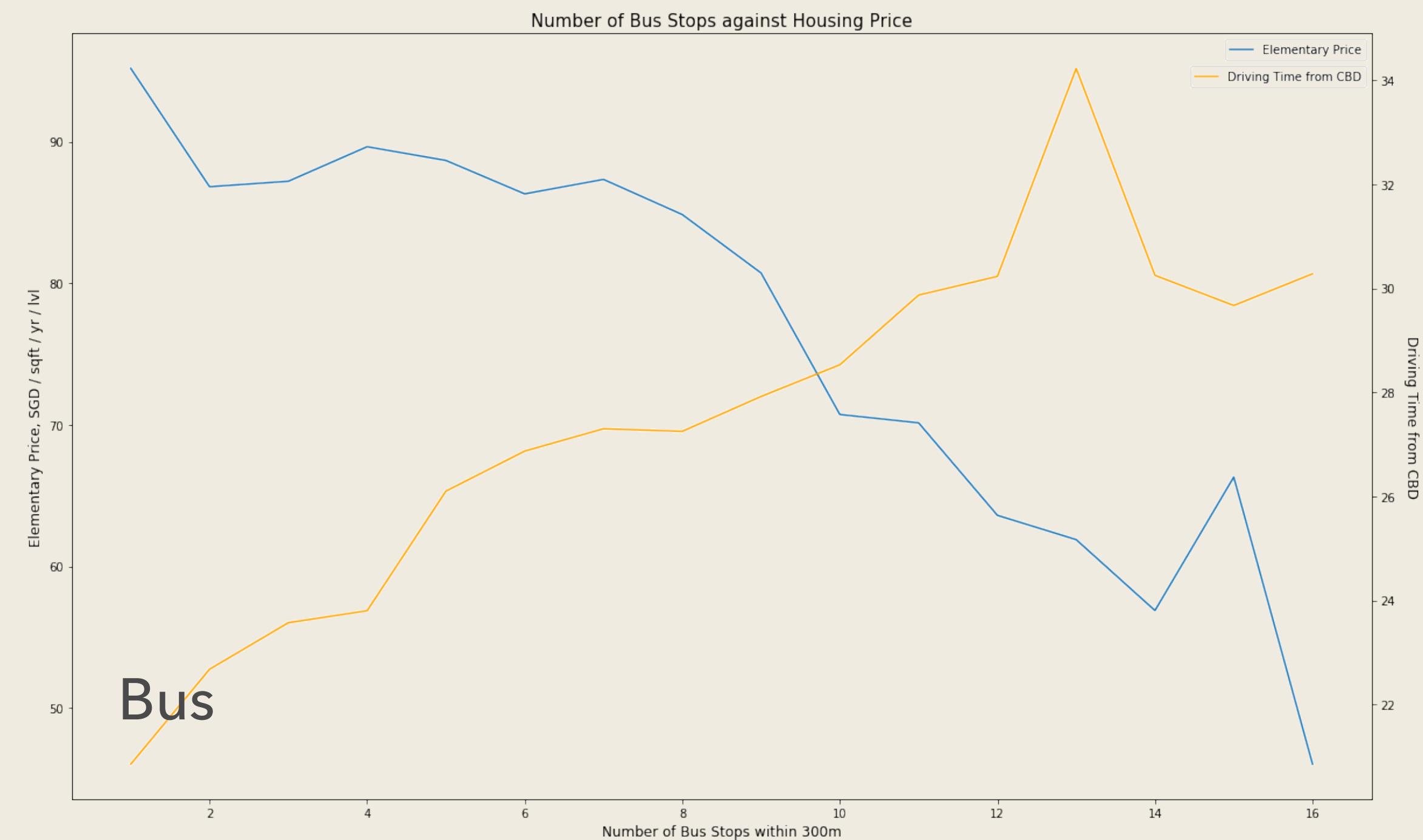
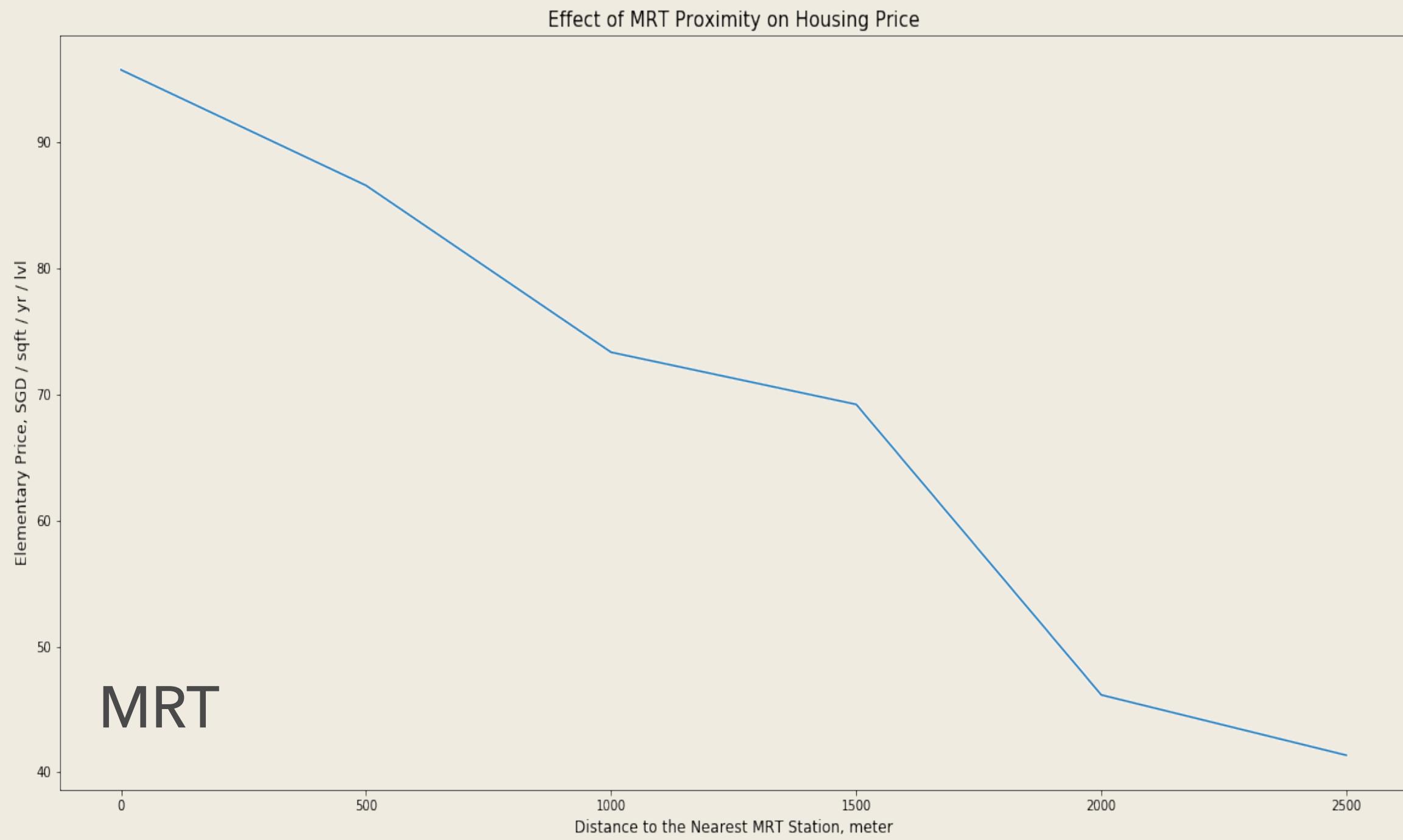
Remaining Lease



Floor Level

- There are 3 key features that are almost linearly proportional to the resale price
- To study the effect of features, different price metrics were created
- Strong indicators, able to explain 55% of variance
  - Price per square feet - Simplest form, commonly used in market ( SGD / ft<sup>2</sup> )
  - Unit Price - Take away the effect of remaining lease ( SGD / ft<sup>2</sup> / year )
  - Elementary Price - Divide by all 3 features ( SGD / ft<sup>2</sup> / year / level )

# 1 - Public Transport Connectivity

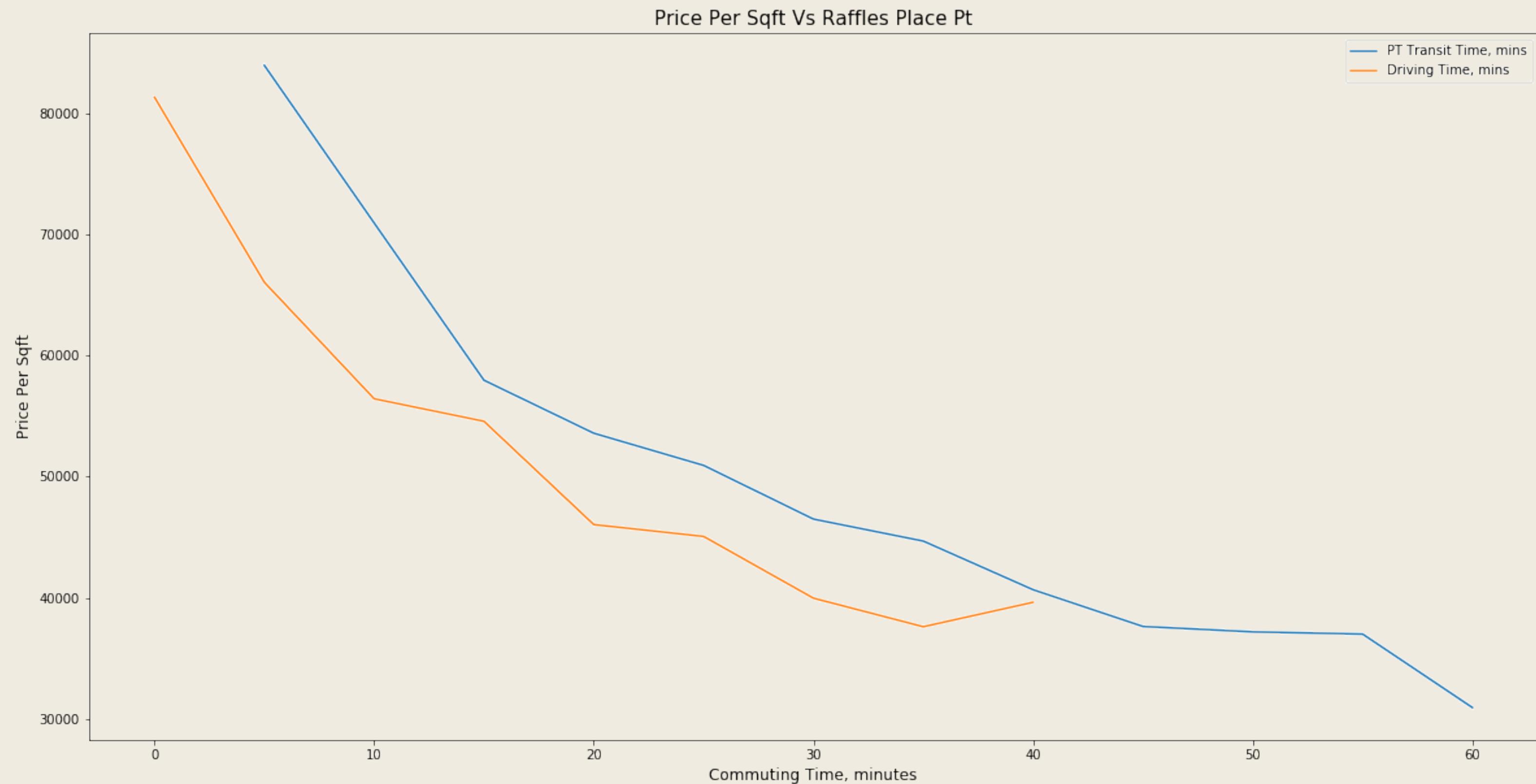


- Almost-linear negative relationship between MRT station distance and property price
- One of the key indicator of property prices

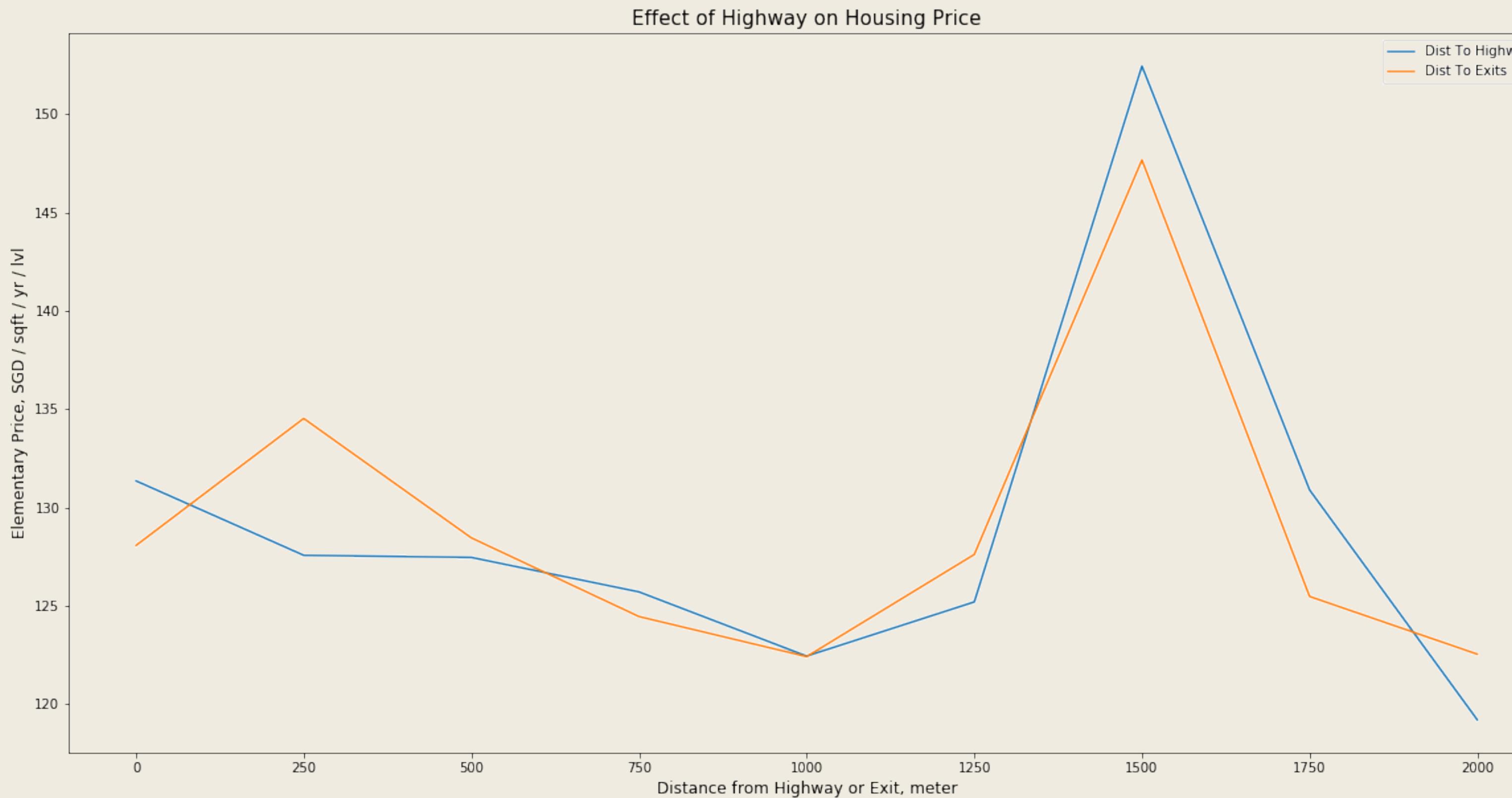
- Negative relationship between number of bus stops within 300m and property price
- However, suburban further away from city center tend to have more bus stops

## 2 - Proximity to Employment Hub

- 5 employment hubs are studied:
  - Raffles Place - CBD
  - one-north - Innovation district
  - Jurong East - Business park
  - Orchard - Commercial Hub
  - Changi - International Airport
- Driving and Public Transport have almost the same effect
- Raffles Place shows the most profound effect

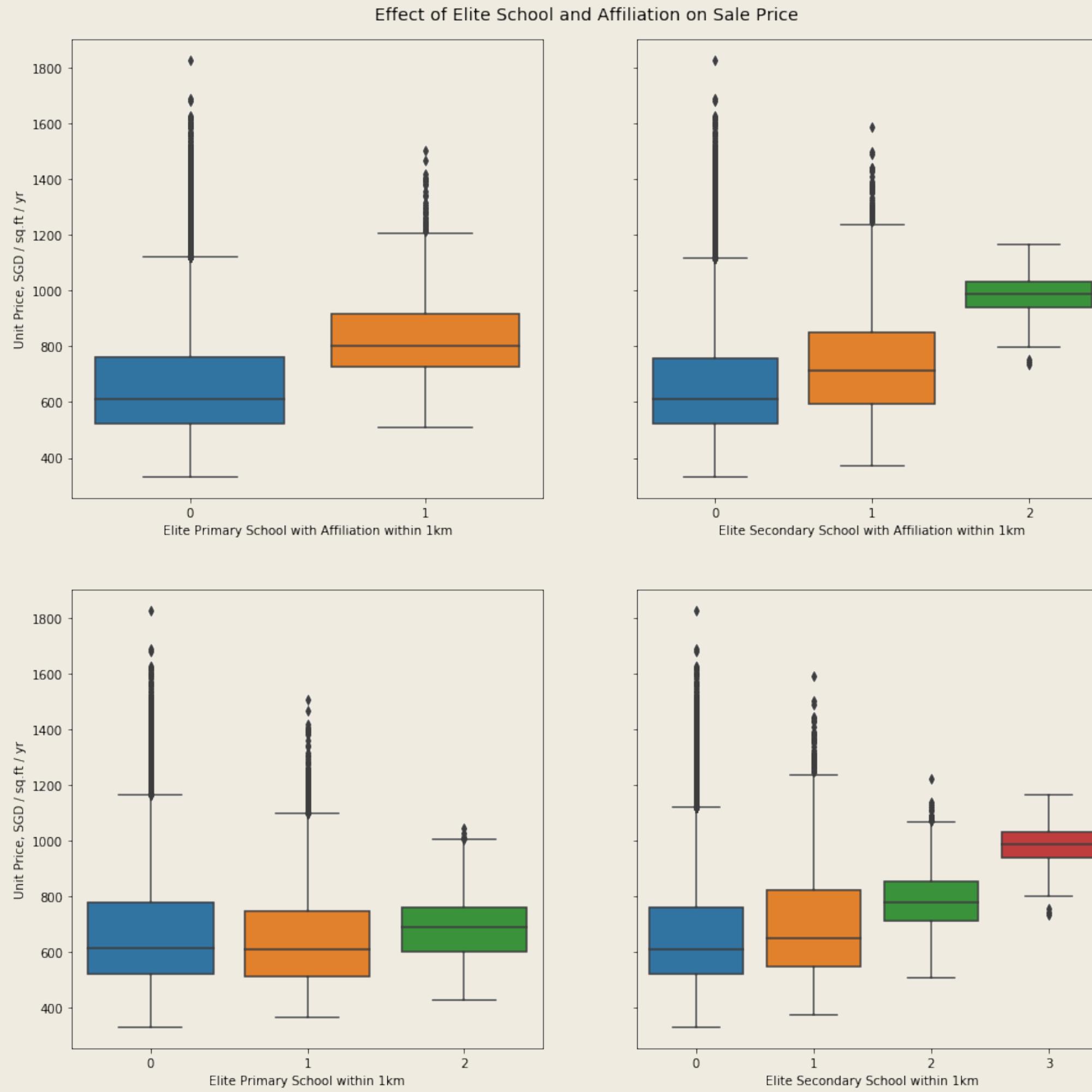


# 3 - Highway & Exit Ramps



- Noises from highway traffic are undesirable in general
- Housing prices peaked at 1.5km from nearest highway or ramps
- Fell again after 1.75km
- Optimal when flat unit is away from noises but close enough for convenience
- Highway and ramps have almost the same pattern
- Highly collinear variable

# 4 - Elite School and Property Price



- Highly competitive education system
- Priority is given to residents who live closer to the preferred school when balloting
- In the following order:
  - Resident living within 1km radius
  - Resident living within 1km to 2km radius
  - Resident living more than 2km radius
- Assumed top 20 schools classified as "Elite School" in term of either vacancy or cut-off-points
- Elite secondary school affect prices more than primary school
- Elite school with affiliation raise premium further

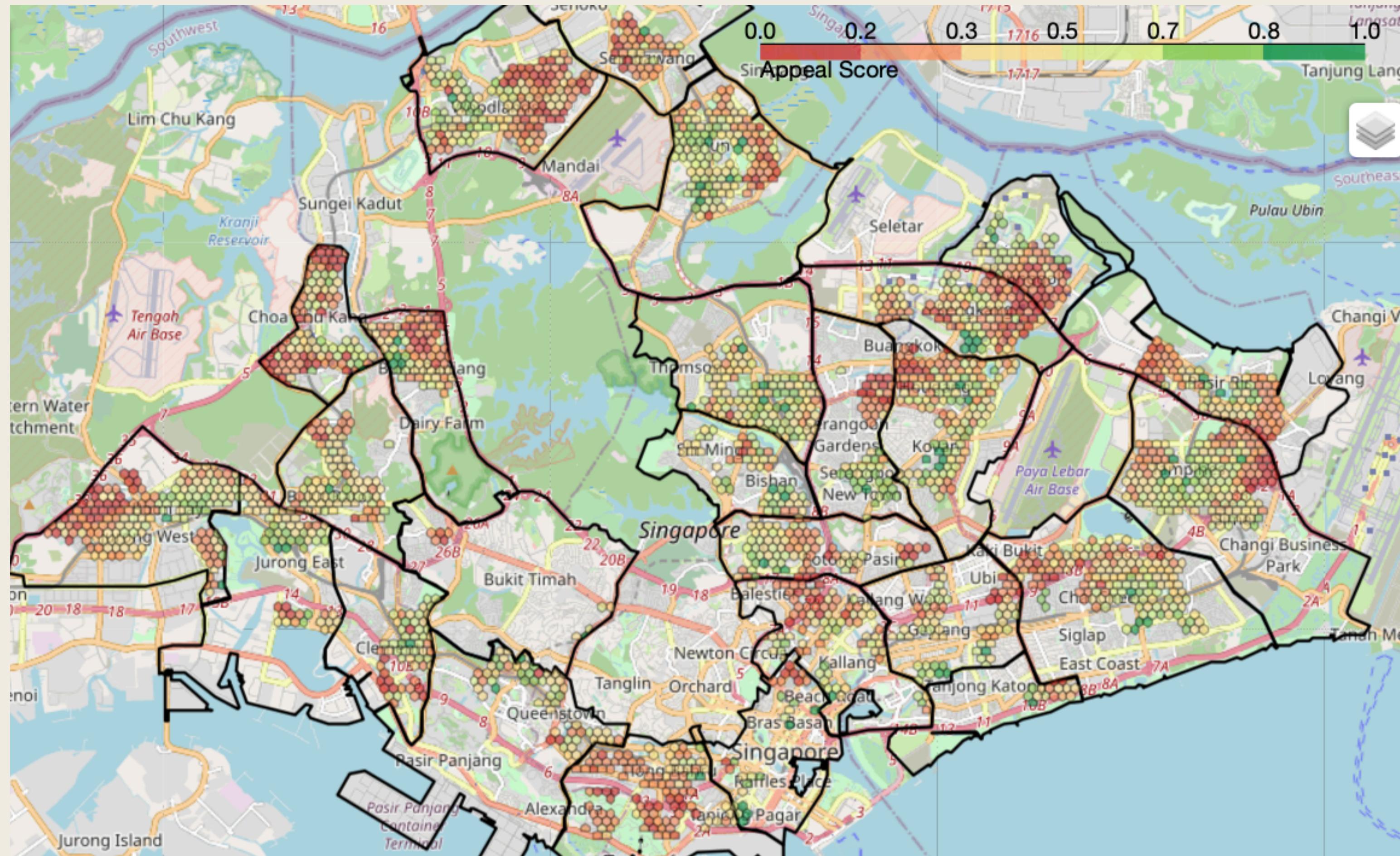
# Price per Sq-ft across Singapore



# Regional Appeal Score

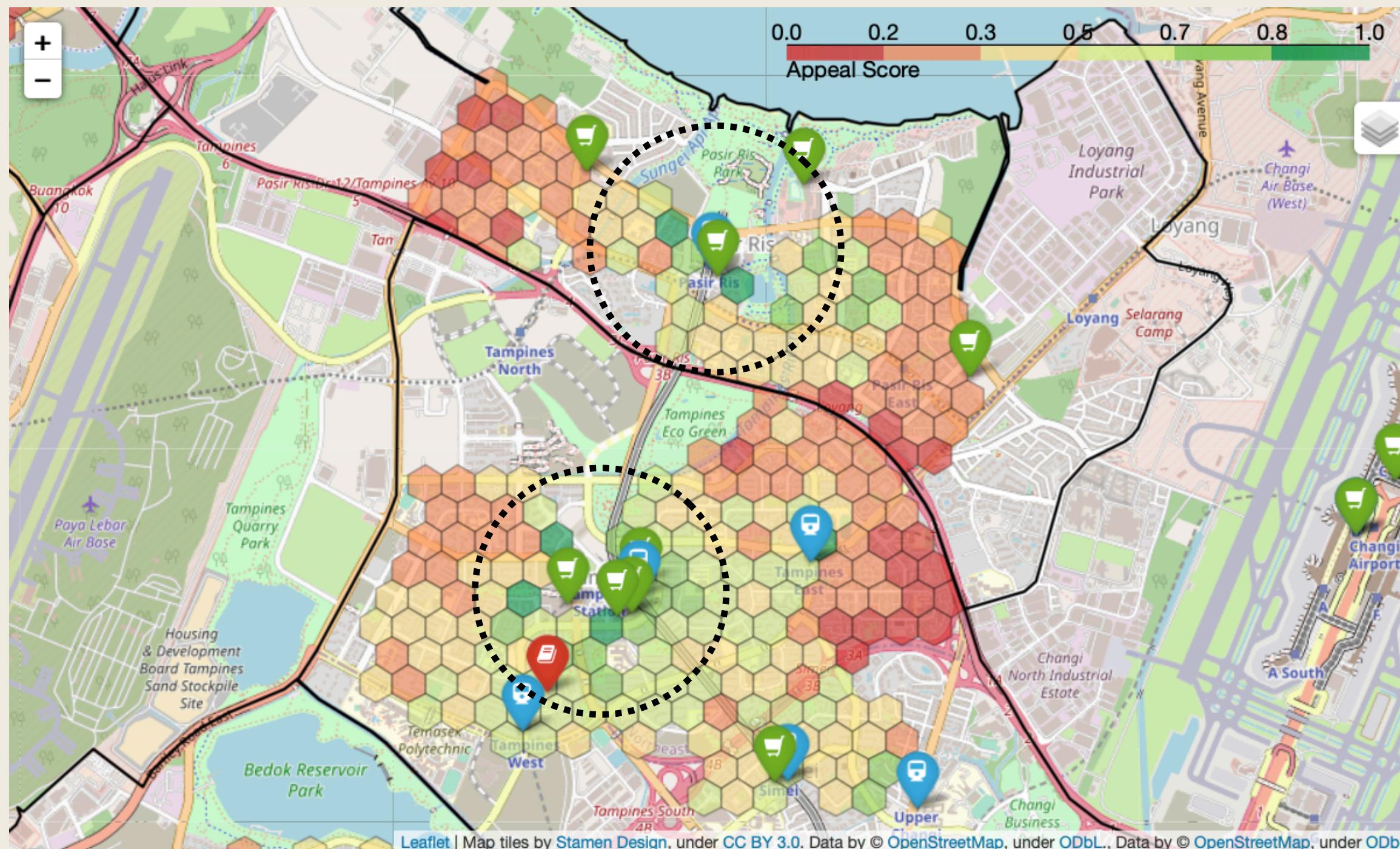


# Regional Appeal Score

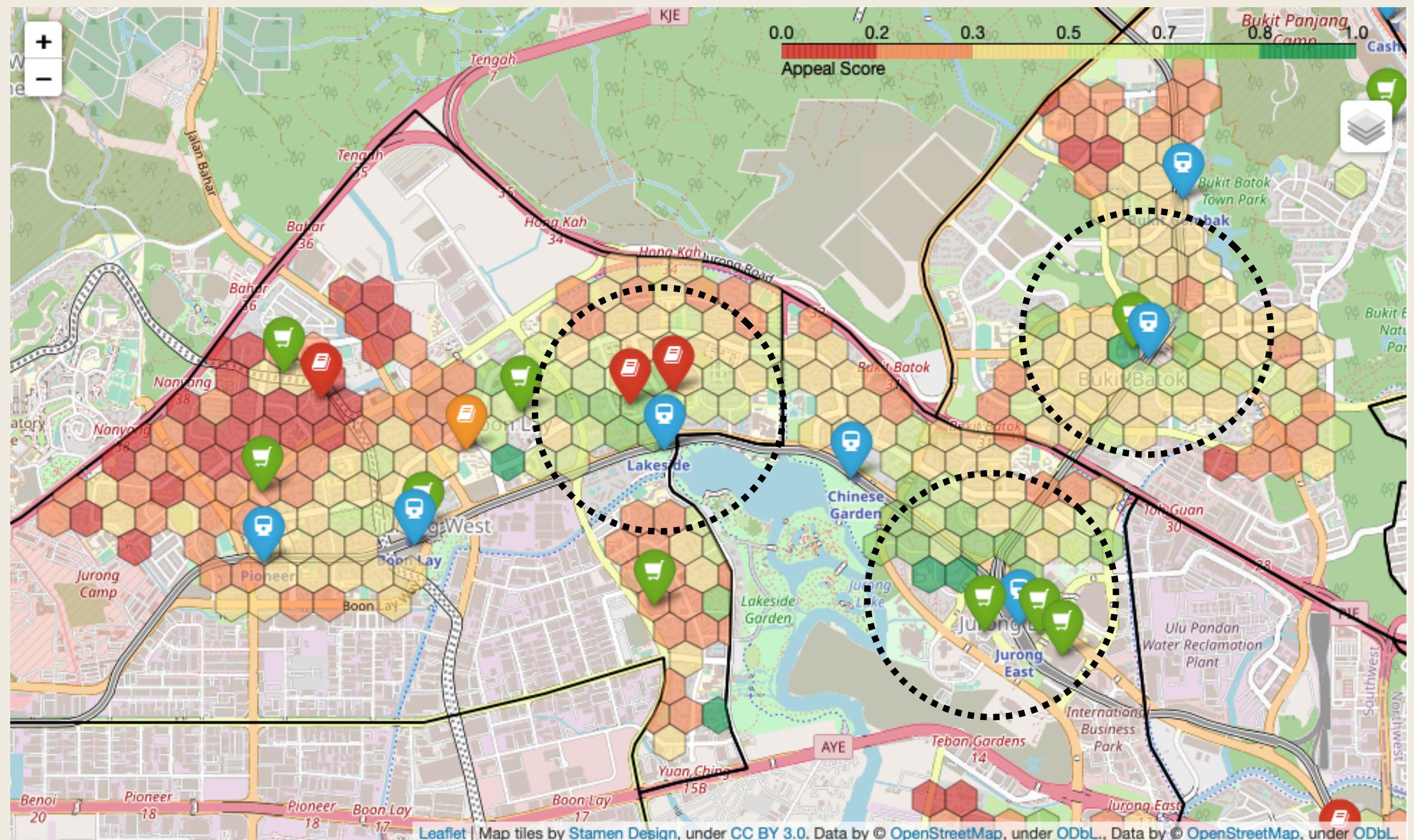


# Effect of MRT on Property Pricing

## Tampines & Pasir Ris



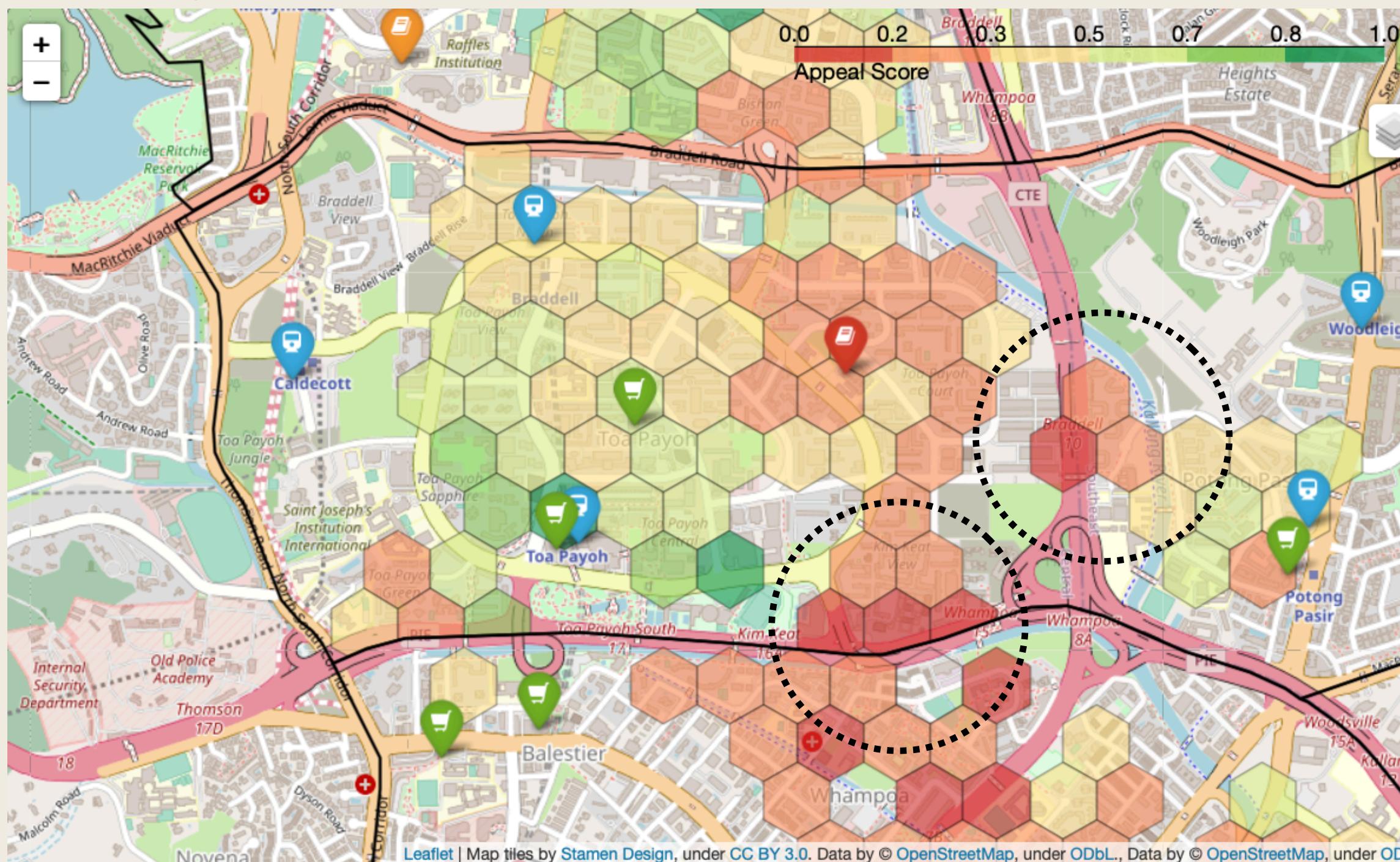
## Jurong & Bukit Batok



- Property prices almost always has a local maximum near MRT Station

# Other Effects

## Toa Payoh



## Marine Parade



- Adverse effect from Highway

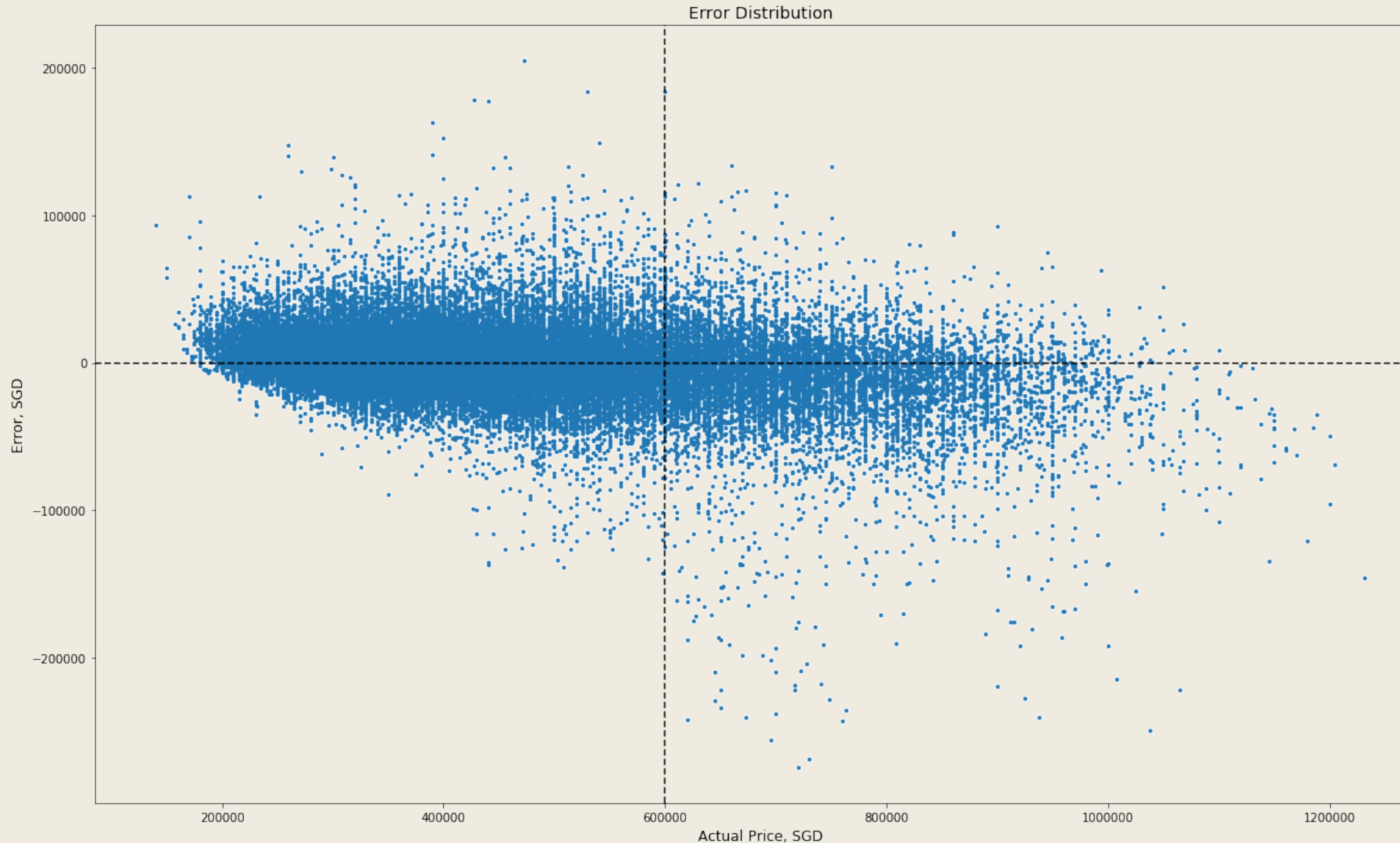
- Positive effect of Commercial Center
- Especially when MRT is absent

# Model Performance

MODEL / TECHNIQUES	Train-set R <sup>2</sup> Score	Test-set R <sup>2</sup> Score	Test-set RMSE
Linear Regression	0.88	0.87	\$54,630
Lasso Regression	0.88	0.87	\$54,629
Random Forest	<b>0.97</b>	<b>0.95</b>	<b>\$32,982</b>
Neural Network	0.95	0.93	\$39,474

- 3 types of approaches were used for modelling:
  - Multiple Linear Regression
    - With or Without Regularisation
  - Tree-Based Ensemble Technique
    - Random Forest
  - Neural Network
    - 4 hidden layers with ReLu activation
    - With Early Stopping mechanism

# Model Limitation



- Model very accurate in general
- Multiple property above 600,000 SGD undervalued significantly
- Model unable to capture condition of property during transaction
- Furnishing quality and general condition of the interior defines pricing heavily as well
- However, for most part of the sample, error (RMSE) only up to 23,000 SGD

# Conclusion

- Random Forest model able to achieve an accuracy (R2) score of up to 95%
- Key indicators for price prediction are
  - Floor Area
  - Remaining Year Lease
  - Storey Range
  - Township where the property is situated
- Although HDB is in highly controlled environment, this project still serve as a window to look into home buyer's psychology and how premium is justifiable by numerical features or attributes
- Same methodology could still be applied to Condominium sales in Singapore or real estates in other City States that resembles Singapore
- Some of the key features could still be applicable and worth researching for

---

THANK YOU!

---

Q & A SESSION