

Project 2

HDB Resale Prices

vivalamuffin





TABLE OF CONTENTS

01

Context

Who are we?

02

MARKET ANALYSIS

Understanding the market

03

Problem

What is the issue?

04

Research findings

Trends

05

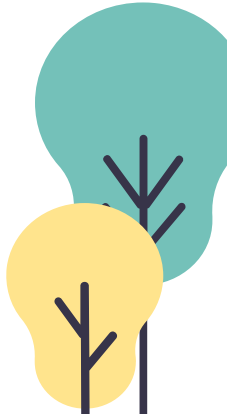
Recommendations

Our solution

06

Future work

Limitations



Definitions of Key Terms

Non Performing Loan (NPL)

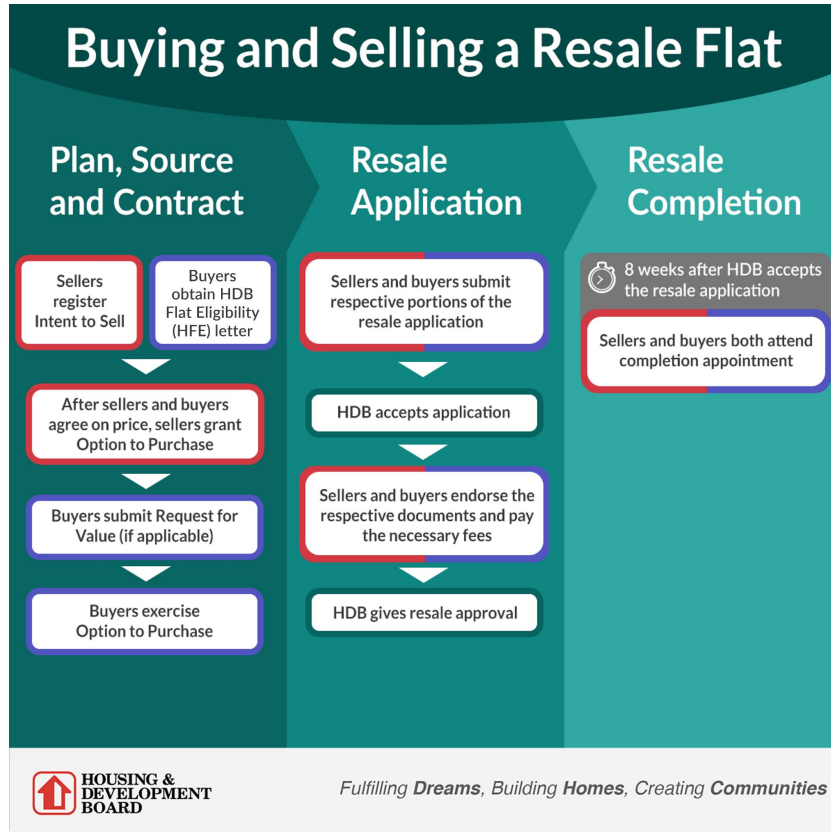
Loans that are **unlikely** to get repaid. Poses a **major threat** to profitability.

Default Risk

The risk a that a borrower will **not make** the required payments on a debt obligation.



Resale Process Overview



- HDB's information on resale process is **comprehensive** and has a detailed guide on the procedure and documentation required.
- There is a wealth of information and resources available to **multiple stakeholders**.
- Conclusion: the process is **easy to follow**, and information is **freely and readily** available.

Source: Infographic from - <https://www.hdb.gov.sg/residential/buying-a-flat/buying-procedure-for-resale-flats/overview>

The Situation in Singapore

40%

Of Singaporeans are facing mortgage stress.

3.85%

Fixed home loan rates have spiked to levels not seen in years.

\$32 million

NPL amount of top 3 SG banks (2022).

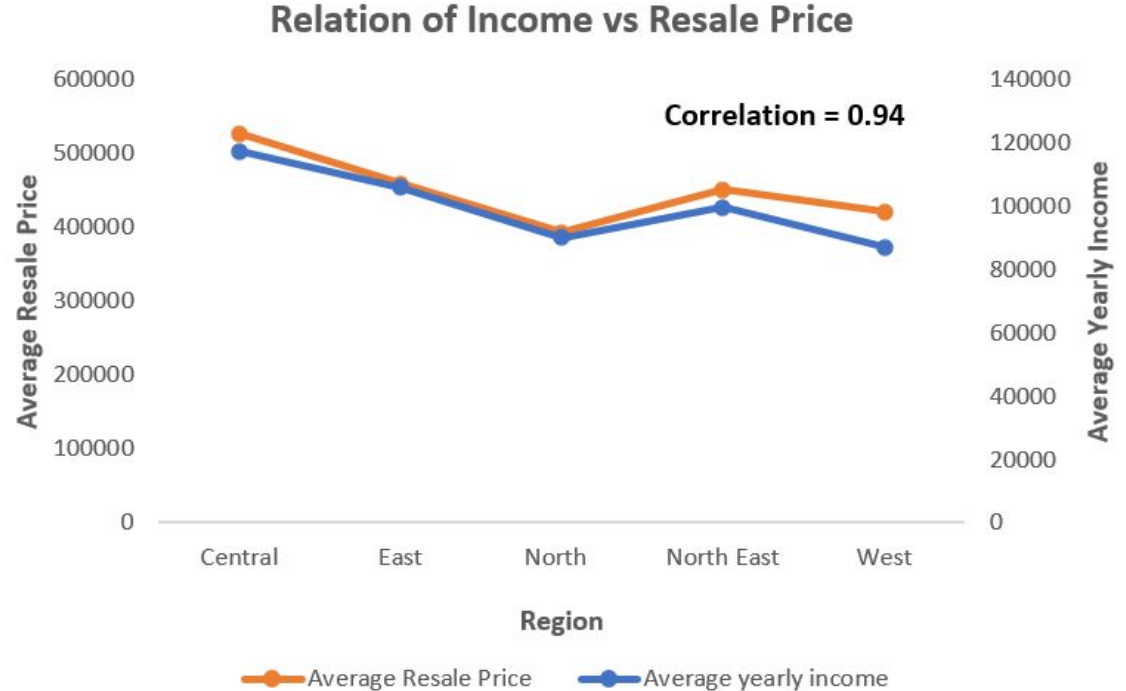


The Relation of Income and Resale Price

Property price and income are two **critical factors** that lenders consider while assessing loans. **Easy access** to income, but **not to** resale price.

Both factors have a **strong positive correlation** and will give a **holistic indication** of affordability.

They are thus ideal to create a **simple metric** that can be an **early-stage risk assessment** tool.



Source: Income statistics from - <https://blog.seedly.sg/monthly-household-income-singapore-comparison-district-where-you-stay/>

Source: What FIs look for - <https://www.homeloanexperts.com.au/home-loan-articles/7-factors-that-affect-your-home-loan-eligibility/>

“Is it possible to predict HDB resale prices using indicators such as flat attributes and location? How can a predictor help as an early warning risk indicator to Banks?”



Problem Statement

Research findings

Price of resale flats in the years period: 2013 to 2020

Average price of resale flat

\$449 162

Cheapest resale flat

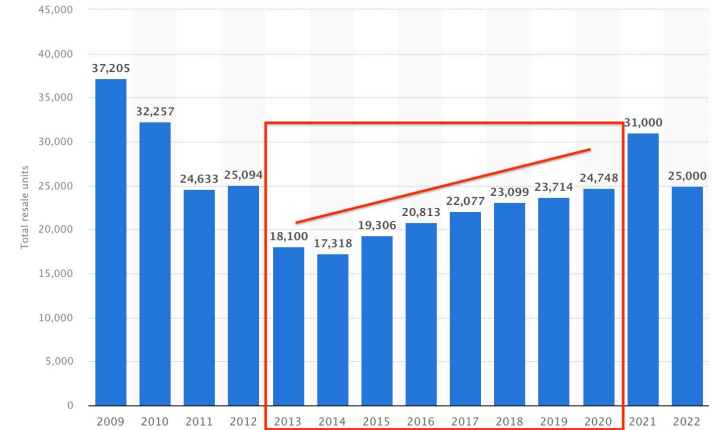
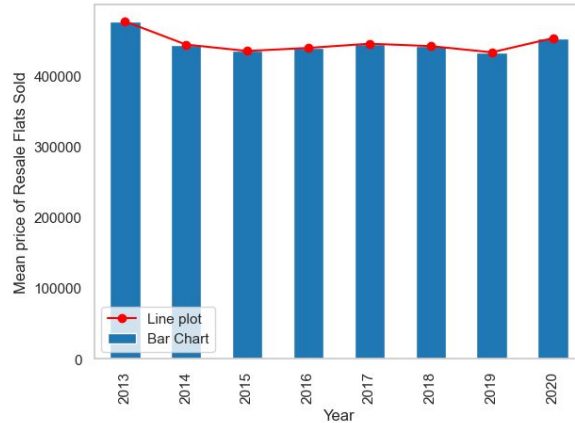
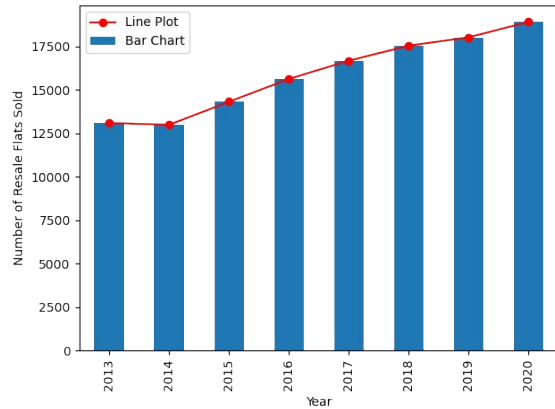
\$150 000

Most expensive resale flat

\$1 258 000

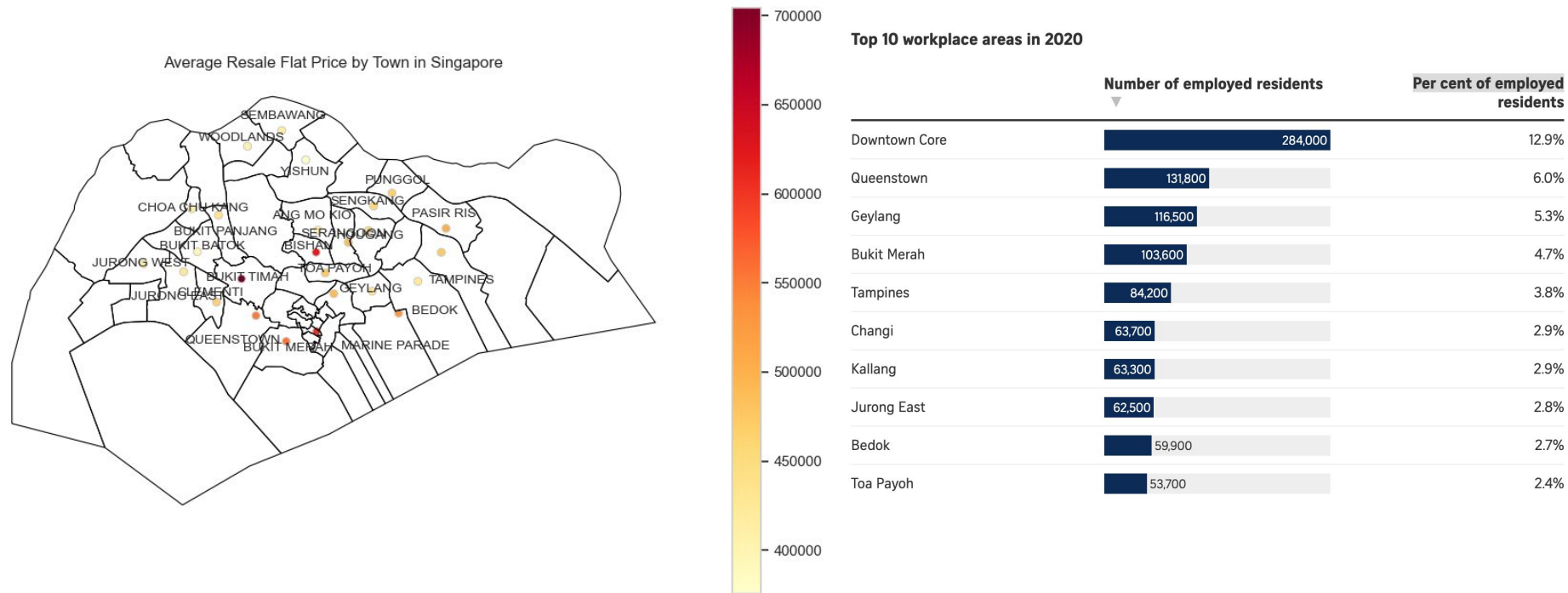


Is there a consistent upward trend in the pricing of residential flats over the years?



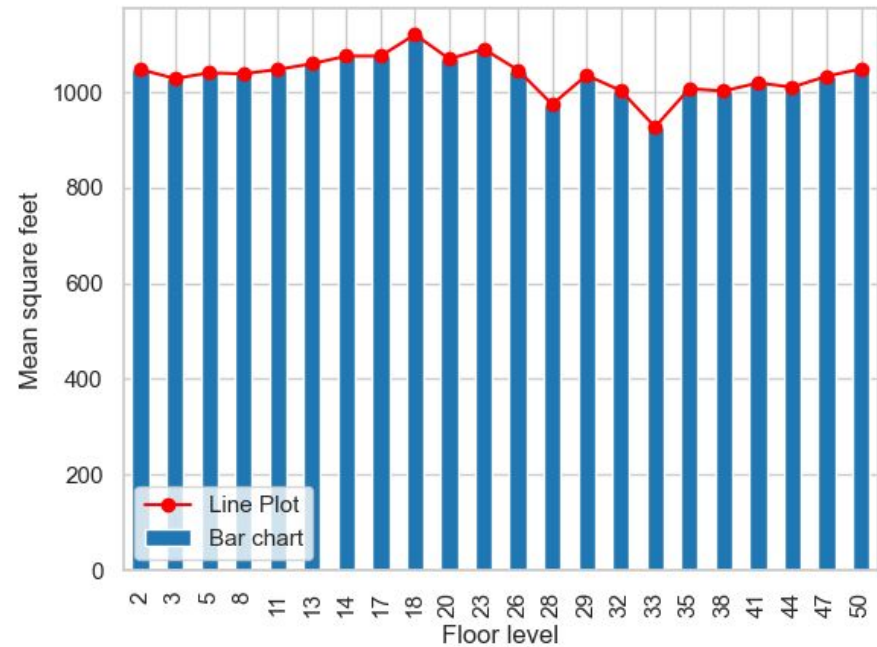
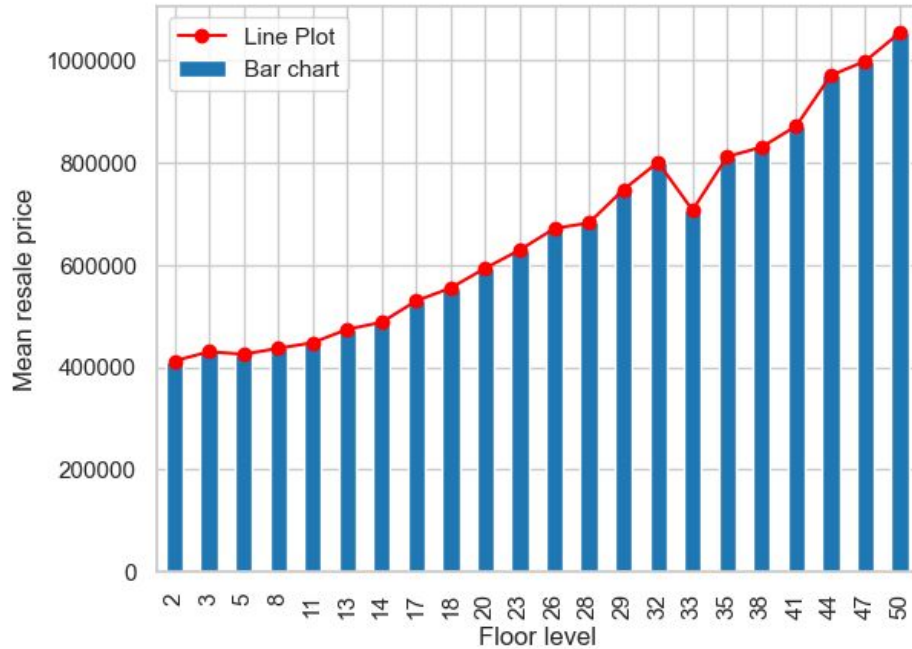
The number of resale flats being sold increases over the years but the average price remains about the same.
This could be due to an increased amount of supply of resale flats.

Resale flats in the central part of Singapore are more expensive



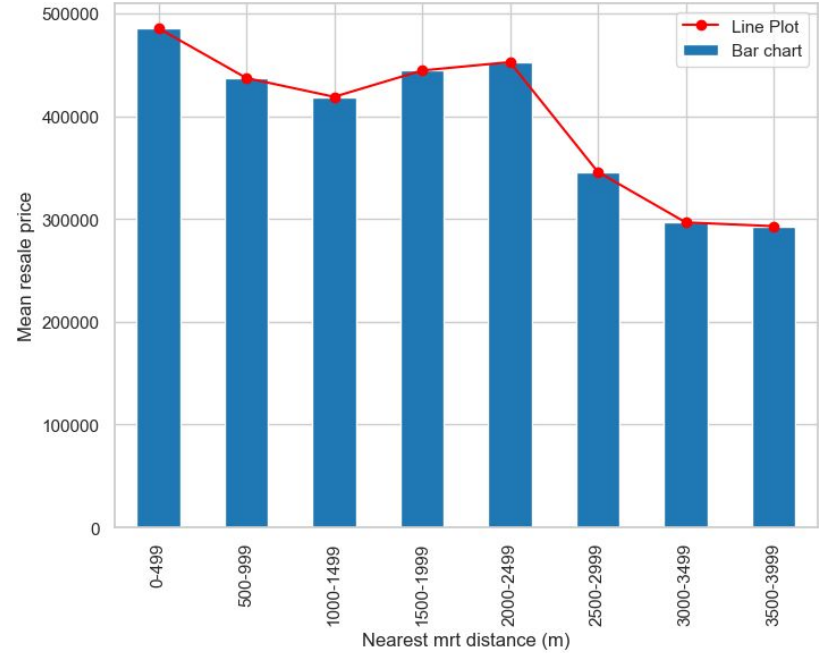
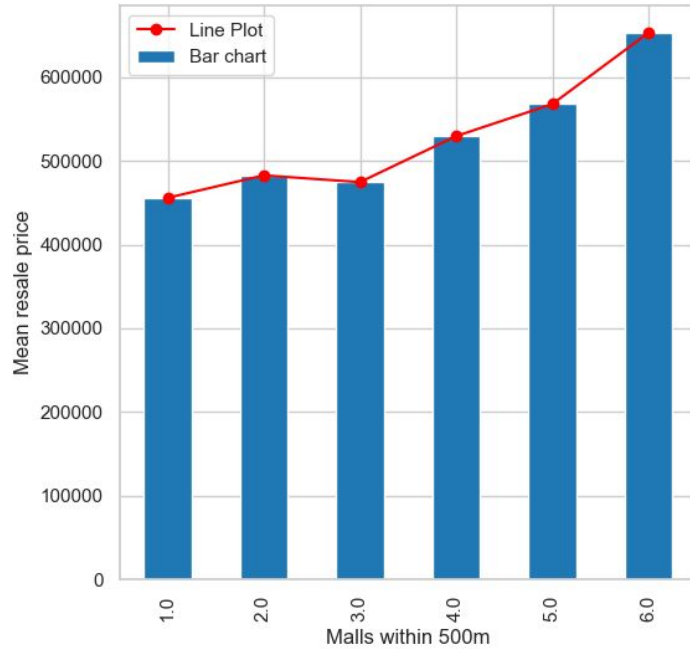
The central area enjoys a premium price due to its proximity to the central business district and the abundance of amenities like shopping malls.
The Downtown Core area is composed of 8 eight subzones – Tanjong Pagar, Marina Centre, Raffles Place, etc

The higher the floor, the more expensive the flat



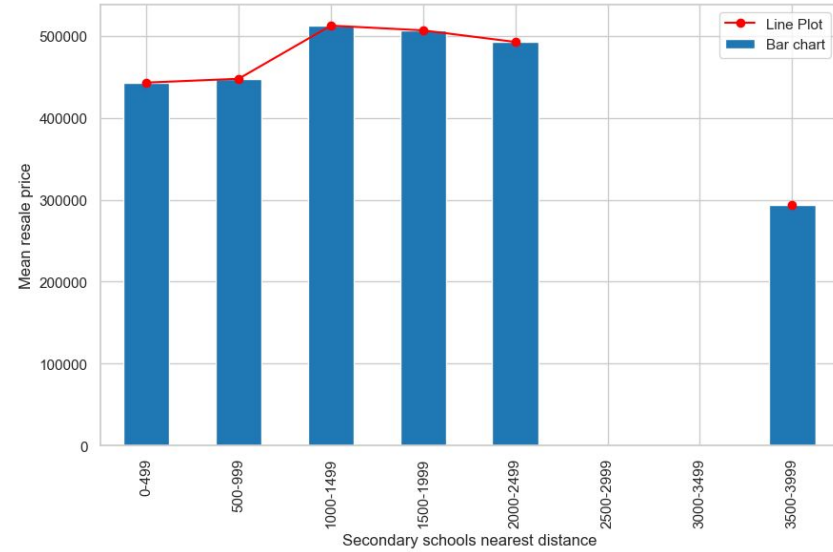
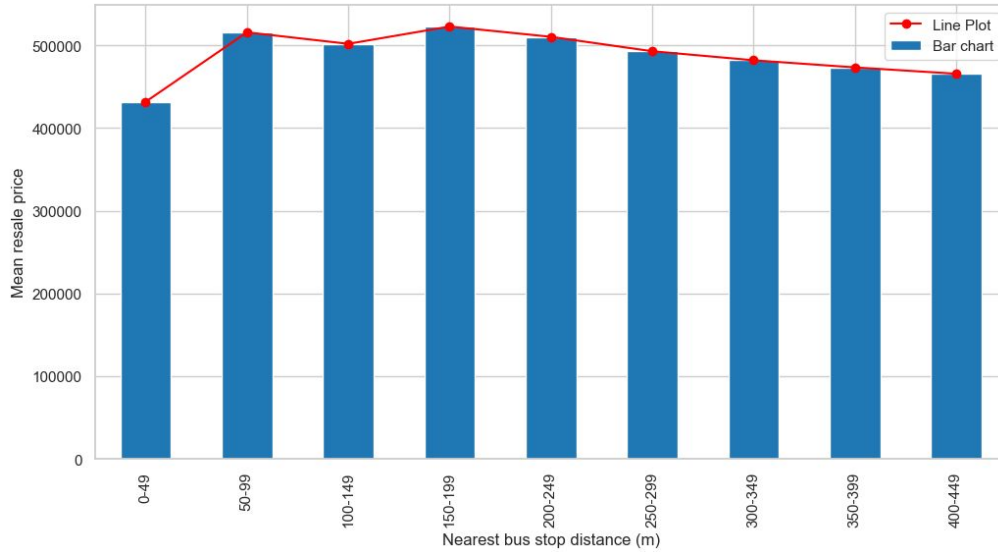
Flats on the higher floors do not necessarily have more area.
Flats on higher floors offer better views and more privacy.

The more amenities that are nearby, the more expensive the flat is



The more malls within 500m, the more expensive the flat.
The highest average resale price of a flat is found when an mrt is within 500m.

This trend is not seen for the other amenities



Flats with bus stops and schools that were a little further away had an impact on the resale price as well. It is a congruence of these factors that affects the resale price.



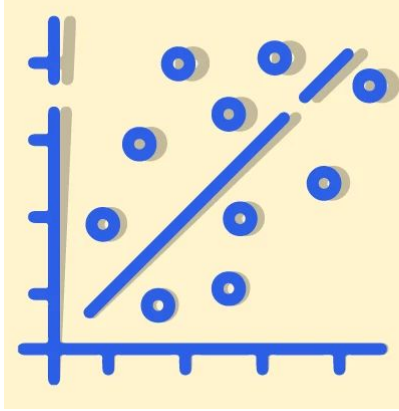
How we dealt with missing values

Missing values	Number of missing values
Mall_Nearest_Distance	829
Mall_Within_500m	92789
Mall_Within_1km	25426
Mall_Within_2km	1940
Hawker_Within_500m	97390
Hawker_Within_1km	60868
Hawker_Within_2km	29202

We used the K Nearest Neighbor method to impute the missing data and chose k to be 5.

Feature Selection & Model Performance

Main Metrics for Selection of Features and Models



Feature Selection

Feature Indicator

R-Value:
-1 to 1



Model Performance

Model Indicator

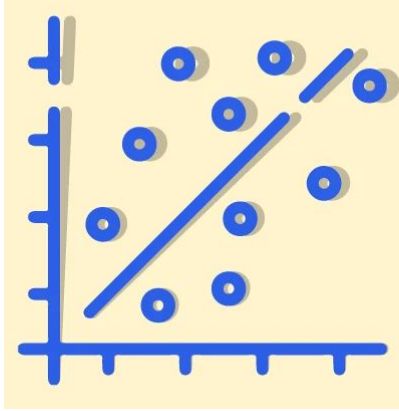
R^2 -Value:
0 - 100%

Model Indicator

RMSE Value:
Price Prediction

RMSE - Root Mean Square Value

Main Metrics for Selection of Features and Models



Feature Selection

R-Value

Strength of
Linear Relationship



Model Performance

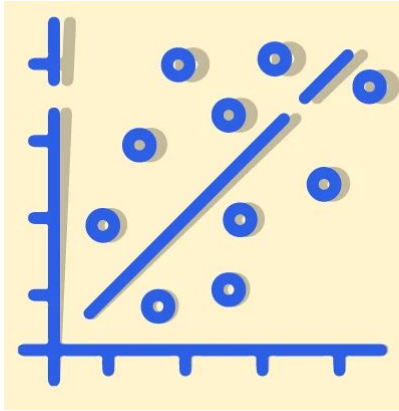
R^2 -Value

Variance
Explainability

RMSE Value

Accuracy of
Prediction

Main Metrics for Selection of Features and Models



Feature Selection

R-Value

Strength of
Linear Relationship



Model Performance

R^2 -Value

Variance
Explainability

RMSE Value

Accuracy of
Prediction

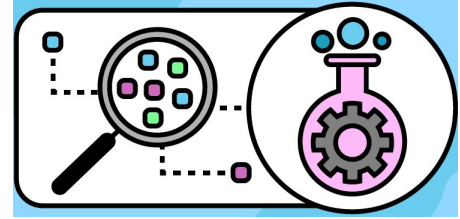


Correlation

What is the strength of the linear relationship between a factor and the resale price?

Domain Research

What do buyers and domain experts say about this subject?



Engineering

Based on Correlation and Domain Research, can we create features that show a stronger relationship?



Rationale for Feature Selection

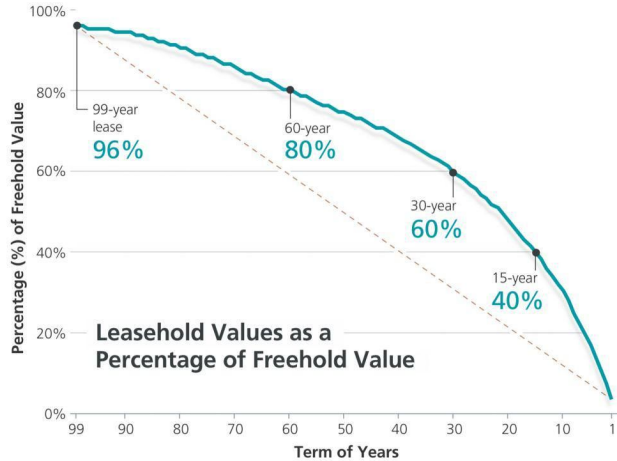
PropertyGuru Group, said that while being close to nature reserves and national parks may be an advantage, it may not be as critical as other factors such as proximity to schools, shopping centres and MRT stations.

Flats in mature estates such as Ang Mo Kio, Bedok, and Queenstown tend to command higher prices. The proximity of

If you're eyeing a rare type of resale flat, such as an HDB maisonette or jumbo flat, be prepared to fork out a higher COV. As these flat types are no longer in production, they're highly in demand because of their rarity.

Feature Selection

Homes located on **higher storeys tend to command higher premiums** In fact, it's not uncommon to see a price difference of thousands of dollars based on the floor level alone. As such, you can sell (or



Because the PSF is directly affected by the size of the property, it is not the most accurate indicator of affordability. It's like comparing apples and oranges; the playing field is not exactly equal. Though you



<https://www.propertyguru.com.sg/property-guides/high-floor-vs-low-floor-unit-which-is-better-45449>

<https://ohmyhome.com/en-sg/blog/psf-doesnt-tell-you-enough-heres-what-else-you-can-focus-on/#:~:text=In%20contrast%2C%20with%20a%20larger,hence%20the%20higher%20overall%20quantum>

HDB Resale Statistics (Q1 2023)



TOWNS	1-ROOM	2-ROOM	3-ROOM	4-ROOM	5-ROOM	EXECUTIVE
ANG MO KIO	-	*	\$380,400	\$520,000	\$697,500	*
BEDOK	-	*	\$367,000	\$518,000	\$684,000	*
BISHAN	-	-	*	\$690,000	\$880,000	*
BUKIT BATOK	-	\$340,000	\$385,000	\$590,000	\$758,000	*
BUKIT MERAH	*	*	\$455,000	\$838,000	\$875,400	-
BUKIT PANJANG	-	-	\$388,000	\$493,000	\$608,500	*
BUKIT TIMAH	-	-	*	*	*	*
CENTRAL	-	*	*	\$894,000	*	-
CHOA CHU KANG	-	*	*	\$495,000	\$603,000	\$730,000
CLEMENTI	-	*	\$390,400	\$597,500	*	*
GEYLANG	-	*	\$332,500	\$668,000	*	*
HOUGANG	-	*	\$382,900	\$525,000	\$670,000	*
JURONG EAST	-	*	\$365,000	\$492,000	*	*
JURONG WEST	-	*	\$345,000	\$476,000	\$589,000	\$727,500
KALLANG/WHAMPOA	-	*	\$383,000	\$778,900	\$865,000	*
MARINE PARADE	-	-	\$425,000	*	*	-

[https://www.hdb.gov.sg/residential/selling-a-flat/overview/resale-statistics#:~:text=Resale%20Price%20Index%20\(RPI\)&text=This%20index%20can%20be%20used,%2C%20flat%20types%2C%20and%20models.](https://www.hdb.gov.sg/residential/selling-a-flat/overview/resale-statistics#:~:text=Resale%20Price%20Index%20(RPI)&text=This%20index%20can%20be%20used,%2C%20flat%20types%2C%20and%20models.)

Estate

Does the town area matter?

Floor

Does a higher floor matter?

Model

What is considered valuable?

Convenience

Do these amenities fall within
1km?

Inflation

Do prices increase year on
year?

Distance

How far are these amenities?

Observation:

-
- A triangle correlation heatmap titled "Triangle Correlation Heatmap of Engineered HDB Features". The heatmap displays the pairwise correlation coefficients between 18 features. The features listed on both the x-axis and y-axis are: resale_price, mid_storey, floor_area_sqm, Tranc_Year, Tranc_Month, max_floor_lvl, hdb_age, vacancy, Hawker_Within_1km, Mall_Within_1km, hawker_food_stalls, hawker_market_stalls, Mall_Nearest_Distance, Hawker_Nearest_Distance, mrt_nearest_distance, bus_stop_nearest_distance, pri_sch_nearest_distance, and sec_sch_nearest_dist. The color scale on the right ranges from -0.6 (dark blue) to 0.6 (dark red), with 0.0 being white. The diagonal elements are all 1.0 (white). Notable correlations include a strong positive correlation (red) between resale_price and floor_area_sqm, and a strong negative correlation (dark blue) between max_floor_lvl and hdb_age.

Triangle Correlation Heatmap of Engineered HDB Features

Correlation of Referenced Variables

Evaluation:

Floor Level



Amenity Distance



Convenience



R-values:

0.5

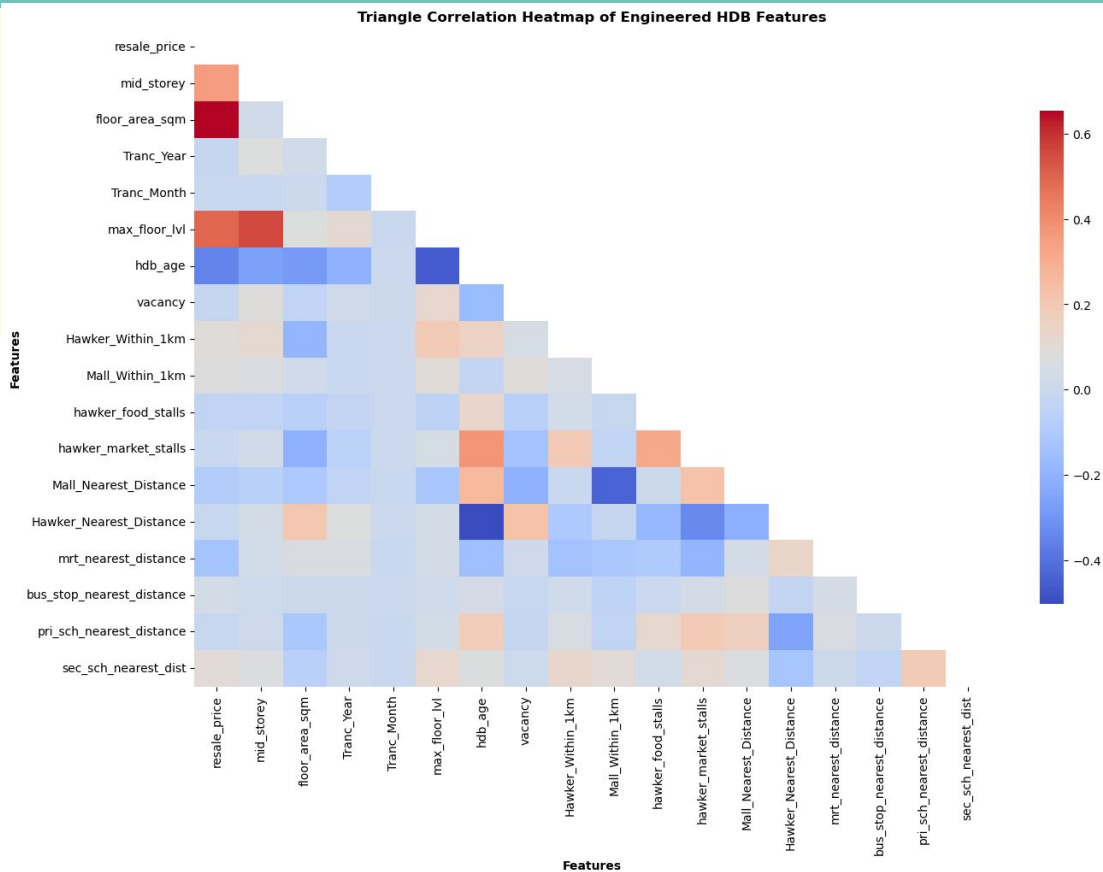
~0.05

~0.1

Correlation of Referenced Variables

Insight:

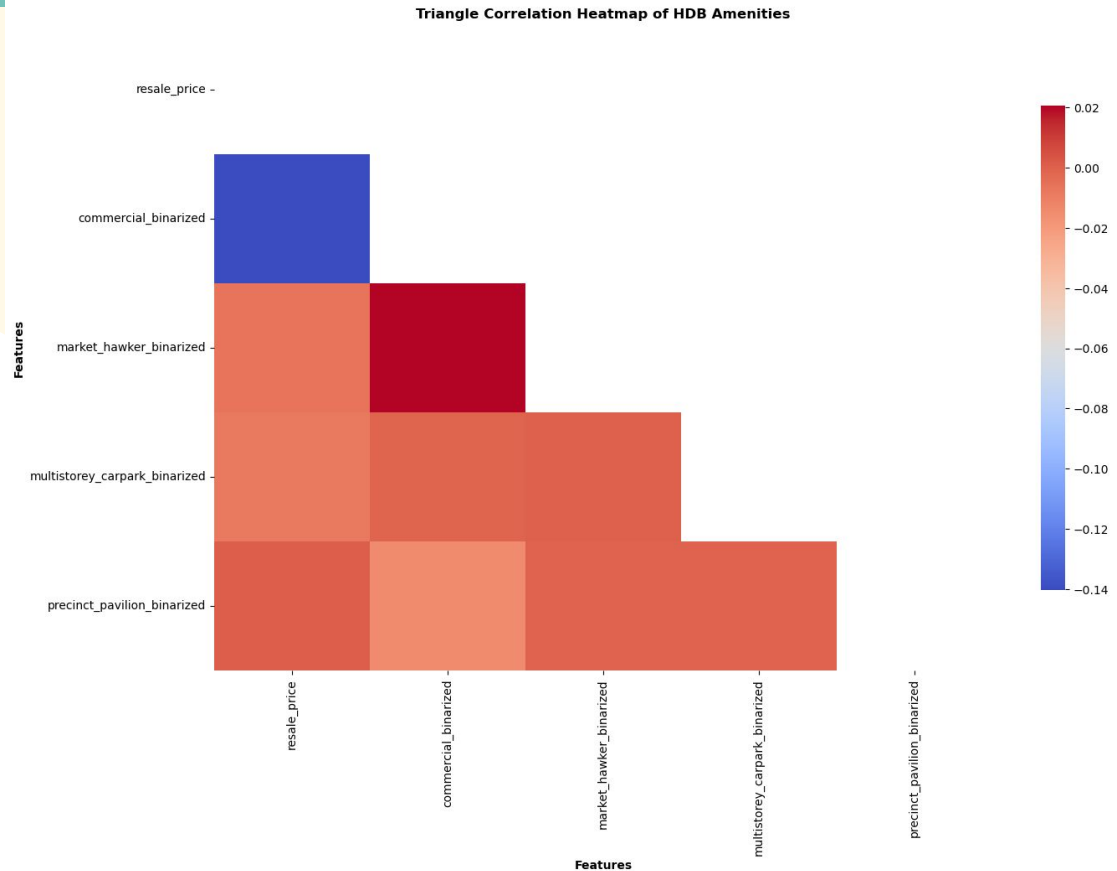
- Include HDB Block Height in modelling in modelling
- Group similar features to reveal stronger relationships



Correlation of Referenced Amenities

Observation:

- Individual Amenities do not contribute significantly to price
- Having access to 1 amenity does not raise prices





Additional Hypothesized Features

Primary School

Is proximity to a good school
a selling point?

Mature Estate

Do mature estates command
a higher price?

Town Region

Does the town region affect
demand?

HDB Model Type

What about these rarer types?

Floor Proportion

How high is their floor?

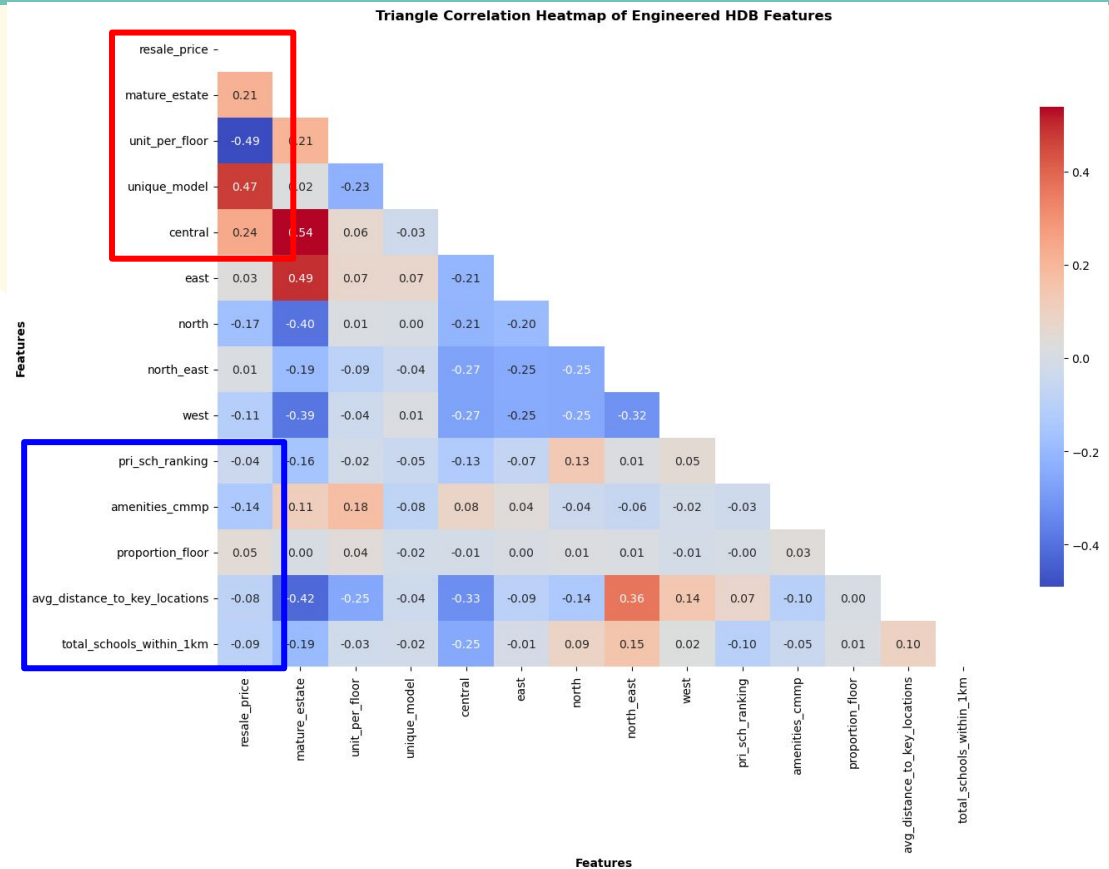
Corridor Length

Does it affect price?

Correlation of Hypothesized Features

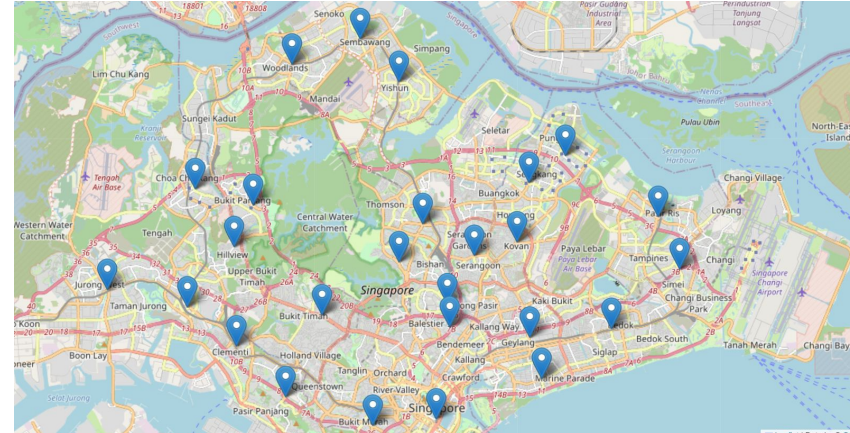
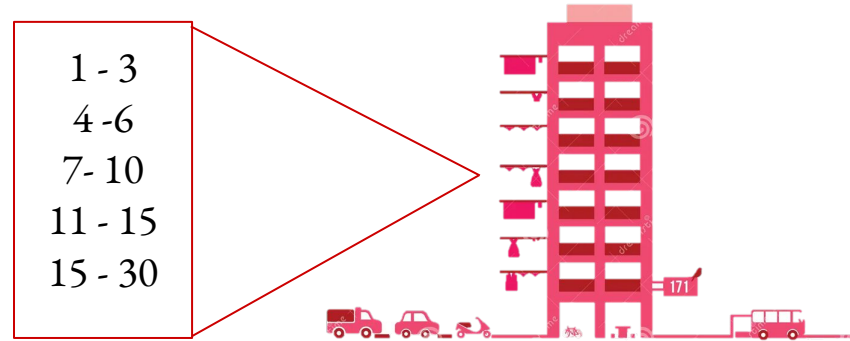
Insight:

- Many hidden relationships are significant
- e.g **Units per Floor**
- Certain variables that seemed obvious are not strong e.g **School Ranking**

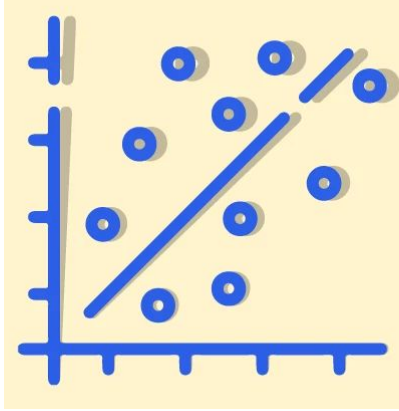


Total Number of Features

Features	Number
Numerical Features	26
Categorical Features	2
Total	28



Main Metrics for Selection of Features and Models



Feature Selection

R-Value

Strength of
Linear Relationship



Model Performance

R^2 -Value

Variance
Explainability

RMSE Value

Accuracy of
Prediction



Main Metrics for Selection of Features and Models

Linear Regression Model

Test Scores

R^2 -Value: 93.211%

RMSE: \$37,342

Ridge Regression Model

Test Scores

R^2 -Value: 93.208%

RMSE: \$38,349

Lasso Regression Model

Test Scores

R^2 -Value: 93.206%

RMSE: \$37,353



Main Metrics for Selection of Features and Models

Linear Regression Model

Test Scores

R^2 -Value: 93.211%

RMSE: \$37,342

Ridge Regression Model

Test Scores

R^2 -Value: 93.208%

RMSE: \$38,349

Lasso Regression Model

Test Scores

R^2 -Value: 93.206%

RMSE: \$37,353



Main Metrics for Selection of Features and Models

Linear Regression Model

Test Scores

R^2 -Value: 93.211%

RMSE: \$37,342

Ridge Regression Model

Test Scores

R^2 -Value: 93.208%

RMSE: \$38,349

Lasso Regression Model

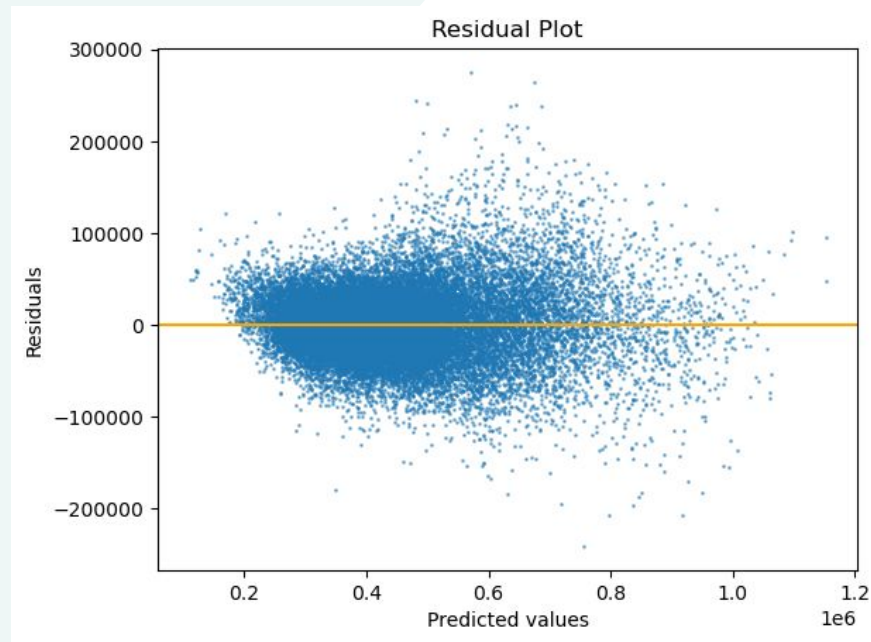
Test Scores

R^2 -Value: 93.206%

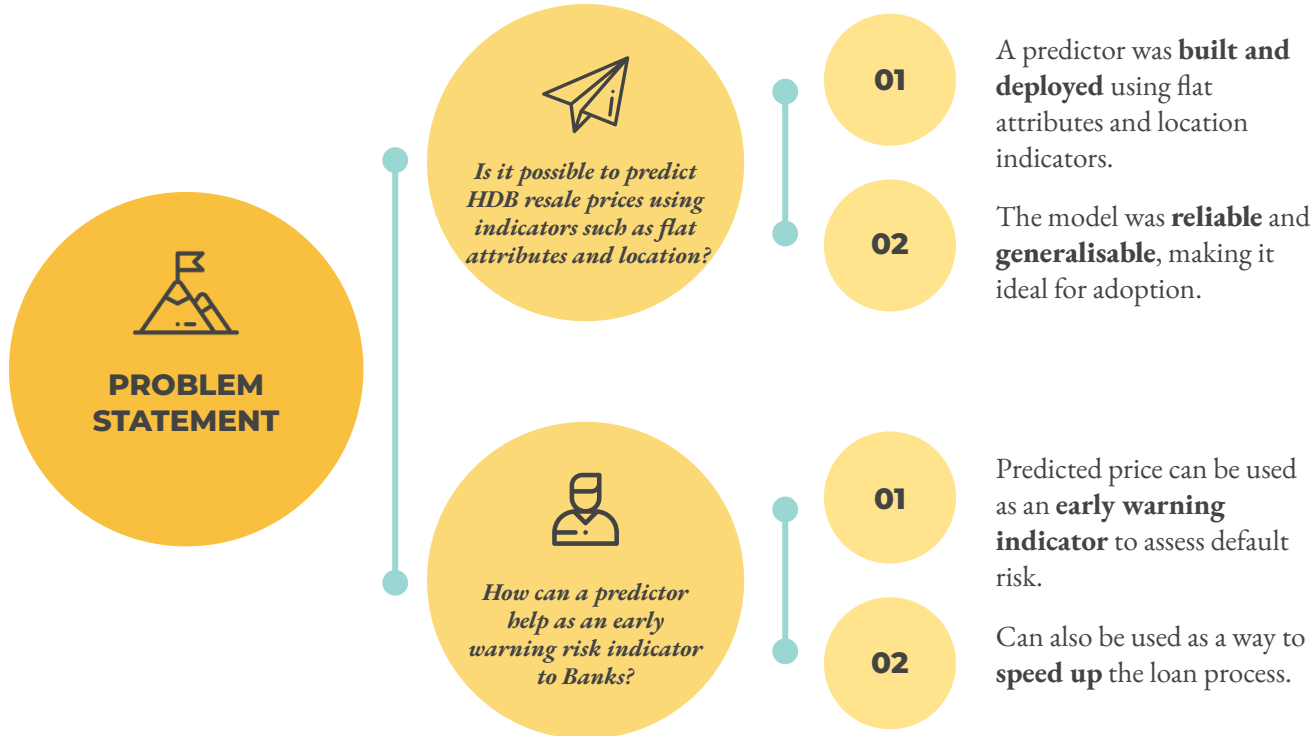
RMSE: \$37,353

Linear Model Residual Plot

```
count    30127.000000  
mean      -5.903119  
std       37342.479250  
min       -241464.828125  
25%       -22458.968750  
50%        -1128.328125  
75%        20821.945312  
max        275192.078125  
Name: resale_price, dtype: float64
```



Addressing the Problem



Model Demonstration

Creating a Metric to assess Loan Default Risk

Resale Price-to-Income (RPI) Ratio

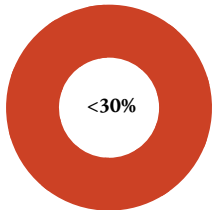
RPI Ratio = (Applicant's Annual
Income/Predicted Resale Price)
×100

Interpretation

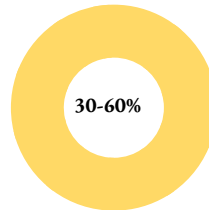
- The higher the RPI ratio, the less likely the applicant is to default on payments.
- Thresholds can be customised to suit the needs of the bank.

Example Thresholds

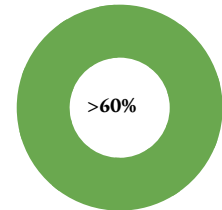
High Risk



Moderate Risk



Low Risk



Recommendations and Impact

PROFIT DRIVEN

1

The RPI Ratio can be used as an **early stage risk assessment** tool, thus **reducing Default Risk**.

2

Banks can **customise** their interest rates based on the RPI Ratio.

SERVICE DRIVEN

3

The model can be provided as a **white-label app** to banks to include in their **services portfolio**.

4

Using predicted resale price can **speed up** the loan process and **personalise** each loan.



Future Scope and Developments

- The model can also include prices for private properties (like condominiums) to give a more comprehensive outlook of SGs property market.
- The app can include a comparison of predicted price vs average price (say by region) to give an indication of heating/cooling of the market.
- The model can be adjusted for inflation to make predictions more realistic.

WHAT DO YOU MEAN YOU



"SKIPPED" LAST MONTH'S PAYMENT?