



Five & Furious Agency

# HDB PRICES

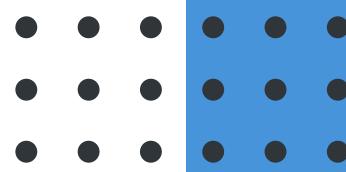
**Prepared by:**

Ong Hao Ting Aiken

Win Lae Yee

Bae Jongseong

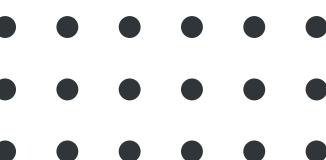
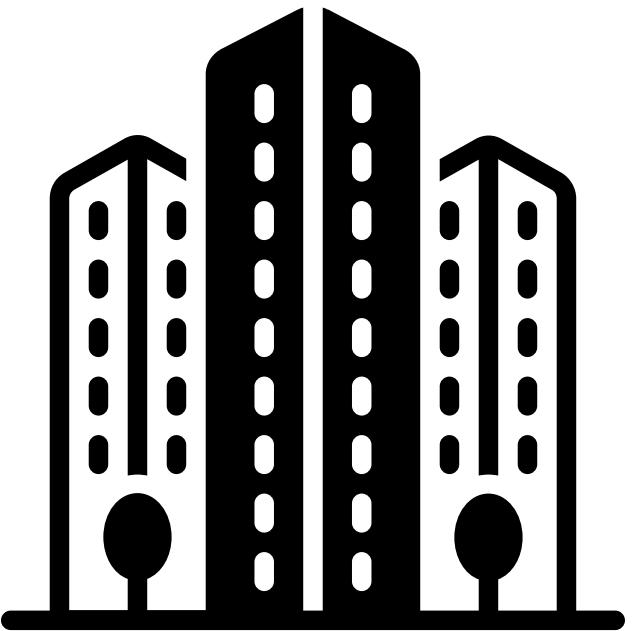
Tan Ke Wei Nicholas



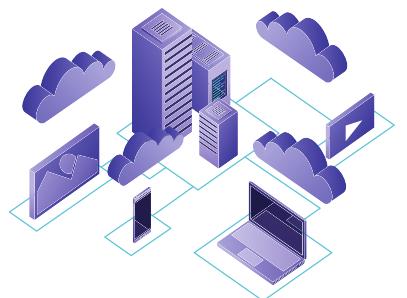
**28th Sept 2022**

# PROBLEM STATEMENT

You're working for a **real estate** company who wishes to create an easy tool for customers to instantly **value their HDB flats**. They have asked you to come up with an ML algorithm that will be used as the basis of this tool.

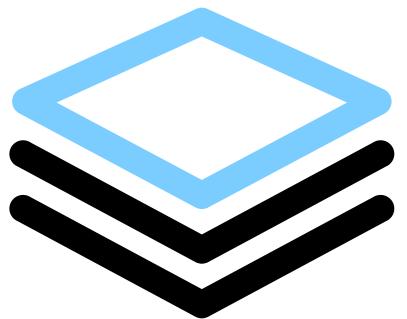
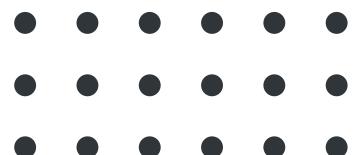


# INTRODUCTION TO DATASET



## DATASET ANALYZED:

hdb-resale-flat-prices.csv

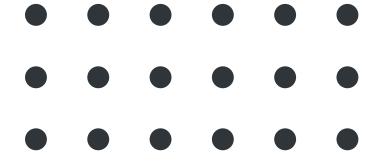


## DATASET DIMENSIONS:

135,568 rows, 11 columns

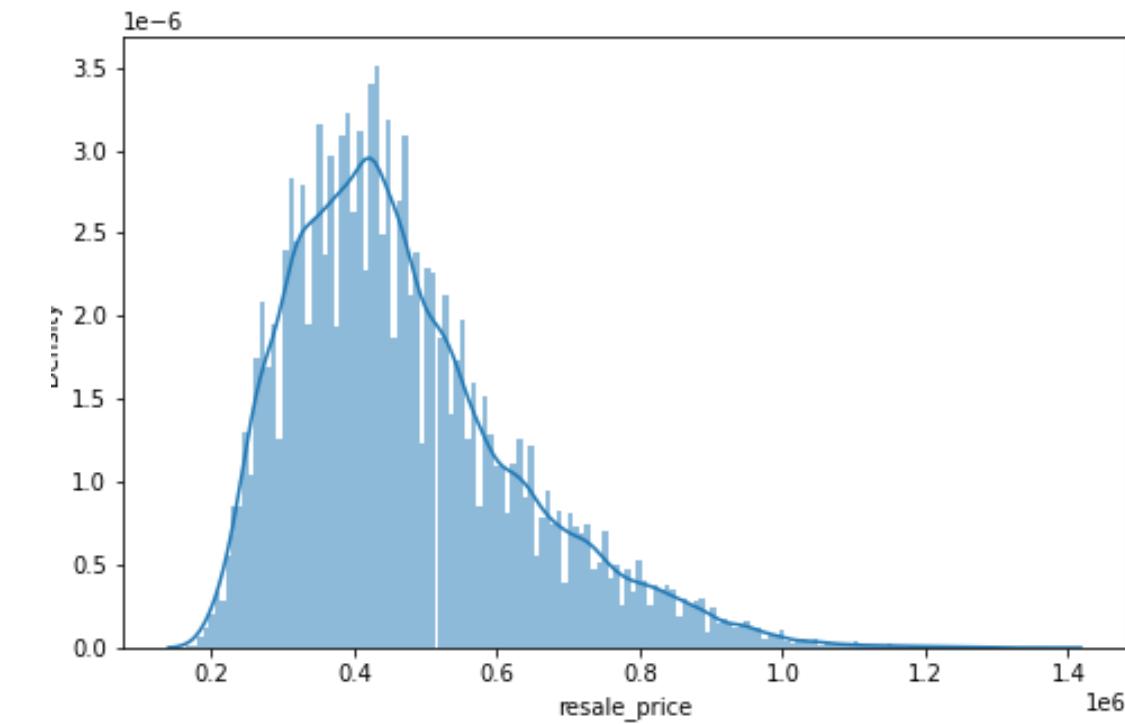
	month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	remaining_lease	resale_price
0	2017-01	ANG MO KIO	2 ROOM	406	ANG MO KIO AVE 10	10 TO 12	44.0	Improved	1979	61 years 04 months	232000.0
1	2017-01	ANG MO KIO	3 ROOM	108	ANG MO KIO AVE 4	01 TO 03	67.0	New Generation	1978	60 years 07 months	250000.0
2	2017-01	ANG MO KIO	3 ROOM	602	ANG MO KIO AVE 5	01 TO 03	67.0	New Generation	1980	62 years 05 months	262000.0
3	2017-01	ANG MO KIO	3 ROOM	465	ANG MO KIO AVE 10	04 TO 06	68.0	New Generation	1980	62 years 01 month	265000.0

# EXPLORATORY DATA ANALYSIS



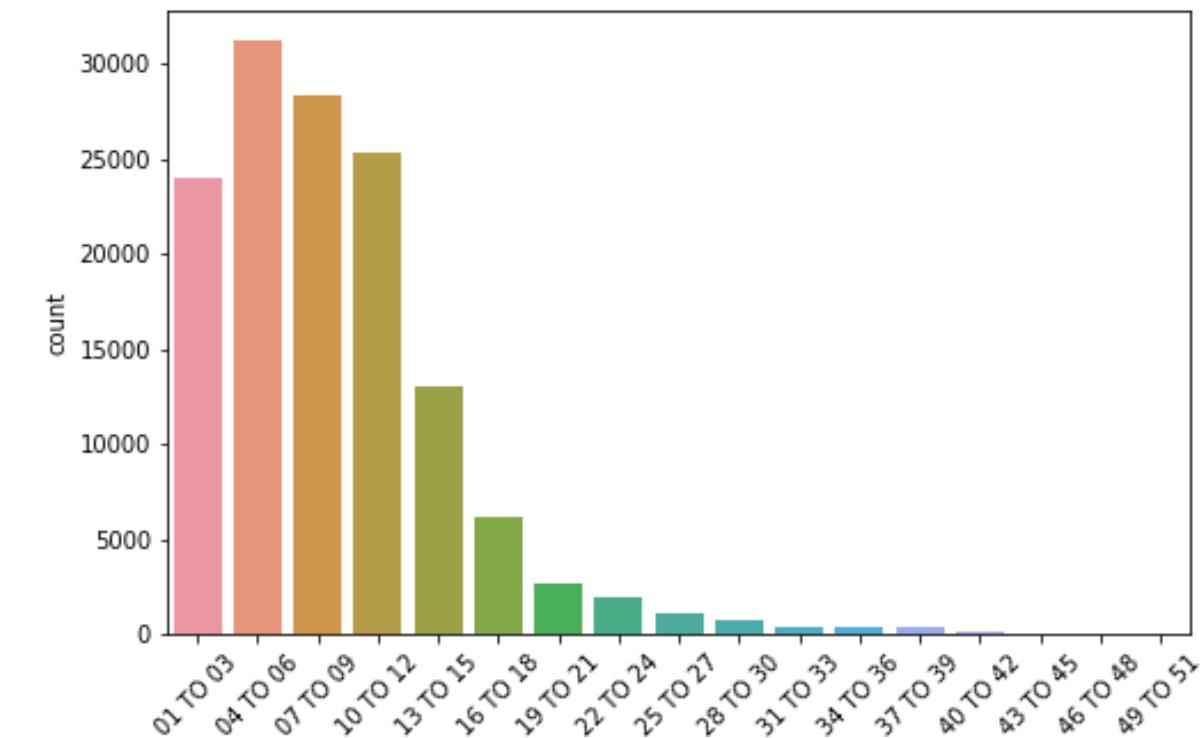
## — DISTRIBUTION OF LISTINGS' PRICES

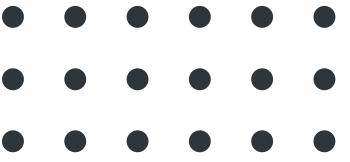
The most common resale prices of listings lies within the 400k to 500k range.



## — DISTRIBUTION OF LISTINGS' STOREY

The listings mainly fall into the 4 ranges from 01 to 12 . This may be due to the common maximum height of HDBs.





# DATA PREPROCESSING

- UPDATED REMAINING LEASE TO FLOAT DATA TYPE AND CALCULATED ITS VALUE

e.g. 61.3 years from 61 years and 04 months

- DROPPED SOME INDEPENDENT VARIABLES TO PREVENT MULTICOLLINEARITY

e.g. street\_name is a subset of town. Hence, the 2 independent variables are highly correlated with each other and share a large proportion of variance. This leads to inflated estimates of errors and unstable models.

- DROPPED REGRESSORS THAT HAVE LOW CORRELATION WITH THE DEPENDENT VARIABLE

e.g. month, block

# MODEL TRAINING



## TRAIN-TEST SPLIT RATIO

We adopted the typical train-test split ratio of  
**80:20**

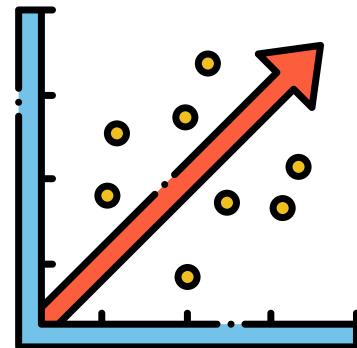


## FEATURE ENCODING

We used **one-hot encoding** to transform our categorical independent variables (`flat_model`, `flat_type`, `store_range`, `town`) into numbers



# MODEL SELECTION & EVALUATION



## LINEAR REGRESSION

We analyzed the dataset using **supervised multiple linear regression** (as there are multiple independent variables) to predict HDB resale price (continuous variable)



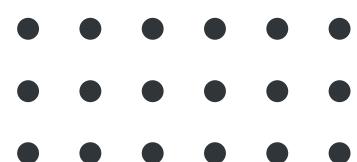
## PERFORMANCE EVALUATION

**R Squared (R<sup>2</sup>)** = 0.81594322461061

**Mean Squared Error (MSE)** = 4975349908.485722

**Root-mean-square deviation (RMSE)** = 70536.16029020662

**Mean absolute error (MAE)** = 55803.198585415215

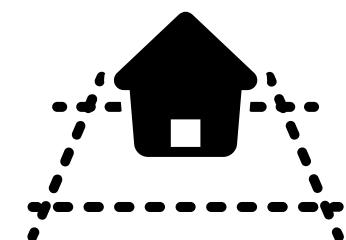


# MODEL INTERPRETATION



## REMAININGLEASE

- Coefficient: Coefficient of remaining\_lease is **4540.0588**. This means that a change in 1 unit of remaining\_lease cause HDB resale price to increase by **\$4540.06 (2.dp)**
- p-value: The p-value of remaining\_lease is 0.000 (which is less than 0.05). At 5% significance level remaining\_lease is **statistically significant** in determining HDB resale price



## FLOORAREA

- Coefficient: Coefficient of floor\_area is **3309.8357**. This means that a change in 1 unit of floor\_area cause HDB resale price to increase by **\$3309.84 (2 d.p)**
- p-value: The p-value of floor\_area is 0.000 (which is less than 0.05). At 5% significance level floor\_area is **statistically significant** in determining HDB resale price

# MODEL INTERPRETATION

## — FLAT\_MODEL

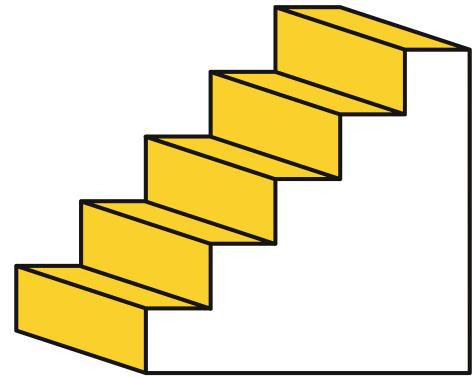
- Dummy variable of flat\_model categorical variable is '2-room'.
- *Highest Positive Coefficient:* Coefficient of terrace is the highest and positive at SDG 338400, comparing against other flat models. This means that **resale price** of the HDB flats when it is a **terrace, compared to when it is a 2-room, is higher by SDG 338400**.
- *Highest Negative Coefficient:* Coefficient of improved HDB flat is at -SDG 47460 comparing against other flat models. This means that resale price of the HDB flats when it is **improved, compared to when it is a 2-room, is lower by SDG 47460**.

## — FLAT\_TYPE

- Dummy variable of flat\_type categorical variable is '1 ROOM'.
- *Highest Positive Coefficient:* Coefficient of '5 ROOM' HDB flat is the highest and positive at SDG 138600, comparing against other flat types. This means that resale price of the HDB flats when it is a **5 room, compared to when it is a 1 room, is higher by SDG 138600**.
- *Lowest Positive Coefficient:* Coefficient of '2 ROOM' HDB flat is the lowest and positive at SDG 37410, comparing against other flat types. This means that resale price of the HDB flats when it is a **2 room, compared to when it is a 1 room, is higher by SDG 37410**.

# MODEL INTERPRETATION

## STOREY\_RANGE



- Dummy variable: '01 TO 03'.
- *Highest Positive Coefficient:* Coefficient of "49 TO 51" is SGD \$221,800. This means that **resale price** of the HDB flats in the "49 TO 51" range, **compared to a flat in the "01 TO 03" range, is higher by SGD \$221,800.**
- *Lowest Positive Coefficient:* Coefficient of "04 TO 06" is at SGD \$16,960. This means that resale price of the HDB flats in the "04 TO 06" range, **compared to a flat in the "01 TO 03" range, is higher by SGD \$16,960.**

# MODEL INTERPRETATION



## TOWN

- Dummy variable: 'ANG MO KIO'.
- *Highest Positive Coefficient:* Coefficient of "BUKIT TIMAH" is at SGD \$191,600. This means that **average resale price** of the HDB flats in "BUKIT TIMAH" region, **compared to a flat in the "ANG MO KIO" region**, is **higher by SGD \$191,600**.
- *Highest Negative Coefficient:* Coefficient of "SEMBAWANG" is at SGD \$180,300. This means that **average resale price** of the HDB flats in the "SEMBAWANG" region, **compared to a flat in the "ANG MO KIO" region**, is **lower by SGD \$180,300**.
- **The p-values of all regions is 0.000** (which is less than 0.05). **At 5% significance level**, all Town regions are **statistically significant** in determining HDB resale price.

# BUSINESS INSIGHTS

- REAL ESTATE COMPANY CAN LOOK INTO HDB TERRACE AND 5-ROOM APARTMENTS FOR **HIGHER PROFIT MARGINS**. THEY CAN ALSO ADVISE THEIR CUSTOMERS TO BUY A 1-ROOM IMPROVED HDB FLAT IF THEY ARE ON A TIGHT BUDGET.
- THE COMPANY CAN PROMOTE HOUSING WITH LONGER LEASE YEARS & GREATER FLOOR AREAS TO FAMILIES/PEOPLE WHO WANTS TO SETTLE DOWN IN SINGAPORE **LONG-TERM** WHILE OFFERING AFFORADABLE **SHORT-TERM** HOUSING WITH SHORTER LEASE YEARS TO OTHERS (EG. STUDENTS OR PERMANENT RESIDENT WORKING HERE)
- THE AVERAGE RESALE PRICE OF HDB FLATS SIGNIFICANTLY INCREASES WITH INCREASE IN STOREY RANGE.
- THE 3 MOST EXPENSIVE HDB TOWNS ARE BUKIT TIMAH, MARINE PARADE AND QUEENSTOWN. IN ADDITION TO SCHOOLS, SHOPPING CENTRES, AND NATURE PARKS THAT THESE AREAS HOUSE, THESE AREAS ARE DIRECTLY CONNECTED WITH THE DOWNTOWN MRT LINE AND EAST WEST LINE THAT CONNECT DIFFERENT AREAS OF THE TOWN.
- THE 3 MOST AFFORDABLE HDB TOWNS ARE SEMBAWANG, CHOA CHU KANG AND WOODLANDS. WHILE THESE AREAS MAY NOT BE AS WELL-EQUIPPED AS OTHER POPULAR HDB ESTATES, REGIONS LIKE SEMBAWANG AND WOODLANDS HAVE RECENTLY BUILT MORE MRT STATIONS SUCH AS THE THOMSON-EAST COAST MRT LINE, WHICH CUSTOMERS CAN TAKE INTO ACCOUNT WHEN VALUING THEIR FLATS.



"THANK YOU"