Expressway impact on HDB Resale Price

SCENARIO 1| LOH SI JUN SHAUNA

Executive Summary

- Problem Statement

- The Dataset

- Exploratory Data Analysis

- Feature Engineering

- Feature Selection

- Regression Model

- Model Evaluation

- Conclusion and Recommendations

Problem Statement

Some forum posters have complained that the value of their HDB flats suffer because they are near expressways, which are very noisy. Others say expressway proximity is good, due to the unblocked view (at least for higher floors).

Analyse if the HDB resale prices are indeed negatively affected due to close proximity from an expressway.

Dataset

- **Data.gov.sg** | HDB flat details with resale prices from 2017 onwards

- **National Map Line** | Locations of all expressways and sliproads in Singapore

- **Onemap** | Onemap API to query HDB locations

What affects HDB prices?

- **Transaction time** | HDB prices can be affected by economic factors

- **Location** | HDB prices differ based on town, district and if it's in an urban/sub-urban area

- **Flat-type** | HDB prices can be affected by demand of a certain flat-type

- **Storey-range** | HDB prices can be affected by preference for higher, unblocked view units

- **Floor-area-sqm** | HDB prices are determined by $ per sqm/per sq ft

- **Flat-model** | HDB prices can be affected by demand of a certain flat-model

- **Lease commence date/remaining lease** | HDB prices depreciate over time

What affects HDB prices?

- **Distance to Expressway** | How are HDB prices affected by their proximity to expressway?

Approach to determining HDB distance from expressway:

- Obtain latitude, longitude of all major expressways and sliproads
- Obtain latitude, longitude of all HDBs with its address (street name and block no.)
- Calculate nearest geodesic distance each HDB is away from an expressway/sliproad

|	
— Pan Island Expressway	
— East Coast Parkway	
— Ayer Rajah Expressway	
— Central Expressway	
— Tampines Expressway	
— Seletar Expressway	
— Kranji Expressway	
— Bukit Timah Expressway	

Exploratory Data Analysis

Heatmap:

- Strong positive correlation between floor area sqm and resale price – larger homes command higher selling price

- Positive correlation between remaining lease and resale price – home buyers prefer newer infrastructure and longer ownership period

- Positive correlation between floor and resale price – higher the floor, the better the view/unblocked view, hence commanding higher resale value

- Interestingly, distance to expressway does not seem to have a strong impact on the resale price.
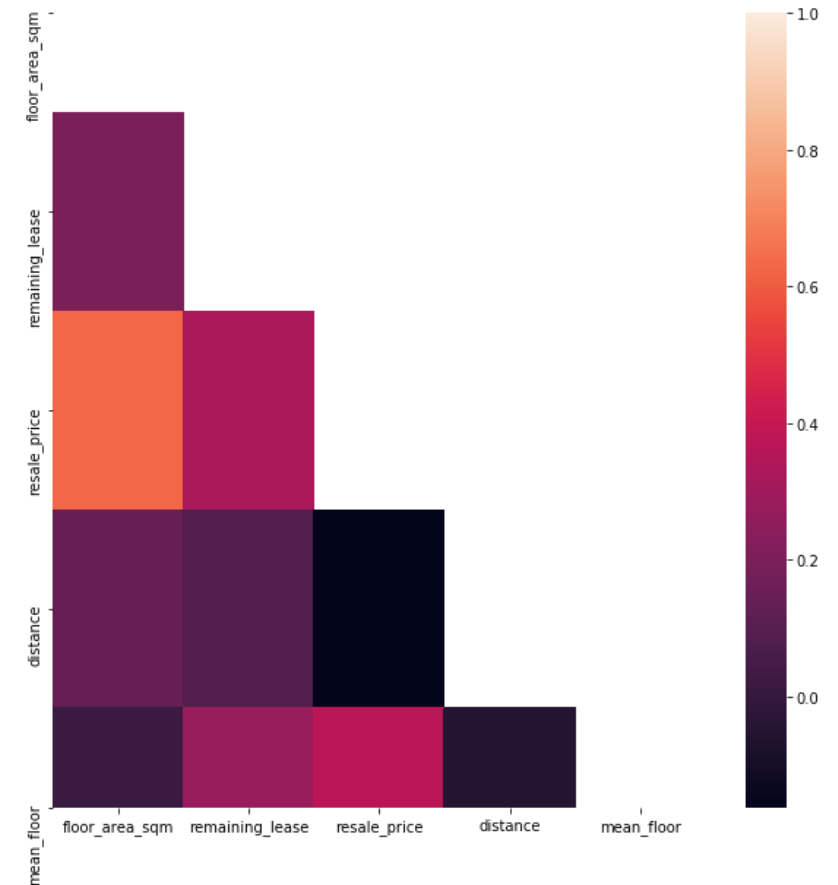
Exploratory Data Analysis

**Pairplot: Relationship between numeric variables and resale price**
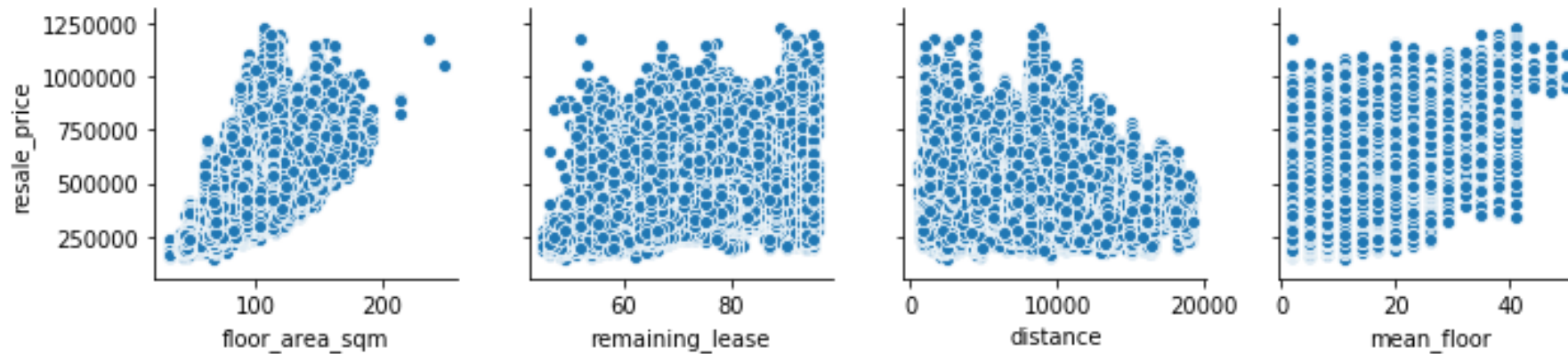
Strongest positive correlation between floor area sqm and resale price. HDB size is an important factor in the determination of HDB resale prices.

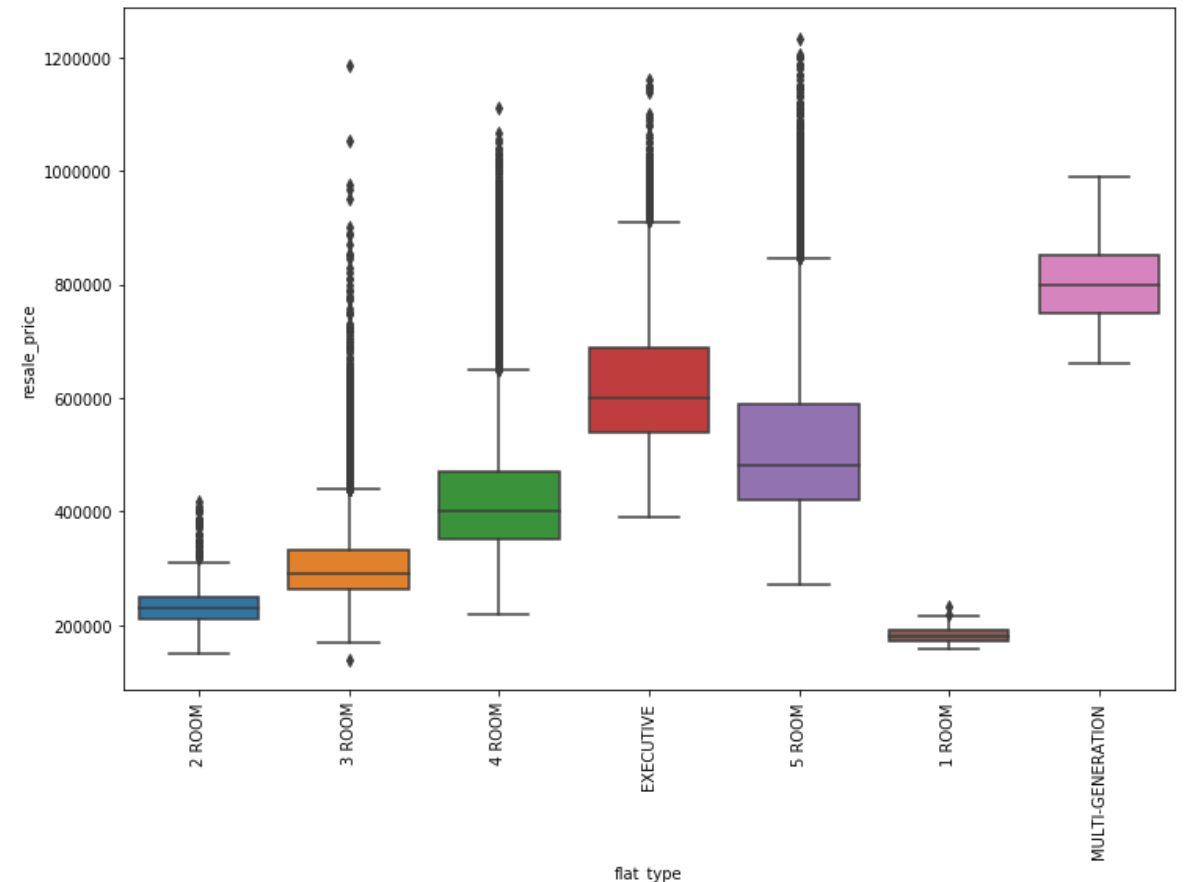# Exploratory Data Analysis

Boxplot: Resale prices across towns

• Highest HDB resale prices are in the known prime districts i.e. central area and bukit timah. Towns which are nearer CBD also fetch higher resale prices e.g. queenstown, bishan, toa payoh. However, these towns also have higher variation in price (larger interquartile range)

• Cheapest housings are those nearer the north, such as yishun, woodlands and sembawang. Interestingly, these are also the houses which have the least variation in resale prices.
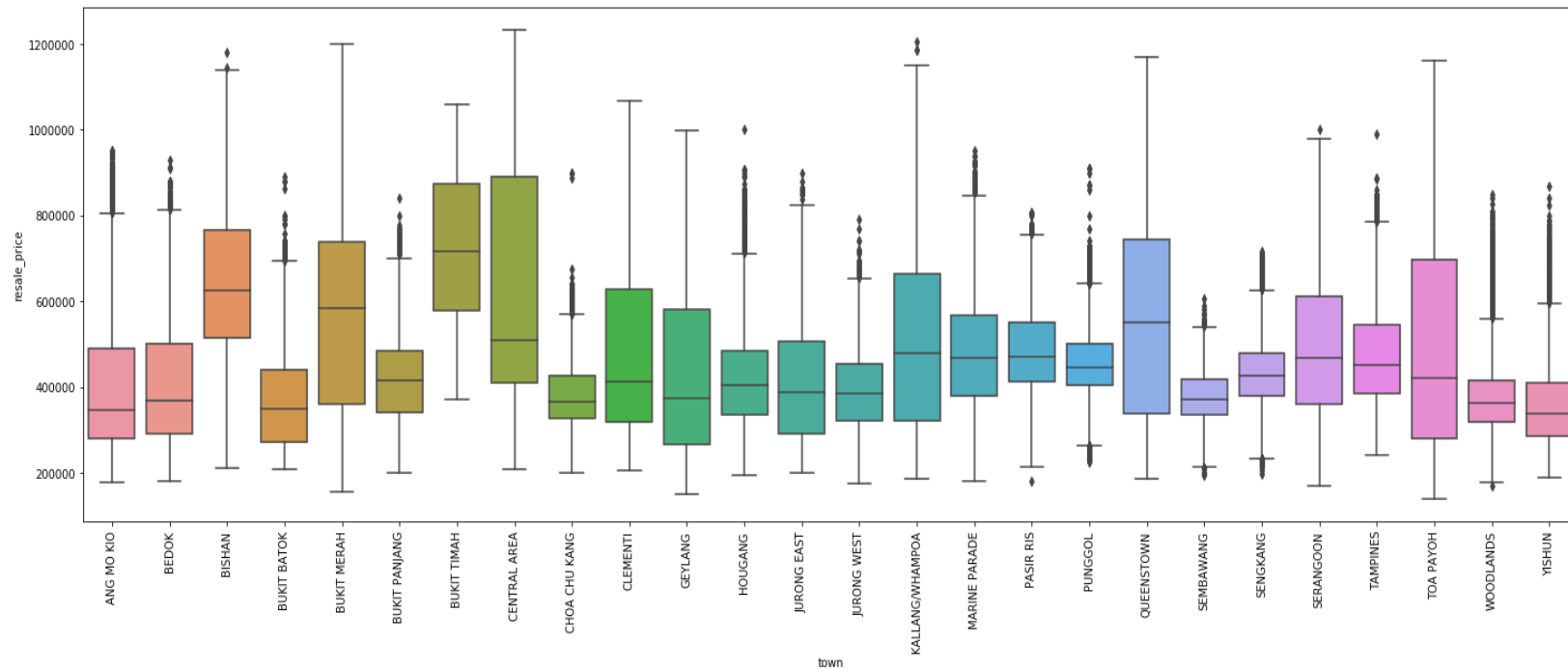
Exploratory Data Analysis
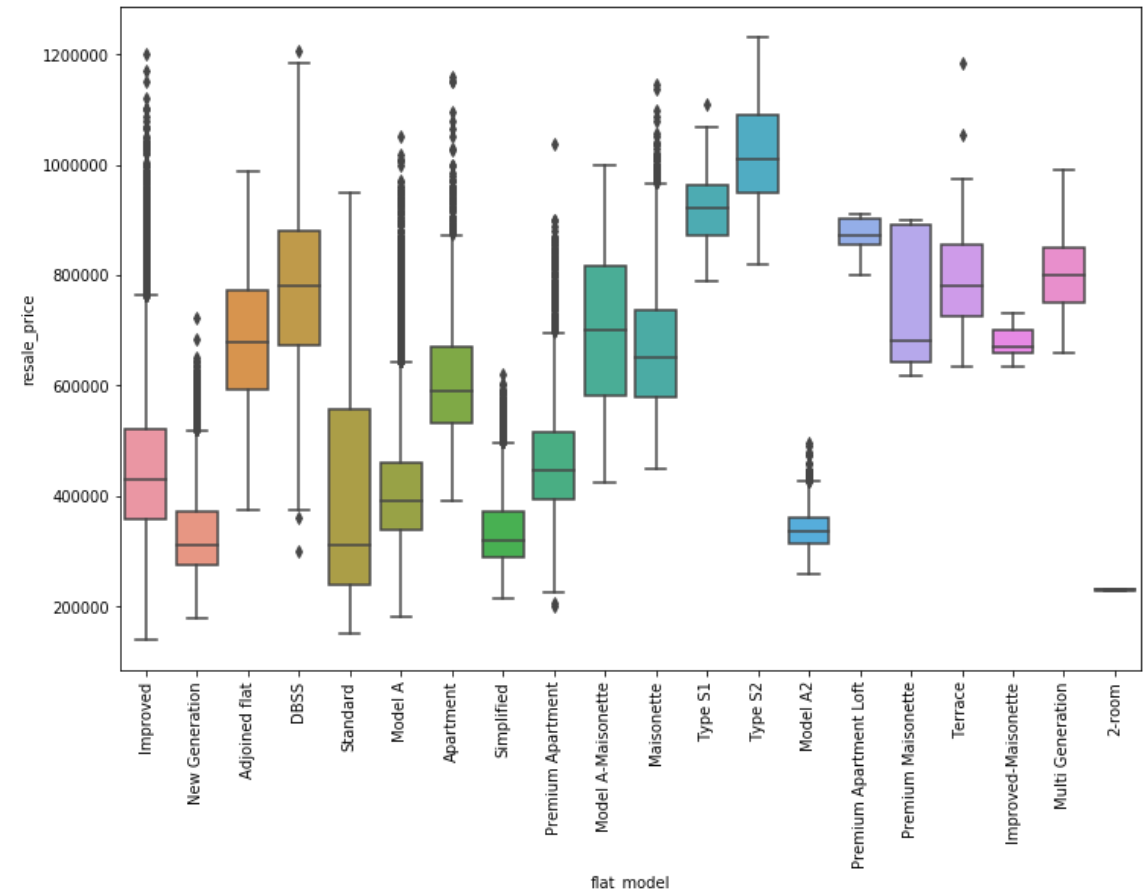
Boxplot: Resale prices across flat types

- The flat types are actually representative of the size of the HDB e.g. 1-room flat is the smallest while multi-generation is the largest. The smaller the flat, the lower the resale value.

- The more popular flat types such as 4-room, 5-room and executive have a wider variation in resale prices.

- Multi-generation flats have lower variation in their home prices compared to 4/5 room and executive flats. This could be because there is little room to increase the price for multi-gen units as home buyers may choose the next tier upgrade i.e. Executive Condominums.

Exploratory Data Analysis

Boxplot: Resale prices across flat models

- Type s2 housing (5-room) has the highest resale price, perhaps owing to its popularity among young couples, small families and even multi-generation families.

- DBSS and Premium Maisonette also have high resale prices, perhaps due to their limited quantity and unique built.

Exploratory Data Analysis

Boxplot: Resale prices across floor height

- Higher floors command higher resale prices due to unblocked and hence better view.
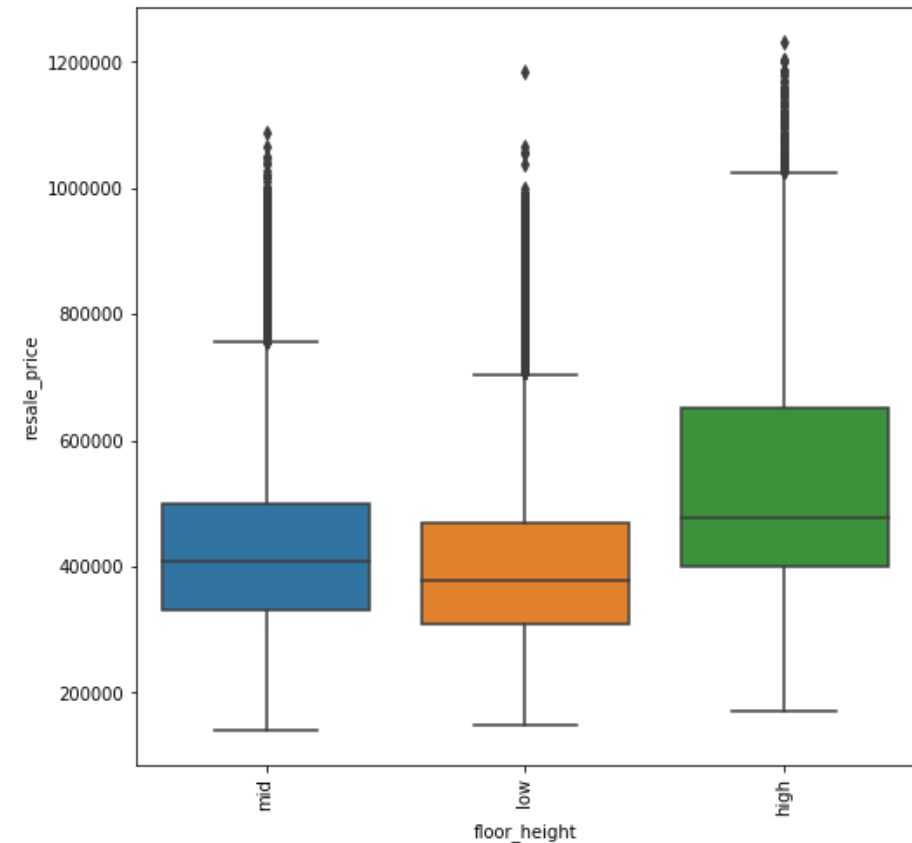
Feature Engineering

Distance	Calculated distance from each HDB to their nearest expressway/sliproad based on features latitude, longitude.
Storey-range	Mapped each storey-range to its mean value Created new feature 'floor-height' to categorize HDBs based on low-rise, mid-rise, high-rise*
Remaining-lease	Remaining-lease is taken to be the difference between the transaction date and lease commencement date

*Note: Floor-height categories should be relative to each HDB development. In this context, the following assumptions were made: Low-rise : Storey 1 to 6, Mid-rise: Storey 7 to 12, High-rise: Storey 13 and above

Feature Selection

- Discard categorical variables which exhibit dependency via **Chi-Square test.**

    H0: The features **x1** and **x2** are **independent** (which means they are not associated).
    H1: x1 and **x2** are **not independent** (which means they are associated).

    (p value < alpha, we **reject** the Null Hypothesis and if **p value > alpha** we **do not reject it**.)

- Categorical features retained: town, flat-model, flat-type (mapped to ordinal), floor-height (mapped to ordinal)

- Numeric features retained: floor-area-sqm, remaining-lease, distance, mean-floor*

*Note: Given more time, could have explored variance inflation factor to measure collinearity among the numeric features

Regression Model

- **Dummy features**: town, flat-model

- **Train-test-split**: Train set 80%, Test set 20%

- **Multi-linear regression**
  - Accuracy score: 0.866
  - Cross-validation score: 0.859
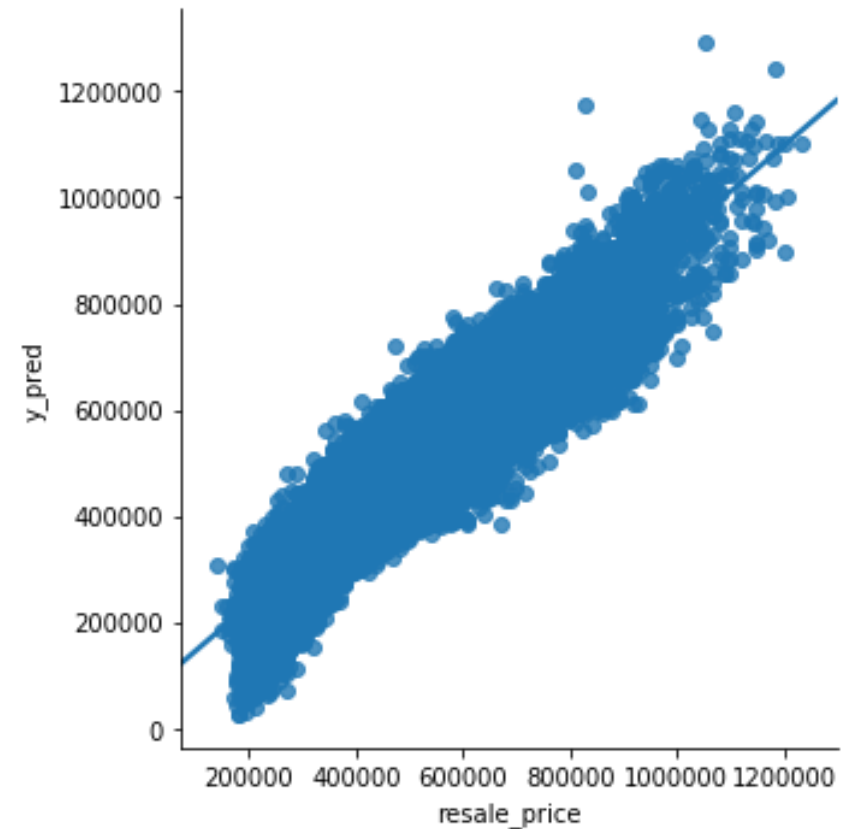
Model Evaluation

Metric	Value	Remarks
Mean Absolute Error (MAE)	43144.21	MAE gives us the average absolute value of our actual data from the model's prediction.
Mean Squared Error (MSE)	3138722873.95	The model's prediction will miss the actual value of the property by $3138722873.95 on average squared.
Root Mean Squared Error (RMSE)	56024.31	The model's prediction will miss the actual value of the property by $56024.31 on average.

- Potential improvements to model: Regularization (RidgeCV, LassoCV)

- Could explore other regression models (e.g. Gradient Boosting Regression – provides weightage to the features for modelling)

Conclusion and Recommendations

Ordinary Lease Squares (OLS) Summary Table:

R-Squared — Our model can explain only about 0.865 or 86.5% of the variance in our predictions.

Statistical Significance of our Coefficients — Our p-value in the model is small (<0.05) hence we can reject the null hypothesis and conclude that our variables has a relationship with resale price.

Coef – it tells the strength of the relationship between the feature and the target variable. Features floor area sqm and remaining lease have a strong positive correlation with resale prices.

'Distance' has a weak negative relationship with resale price. When distance decreases (i.e. home is nearer the expressway), the resale price increases. However, this relationship is not significant.

Conclusion and Recommendations

Why does living near an expressway actually result in an increase in HDB resale price?

For those who drive, living near expressway provides greater accessibility, makes commuting more convenient and reduces the time taken to travel to and fro home. This is a likely reason why some buyers prefer to buy property near expressway, hence house prices near expressway may be even higher than homes that are not.

• Similar to living near an MRT, which generates noise every time a train passes by the station, home owners have circumvented the noise issue by installing noise cancelling panels on windows. If the problem can be reduced or mitigated, home owners may not necessarily find living near an expressway a bane.

Why does living near an expressway not have a significant impact on the HDB resale price?

• There are many other important factors which play a role in influencing resale prices, such as the size of HDB and its floor height.

Conclusion and Recommendations

To further improve the analysis:

- Take into account the orientation of the property. A HDB development may be built close to the expressway, but some units may not face the expressway. Hence, these unit owners enjoy the advantage of living near the expressway but are not affected by the expressway noise.

- Find out the resale price characteristics of living near an expressway vs a sliproad. A sliproad may experience fewer traffic and hence less noise. This may affect the resale price.

- Consider the amenities located near the HDB e.g. childcare, schools, markets, hospitals etc. and town planning/upgrades. All these factors may play a role in determining the resale price and outweigh the impact of distance to an expressway.

Thank you