DATA-DRIVEN PRINCIPLES

FOR UNDERSTANDING

HDB RESALE DEMAND

PROPERTY AGENCY TRAINING COURSE LECTURE

RGNT CONSULTING

Table of Contents

01 Introduction and Relevance of Topic

7min

- Introduction and the Problem Statement

02 Exploratory Data Analysis of HDB Resales

10min

- Exploratory Data Analysis (EDA)
- Feature Selection

03 Introduction of Machine Learning (ML) Model

7min

- Model Explanation, Demonstration and Comparison

04 Recommendations on our ML Model

5min

- Recommendations and Future Actions

1 Introduction

The HDB Resale market in Singapore is a dynamic and complex landscape.

- Continually changed and challenged by:
  - economic conditions,
  - government policies,
  - demographic trends.

State of the HDB Resale Market

Big market. Important to get ahead in one's planning when looking to buy/sell a HDB resale flat

$75B

Dollar amount of transactions

Total value of HDB resale transactions from 2017 to 2023 to date[1]

160K

Number of transactions

Number of HDB resale units sold from 2017 to 2023 to date[2]

4.3

Median years

Average waiting time for a fresh Build-to-Order (BTO) HDB flat to be completed[3]

Sources:
1 and 2: data obtained from data.gov
3: Ng, 2023. https://www.straitstimes.com/singapore/housing/how-are-bto-flats-built-and-why-do-waiting-times-vary-so-much

Objectives of the Property Agent

- Negotiation and Closing Deals
- ↑
- Marketing and Exposure
- ↑
- Property Valuation
- ↑
- Updated Market Insights and Knowledge

The Problem for Buyers/Sellers

Money wasted

Buyer loses the option fee if there is a fallout in any later part of the resale process

Time wasted

Buyer, seller and agent waste time if buyer backs out of the deal

Planning failure

You fail to provide your clients a seamless experience

Why is HDB Resale Value Important?

Cash-Over-Valuation (COV) - the amount the buyer ‘overpays’ for an HDB resale flat.

The COV Process:

1. Buyer found the flat of their dreams in Queenstown after a long search and multiple house visits

2. The seller quotes **S$600,000**. Buyer likes the price

3. After buyer pays **S$1,000** for the Option to Purchase (**OTP**) and request the **valuation report from HDB**,
it turns out that HDB values the flat at S$500,000.

4. The COV is thus S$600,000 – S$500,000 = **S$100,000**.

5. If buyer proceed with the purchase, the **home loan** will be based on **S$500,000** as the total amount.
Buyer has to pay the **S$100,000** COV in cash — and that’s on top of the downpayment

6. Otherwise, buyer can back out of the deal and resume their HDB hunt. But **the S$1,000 option fee** is **forfeited**

The Value of Data

Integrating data-based guidelines and predictive models can significantly enhance service.

- Enabling them to offer more accurate and reliable advice to their clients.
- Streamlines workflow, saving them time and effort in market analysis and property valuation.

Our Dataset

Taken from data.gov.sg

Compiled HDB Resale data from 2012 to 2021

Information collected on the details of each unit and its sale, describing a total of 78 features of every single unit

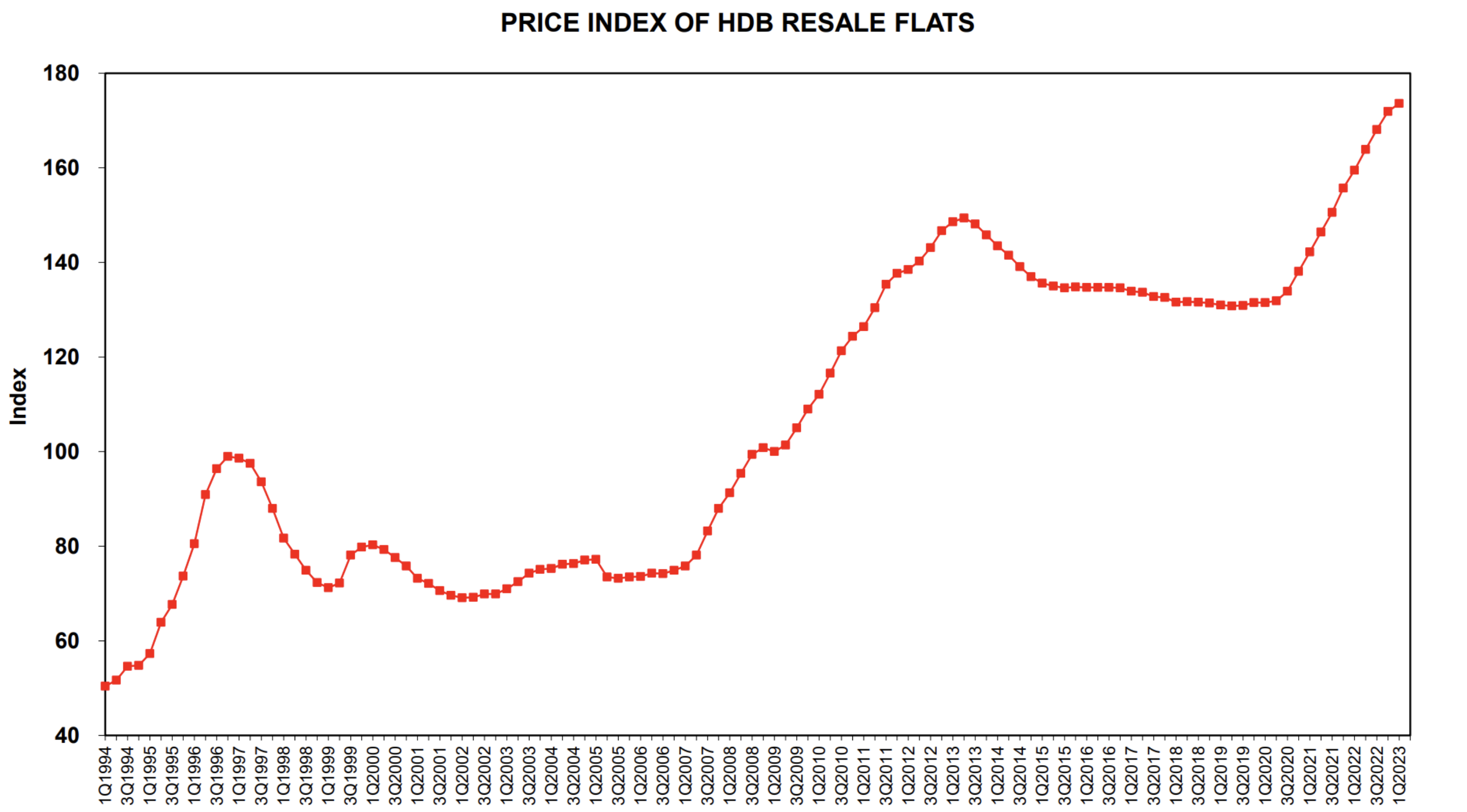
Our Dataset

Analysed using Python, Pandas, Numpy, and Scikit-Learn library packages

Visualized using Seaborn and Matplotlib to generate graphs and charts

Exploratory Data Analysis (EDA)

HDB Resale Flat Prices on an Upward Trend



Notes:
1) 1Q2009 is adopted as the new base period with index at 100.
2) The index from 1Q1990 to 3Q2014 is rebased to the new base period at 1Q2009.
3) The index for 4Q2014 onwards is computed using the stratified hedonic regression method.

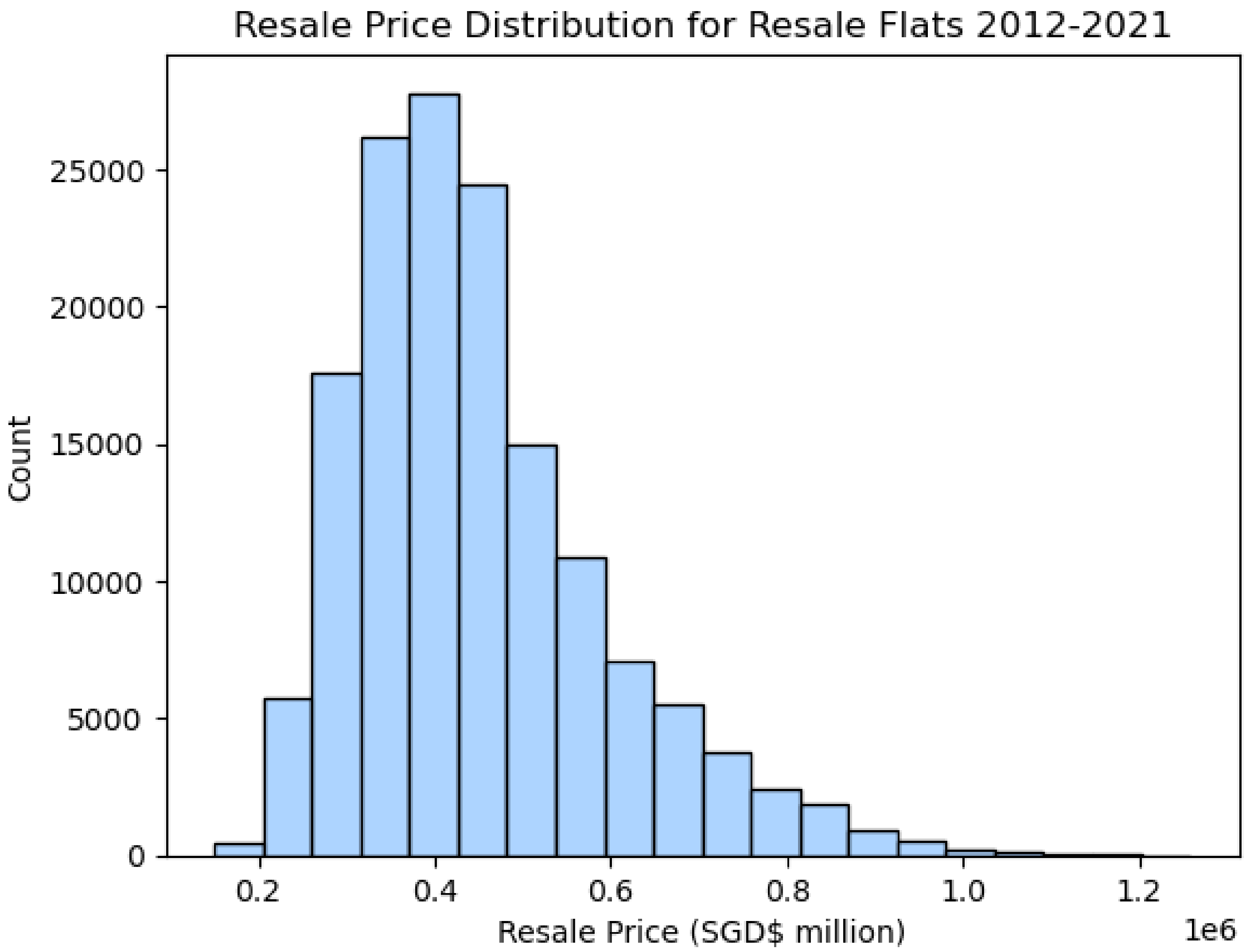
- HDB resale flat prices have been on the rise
- Takes up significant proportion of buyer's finances

Source: Housing Development Board
https://www.hdb.gov.sg/cs/infoweb/-/media/doc/EAPG-CSC/1Q2023-RPI-Big-Chart.ashx

Resale Price Distribution

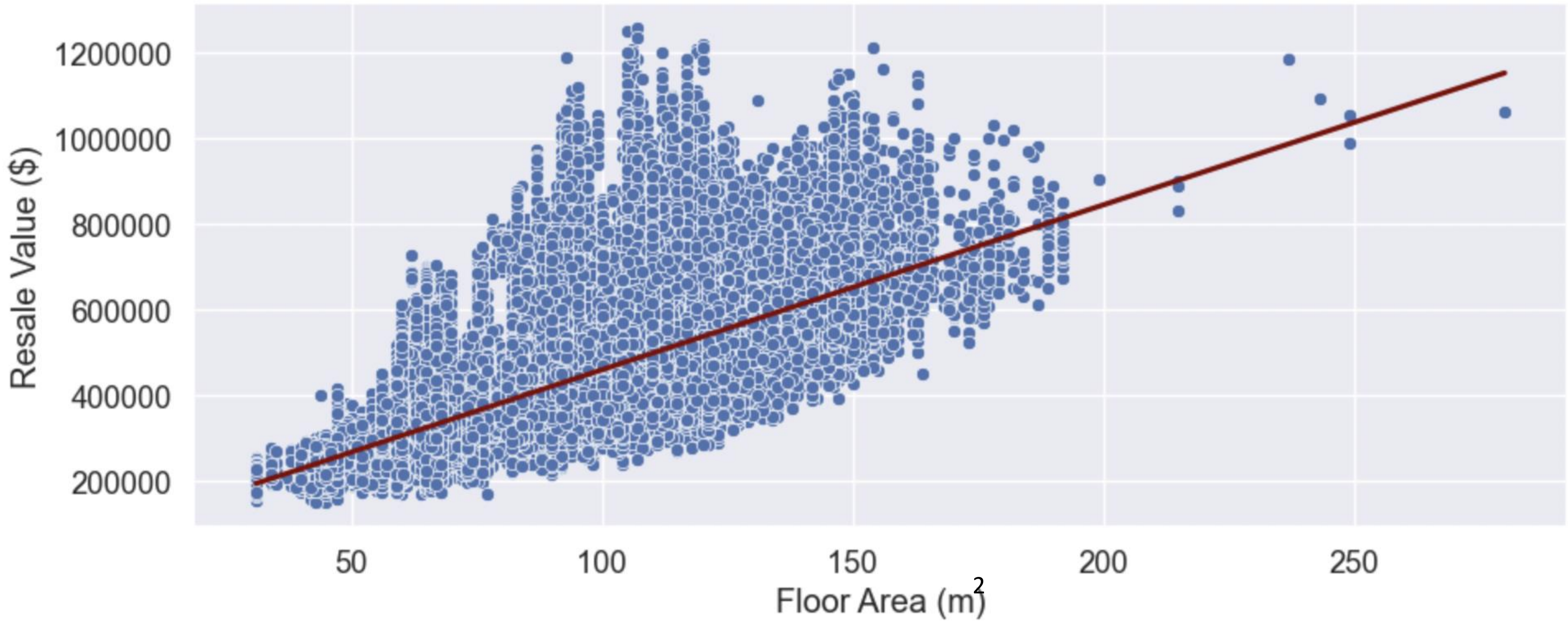
Highest frequency of resale transactions occurred near $400,000 SGD

Skewed towards early hundred thousands SGD

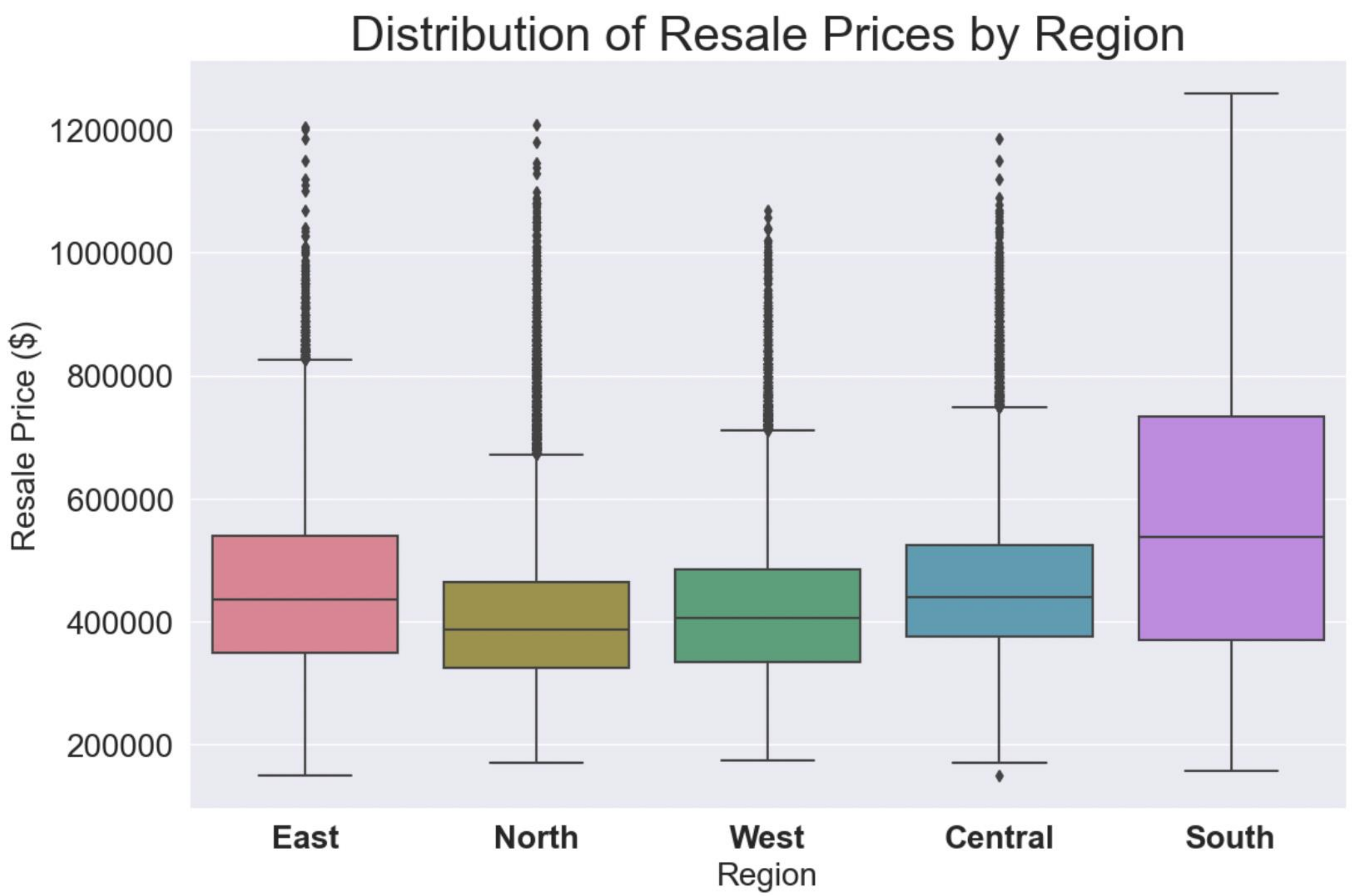


Floor Area vs Resale Price

Floor Area is the strongest predictor of Resale Price



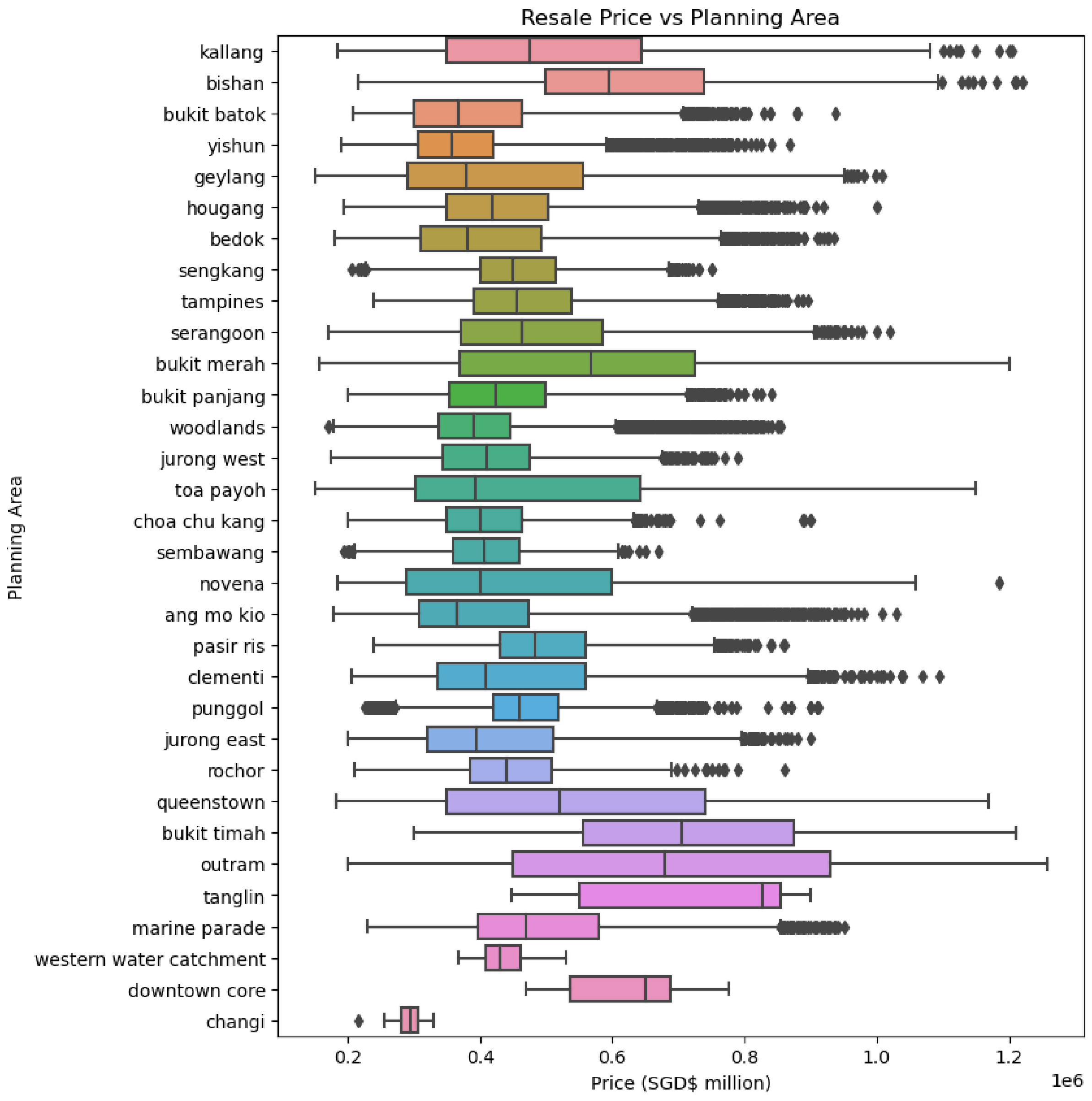
How expensive is each region?



Most expensive resale flats

- We can see that the most expensive locations for resale flats are in Tanglin, Bukit Timah, and Outram

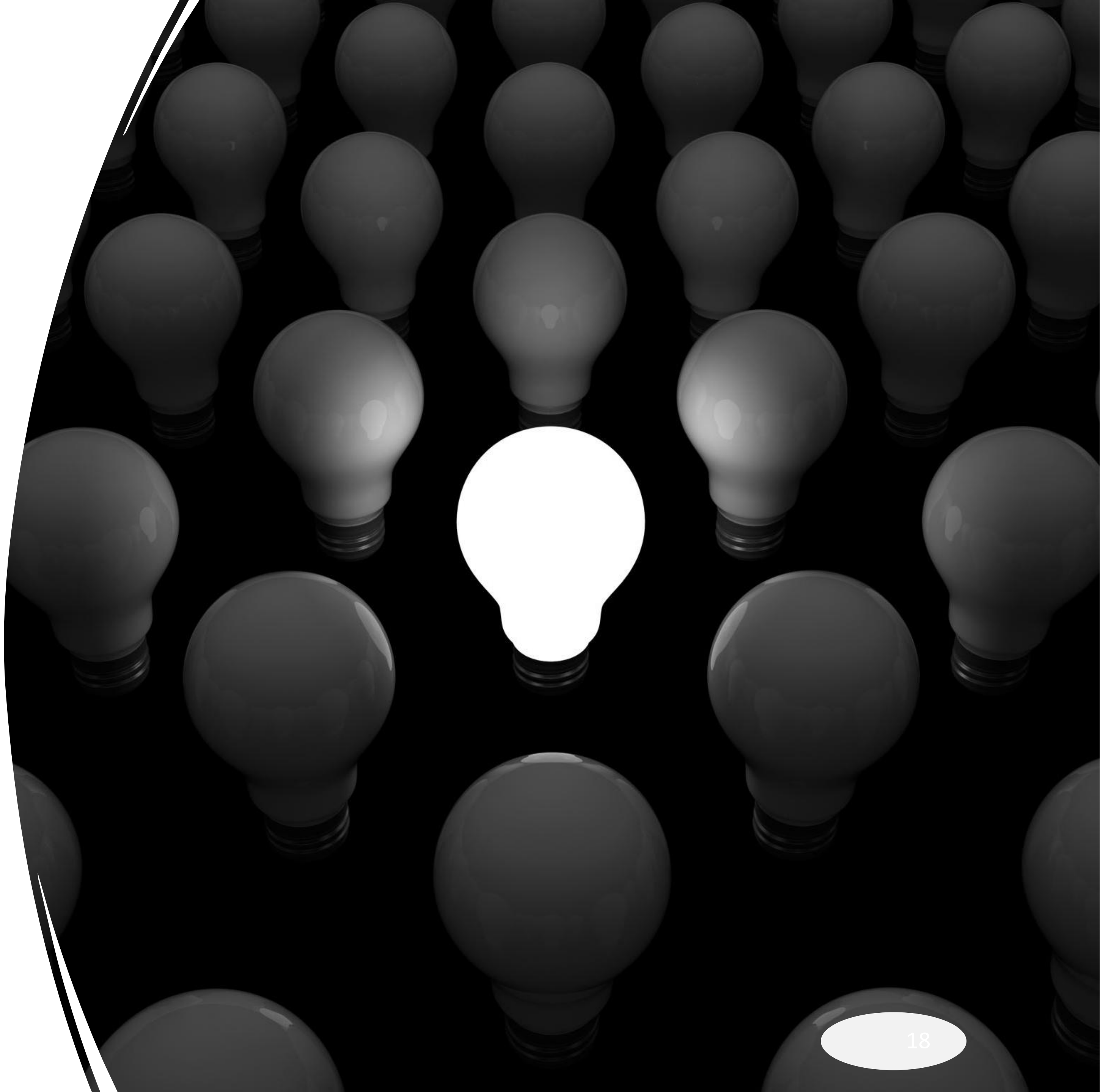
- Cheapest resale units are Changi, Yishun, Bukit Batok and Woodlands



Feature Selection

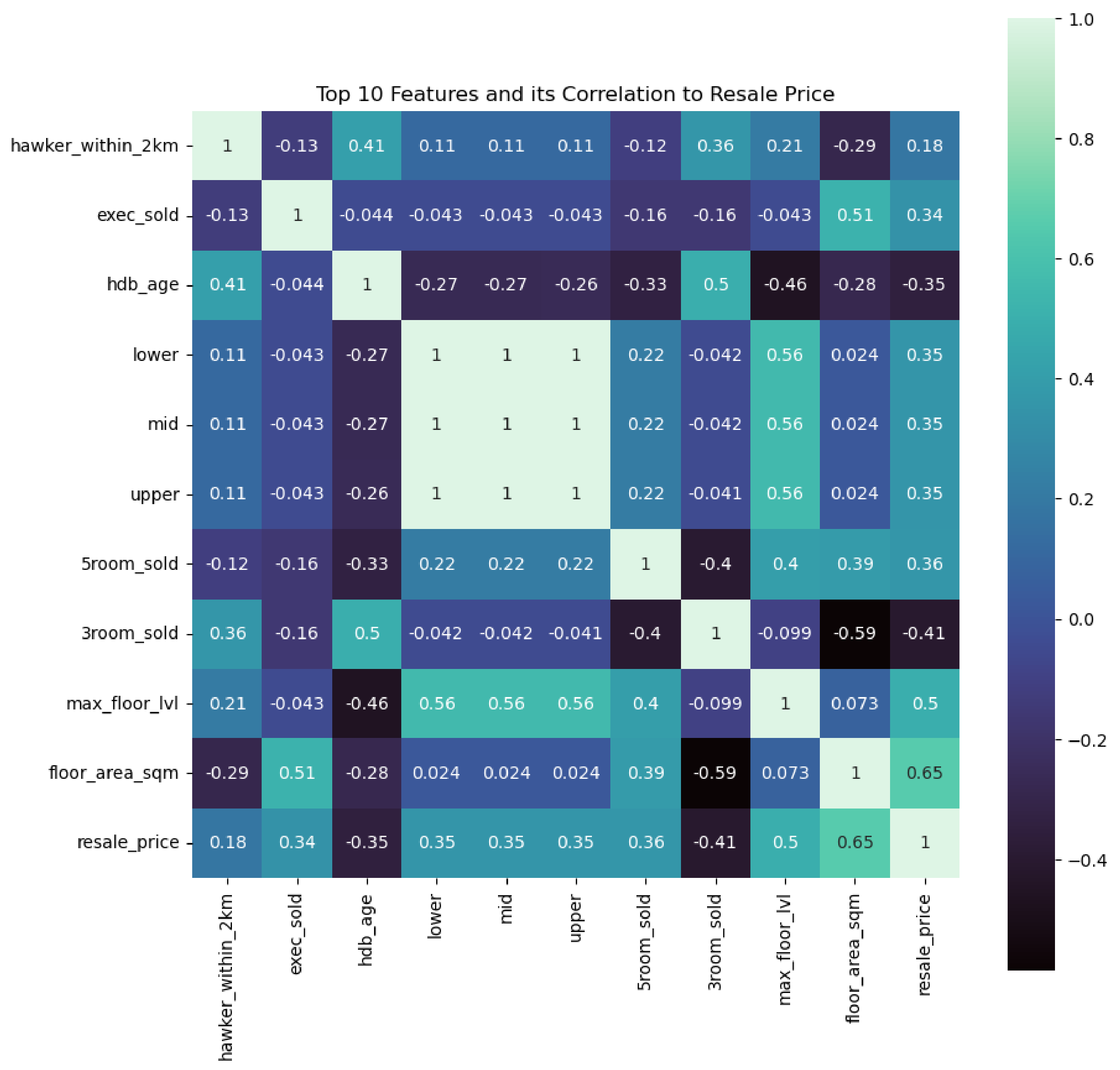
WHAT IS IT?

- **Choosing features that offer greater predictive value**



Exploration & Analysis

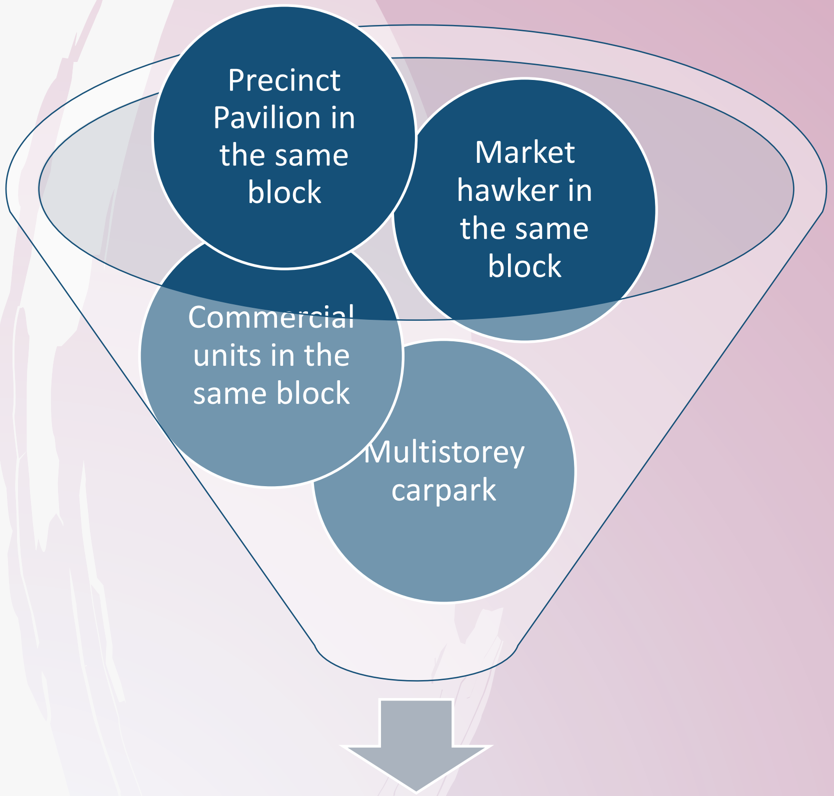
We made a heatmap to highlight statistically important features that may be useful



Exploration & Analysis

Certain individual features have only 2 unique values, which may not be useful to build a model

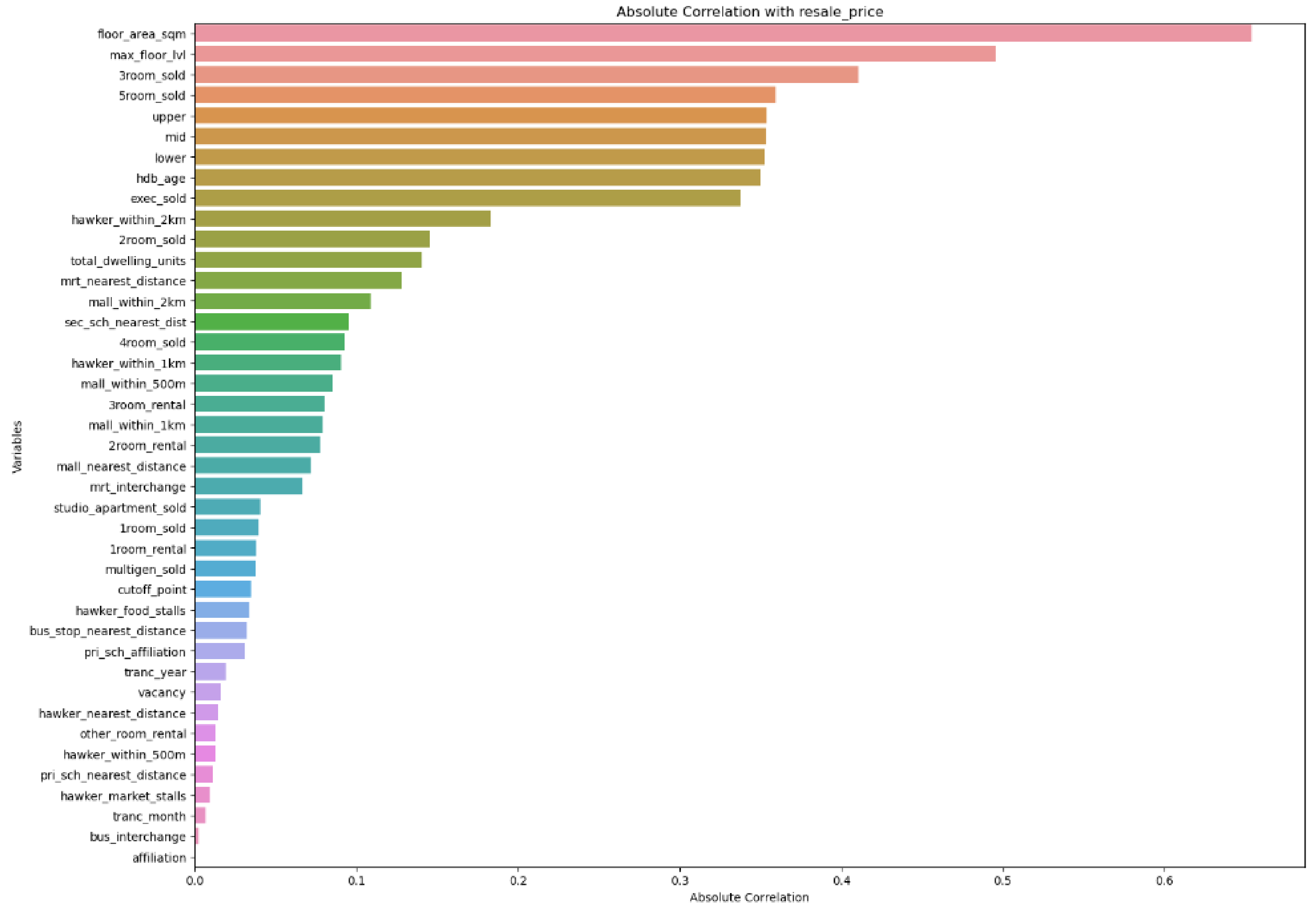
We used statistics to check if these features significantly affect resale price



Precinct Pavilion dropped

Which quantitative factors impact resale price the most?

- Floor area
- Floor level
- Number of dwelling units
- MRT distance
- HDB Age



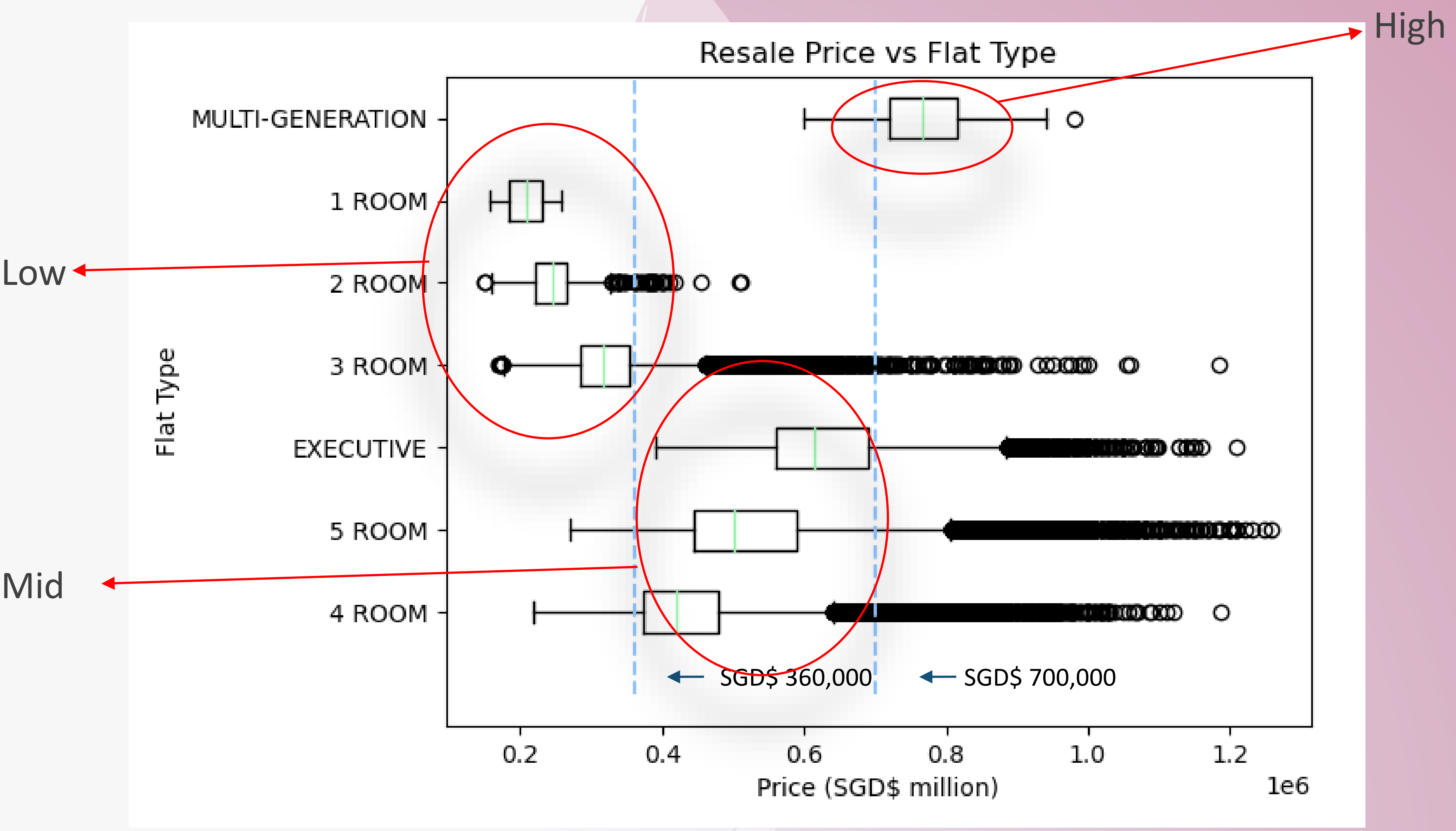
Feature Selection - Domain Knowledge

We used domain knowledge to pick out factors that affect resale prices:

- Distance to Malls
- MRT Name
- Flat Age
- School Distance
- Distance to MRT
- Floor Level
- Flat Model
- Hawker Centre Distance
- Planning Area
- Floor Area
- Secondary School
- Primary School

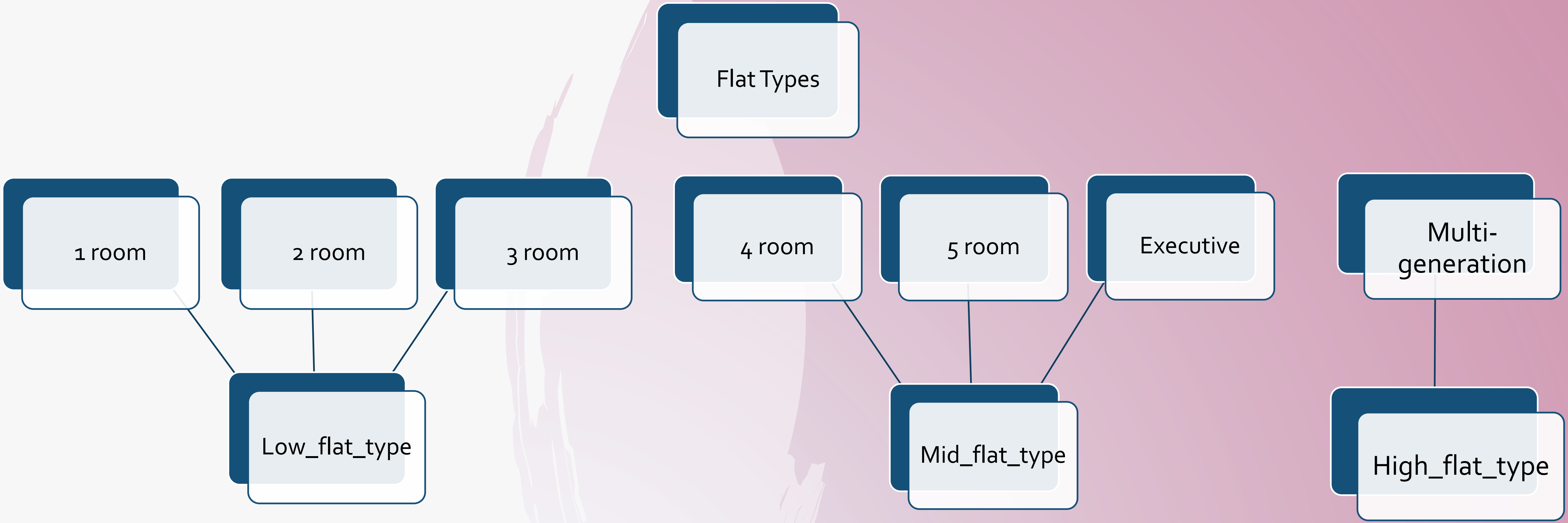
Feature Selection - Ordinal Encoding

We look for distinctions between the types, and group them together



Feature Selection - Ordinal Encoding (cont'd)

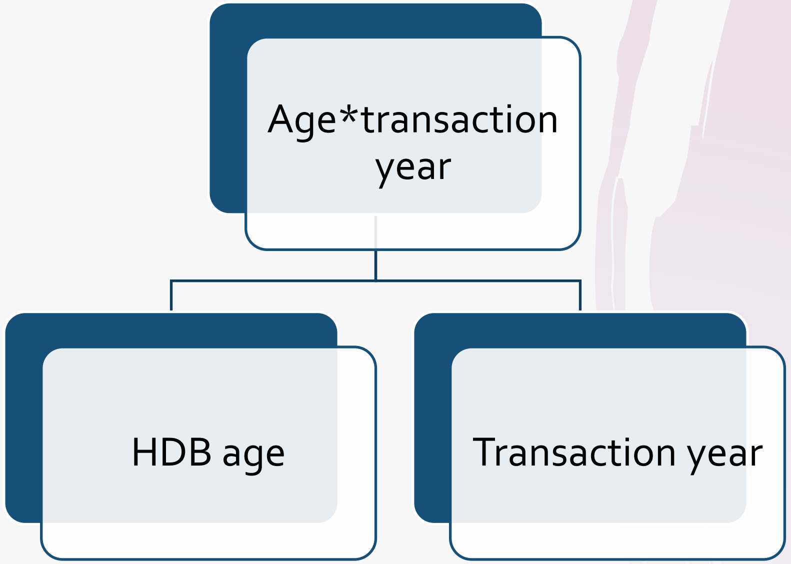
To minimise complexity of our model and improve accuracy, we perform ordinal encoding. For example:

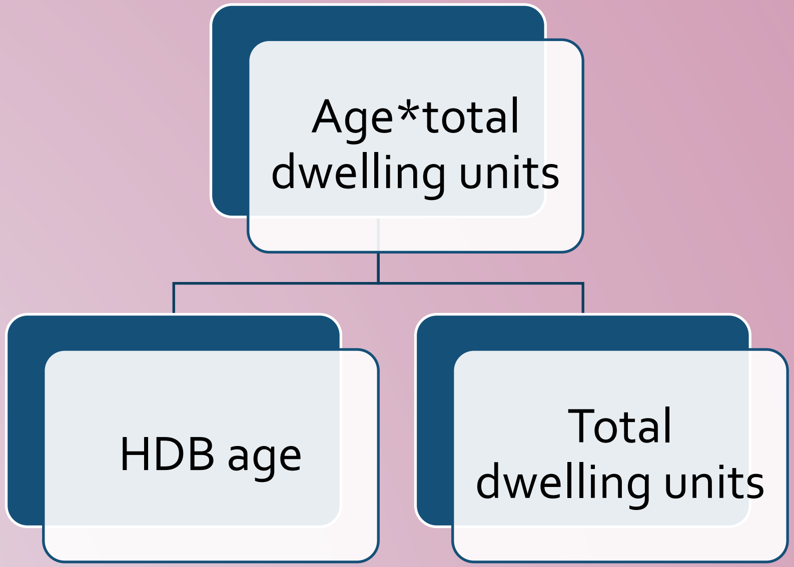


We then do the same for flat model, planning area, storey range, mrt name, primary school and secondary school

Feature Selection - Creating Interaction Terms

We combine certain categories to be used as features in our model. For example:





Final Features

floor_area_sqm
pri_sch_name
max_floor_M
flat_model_low
sec_sch_name_low
hdb_age
hawker_within_2km
mrt_nearest_distance
flat_model_mid
mrt_name
mid
planning_area_low
sec_sch_nearest_dist
mall_within_2km
sqm_year_max_floor
max_floor_5room
age_3room
floor_hawker2km
age_totalunit
year_floor
maxfloor_secsch
year_age
age_execsold
age_pri_sch
floor_hawker
floor_mall1km
storey_range
which_floor
floor_maxfloor

Model Explanation, Demonstration and Comparison

Models

Linear Regression

- Models the relationship between a target variable and one or more predictors
- Assumes a linear relationship between the predictors and the target
- Tries to find the best fitting line through the data points

Lasso

- Predictive model that helps simplify complex models by 'penalizing' them for using too many features
- 'Shrinks' the impact of less important features, making them easier to interpret
- Benefits of Lasso:
  - Makes the model simpler and more interpretable
  - Can improve accuracy by excluding irrelevant features

Ridge

- Minimizes prediction errors and prevents overfitting,
- "Penalizes" or reduces the effect of less important features, causing their impact to shrink but not disappear
- Benefits of Ridge:
  - Helps to prevent overfitting.
  - Balances the importance of all features.

How the Model is Built

Split data:

75% to train our model

25% to test our model

Train, assess and refine our model with the selected features

Make predictions

Model Comparison

We used R^2 score and Root Mean Squared Error (RMSE) to evaluate our models

Baseline Model: Selecting the highest correlated category (floor area) from the original data from HDB

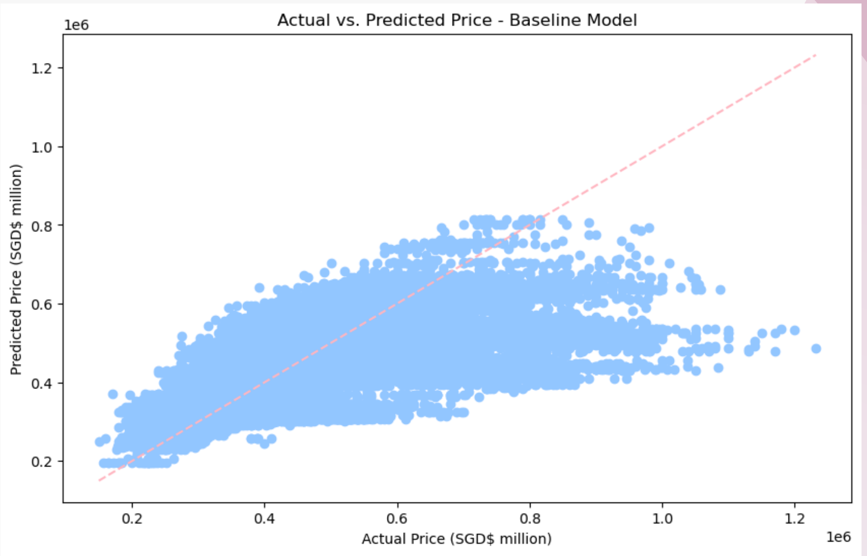
Baseline Model	Performance
Linear Regression	R^2: 0.42630236 RMSE: 108,541.94

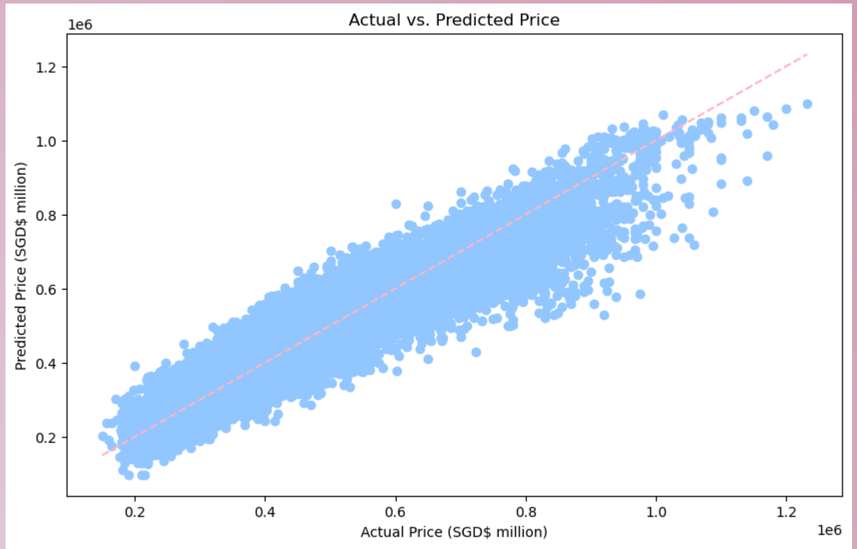
Best Model: Perform ordinal encoding on certain categories, and feature engineering on interaction between categories

Best Model	Performance
Linear Regression	**R2: 0.86526362** **RMSE: 52,601.55**
Lasso Regression	R^2: 0.85001801 RMSE: 55,497.78
Ridge Regression	R^2: 0.85077862 RMSE: 55,356.88

Model Comparison

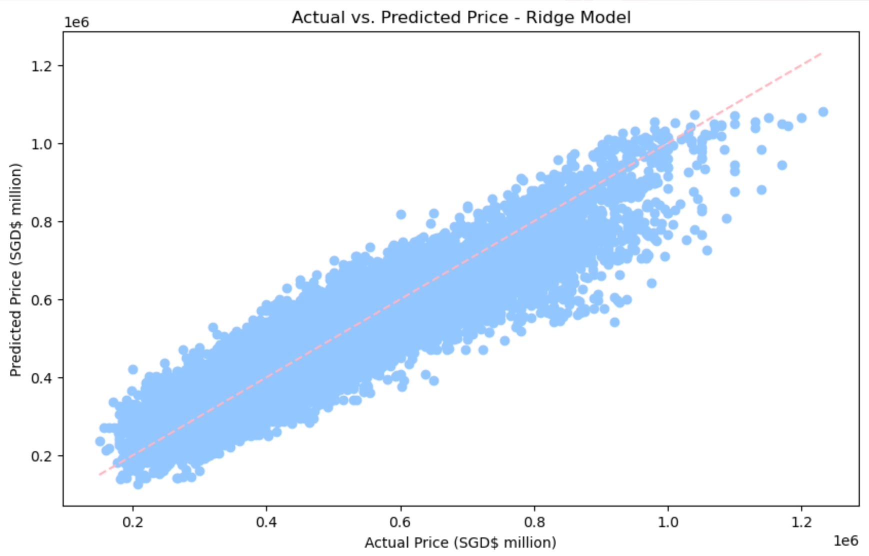
Baseline Model – Linear Regression

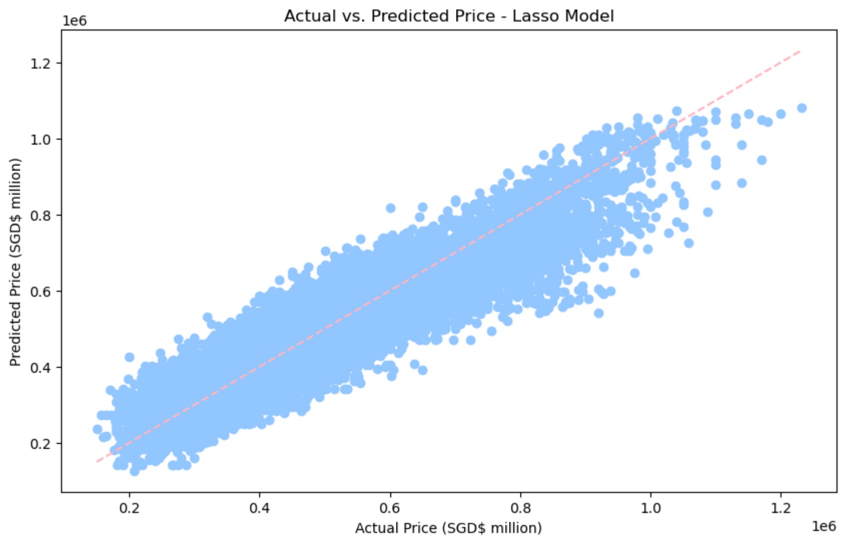




Best Model – Linear Regression

Ridge Regression





Lasso Regression

Top 5 Features

Flat Age

Older flats tend to see lower resale price

Floor Area

Bigger floor areas tend to see higher resale price

Floor Level of Unit

Higher floors tend to see higher resale price

Transaction Year

Recent transactions tend to see higher resale price

Distance to Nearest MRT

Shorter distances to MRT tend to see higher resale price

Recommendations

Our Model

The powerful combination as a trusted and innovative player in the market

Enhance Accuracy and Predictive power

Increased trust and confidence

Reduce potential for disputes or dissatisfaction.

Time & Cost Efficiency

Streamlined approach

Automates process

Scalability & Adaptability

Easily updated and fine-tuned

Stay ahead of market fluctuations

Future Work

A Dynamic and Evolving tool

Optimisation

With more machine learning models for increased accuracy.

Check the model for alignment with linear regression assumptions

Trend Adjustment

Factor in inflation and projected year-on-year trend to adjust prediction

Feed Recent Data

Continue to train our model with more recent resale transaction data

Streamlit Demo

THANK YOU

RGNT Consulting

bd@rgntconsulting.com