# DREAMHOUSE

# Project 2
# Singapore Housing Data

# **Table of** contents

**DREAM**HOUSE
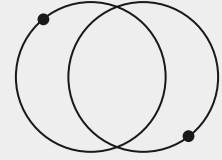
## 01
**Problem Statement**
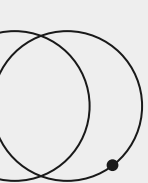
## 02
**Data used**

## 03
**Findings**

## 04
**Model results**

## 05
**Summary**

* * * *

# Problem Statement

* * * *

- **House pricing affects the decision making** process of buyers in their assessment of the unit.

- The goal is to **have the model accurately predict the sales price** of the houses in the test set, which will be evaluated based on common evaluation metrics such as R2 and RMSE.

- This will **give those who are impacted by housing prices**, e.g. owners, buyers and agents **additional information** to consider in their decision making process.

**150,000**

Data points
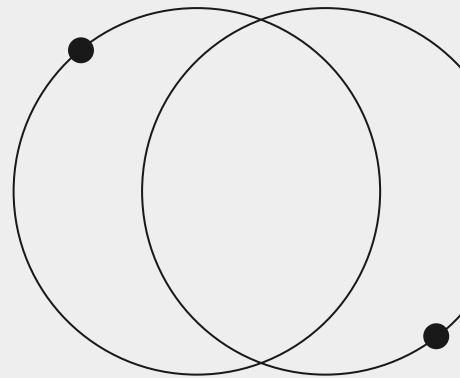
**2012 –2021**

Period

**29**

Variables used

# Price

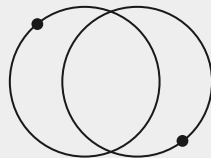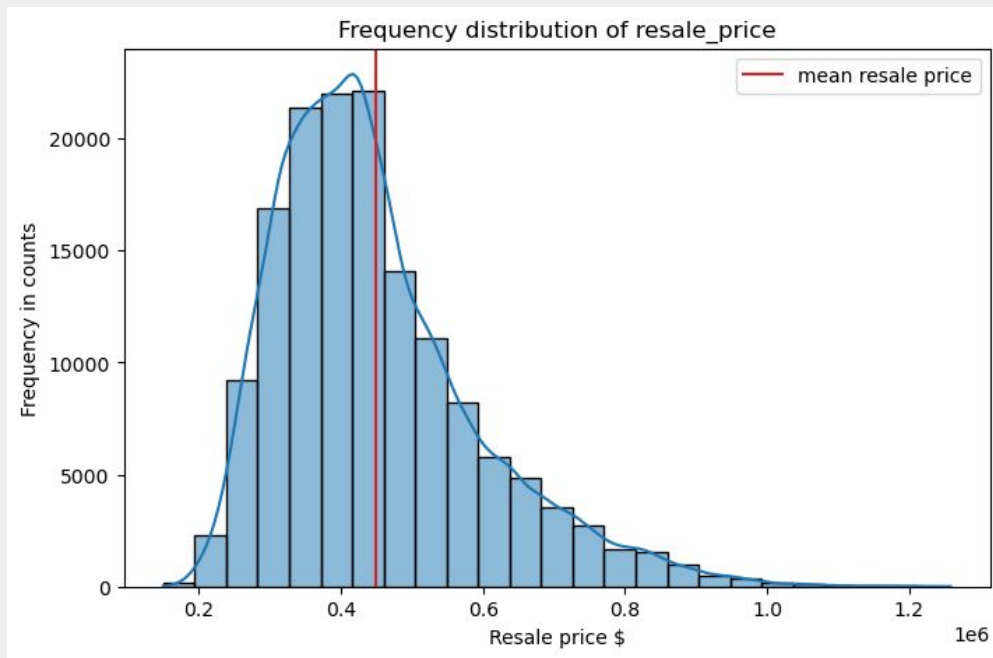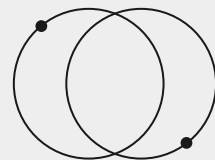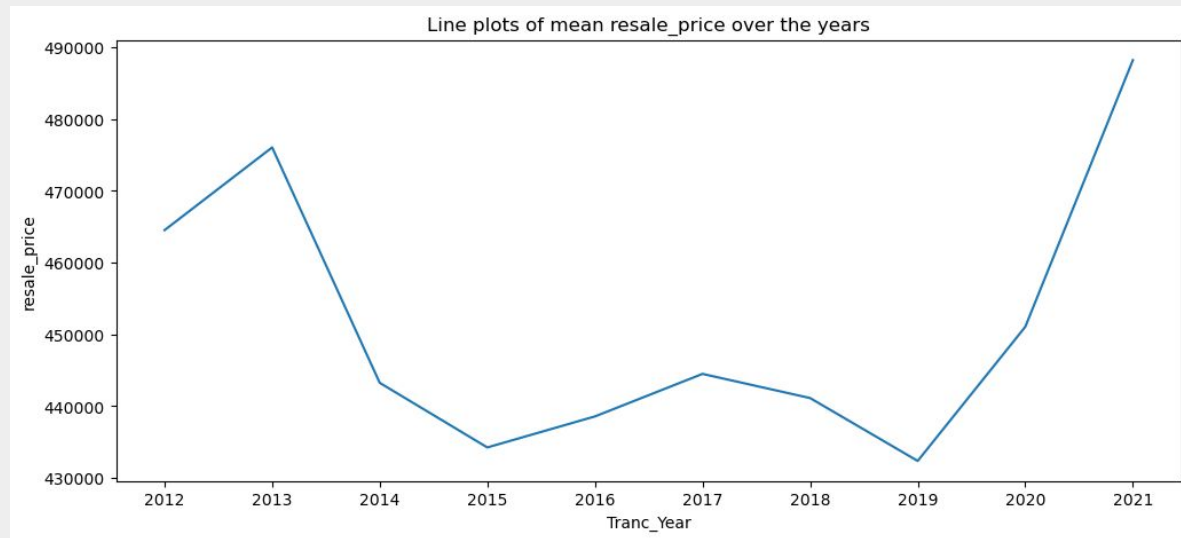## Price
## (in $ hundred
## thousands)

- Most transactions are below mean price

  - $448,661

- High price flats pulling up the average.



Frequency distribution of resale_price

# Resale price by year

- Not constant

- Increase of 60k

- Increasing after 2019
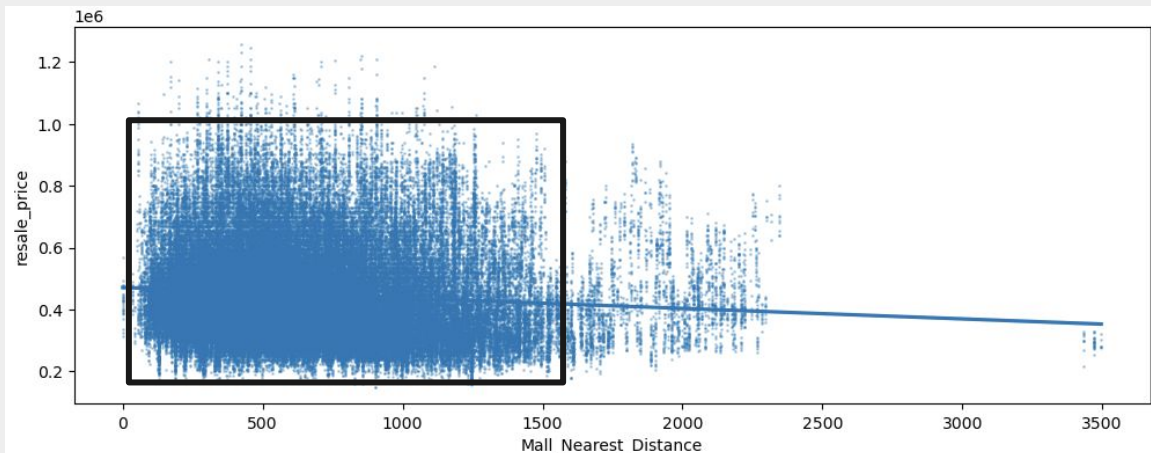  - Covid-19



Line plots of mean resale_price over the years

# Resale price by month

- Not constant

- Swing of 6k

- Lowest in June and July
  - Best to Buy

- Highest between Feb - Apr
  - Selling Spree



Line plots of mean resale_price over the months
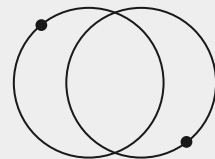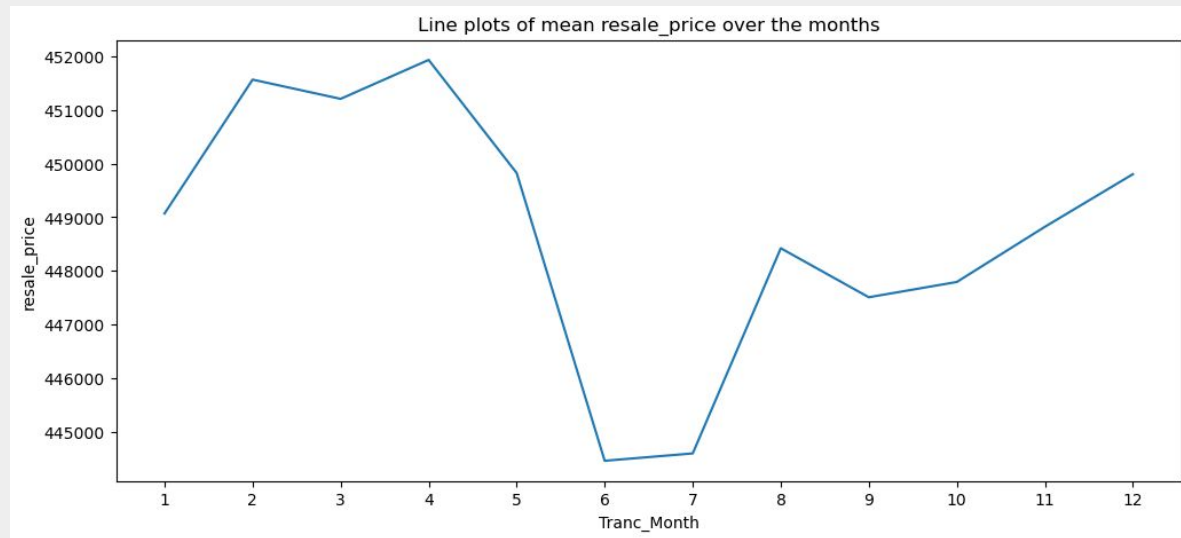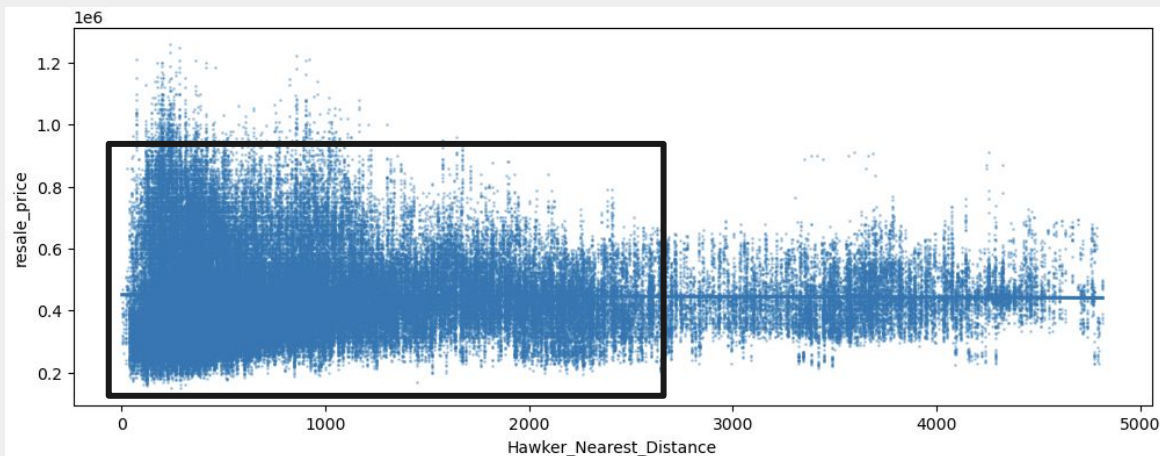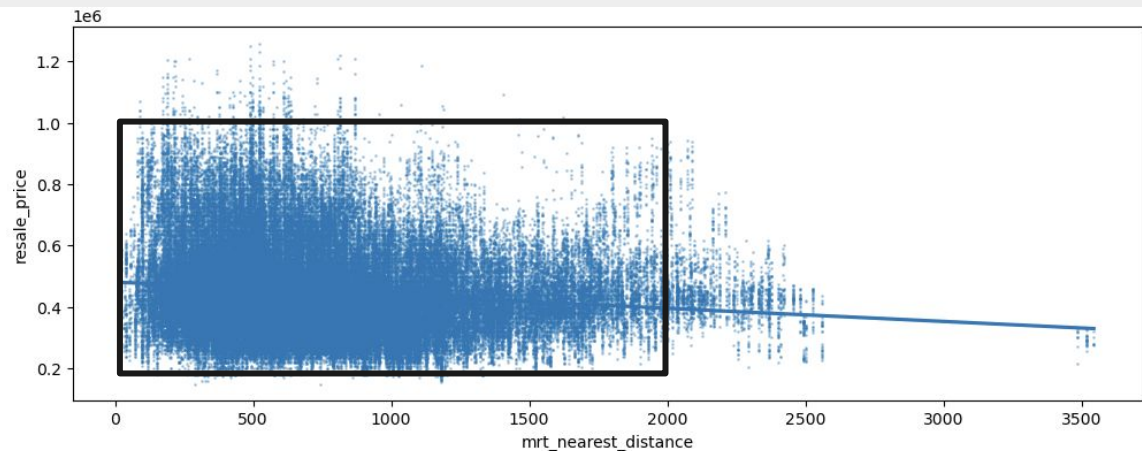
# Hawker and Mall distance

- Flat line shows poor relation with price
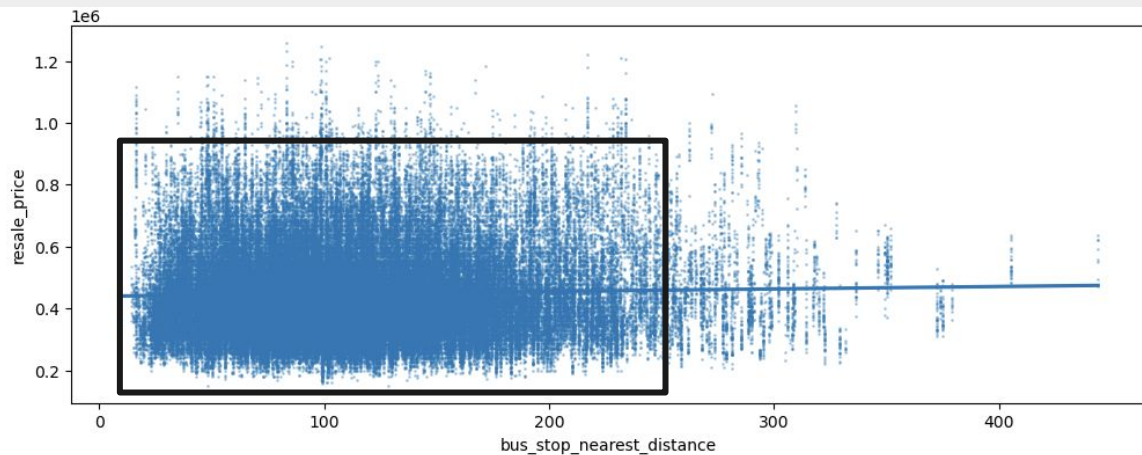
- Clustering on the left
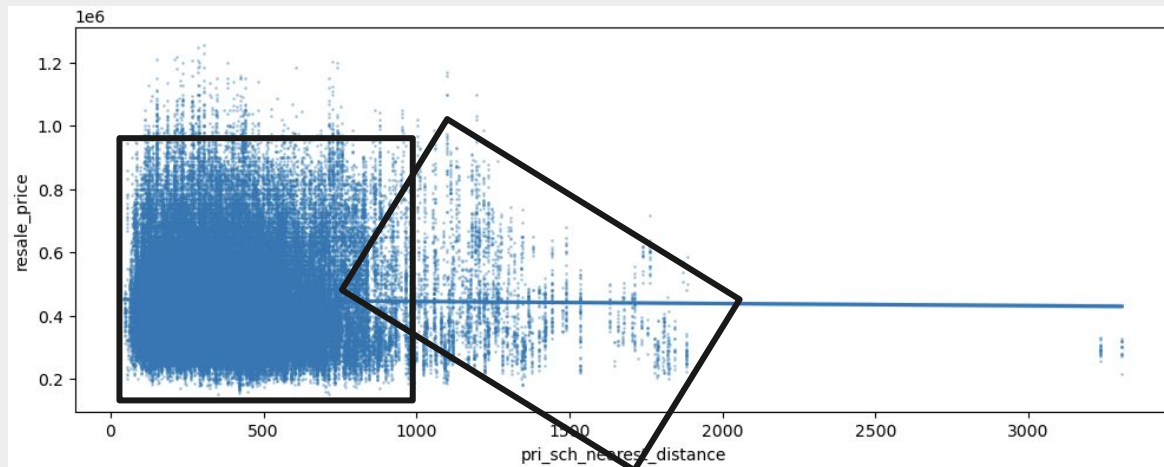
- Nearer is better

# MRT and Bus stop distance

✳✳✳✳



- Flat line shows poor relation with price

- Clustering on the left

- Good transport network
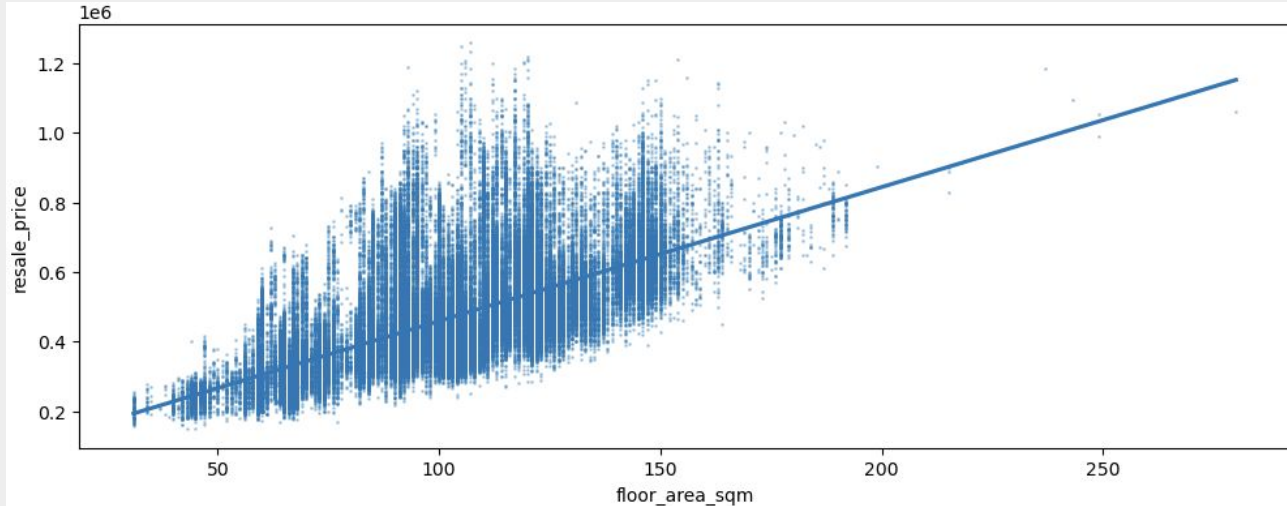
- Access is important

# Primary school distance

✳✳✳✳



- Tight clustering on the left

- Important to be within 1km for primary school

✦
✦
✦

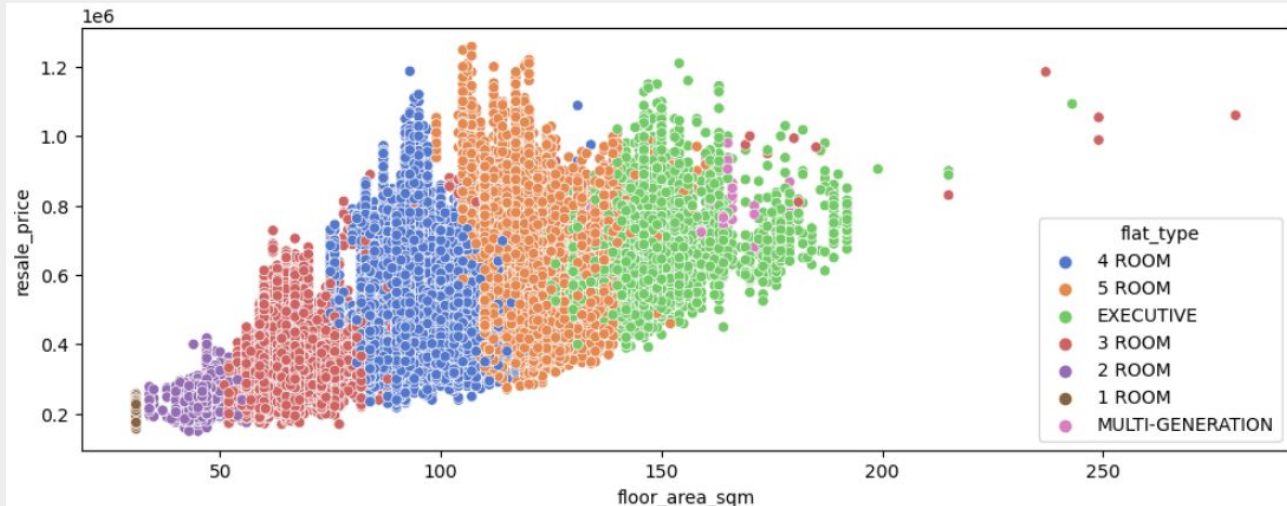# Floor area (sqm) and flat type

✳ ✳ ✳ ✳



- Diagonal line shows **strong association** with price

- Prices follow **floor area** which follow **flat type**.
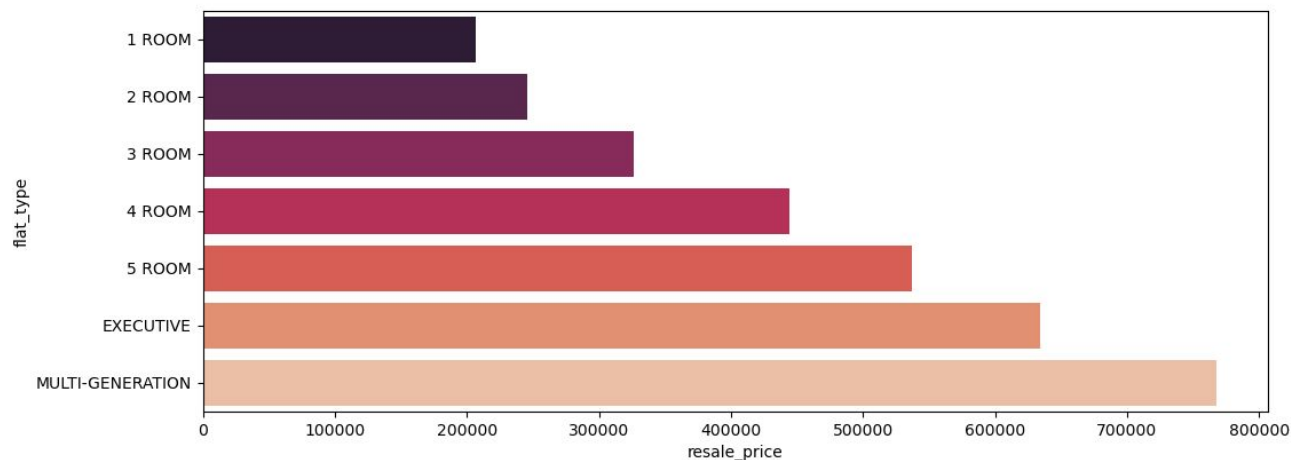
✦
✦

# Floor area (sqm) and flat type

✳ ✳ ✳ ✳



- Diagonal line shows **strong association** with price

- Prices follow **floor area** which follow **flat type**.
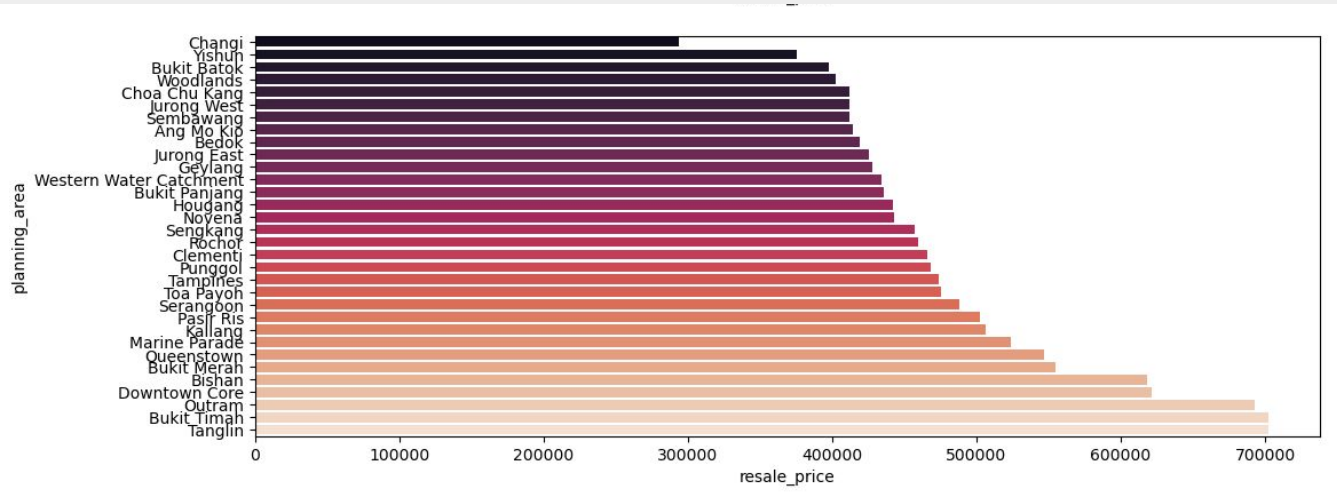
# Flat type and resale price

✳ ✳ ✳ ✳



- The **larger** the flats, the **higher** the price
  - Paying for space

✦
✦
✦

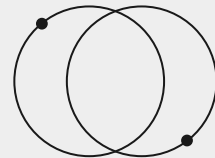# Planning area and resale price

\* \* \* \*



- Central is more expensive

- Fringe is cheaper

# Model finding

- Ridge Regression
  - Linear Model
- RMSE is a model scoring metric
- Represents the **price variation**
- **On unknown data $41,168**

| Activity | Start date | Ridge Regression |
|---|---|---|
| V1 | Train RMSE | 38,916 |
| V1 | Test RMSE | 39,284 |
| V2 | Train RMSE | 38,916 |
| V2 | Test RMSE | 39,284 |
| V3 | Train RMSE | 39,090 |
| V3 | Test RMSE | 39,443 |
| V4 | **Train RMSE** | **41,114** |
| V4 | **Test RMSE** | **41,296** |

# Summary
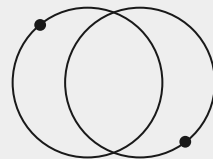
## Timing

- Resale price increasing since 2019
- Mid year best to buy
- Start of year best to sell

## Amenities

- Nearer is better
- Not necessarily impact price.

## School

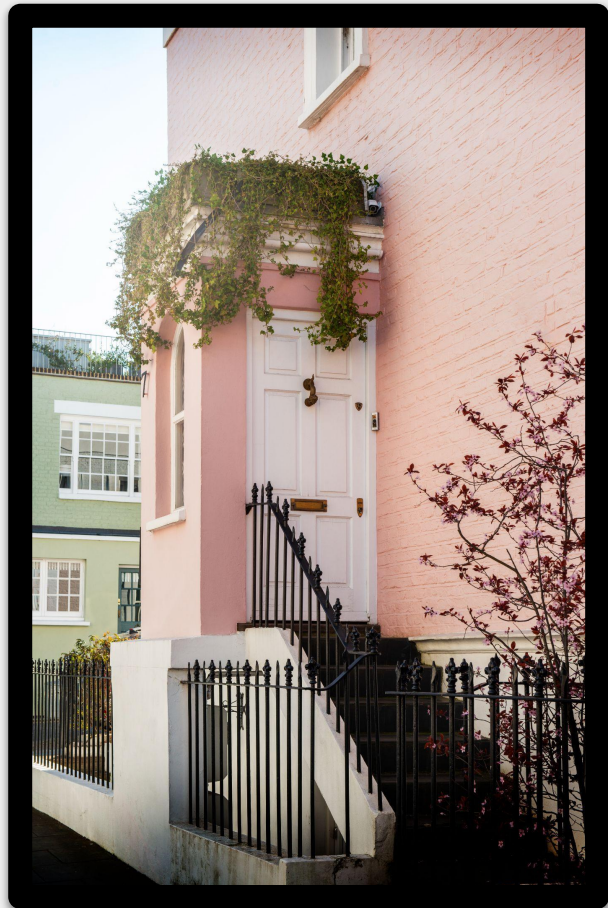- Nearer is better up to 1km

## Flat size

- Linked to flat type
- Bigger = More expensive

## Location
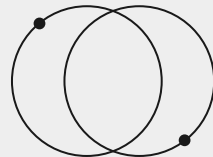
- Central is more expensive
- Fringe is cheaper

**DREAM**HOUSE

# Thanks!

## Do you have any questions?