Resale Done Right – Predictive Analytics for Housing Decisions

---

Analytics and Business Report

Table of Content

1. Executive Summary **3**
2. Business Problem **4**
2.1 Prevalence of Unstable HDB Resale Prices 4
2.1.1 General Trend 4
2.1.2 Affordability Metrics 4
2.2 Significance of Fluctuating HDB Prices 4
2.2.1 Implications of Insufficient Information on House Buyers 4
2.2.2 Implications of Insufficient Information on House Sellers and Property Agents 5
2.3 Problem Statement 5
3. Datasets **7**
3.1 Data Preparation 7
3.1.1 Analysis of Dataset 7
3.1.2 Data Cleaning 7
3.1.3 Inflation Adjustment 7
3.2 Exploratory Data Analysis 8
3.2.1 Time Series Analysis of Price 8
3.2.2 Spatial Price Analysis 8
3.2.3 Distribution Analysis - Numerical 9
3.2.4 Distribution Analysis – Categorical 10
3.3 Examining Factors Affecting Price 11
3.3.1 Floor Area 11
3.3.2 Town Area, Flat Models and Storey Range 11
4. Our Models **13**
4.1 Market Demand Analysis 13
4.1.1 K-Means Clustering 13
4.1.2 PCA & Hierarchical Clustering 14
4.2 Price Prediction Models 15
4.2.1 Linear Regression model 15
4.2.2 CART Regression and Random Forest 16
4.2.3 XGBoost 18
5. Our Solutions **21**
6. Conclusion **24**
7. Bibliography **25**
8. Appendix **31**

1. Executive Summary

This project aims to address the volatility of resale prices in Singapore's Housing and Development Board (HDB) market, which may affect housing affordability and accessibility. By leveraging data analytics, we examined the factors influencing resale prices, including flat type, location, floor area, proximity to amenities (eg.MRT stations), and remaining lease years. Our goal is to create data-driven solutions that would enhance the decision-making process for house buyers, house sellers, and property agents, while improving transparency in the HDB resale market.

Key insights into market demand were obtained from models like K-means clustering and Principal Component Analysis (PCA), which revealed how the location of HDB flats may influence its price. Moreover, statistical models like Linear Regression, CART and XGBoost revealed that factors such as proximity to the Central Business District (CBD), floor area, and lease duration play significant roles in determining resale prices. For example, properties that are closer to the CBD with larger floor areas and longer remaining leases tend to command higher prices. Geographic patterns were also observed, with areas closer to amenities or central hubs displaying elevated demand and higher resale values.

Our predictive models were integrated into a web-based application, "Resale Done Right," which provides actionable insights tailored for different stakeholders. The application features personalised dashboards and interactive maps that highlight price trends, helping home buyers to locate affordable properties, sellers to set competitive prices, and agents to enhance advisory capabilities. Specifically, the tool’s clustering and price prediction functions enable buyers to benefit from simplified property searches based on budget and preferred location, sellers to gain insights into market demand for price optimization, and property agents to access a suite of tools for demand and price forecasting, ultimately enhancing client engagement and service efficiency.

By aligning property valuation with real-time data trends and predictive analytics, we hope to empower home buyers, sellers and agents to make smarter, more confident housing decisions and foster greater efficiency and transparency in the dynamic HDB resale market.

2. Business Problem

With a mission to "provide affordable homes of quality and value", the Singapore government and the Housing Development Board (HDB) have committed themselves to making home ownership accessible through strategic measures (Housing & Development Board, 2018) such as the Additional Buyer's and Seller's Stamp Duties, and financial assistance schemes like the Additional CPF Housing Grant (Channel News Asia, 2023). However, HDB resale prices remain subject to fluctuations due to a complex array of factors, ranging from location-specific demand and supply to broader economic shifts (Sit, 2024). This makes it challenging for buyers, sellers and property agents to assess fair market prices, highlighting the need for data-driven tools to navigate the market effectively.

2.1 Prevalence of Unstable HDB Resale Prices

2.1.1 General Trend

HDB resale prices have been showing a continuous upward trend for 10 consecutive quarters, increasing significantly by 10.4% in 2022 alone (Wong, 2024). Heartland and suburban towns like Sembawang and Woodlands have experienced relatively steeper increases in prices, with median resale prices in some areas rising over 40% since 2020 (Louisa, 2022).

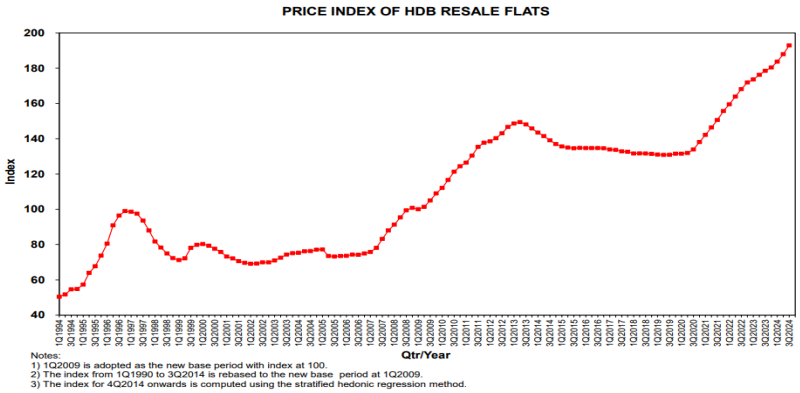


Figure 1: Trend of resale price index from 1994 to 2024 (HDB | Resale Statistics, n.d.)

2.1.2 Affordability Metrics

Fluctuating HDB resale prices in Singapore have significantly impacted affordability for prospective buyers (Wendell, 2022). According to the Demographia International Housing Affordability Report, the median price-to-income ratio for HDB resale flats in 2022 was 5.3, classifying it as "severely unaffordable" (see Appendix 2). To address this issue, government grants and subsidies were provided, reducing the ratio to 3.8 (see Appendix 2). Yet, housing remains "moderately unaffordable" by global standards, highlighting affordability concerns (Chong, 2024).

Therefore, it is imperative to develop statistical models that can reliably predict HDB resale prices. Such models will enable stakeholders to analyse price trends and demographic factors, allowing them to better anticipate market shifts and make informed decisions in the volatile HDB resale market.

2.2 Significance of Fluctuating HDB Prices

2.2.1 Implications of Insufficient Information on House Buyers

(a) Greater Financial Burden

Research showed that uninformed house buyers tend to pay more than informed buyers for the same property (Li & Chau, 2022), as they often lack insights into market conditions, property values, and neighbourhood trends due to high search costs (Turnbull & van der Vlist, 2022). Consequently, they may rely on their personal knowledge and emotions over objective market analysis when making property purchases (Bartkowiak et al., 2018).

Furthermore, limited understanding of market dynamics also weakens buyers' negotiation power, giving sellers and agents the opportunity to highlight appealing aspects while downplaying less favourable attributes (Cheung et al., 2021). This knowledge gap can lead buyers to overestimating and inflating property values, thereby increasing their financial burden (Martimort et al., 2021).

(b) Increased Financial Risk

Home buyers lacking market knowledge may face increased financial risks, as they may not fully understand mortgage obligations and long-term housing costs. Hence, they may overpay for properties, which risks financial strain if HDB resale values fall below the mortgage balance (Turnbull & van der Vlist, 2022). Such situations can limit their ability to sell or refinance without losses, impacting long-term financial flexibility and security.

2.2.2 Implications of Insufficient Information on House Sellers and Property Agents

(a) Inefficient Property Valuation

House sellers and agents traditionally estimate property values based on experience and nearby comparisons (Tanguturu et al., 2023). This inefficient, time-consuming approach requires manual data collection, delaying listings and reducing market competitiveness (Hiscock, n.d.).

(b) Inaccurate Estimation of Property Prices and Buyer Behaviour

The reliance on subjective judgement by sellers and agents also leads to inaccurate price and timing estimates due to limited information on housing markets (Li et al., 2022; Crosby, 2000). This may result in overpricing, which reduces competitiveness, or underpricing, which restricts profits. Thus, accurate price prediction is critical in property sales, especially for resale flats, where market dynamics can affect buyer behaviour and investment outcomes (Dhanorkar, 2024).

In addition, limited information also restricts sellers' and agents' ability to predict buyer preferences, reducing the effectiveness of property positioning and targeted marketing (Broxterman & Zhou, 2022). Without sufficient data, agents may be unable to forecast price drops and demand spikes, reducing their advisory capacity and causing missed opportunities for their clients (Li et al., 2022).

2.3 Problem Statement

As outlined in Sections 2.1 and 2.2, fluctuating HDB resale prices create uncertainty for buyers seeking affordability and sellers/agents aiming for optimal returns. Our project aims to address this by using AI and Machine Learning techniques to develop predictive models that offer the following benefits:

1. To empower homebuyers to identify affordable and fair housing options

2. To support home sellers in competitive price setting and targeted marketing to maximise profits
3. To develop comprehensive market tools that allow agents to provide tailored advice for their clients' specific needs

By implementing these predictive models, these stakeholders can forecast resale prices more accurately and gain a clearer view of the complex interplay of factors influencing resale price, placing them in a better position to make informed decisions.

3. Datasets

We leveraged on three key datasets specified below.

No.	Dataset	Description	Source	Rows	Columns	Time Period
1	HDB Pricing and Associated Variables	HDB resale transaction data with potential pricing factors	Kaggle	150,443	22	2017 - 2023
2	CPI Data	Yearly Consumer Price Index data	data.gov.sg	152	64	1961 - 2023
3	GeoJson Data	Map of Singapore with town boundaries	Kaggle	NA	NA	NA

Table 1: Table of datasets used and their descriptions

3.1 Data Preparation

3.1.1 Analysis of Dataset

Year	2017	2018	2019	2020	2021	2022	2023
Number of records	20509	21561	22186	23333	29087	26725	7042

Table 2: Number of records from 2017 - 2023

Although the record count for 2023 is lower due to incomplete data collection, the dataset remains valuable for analysis as HDB resale entries were consistently collected from the same areas (Appendix 3, Figure 25). Hence, this consistency makes Dataset 1 reliable for analysing HDB resale price trends over time.

3.1.2 Data Cleaning

To ensure reliable model implementation, we cleaned the dataset to remove missing values and inconsistencies, allowing us to better identify key variables in resale price prediction while reducing statistical noise.

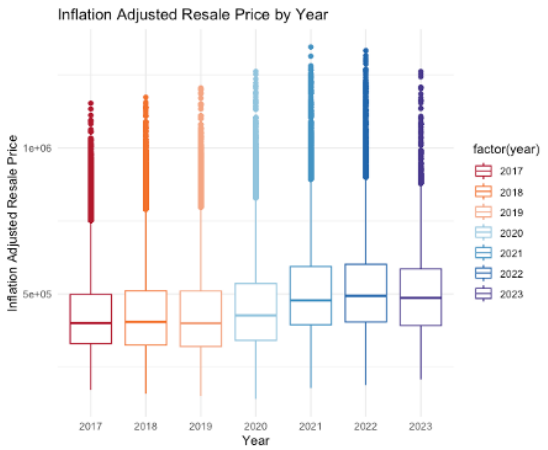
Through data exploration, 6 observations with missing values in the `resale_price` variable of Dataset 1 were found. Hence, to maintain data integrity, we decided to remove those rows, leaving us with a final dataset with 150,437 rows.

3.1.3 Inflation Adjustment

To accurately reflect the purchasing power of house buyers, HDB resale prices were adjusted for inflation by using the Consumer Price Index (CPI). The base year is set as 2019, as it reflects a period of economic stability and growth free of disruptions caused by COVID-19 (Ministry of Trade and Industry, 2020). By accounting for the cumulative effects of inflation and comparing real values of HDB resale prices instead of nominal values, it ensures that the analysis reflects the intrinsic changes in property values and not the impacts of inflation (Cornell et al., 2021).

3.2 Exploratory Data Analysis

3.2.1 Time Series Analysis of Price



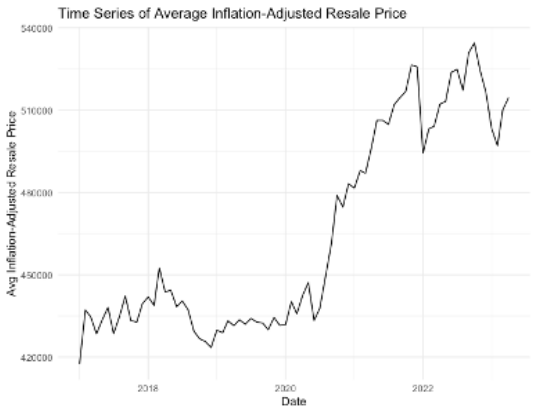


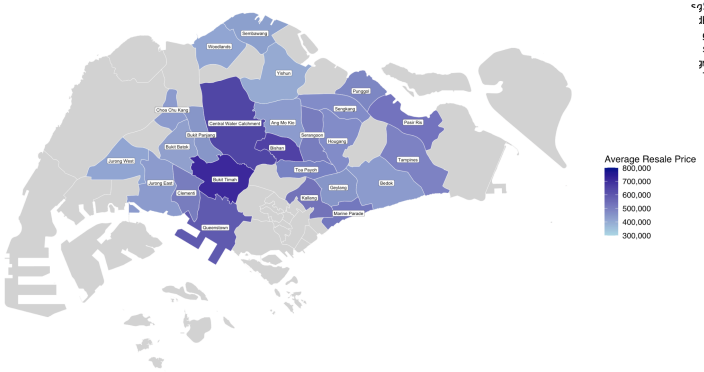
Figure 4:
Boxplots illustrating distribution of price 2017 - 2023 the months in 2018 - 2022

Figure 5:
Average HDB resale prices against

Figure 4 shows a significant price rise from 2019 to 2021 before stabilising from 2021 to 2023. The slight decline in 2023 is likely due to the implementations of cooling measures, increased BTO supply, and higher interest rates, resulting in a decreased demand for resale flats (Ng, 2024). The time series analysis in Figure 5 highlights the steepest price rise from 2020 to 2021, with median prices surging across HDB towns (Appendix 1). This spike can be attributed to pandemic-driven lifestyle shifts which redefined people's home needs, increasing the demand for larger homes. Furthermore, manpower shortages and supply chain disruptions due to the pandemic also extended BTO wait times, pushing buyers to resale markets (Tang, 2022). Lastly, low interest rates and rising private property prices also increased the appeal of resale flats, boosting its demand (Tan, 2021).

3.2.2 Spatial Price Analysis

For better visualisation of HDB resale prices, we mapped the average resale prices of flats in the different regions onto the map.




```
sg <- read_sf("/Users/fooyockhaw/Desktop/sg.geojson")
hdb <- read.csv("/Users/fooyockhaw/Desktop/hdb2017-2023.csv", stringsAsFactors = FALSE)
sg$planning_area <- as.factor(sg$planning_area)
 hb_avg_price <- hdb %>%
  group_by(town) %>%
  summarize(avg_resale_price = mean(inflation_adjusted_resale_price, na.rm = TRUE))
 merge <- sg %>%
  left_join(hdb_avg_price, by = c("planning_area" = "town"))
```

Figure 6: Spatially distributed data for Inflation Adjusted Resale Price of HDBs in Singapore

Each area is shaded in different tones of blue, indicating varying price levels, with darker shades representing higher average resale prices. This spatial distribution allows for a quick assessment of which regions are “hotspots” with higher resale values, and which are more affordable.

3.2.3 Distribution Analysis - Numerical

Next, we had a look at the distribution of our variables.

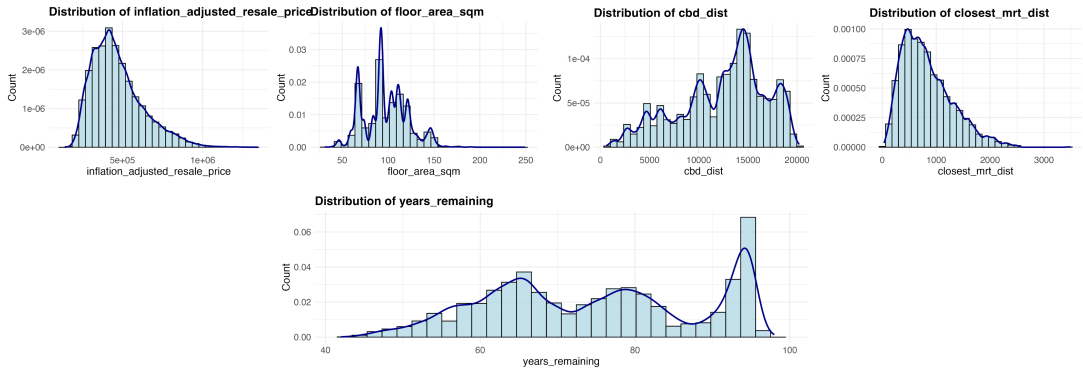


Figure 7: Histograms illustrating distribution of numerical variables

The distributions are mostly unimodal and not normal, suggesting that transformations may be needed for linear models to improve prediction reliability.

(a) Inflation-Adjusted Resale Price

The right-skewed distribution of inflation-adjusted resale prices suggests that most transactions are at lower prices, with fewer high-priced sales. This skewness indicates both affordable options and some high-value properties in the resale market.

(b) Floor Area (sqm)

The floor area distribution shows multiple peaks, likely reflecting different flat types (e.g., 3-room, 4-room, 5-room). This variable is key in assessing the impact of size on resale prices, as larger units generally fetch higher prices (Suen & Tang, 2002).

(c) Distance to Central Business District (CBD)

The distance to the CBD is multimodal, with various peaks indicating clusters of HDB properties at different distances from the city centre. This variable helps assess how proximity to the CBD influences pricing, as closer proximity is typically associated with higher demand (Fesselmeyer & Haoming, 2019).

(d) Distance to Closest MRT Station

The histogram of the distance to the nearest MRT station shows a right-skewed distribution, with most HDB flats located close to public transport. This is a significant variable, as easy access to MRT stations is often a desirable feature in urban areas (Fesselmeyer & Haoming, 2019).

(e) Years Remaining on Lease

The distribution is left-skewed, with most HDB flats having longer remaining leases, though some have fewer years left. Lease duration significantly influences resale prices, as properties with longer leases are generally more valuable.

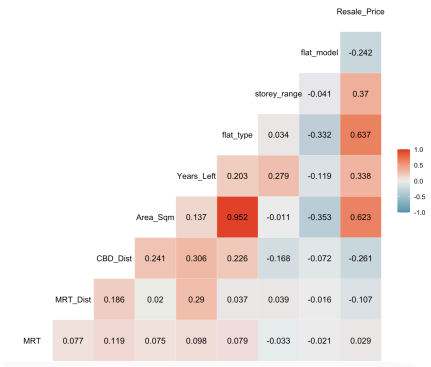


Figure 8: Heatmap showing relationship between variables and resale price

Figure 8 further supports insights (b) to (c), indicating that increasing distance from CBD and MRT are negative drivers of HDB resale value, while bigger floor area and longer lease remaining are positive drivers of HDB resale value. For further elaboration, please refer to Appendix 3.1.

3.2.4 Distribution Analysis – Categorical

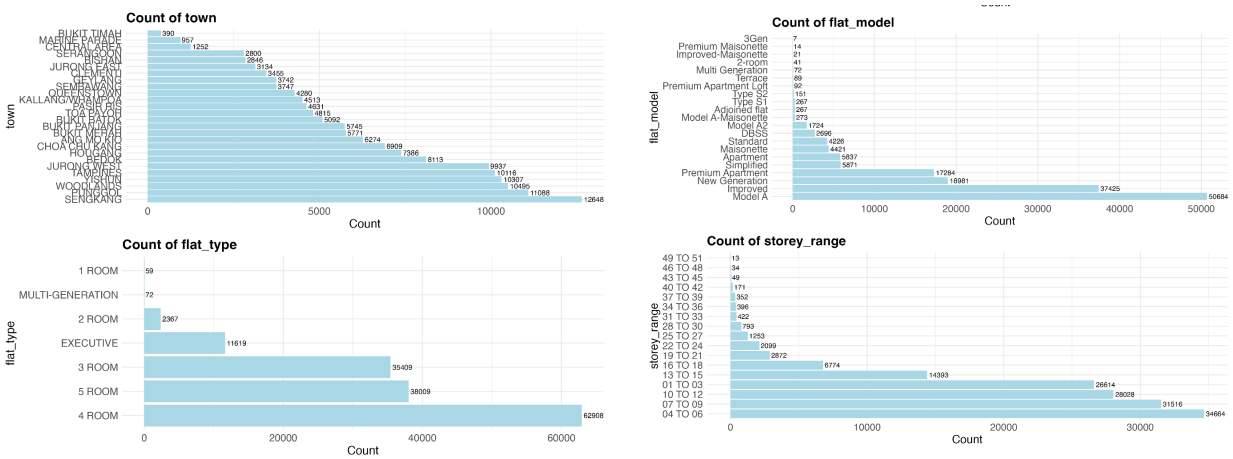


Figure 9: Barcharts illustrating distribution of categorical variables

Our categorical variables are also disproportionately distributed.

(a) Count of Towns: Higher counts in towns like Seng Kang and Punggol indicate more resale flats. This distribution is useful for understanding demand and supply trends across the region.

(b) Count of Flat Types: 4-room and 5-room flats are the most common, showing buyers' preference for mid-sized units, which can be useful when strategising pricing models.

(c) Count of Flat Models: Models like Standard and Model A are common, reflecting architectural diversity and older designs.

(d) Count of Storey Range: Lower floors are most common, with fewer counts for higher levels. This distribution may influence pricing as higher floors often command premium prices due to better views and reduced noise.

3.3 Examining Factors Affecting Price

3.3.1 Floor Area

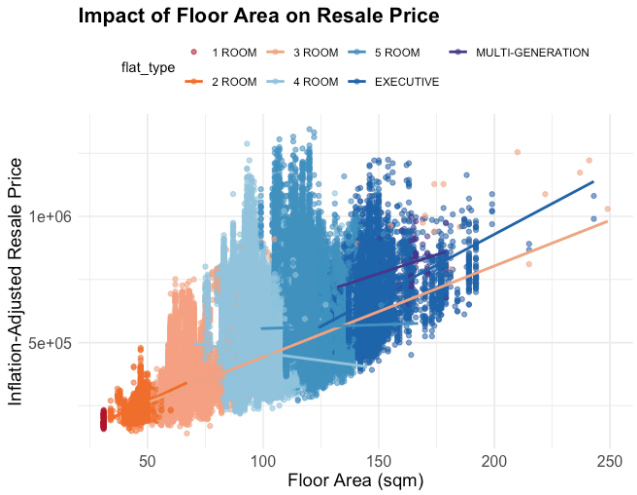


Figure 10: Relationship between floor area and resale price

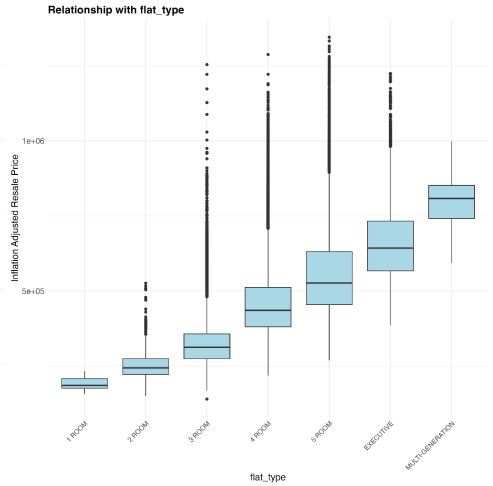


Figure 11: boxplot between flat type and resale price

From Figure 10 and 11, it can be seen that as floor area increases, resale prices also increase, with a distinct price range for each category of floor area. Flats with unique designs and larger floor areas, such as Executive flats and Multigenerational flats, also show higher prices, as compared to regular HDB flats. This highlights that floor area is a significant factor influencing both demand and price.

3.3.2 Town Area, Flat Models and Storey Range

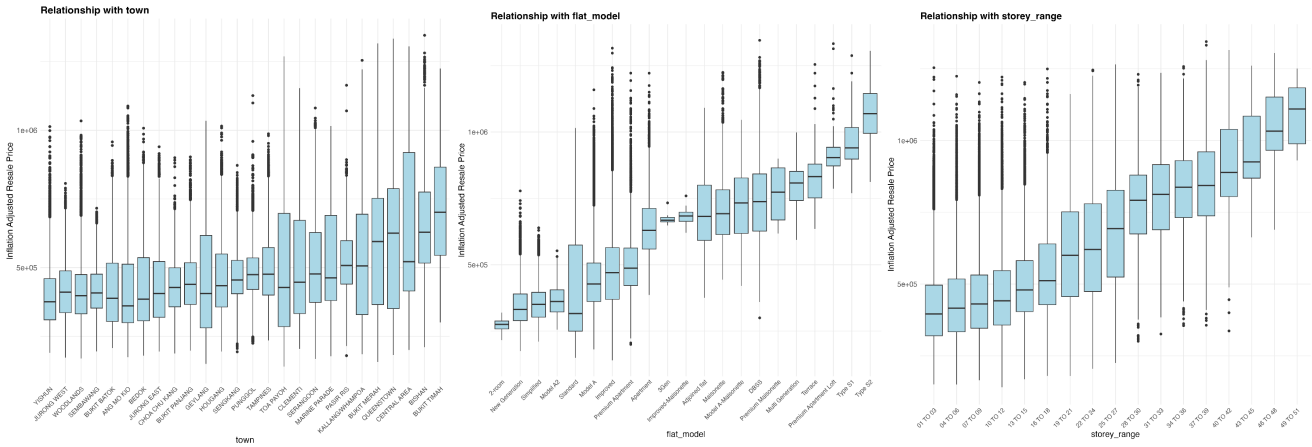


Figure 12: Boxplots illustrating relationship between town area, flat model, storey range and resale price

As shown in Figure 12 (left), HDBs located in areas like Bukit Timah and Bishan have the highest resale value while those in Yishun, Jurong West and Woodlands have the lowest resale value. Furthermore, it can be seen that certain flat models warrant higher resale prices, with an increasing trend for resale prices as the floors of units increase, suggesting that higher floors command higher prices.

4. Our Models

4.1 Market Demand Analysis

K-Means Clustering, Principal Component Analysis (PCA) and Hierarchical Clustering will be used to analyse the market demand for resale flats.

4.1.1 K-Means Clustering

Since HDB has attributed higher resale prices to higher demand, analysing regions with higher resale prices can provide us with insights on which regions experience a greater demand for resale flats (Yang, 2024). This was achieved through the creation of a graph comparing the within-cluster sum of squares (WCSS) for different numbers of clusters, whereby a lower WCSS indicates better clustering.

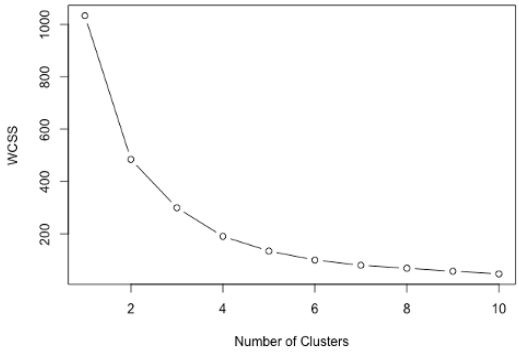


Figure 13: Graph of WCSS against number of clusters

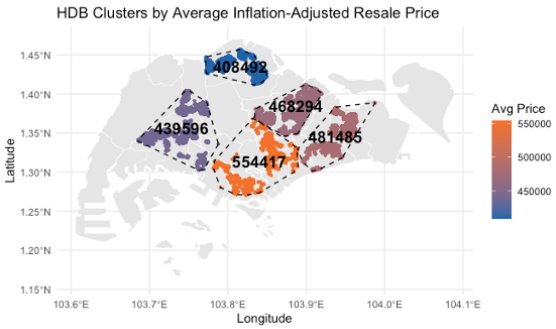


Figure 14: Average resale price grouped by region

We divided HDB areas into five clusters — Central, North, North-East, East, and West Regions— aligning with the Urban Redevelopment Authority's regional classification (Urban Redevelopment Authority, 2024). As seen in Figure 13, this clustering achieves a low WCSS, providing a balanced market view that highlights differences like proximity to amenities, the CBD, and public transport. This approach also aids in understanding neighbourhood dynamics, revealing trends in housing demand and the resale market.

Figure 14 shows the highest resale HDB prices in the Central Region, indicating the strongest demand, while the Northern Region has the lowest prices, suggesting weaker demand. This may stem from the Central area's proximity to employment hubs like the CBD, extensive public transportation, and amenities such as malls, hospitals, and schools. Limited housing supply and strong investment potential in the Central area further boost demand and prices (Noonan, 2020).

4.1.2 PCA & Hierarchical Clustering

PCA and Hierarchical Clustering aid resale market analysis by simplifying complex data to highlight influential factors across towns. PCA reduced dimensionality, capturing key variables in a few components, which reveals patterns across neighbourhoods.

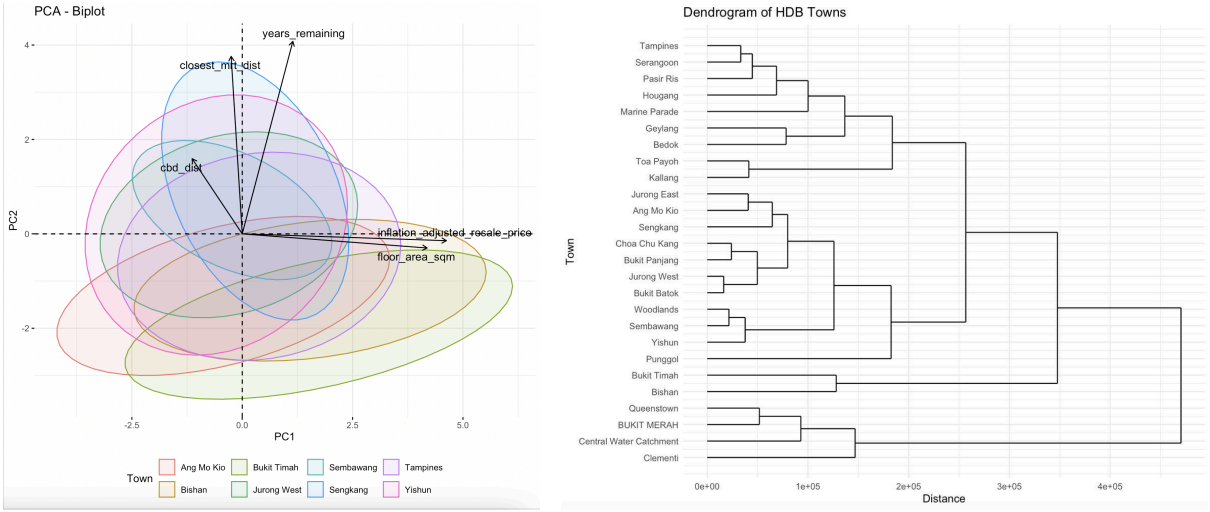


Figure 15: PCA Clustering analysis on left reveals differences between towns based on ellipses. Dendrogram on right groups towns based on their location and structural characteristics.

Figure 15 (left) shows substantial differences in buyer demand across towns like Bukit Timah and Sengkang. The first two principal components (PCs) accounted for 65% of the variance, consolidating related variables such as resale price and floor area.

The dendrogram (Figure 15, right) from hierarchical clustering groups towns by location and structural variables, with distances indicating differences—for instance, Clementi and Tampines are most dissimilar, while Bishan and Bukit Timah are closely related.

These visualisations help potential buyers identify neighbourhoods that match their preferences based on key HDB features in each town.

4.2 Price Prediction Models

4.2.1 Linear Regression model

(a) Rationale for Price Predictors

The model predicts HDB prices using factors like distance to the nearest MRT, proximity to the CBD, HDB town, flat type, storey range, floor area, flat model, and remaining lease. Variables were selected using the Akaike Information Criterion (AIC) to minimise prediction error.

The dataset was split into training and test sets, and diagnostic plots checked for assumption violations. Due to the right-skewed distribution of inflation-adjusted resale prices (Section 3.2.3(a)), a log transformation was applied to reduce variance and improve normality, enhancing the model's reliability across variable combinations.

(b) Analysis of Linear Regression Model

```
Response: log(inflation_adjusted_resale_price)
                    Df Sum Sq Mean Sq   F value    Pr(>F)
closest_mrt         95 2069.8   21.79  1515.858 < 2.2e-16 ***
closest_mrt_dist     1  109.2  109.20  7597.249 < 2.2e-16 ***
cbd_dist             1    0.2    0.18    12.562 0.0003937 ***
town                18  208.1   11.56   804.497 < 2.2e-16 ***
flat_type            6 5688.9  948.15 65966.285 < 2.2e-16 ***
storey_range        16  544.5   34.03  2367.552 < 2.2e-16 ***
floor_area_sqm       1  132.0  132.03  9185.720 < 2.2e-16 ***
flat_model          19  588.9   31.00  2156.586 < 2.2e-16 ***
years_remaining      1  556.5  556.47 38715.476 < 2.2e-16 ***
Residuals       105131 1511.1    0.01
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 16: Summary of linear regression model

Figure 16 shows that distance to amenities (e.g., MRT and CBD), HDB town, flat type, storey range, floor area, flat model, and remaining lease duration are significant predictors of resale price.

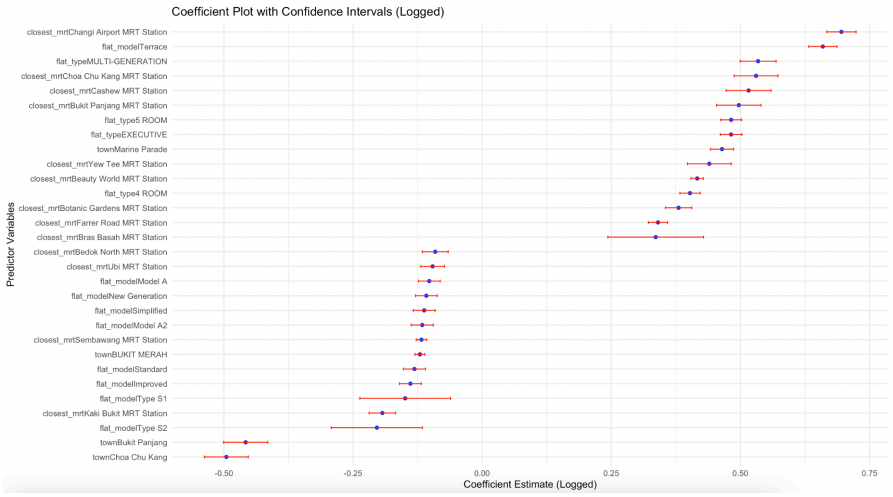


Figure 17: Significance of predictors of resale price

Some variables have a greater impact on resale price prediction than others. The coefficient plot shows that HDB town and flat model significantly influence resale price. For instance, Figure 17 highlights that flats near Changi Airport and terrace flats command a premium, whereas locations like Bukit Panjang and Choa Chu Kang have lower prices.

```
---------------------------------------------------------------------------
Observations                                          105,445
R2                                                      0.867
Adjusted R2                                             0.867
Residual Std. Error                           0.120 (df = 105286)
F Statistic                             4,342.527*** (df = 158; 105286)
===========================================================================
```

	Dataset	RMSE
1	Train	0.1197966
2	Test	0.1204272

Figure 18: Evaluation of linear regression model

Our linear regression model demonstrates strong predictive accuracy, achieving an R^2 of 0.867 and low RMSEs for both train and test sets. This indicates a good fit for predicting HDB resale prices. For buyers, the model highlights desirable attributes that contribute to long-term appreciation and higher resale value. Sellers on the other hand can pinpoint desirable attributes to showcase, helping them understand buyer preferences and find budget-friendly areas that still offer essential amenities.

(c) Limitations and Recommendations

Despite high prediction accuracy, some variance inequality persists, indicating imperfectly met assumptions for linear regression (Appendix 4.1). Thus, non-parametric models like CART and Random Forest may perform better with fewer assumptions, as discussed next.

4.2.2 CART Regression and Random Forest

Decision trees visualise the importance of factors influencing HDB resale prices. The same predictors from the linear regression model were used here, as they consistently demonstrate relevance across modelling approaches, confirming their importance in predicting HDB resale values.

(a) Analysis of CART Regression and Random Forest model

A key advantage of decision trees is their visual clarity in showing how each factor influences resale prices. In the tree plot below, floor area and CBD proximity are primary factors, with decision nodes highlighting splits in observations (Figure 19). This helps buyers and agents identify essential features and see the interaction of property attributes, aiding decision-making.

Although it does not predict the exact resale prices, the model provides a price range, offering buyers budgeting flexibility, sellers a competitive benchmark, and agents data-driven guidance.

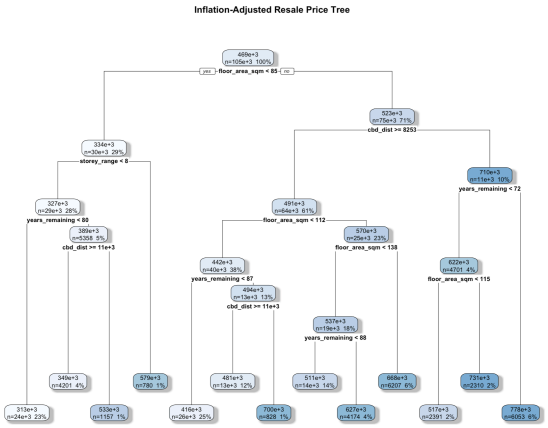


Figure 19 : Simplified CART decision tree

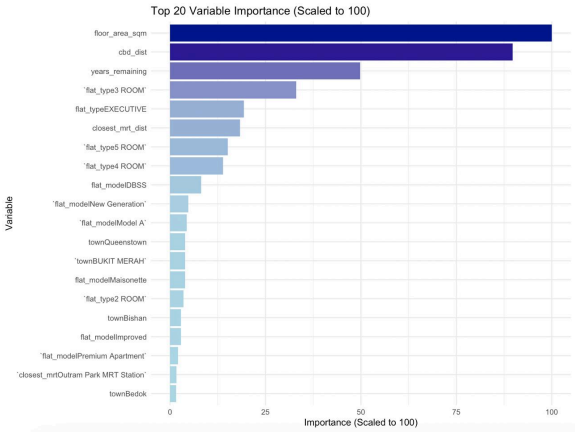


Figure 20: Top 20 important variables in predicting resale price

Notably, CART and Random Forests models achieved higher R^2 values than linear regression (Table 3), suggesting a better fit for predicting resale prices. These models handle mixed data types without assuming normality, supplementing linear regression by enhancing understanding of resale price factors and accommodating unique property features and market dynamics.

Model	Complexity	Trainset RMSE	Testset RMSE	Pseudo R^2 value
CART	5538 Terminal nodes	26770.09	37604.16	0.945
Random Forest	50 best predictors	34278.47	34443.82	0.954

Table 3: Summary of evaluation metrics for accuracy of CART model

(b) Limitations and Recommendations

From Table 3, the CART model displays a highly complex structure with 5,538 terminal nodes. This, along with a relatively lower RMSE for the training set compared to the test set RMSE, suggests potential overfitting. Such overfitting can mislead buyers, sellers, and property agents, leading to inaccurate price predictions that do not reflect actual market conditions, which can result in skewed property valuation. Additionally, tree models are computationally intensive, requiring more resources and time for training and predictions, which can hinder real-time price forecasting.

To improve the model, we can apply cross-validation for hyperparameter tuning to identify optimal parameters. This will help minimise overfitting while ensuring that the CART model maintains an appropriate level of complexity for both the training and test sets.

4.2.3 XGBoost

XGBoost uses gradient boosting to build a sequential ensemble of weak learners, with each new tree correcting errors from the previous one, improving overall performance. Our model included a comprehensive set of features to predict HDB resale prices (Table 4):

Location-Based Features:	Closest MRT station and distance
	Distance to the CBD
	Town

Property-Based Features:	Flat type
	Street name
	Storey range
	Floor area (sqm)
	Flat model
	Remaining lease
	Year of sale
	Years remaining

Table 4: Summary of Location and Property-based features used in XGBoost

The dataset was first divided into training and testing sets to train and evaluate the model's performance. The XGBoost model was then trained on the training set, with various hyperparameters optimised to minimise prediction errors.

(a) Analysis of XGBoost

(i) Price Segmentation by Location

This model focuses on geographic and property characteristics to segment HDB resale prices. Key features include geographic coordinates (latitude and longitude), distance to the nearest MRT station, distance to the CBD, floor area, flat type, storey range, flat model and remaining lease.

Metrics	Value
RMSE	48,481.52
R^2	0.9155

Table 5: Evaluation of location price segmentation in price prediction

The model achieved a R^2 score of 0.9155, demonstrating its ability to capture a substantial portion of the variance in resale prices. Moreover, its RMSE of 48,481.52 indicates that on average, the model's predictions are accurate to within approximately S$48,000 (Table 5).

(ii) Price Segmentation by Property-Based Features

Metrics	Value
RMSE	34,068.72
R^2	0.9583

Table 6: Evaluation of XGBoost in predicting resale prices

The model achieved an R^2 of 0.9583, indicating it explains a significant portion of variance in HDB resale prices, effectively capturing the relationship between selected features and prices. An RMSE of 34,068.72 suggests an average prediction accuracy within approximately S$34,000, an acceptable error margin given property values (Table 6). This strong predictive power provides buyers, sellers, and agents confidence in informed pricing strategies and market assessments.

In summary, XGBoost effectively predicts HDB resale prices by combining location- and property-based features, capturing market trends and patterns. The high R^2 and low RMSE confirm the model's reliability in providing accurate price estimates.

(b) Limitations and Recommendations

The model's performance relies heavily on the dataset's quality and scope. A more comprehensive dataset with additional variables reflecting varied market conditions could enhance accuracy. The model currently focuses on selected internal variables, potentially missing influential factors like economic conditions (e.g., unemployment rates, economic growth), housing policy changes, and global events.

To maintain accuracy and relevance, regular updates with new data are essential. Additional variables, such as proximity to schools, could be beneficial, as research shows that homes near reputable schools often have higher prices (Chew, 2016). However, these factors were limited by our dataset.

5. Our Solutions

Based on our models and insights gained from the housing market, we have come up with solutions that will cater to the specific needs of home buyers, sellers and property agents respectively.

5.1 “Resale Done Right” Web Application

The “Resale Done Right” web app is a comprehensive platform that meets diverse stakeholder needs, offering insights from price predictions to demand forecasting based on models in Section 4. Home buyers find affordable flats, while sellers and agents set competitive prices. The app allows sellers to use buyer trends and agents to access tools for pricing, market trends, and buyer sentiment, streamlining data management for efficiency (Shen et al., 2022). To meet data needs, Role-Based Access Control (RBAC) ensures security by restricting access based on roles (Lindemulder & Kosinski, 2024), with personalised dashboards showing relevant data (SafetyCulture, 2024). Details are in Sections 5.1.1–5.1.3.

5.1.1 User Interface for Home Buyers

As the primary concern of home buyers is to identify affordable housing options, their dashboards will contain insights generated from our models described in Section 4, and a thorough analysis of housing trends to facilitate the search process.

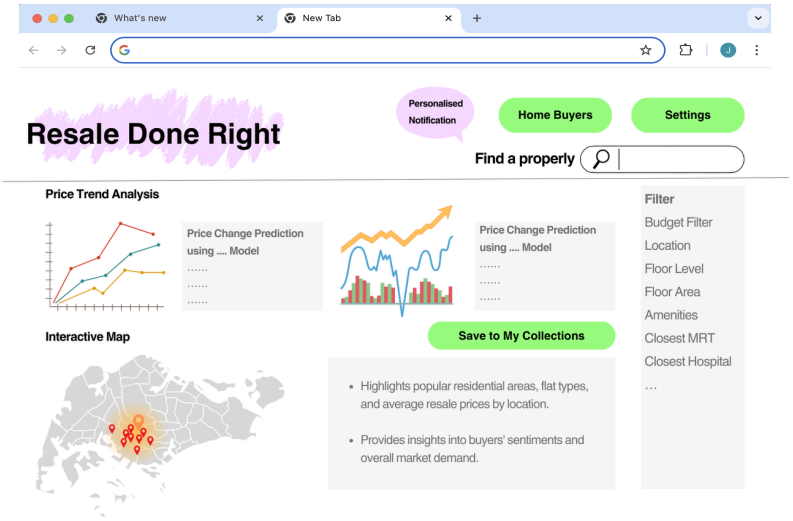


Figure 21: User interface for home buyers

The “Find a Property” search bar lets buyers filter listings by location, floor level, amenities, and budget, showing only relevant properties.

The dashboard provides price trend analysis with visuals powered by models like linear regression, CART, and random forest, displaying inflation-adjusted trends to help users anticipate price fluctuations. High R^2 values give buyers confidence in market insights (Gandhmal & Kumar, 2019).

Interactive maps highlight popular areas, flat types, and average prices, and personalised notifications alert users to new listings and price drops, streamlining the search process for a seamless, cost-effective experience.

These features streamline the property search process, providing buyers a seamless and cost-effective experience.

5.1.2 User Interface for Home Sellers

Home sellers aim to set competitive prices and understand market demand to maximize sale potential. Our dashboard includes specific features tailored to these needs, detailed below.

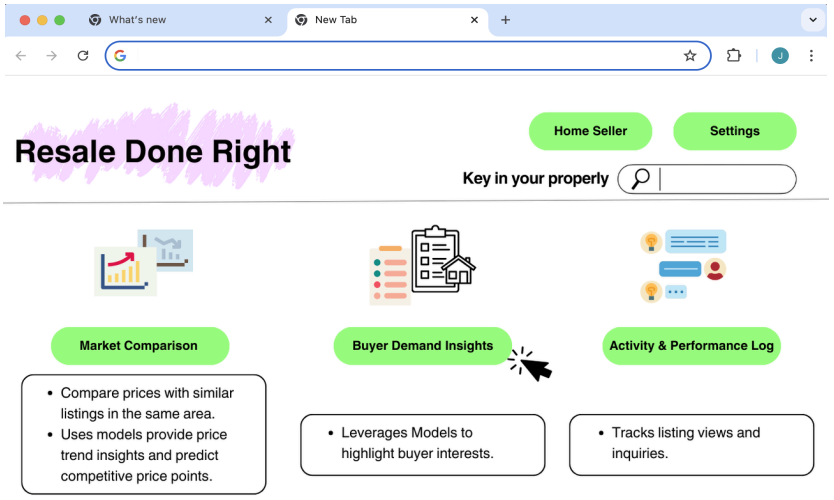


Figure 22: User interface for home sellers

Our application's market comparison tools, using statistical models, offer sellers insights for effective pricing. Sellers can compare their property's price to similar listings, with models like linear regression and random forest revealing price trends, while XGBoost enhances accuracy by analysing non-linear relationships.

The app also provides buyer demand insights using CART and K-means clustering, focusing on location, floor level, and amenities to help sellers align listings with market trends. An activity log tracks views and inquiries, supporting direct communication and increasing sales likelihood (Qualtrics, 2024).

5.1.3 User Interface for Property Agents

Property agents, serving both buyers and sellers, need comprehensive market data, pricing tools, and marketing support. Their dashboard will include key functionalities like property comparison tools, advanced price predictions, and buyer demand insights, as detailed in 5.1.2. Additionally, a Client Relationship Management (CRM) tool will allow agents to track client preferences, manage recommendations, and streamline follow-ups. This centralised interface enhances data access and simplifies client interactions, enabling agents to provide more efficient and personalised service.

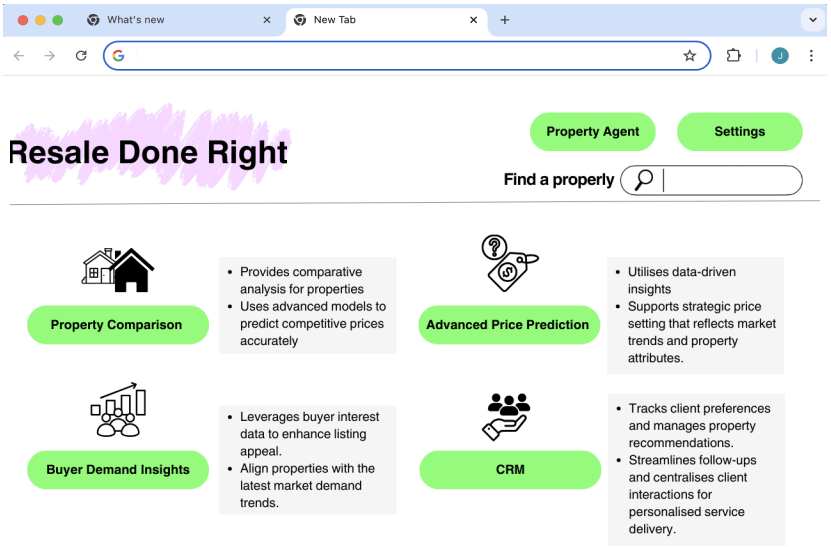


Figure 23: User interface for property agents

5.2 Education for Property Agents on Predictive Modelling

Another solution is to train property agents in basic modelling techniques and predictive analytics, enabling them to make data-driven decisions tailored to clients' needs. A structured program could teach foundational models like linear regression and CART, which are beginner-friendly and offer high accuracy in forecasting prices and capturing market trends (Morgan, 2014).

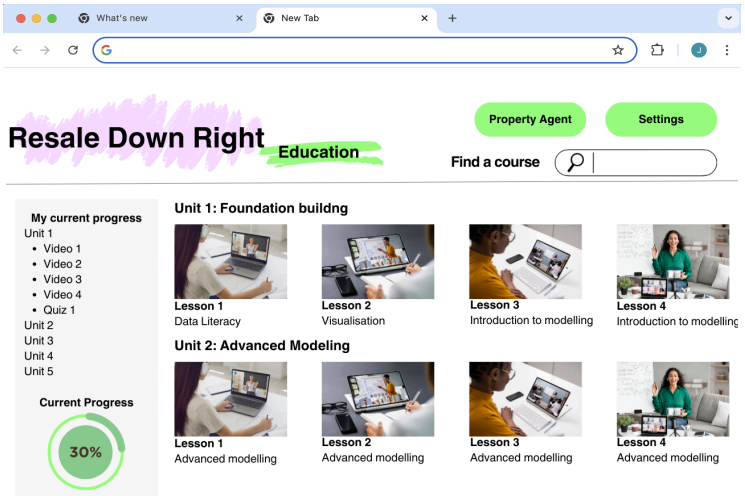


Figure 24: Predictive analytics education for property agents

The curriculum could begin with basic data literacy and visualisation techniques, followed by hands-on sessions where property agents learn to interpret model outputs, identify key variables, and apply insights to pricing and marketing. User-friendly platforms like R and Tableau can be introduced to ease learning and encourage real-world application.

Equipping agents with predictive analytics knowledge can reduce lead times, shorten sale cycles, and improve accuracy in matching properties to buyer preferences (Mathotaarachchi et al., 2024), promoting a more transparent resale market that benefits both sellers and buyers.

6. Conclusion

In conclusion, the development of predictive models for HDB resale prices marks a significant milestone in navigating the intricate and volatile resale market. By using predictive modelling techniques and analysing key factors influencing resale prices, this project aims to equip house buyers with insights needed to make informed purchase decisions, enable house sellers to price their properties competitively, empower property agents in offering more nuanced and data-driven advice to their clients. As such, this project not only promotes greater transparency within the HDB resale market, it also enhances the efficiency and effectiveness of the real estate ecosystem, fostering better outcomes for all stakeholders involved.

Together, we are paving the way for a smarter and more responsive housing market that meets the evolving needs of buyers, sellers and property agents alike.

7. Bibliography

Bartkowiak, P., Potrawiak, P., & Pavlenko, J. G. (2018). *Psychological factors affecting purchasing decisions on the real estate market.* Research Gate. https://www.researchgate.net/publication/332389864_Psychological_Factors_Affecting_Purchasing_Decisions_on_the_Real_Estate_Market

Broxterman, D., & Zhou, T. (2022). Information Frictions in Real Estate Markets: Recent Evidence and Issues. The Journal of Real Estate Finance and Economics. https://doi.org/10.1007/s11146-022-09918-9

Channel News Asia. (2023). *Property cooling measures: Additional buyer's stamp duty (ABSD) doubled for foreigners.* https://www.channelnewsasia.com/singapore/property-cooling-measures-absd-additional-buyers-stamp-duty-doubled-foreigners-increase-3446376

Cheung, K. S., Chan, J. T., Li, S., & Yiu, C. Y. (2021, August 27). *Anchoring and asymmetric information in the real estate market: A machine learning approach.* Journal of Risk and Financial Management, 14(9), 423. https://www.mdpi.com/1911-8074/14/9/423

Chew, C. H. (2016, May 13). Are properties near “famous” schools more expensive? PropertyGuru. https://www.propertyguru.com.sg/property-management-news/2016/5/124914/are-properties-near-famous-schools-more-expensive

Chong, X. W. (2024, October 2). *Singapore public housing rated as highly attainable, affordable: Reports.* Business Times. https://www.businesstimes.com.sg/property/singapore-public-housing-rated-highly-attainable-affordable-reports

Cornell, B., Gerger, R., Jarrell, G. A., & Canessa, J. L. (2021). Inflation, Investment and Valuation. *Journal of Business Valuation and Economic Loss Analysis*, 16(1), 1–13. https://doi.org/10.1515/jbvela-2020-0029

Cox, W., & Pavletich, H. (2022). *Demographia International Housing Affordability: 2022 Edition*. Demographia. Retrieved from https://www.newgeography.com/files/Demographia-International-Housing-Affordability-2022-Edition.pdf

Crosby, N. (2000). *Valuation accuracy, variation and bias in the context of standards and expectations*. *Journal of Property Investment & Finance, 18*(2), 130-161. ProQuest https://www.proquest.com/docview/212975582?parentSessionId=bX4dq2K5i9%2BA8paFghuu2b5ClUtqTRhknreWrWtCN4M%3D&pq-origsite=primo&sourcetype=Scholarly%20Journals

Dhanorkar, S. (2024, October 22). *How to sell your house: Resellers are having a hard time finding buyers; How homeowners can improve the saleability of their property*. *The Economic Times*.

https://economictimes.indiatimes.com/wealth/real-estate/how-to-sell-your-house-resellers-are-having-a-hard-time-finding-buyers-how-homeowners-can-improve-the-saleability-of-their-property/articleshow/114355679.cms?from=mdr

Fesselmeyer, E., & Haoming, L. (2019, April 9). Home Commuting Times Affect HDB Flat Prices [Review of Home Commuting Times Affect HDB Flat Prices]. The Straits Times; National University of Singapore (NUS). https://fass.nus.edu.sg/ecs/wp-content/uploads/sites/4/2020/11/COMMUTING-st-9apr-pA17.pdf

Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. Computer Science Review, 34(100190), 100190. https://doi.org/10.1016/j.cosrev.2019.08.001

HDB | Resale Statistics. (n.d.). https://www.hdb.gov.sg/residential/selling-a-flat/overview/resale-statistics

Hiscock, K. (n.d.). *What are delayed showings and negotiations in real estate?* Rochester Real Estate Blog. RE/MAX Realty Group. https://www.rochesterrealestateblog.com/delayed-showings-negotiations/

Housing & Development Board. (2018). *Vision, mission & shared values.* Housing & Development Board Singapore. https://assets.hdb.gov.sg/about-us/news-and-publications/annual-report/2018/vision-mission-shared-values.html

Housing & Development Board. (2024, August 19). *Measures to cool the HDB resale market and provide more support for first-time home buyers.* https://www.hdb.gov.sg/cs/infoweb/about-us/news-and-publications/press-releases/19082024-Measures-to-Cool-the-HDB-Resale-Market-and-Provide-More-Support-for-First-Time-Home-Buyers

Li, L., Bao, H. X. H., & Chau, K. W. (2022). On the Strategic Timing of Sales by Real Estate Developers: To Wait or To Presell? *The Journal of Real Estate Finance and Economics*. https://doi.org/10.1007/s11146-022-09894-0

Li, L., & Chau, K. W. (2023). Information Asymmetry with Heterogeneous Buyers and Sellers in the Housing Market. The Journal of Real Estate Finance and Economics. https://doi.org/10.1007/s11146-023-09939-y

Lindemulder, G., & Kosinski, M. (2024, August 20). RBAC. Ibm.com. https://www.ibm.com/think/topics/rbac

Martimort, D., Pouyet, J., & Trégouët, T. (2021, December 16). *Contracts as a barrier to entry: Impact of buyer's asymmetric information and bargaining power.* International Journal of Industrial Organization, 79, 102791. https://hal.science/hal-03328387/document

Mathotaarachchi, K. V., Hasan, R., & Mahmood, S. (2024). Advanced Machine Learning Techniques for Predictive Modeling of Property Prices. *Information*, 15(6), 295. https://doi.org/10.3390/info15060295

Ministry of Trade and Industry, Singapore. (2020, February). Economic Survey of Singapore 2019. https://www.mti.gov.sg/-/media/MTI/Resources/Economic-Survey-of-Singapore/2019/Economic-Survey-of-Singapore-2019/FullReport_AES2019.pdf

Morgan, J. (2014). Classification and Regression Tree Analysis. https://www.bu.edu/sph/files/2014/05/MorganCART.pdf

Ng, M. (2024, January 26). HDB resale prices up 4.9% in 2023, lowest increase since 0.1% rise in 2019. *The Straits Times*. https://www.straitstimes.com/singapore/housing/hdb-resale-prices-rise-49-in-2023-lowest-increase-since-01-rise-in-2019

Ng, T. (2023). Singapore District - Planning Area Geojson. Kaggle.com. https://www.kaggle.com/datasets/shymammoth/singapore-district-planning-area-geojson

Noonan, A. (2020, September 18). What really drove Singaporean real estate prices down during Covid-19?. AsiaOne. https://www.asiaone.com/money/what-really-drove-singaporean-real-estate-prices-down-during-covid-19

Ohmyhome (2024, June 3). *Analyzing HDB resale prices: Trends & insights.* https://ohmyhome.com/en-sg/blog/analyzing-hdb-resale-prices-trends-insights/

Qualtrics. (2024). The importance of customer relationship building. Qualtrics. https://www.qualtrics.com/experience-management/customer/customer-relationship-building/

SafetyCulture. (2024, May 1). Create Analytics dashboards. https://help.safetyculture.com/en-US/001997/

Shen, A., Wang, F., Paul, S., Divya Bhuvanapalli, Alayof, J., Farris, A. B., Teodoro, G., Brat, D. J., & Kong, J. (2022). An integrative web-based software tool for multi-dimensional pathology whole-slide image analytics. Physics in Medicine and Biology, 67(22), 224001–224001. https://doi.org/10.1088/1361-6560/ac8fde

Singapore Department of Statistics. Consumer Price Index (CPI). Data.gov.sg; https://data.gov.sg/datasets?query=cpi&page=1&resultId=d_dcb352661fb449c4a4c0ab23aa8d6399

Singapore Government [data.gov.sg] (2024). Consumer Price Index (CPI), 2019 As Base Year, Annual. Retrieved from https://data.gov.sg/datasets/d_dcb352661fb449c4a4c0ab23aa8d6399/view

Sit, D. (2024, January 25). *2024 HDB outlook.* Let's Talk Property. Retrieved from https://letstalkproperty.sg/2024-hdb-outlook/

Suen, W., & Tang, B. (2002). Optimal site area for high-density housing development. Habitat International, 26(4), 539–552. https://doi.org/10.1016/s0197-3975(02)00018-8

Tan, A. (2021, September 1). *Top 5 reasons behind the rise in HDB resale prices*. Redbrick Mortgage Advisory. https://www.redbrick.sg/blog/top-5-reasons-behind-the-rise-in-hdb-resale-prices/

Tang, A. (2023). Singapore Home Resale Prices (HDB). Kaggle.com. https://www.kaggle.com/datasets/talietzin/singapore-hdb-resale-prices-1990-to-2023

Tang, L. (2022, November 6). *How HDB resale prices have risen in every Singapore town in the current property boom.* Channel News Asia. https://www.channelnewsasia.com/singapore/hdb-resale-prices-every-singapore-town-current-property-boom-3043466

Tanguturu, A., Gayathri, C. D. S. L., Singh, A., Gnanacharithra, G., & Mohammed Zabeeulla, A. N. (2023, June 2). *Real estate price prediction using machine learning*. *International Journal of Current Science*, 13(2), 224-237. https://rjpn.org/ijcspub/papers/IJCSP23B1386.pdf

Turnbull, G. K., & van der Vlist, A. J. (2022). *The price of ignorance: Foreclosures, uninformed buyers and house prices.* Journal of Housing Economics, 57, 101844. https://www-sciencedirect-com.remotexs.ntu.edu.sg/science/article/pii/S1051137722000183

Urban Redevelopment Authority. (2024). Urban Redevelopment Authority. https://www.ura.gov.sg/Corporate/Planning/Master-Plan/Master-Plan-2019/Regional-Highlights

Wendell Cox (2022). *Demographia international housing affordability: 2022 edition.* https://www.newgeography.com/files/Demographia-International-Housing-Affordability-2022-Edition.pdf

Wong, Y. (2024, August 21). *Desmond Lee flags psychological factors contributing to higher prices in the HDB resale market.* The Straits Times. https://www.straitstimes.com/singapore/desmond-lee-flags-psychological-factor-contributing-to-higher-prices-in-hdb-resale-market

Yang, W. (2024, October). HDB resale prices up 2.5% in Q3, with more flats sold. The Straits Times. https://www.straitstimes.com/singapore/housing/hdb-resale-prices-up-25-in-q3-with-more-flats-sold

8. Appendix

Appendix 1: Percentage Increases in Median Resale Prices

HDB MEDIAN RESALE PRICES Q1 2020 TO Q3 2022, 5-ROOM

HDB town	Q1 2020	Q3 2022	Difference	% difference
Choa Chu Kang	S$420,000	S$588,000	S$168,000	40.0%
Bukit Batok	S$518,500	S$720,000	S$201,500	38.9%
Sembawang	S$408,000	S$566,000	S$158,000	38.7%
Woodlands	S$405,000	S$560,000	S$155,000	38.3%
Ang Mo Kio	S$590,400	S$800,000	S$209,600	35.5%
Jurong West	S$429,400	S$565,000	S$135,600	31.6%
Jurong East	S$475,000	S$619,000	S$144,000	30.3%
Hougang	S$511,500	S$648,400	S$136,900	26.8%
Toa Payoh	S$688,000	S$870,000	S$182,000	26.5%
Yishun	S$482,500	S$608,000	S$125,500	26.0%
Pasir Ris	S$495,000	S$619,000	S$124,000	25.1%
Punggol	S$493,000	S$612,000	S$119,000	24.1%
Sengkang	S$480,000	S$595,400	S$115,400	24.0%
Bedok	S$550,000	S$680,000	S$130,000	23.6%
Bishan	S$694,000	S$855,000	S$161,000	23.2%
Tampines	S$535,000	S$650,000	S$115,000	21.5%
Kallang/Whampoa	S$698,700	S$840,000	S$141,300	20.2%
Bukit Panjang	S$520,000	S$610,000	S$90,000	17.3%
Bukit Merah	S$817,500	S$875,000	S$57,500	7.0%
Queenstown	S$852,500	S$894,000	S$41,500	4.9%
Serangoon	*	*	NA	NA
Marine Parade	*	*	NA	NA
Geylang	*	S$738,900	NA	NA
Clementi	S$815,000	*	NA	NA
Central	*	*	NA	NA
Bukit Timah	*	*	NA	NA

* Fewer than 20 resale transactions in that quarter. The median prices are not shown as they may not be representative.
Figures are rounded to the nearest hundred dollars.

Infographic: *Rafa Estrada* ***Source:*** *HDB*

cna

Table of percentage increase in median resale prices of 5-room flats from 2020 - 2022 (https://www.channelnewsasia.com/singapore/hdb-resale-prices-every-singapore-town-current-property-boom-3043466)

HDB MEDIAN RESALE PRICES Q1 2020 TO Q3 2022, 4-ROOM

HDB town	Q1 2020	Q3 2022	Difference	% difference
Kallang/Whampoa	S$500,000	S$778,400	S$278,400	55.7%
Sembawang	S$351,000	S$515,000	S$164,000	46.7%
Choa Chu Kang	S$339,000	S$493,900	S$154,900	45.7%
Bukit Batok	S$350,000	S$500,000	S$150,000	42.9%
Woodlands	S$340,000	S$468,000	S$128,000	37.6%
Toa Payoh	S$539,000	S$740,000	S$201,000	37.3%
Jurong West	S$360,000	S$475,000	S$115,000	31.9%
Yishun	S$360,000	S$473,000	S$113,000	31.4%
Ang Mo Kio	S$397,500	S$516,500	S$119,000	29.9%
Hougang	S$400,000	S$506,000	S$106,000	26.5%
Sengkang	S$425,000	S$536,000	S$111,000	26.1%
Tampines	S$422,500	S$530,000	S$107,500	25.4%
Geylang	S$450,000	S$560,000	S$110,000	24.4%
Bishan	S$521,500	S$640,000	S$118,500	22.7%
Jurong East	S$375,000	S$460,000	S$85,000	22.7%
Bedok	S$387,300	S$475,000	S$87,700	22.6%
Serangoon	S$429,000	S$520,900	S$91,900	21.4%
Punggol	S$460,000	S$555,000	S$95,000	20.7%
Bukit Merah	S$640,000	S$765,000	S$125,000	19.5%
Pasir Ris	S$435,500	S$519,000	S$83,500	19.2%
Queenstown	S$728,000	S$860,000	S$132,000	18.1%
Bukit Panjang	S$411,000	S$471,900	S$60,900	14.8%
Clementi	S$644,000	S$555,000	-S$89,000	-13.8%
Bukit Timah	*	*	NA	NA
Central	*	S$680,000	NA	NA
Marine Parade	*	*	NA	NA

* Fewer than 20 resale transactions in that quarter. The median prices are not shown as they may not be representative.
Figures are rounded to the nearest hundred dollars.

Infographic: *Rafa Estrada* ***Source:*** *HDB*

cna

Table of percentage increase in median resale prices of 4-room flats from 2020 - 2022 (https://www.channelnewsasia.com/singapore/hdb-resale-prices-every-singapore-town-current-property-boom-3043466)

Appendix 2: Classification and Evaluation of Housing Affordability

Demographia International Housing Affordability uses a price-to-income ratio to rate housing affordability. The ratio is calculated by taking the median house price divided by the gross median household income (pre-tax). Housing affordability is then rated in four categories, ranging from the most affordable ("Affordable") to the least affordable ("Impossibly Unaffordable").

DEMOGRAPHIA HOUSING AFFORDABILITY RATINGS

Housing Affordability Rating	Median Multiple
Affordable	3.0 & Under
Moderately Unaffordable	3.1 to 4.0
Seriously Unaffordable	4.1 to 5.0
Severely Unaffordable	5.1 & 8.9
Impossibly Unaffordable	9.0 & Over

Median multiple: Median house price divided by median household income

Table of Classification of Housing Affordability Ratings (https://www.newgeography.com/files/Demographia-International-Housing-Affordability-2022-Edition.pdf)

Housing Affordability Ratings by Nation: Totals by Market						
Nation	Affordable (3.0 &Under)	Moderately Unaffordable (3.1-4.0)	Seriously Unaffordable (4.1-5.0)	Severely Unaffordable (5.1 &Over)	Total	Median by Nation
Australia	0	0	0	5	5	8.0
Canada	0	2	0	4	6	6.0
China: Hong Kong	0	0	0	1	1	23.2
Ireland	0	0	0	1	1	5.7
New Zealand	0	0	0	1	1	11.2
Singapore	0	0	0	1	1	5.8
United Kingdom	0	1	9	11	21	5.1
United States	1	9	19	27	56	5.0
TOTAL	1	12	28	51	92	5.2

Figure 2: Comparison of housing affordability in different countries in 2022 (Cox & Pavletich, 2022)

Housing Affordability Ratings by Nation: Totals by Market							
Nation	Affordable (3.0 &Under)	Moderately Unaffordable (3.1-4.0)	Seriously Unaffordable (4.1-5.0)	Severely Unaffordable (5.1 - 8.9)	Impossibly Unaffordable (9.0 &Over)	Total	Median Market
Australia	0	0	0	2	3	5	9.7
Canada	0	1	1	2	2	6	5.6
China: Hong Kong	0	0	0	0	1	1	16.7
Ireland	0	0	1	0	0	1	4.8
New Zealand	0	0	0	1	0	1	8.2
Singapore	0	1	0	0	0	1	3.8
United Kingdom	0	2	12	9	0	23	5.0
United States	0	11	23	17	5	56	4.8
TOTAL	0	15	37	31	11	94	5.0

Figure 3: Comparison of housing affordability in different countries in 2024 (Cox & Pavletich, 2022)

Data Dictionary

HDB Pricing and Associated Variables

Variable	Data Type	Description	Example Values
address	Character	Full address of the HDB property.	"123 Clementi Ave 3"
closest_mrt	Character	Name of the nearest MRT station.	"Clementi MRT"
month	Character	Month and year of the resale transaction in the format YYYY-MM.	"2022-05"
town	Character	Town or estate in Singapore where the property is located.	"Ang Mo Kio"
flat_type	Character	Type of HDB flat (e.g., number of rooms).	"5 ROOM"
block	Character	Block number of the HDB property.	"123"
street_name	Character	Street name of the HDB property.	"Clementi Ave 3"
storey_range	Character	Range of storeys where the unit is located (e.g., "01 TO 03").	"10 TO 12"
flat_model	Character	HDB flat model type (e.g., "Model A", "Improved").	"Improved"

remaining_lease	Character	Remaining lease years for the property, including months.	"73 years 5 months"
latitude	Numeric	Geographic latitude of the property.	1.3521
longitude	Numeric	Geographic longitude of the property.	103.8198
postal_code	Numeric	Singapore postal code of the property.	560124
closest_mrt_dist	Numeric	Distance to the nearest MRT station in meters.	500
cbd_dist	Numeric	Distance to the Central Business District (CBD) in meters.	12000
floor_area_sqm	Numeric	Floor area of the HDB flat in square meters.	95
lease_commence_date	Numeric	Year when the lease commenced.	1995
resale_price	Numeric	Resale price of the HDB flat in Singapore dollars.	480000
year	Numeric	Year of the resale transaction.	2022
years_remaining	Numeric	Remaining lease years as of the resale year.	74
adjustment_factor	Numeric	Adjustment factor to account for inflation or other economic factors affecting prices.	1.02
inflation_adjusted_resale_price	Numeric	Resale price adjusted for inflation, providing a standardized price.	468978

Appendix 3: Data exploration findings

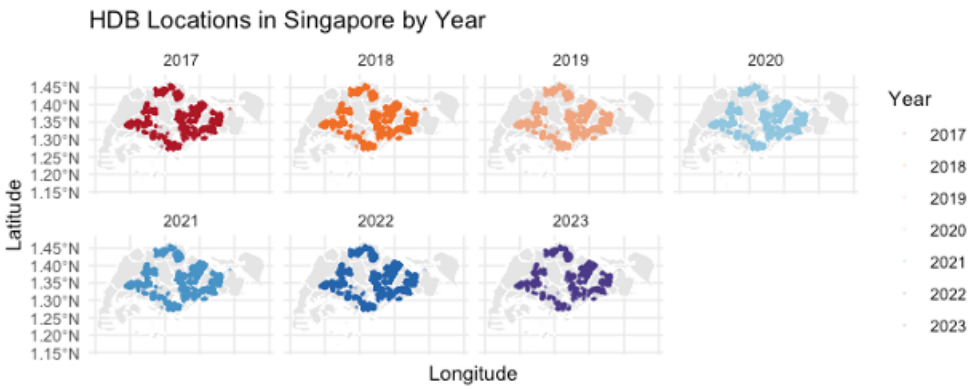


Figure 25: Distribution of data collection across Singapore

Appendix 3.1 Elaboration on Figure 8

This variable has a strong positive correlation with floor_area_sqm (0.62), indicating that larger floor areas are associated with higher resale prices.

It also shows a moderate positive correlation with years_remaining (0.34), meaning properties with more years left on the lease tend to have higher resale values.
Floor Area (sqm):

Floor area has a noticeable correlation with inflation_adjusted_resale_price (0.62), reinforcing that larger units command higher prices.

It has a weak positive correlation with years_remaining (0.24), which may indicate that newer, larger flats are available in the market.

Distance to CBD (cbd_dist):

This variable has a weak negative correlation with inflation_adjusted_resale_price (-0.26) and floor_area_sqm (-0.26). This suggests that properties closer to the Central Business District (CBD) are generally more expensive and may also be smaller in size.

Distance to Closest MRT Station (closest_mrt_dist):

There's a weak negative correlation with inflation_adjusted_resale_price (-0.11), indicating that closer proximity to MRT stations has a minor association with higher resale prices. While not strong, proximity to transport is still a factor buyers may value.
Years Remaining:

This variable has a moderate positive correlation with inflation_adjusted_resale_price (0.34), indicating that properties with more remaining lease years tend to have higher resale prices. This makes sense as flats with shorter leases are typically less desirable due to lease decay.

Factors	**Variables Used**
Location	Distance to nearest MRT (`closest_mrt_dist`), CBD proximity (`cbd_dist`) and town (`town`)
Structural features	flat type(` flat_type`), storey range (`storey_range`), floor area (`floor_area_sqm`), flat model (`flat_model`) and remaining lease duration (`remaining_lease`)

Table 5: Summary of variables used in price prediction using linear regression

Appendix 3.2: Resale Price by Flat Type, model, storey range and Town

```
[1] "Mean Inflation Adjusted Resale Price by flat_model"
# A tibble: 21 x 2
   flat_model                mean_price
   <chr>                          <dbl>
 1 Type S2                     1070993.
 2 Type S1                      958052.
 3 Premium Apartment Loft       923481.
 4 Terrace                      837810.
 5 Multi Generation             806789.
 6 Premium Maisonette           765086.
 7 DBSS                         742835.
 8 Model A-Maisonette           724803.
 9 Maisonette                   706128.
10 Adjoined flat                689632.
11 Improved-Maisonette          681278.
12 3Gen                         675525.
13 Apartment                    643737.
14 Premium Apartment            505856.
15 Improved                     478500.
16 Model A                      452778.
17 Standard                     410049.
18 Model A2                     367287.
19 Simplified                   358598.
20 New Generation               344734.
21 2-room                       272251.
```

```
[1] "Mean Inflation Adjusted Resale Price by flat_type"
# A tibble: 7 x 2
  flat_type        mean_price
  <chr>                 <dbl>
1 MULTI-GENERATION     806789.
2 EXECUTIVE            660081.
3 5 ROOM               561155.
4 4 ROOM               465488.
5 3 ROOM               327879.
6 2 ROOM               250903.
7 1 ROOM               191274.
```

```
[1] "Mean Inflation Adjusted Resale Price by storey_range"
# A tibble: 17 x 2
   storey_range mean_price
   <chr>             <dbl>
 1 49 TO 51       1096233.
 2 46 TO 48       1052038.
 3 43 TO 45        966672.
 4 40 TO 42        908845.
 5 37 TO 39        841886.
 6 34 TO 36        824423.
 7 31 TO 33        805315.
 8 28 TO 30        772793.
 9 25 TO 27        684411.
10 22 TO 24        634304.
11 19 TO 21        607881.
12 16 TO 18        542257.
13 13 TO 15        501899.
14 10 TO 12        466477.
15 07 TO 09        453042.
16 04 TO 06        440955.
17 01 TO 03        422957.
```

Figure 26: Data used to plot boxplots in figure 11 and 12

```
[1] "Mean Inflation Adjusted Resale Price by town"
# A tibble: 26 x 2
   town            mean_price
   <chr>                <dbl>
 1 BUKIT TIMAH        708302.
 2 BISHAN             649896.
 3 CENTRAL AREA       642754.
 4 QUEENSTOWN         590953.
 5 BUKIT MERAH        577432.
 6 KALLANG/WHAMPOA    527090.
 7 PASIR RIS          525801.
 8 MARINE PARADE      524655.
 9 SERANGOON          508899.
10 CLEMENTI           504216.
11 TOA PAYOH          499104.
12 TAMPINES           496508.
13 PUNGGOL            484372.
14 SENGKANG           467917.
15 HOUGANG            463370.
16 GEYLANG            452325.
17 BUKIT PANJANG      451835.
18 CHOA CHU KANG      435651.
19 JURONG EAST        433791.
20 BEDOK              431432.
21 ANG MO KIO         428072.
22 BUKIT BATOK        420752.
23 SEMBAWANG          419931.
24 WOODLANDS          415559.
25 JURONG WEST        415413.
26 YISHUN             397138.
```

Figure 27: Data used to plot boxplots in figure 11

Appendix 4: Models extra information

Appendix 4.1: Sub-optimal diagnostic plots for linear regression model due to slight presence of decreasing variance and non-normality

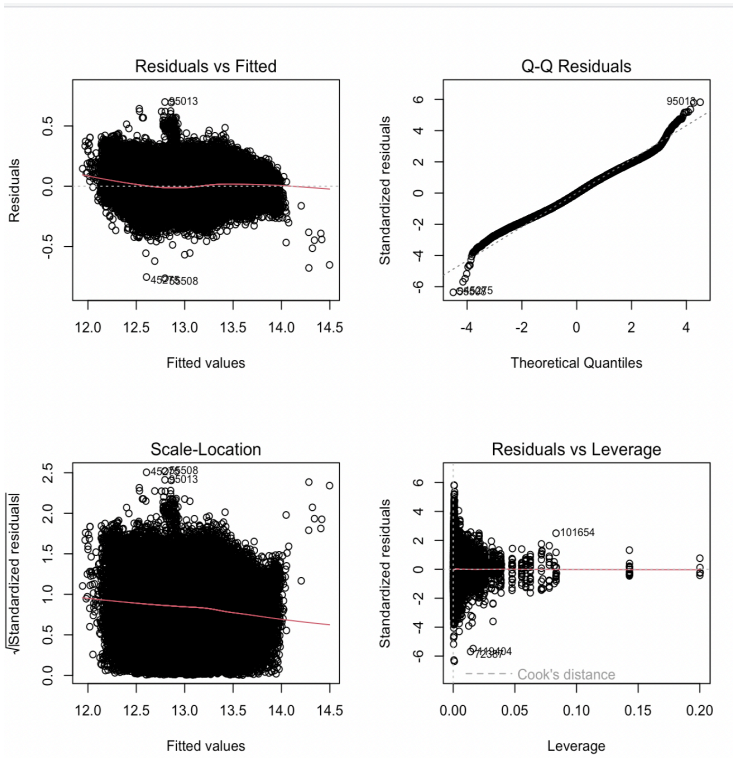


Figure 28: Log-transformed response variable - diagnostic plots for linear regression model. Slight heteroskedasticity in lower left plot

Appendix 4.2: PCA additional information

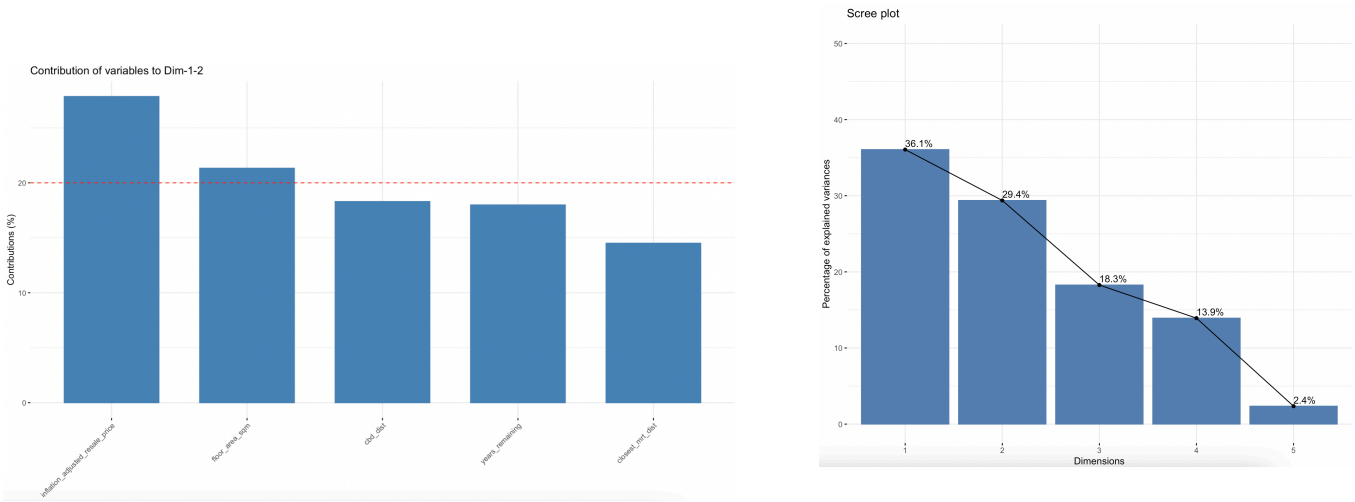


Figure 29: Plot on left shows variables with most contributions with Principal components 1 and 2. Plot on the right shows variance explained by each principal component.

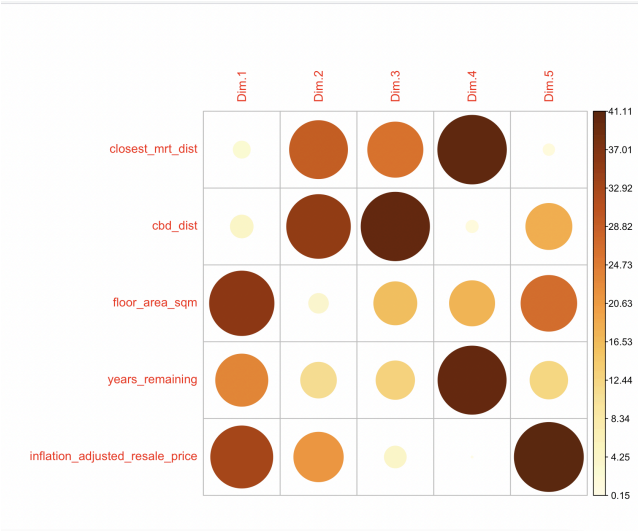


Figure 29: Correlation heat map shows how each variable contributes to each principal component