

HDB Resale Predictor

DSI_39-Project_2-39SIR

15 Sep 2023



Contents

1. Context
 - a. Background
 - b. Problem Statement
 - c. Workflow
2. Data Cleaning
3. Exploratory Data Analysis (EDA)
4. Feature Selection
5. Modelling
6. Demo
7. Conclusion
8. Recommendations

1. Background

Bukit Timah: Not just a rich man's playground



iProperty Singapore

26 DEC 2017

GUIDES SINGAPORE PROPERTY NEWS CONDO DISTRICT HDB NEIGHBOURHOOD

The common perception of the Bukit Timah neighbourhood is an exclusive one, where expensive private homes sit in abundance, their occupants far away from living the typical Singaporean experience.



Isabelle Liew

UPDATED SEP 7, 2023, 5:46 PM SGT ▾

[f](#)
[t](#)
[...](#)

Size or facilities?

SINGAPORE - Prices of Housing Board resale flats inched up by 0.6 per cent in August, with a record 54 units changing hands for at least \$1 million each.

It was the fourth consecutive month that HDB resale flat prices grew by less than 1 per cent, although the increase is higher than July's [0.3 per cent](#), according to flash data from real estate portals [99.co](#) and [SRX](#) on Thursday.

The price growth was mostly driven by larger flats, with prices of five-room flats rising by 1.5 per cent and executive apartments climbing 1.6 per cent, said OrangeTee & Tie senior vice-president of research and analytics Christine Sun.

1. Problem Statement

We are a real estate investors advisory group to **assess the value of a HDB resale flat**, specialized in knowing what helps buy/sell the property.

Use of Machine Learning

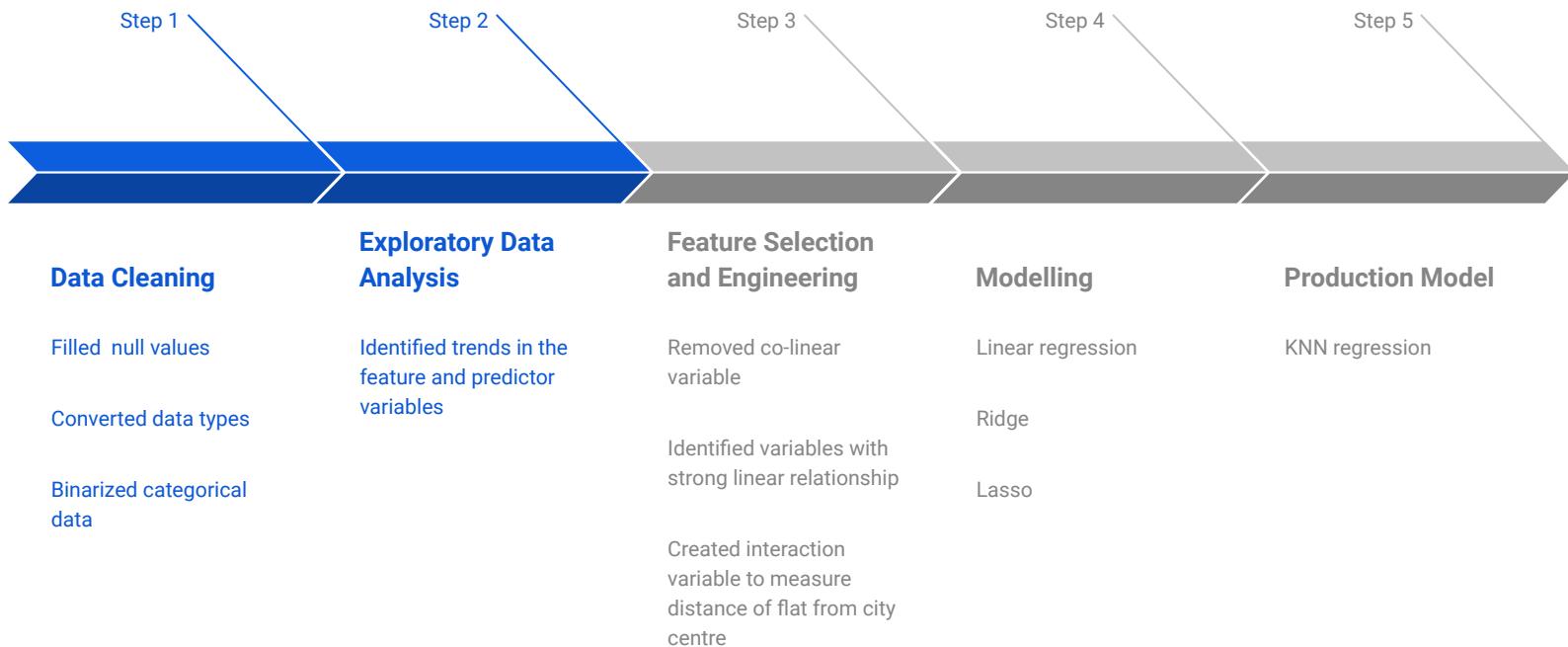
- Reduces the guesswork on the optimal price to buy/sell
- Identify **factors** with greatest **impact** on **resale price**.

Problem Statement

We are a real estate investors advisory group to **assess the value of a HDB resale flat**, specialized in knowing what helps buy/sell the property.

Machine learning (ML) helps to cut down the guesswork on the optimal price to buy/sell. It also helps us understand which **factors** have the greatest **impact** on the **resale price**.

1. Workflow



2. Data Cleaning: Null Nullifier

Feature with null values	No. of null values	Null value nullification means	Explanation
Mall_Nearest_Distance	829	Replaced by mean of column's data	If null value was replaced with 0, then it says mall is on top HDB flat
Mall_Within_500m	92789	Replaced with 0	Null values indicates there's nothing within that zone.
Mall_Within_1km	25426		
Mall_Within_2km	1940		
Hawker_Within_500m	97390		
Hawker_Within_1km	60868		
Hawker_Within_2km	29202		

2. Data Cleaning: Steps

- Variables with Y/N input got **binarized**
- All flats = residential, therefore **residential category dropped**
- **Formatting** all column headers (lowercase, snake_case, no trailing space)

3. EDA: Overview of HDB Data Set



150,634

HDB Resale Transactions



9 Years of Data

Mar 2012 to Apr 2021



74 Variables in Cleaned Data

E.g. Transaction Year/Month, Town, Flat Type, etc

3. EDA: Contrasting Lowest and Highest Resale Flats

\$150,000



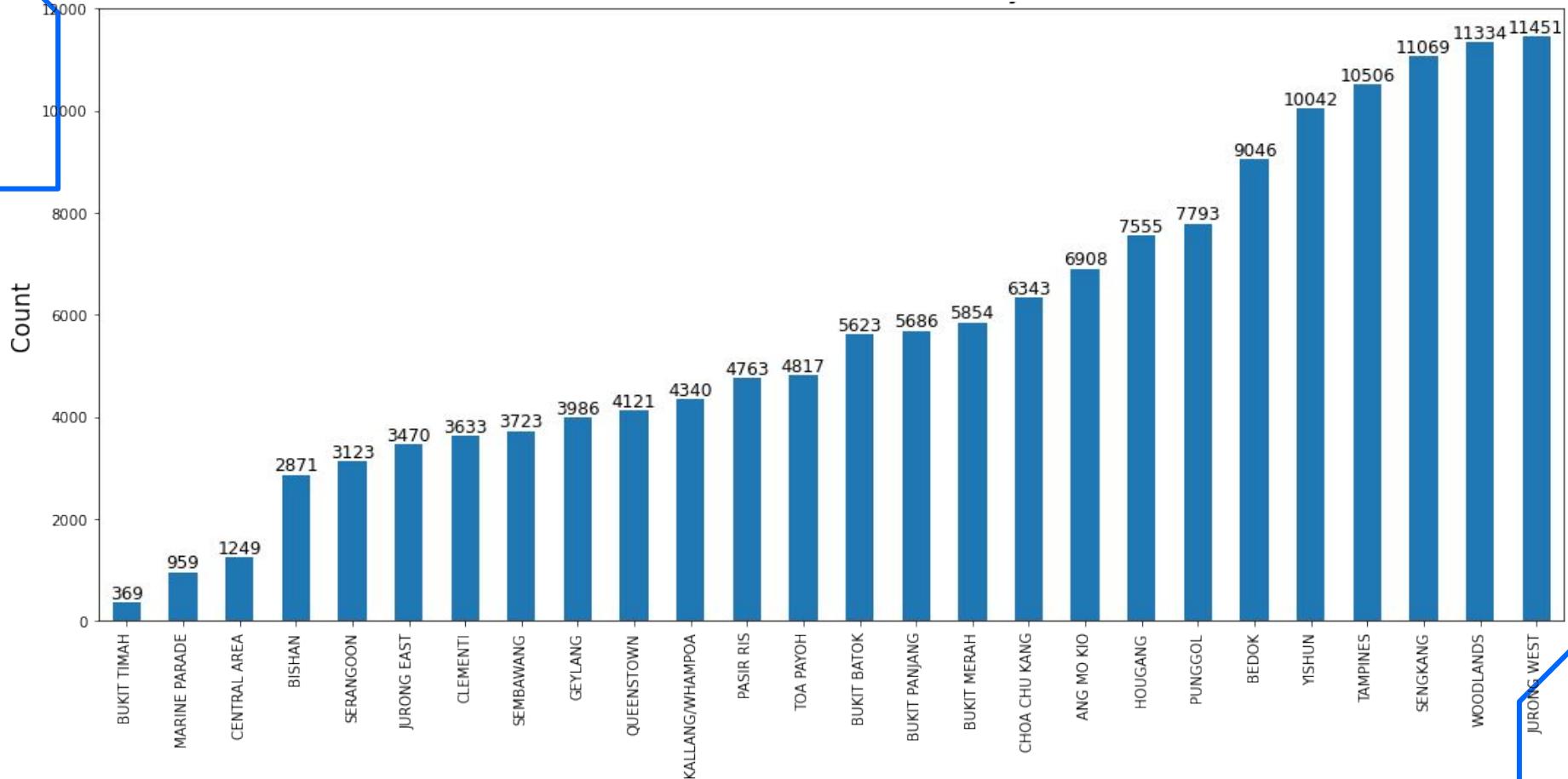
- Geylang
- 2-Room Flat, 45 sqm
- 4th to 6th Storey
- Age of 52 years

\$1,258,000

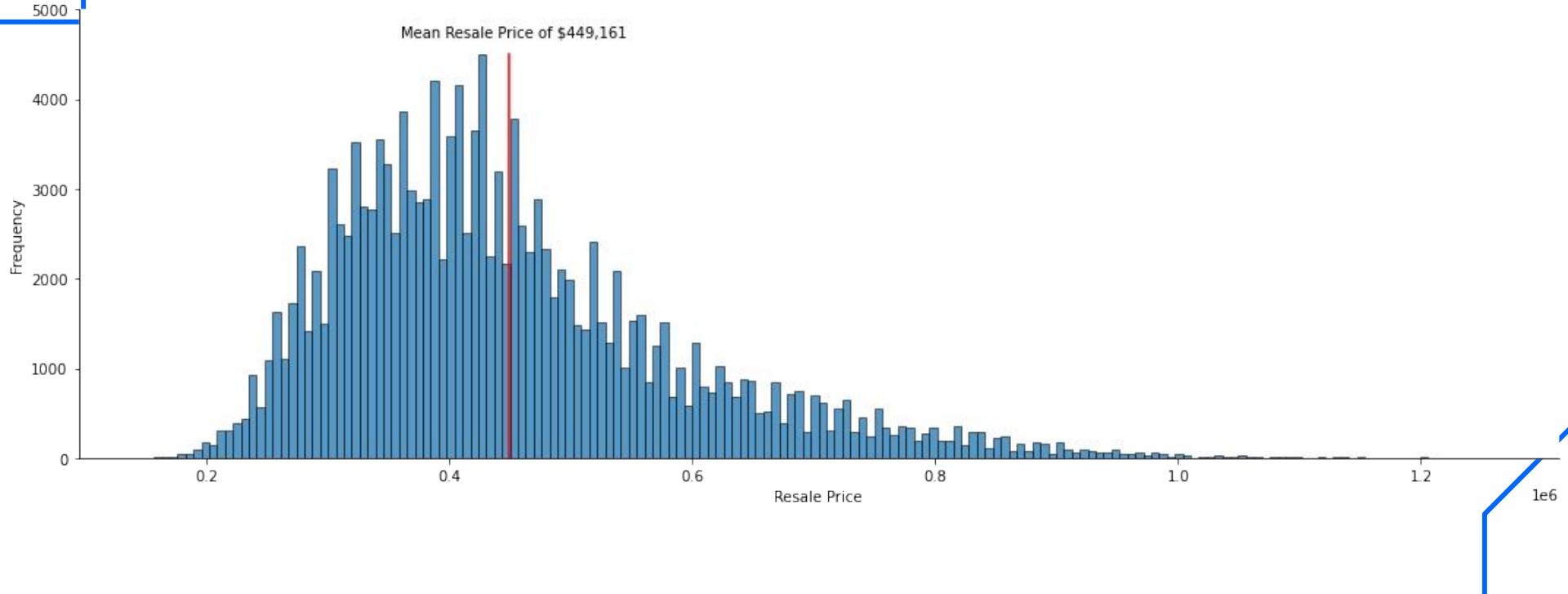


- Central Area
- 5-Room Flat, 107 sqm
- 43rd-45th Storey
- Age of 10 years

3. EDA: Mean Resale Price by Town



3. EDA: Distribution of Resale Prices



4. Feature Selection

Clementi

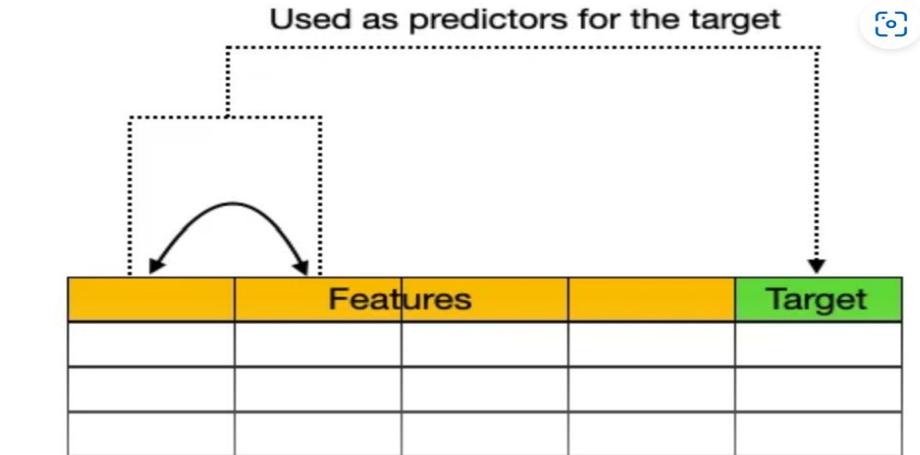
Kallang



4. Feature Selection

Which are the collinear pair(s)?

floor area in sqm
AND
floor area in sqft



4. Feature Selection

lease_commence_date,
Tranc_Year,
hdb_age

mid_storey,
Lower,upper,mid,,

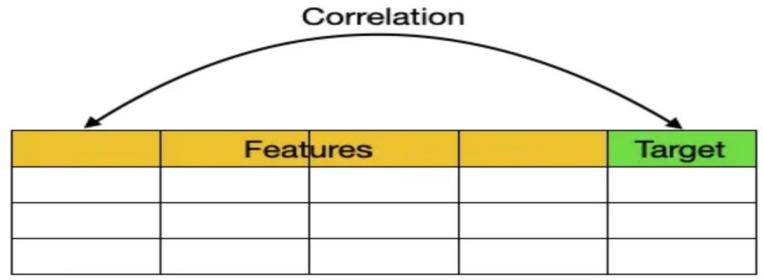
Latitude,Longitude,

bus_stop_latitude,
bus_stop_longitude

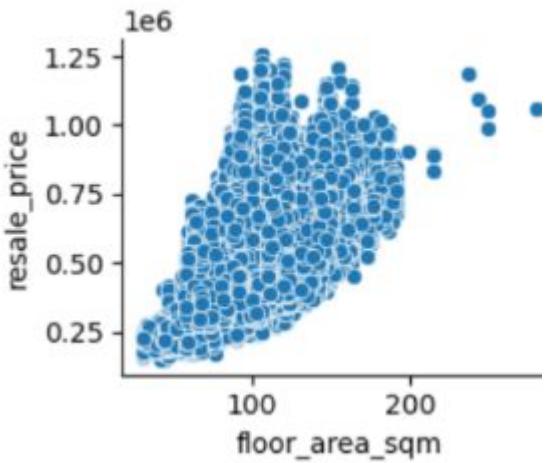
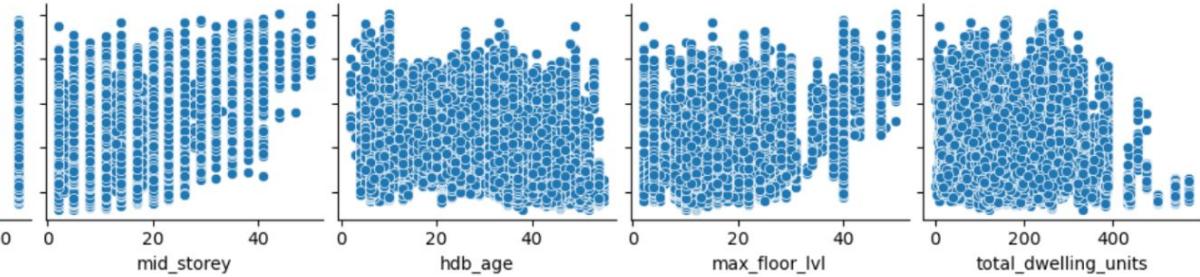
pri_sch_latitude,pri_sch_longitude
sec_sch_latitude,sec_sch_longitude

4. Feature Selection

Is **correlation**
high or low between
remaining features and resale
price?

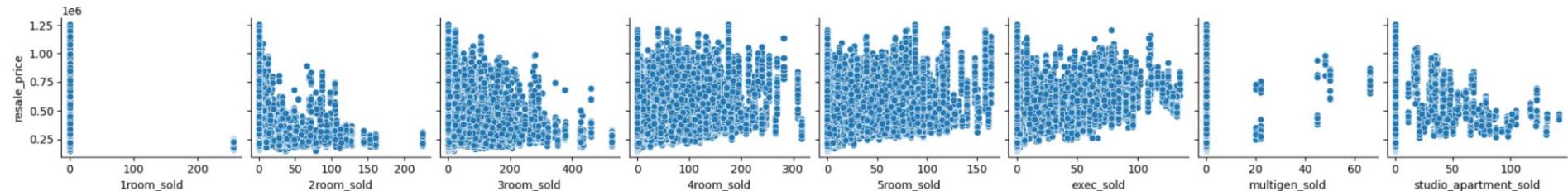


4. Feature Select



3. Feature Selection:

	resale_price
resale_price	1.000000
1room_sold	-0.039488
2room_sold	-0.146455
3room_sold	-0.411948
4room_sold	0.088682
5room_sold	0.360818
exec_sold	0.339819
multigen_sold	0.037608
studio_apartment_sold	0.041589



3. Feature Engineering: “dist”



5. Modelling - 1st Round

	R2 (train)	R2 (test)	RMSE (train)	RMSE (test)
Linear regression	0.834	0.832	58,523	58,464
Ridge	0.795	0.791	58,963	58,992
Lasso	0.800	0.797	58,523	58,468

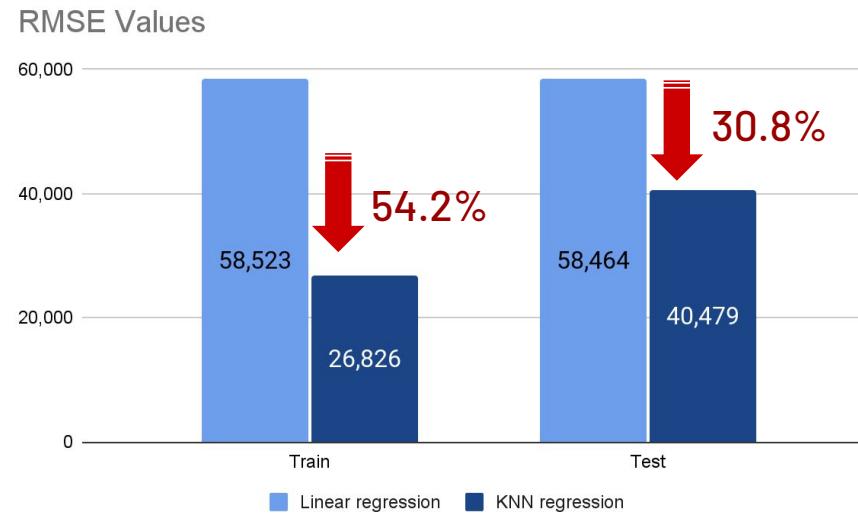
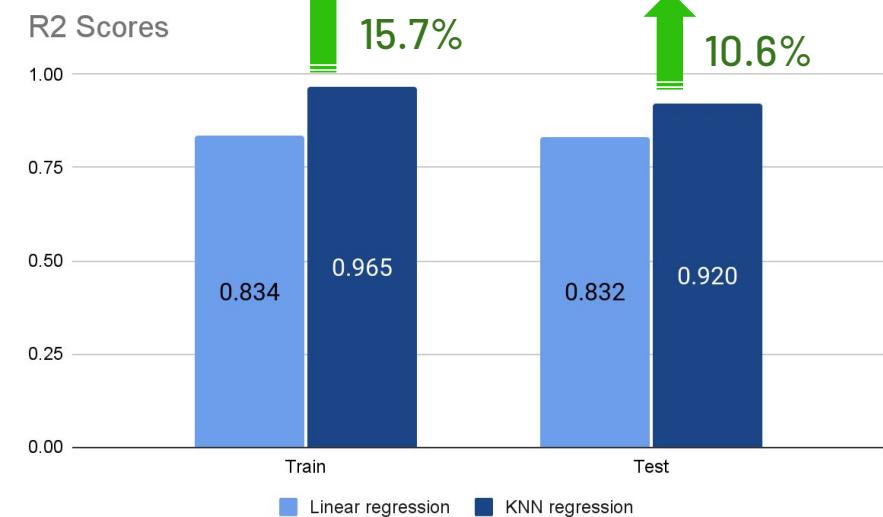
Note:

R2 - The higher the better

RMSE - The lower the better

- Best model: **linear regression**
- **Highest R2 values** on both train and test data
- **Lowest RMSE values** on both train and test data

5. Modelling - 2nd Round



Legend:

- Linear regression - The higher the better
- KNN regression - The lower the better

SelectKBest

Top 5 features:

1. floor area
2. the hdb flat's age
3. height of the hdb flat
4. year completed
5. how many 5 room flats are sold in that flat

6. Demo - App

User Input

Floor Area (sqm)
100 - +

Year of resale transaction
2015 - +

Median value of storey_range
14 - +

Number of years from lease_commence_date to present year
23 - +

Highest floor of the resale flat
18 - +

Year which construction was completed for resale flat
1996 - +

Total number of residential dwelling units in the resale flat
128 - +

of 4-room residential units in the resale

HDB Resale Price Predictor

Predicted Resale Price: \$460,337.42

Model Performance

Mean Squared Error (MSE): 3736872399.67

R-squared (R²): 0.82

Try out the predictor app in the link below:

[Click here!](#)

Made with Streamlit

6. Demo - Context

LATEST NEWS CURRENT AFFAIRS VIRAL LIFESTYLE PEOPLE INSPIRATION VIDEOS

Choa Chu Kang HDB Flat Rented For S\$6,600/Month, Sets New Record For 2023

2,314 sq ft

Another flat in Tampines was rented out for S\$6,550 this month.

Jeremy Lee - 8 Aug 2023, 1:04 am

VIEWED

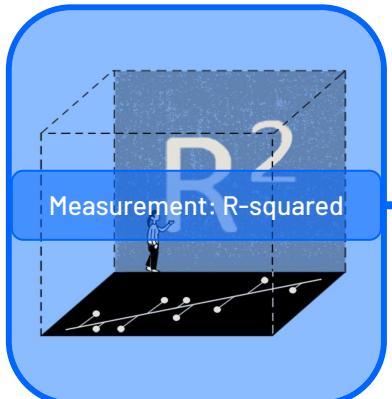
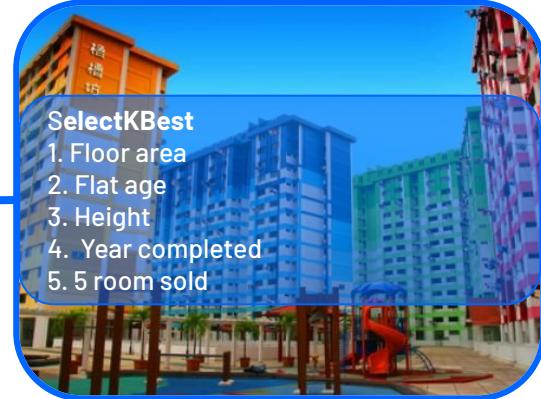
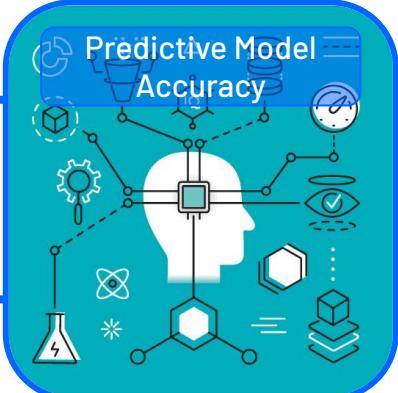
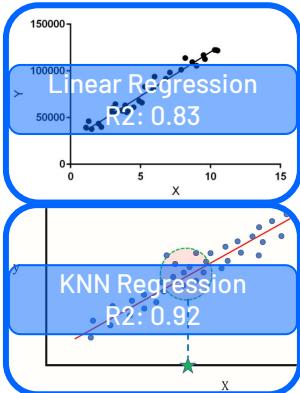
640 Choa Chu Kang Street 64

640 Choa Chu Kang Street 64
S\$ 1,000,000
4 ⚡ 2 ⚡ • 1560 sqft • S\$ 641.03 psf
🚶 3 mins (220 m) to NS5 Yew Tee MRT

HDB Flat 99-year Leasehold Built: 1998

Contact Agent

7. Conclusion



6. Recommendations



Data Enrichment

Non-numerical features
Qualitative indicators of location



Diversification

Across different types of properties,
locations and asset classes

HDB will adopt the Standard-Plus-Prime flat model, here's a guide to
BUYING A HDB FLAT FROM H2 2024

TYPES OF FLATS	STANDARD	PLUS	PRIME
LOCATIONS	Islandwide	Choicer locations e.g. near MRT	Choicest locations e.g. City Centre
SUBSIDIES	Standard	More on top of standard	Most on top of standard
MINIMUM OCCUPATION PERIOD (MOP)	5 years	10 years	10 years
WAIT OUT PERIOD	-	30 months	30 months
SUBSIDY RECOVER	-	Yes lower than Prime	6% of resale price or valuation
RENT OUT FLAT / SPARE ROOMS	Both, after MOP	Rooms only, after MOP	Rooms only, after MOP

find out the restrictions on homeowners and buyers

SeedlyReads

Policy Changes

Thank you

