



# **Apartment Resale Price Prediction Model**

---

**Based on Singapore HDB DataSet**

**Presented by  
Group E**

# Group Participants

---

- Kaung Yarzar
- Thant Zin Bo
- Nay Win Hlaing
- Nang Mo Mo Hein
- Ma Wai Wai Khin



# Problem Statement

---

The project aims at building a prediction model for Housing and Development Board's Apartment resale prices by using the provided dataset. This model should learn from the data and be able to predict the HDB resale price in any town of Singapore.

## **Objective**

Develop a machine learning model to estimate HDB resale prices based on various features.

# Data Source



- HDB resale transaction dataset from [singapore government](https://data.gov.sg/dataset/resale-flat-prices)  
<https://data.gov.sg/dataset/resale-flat-prices>
- Official government data ensures high reliability.
- Data spans from 2017 to 2022.

## Key Features:

- Numerical: floor\_area\_sqm, years\_remaining, closest\_mrt\_dist, cbd\_dist
- Categorical: storey\_range , flat\_model, town

# Sample Data

latitude	longitude	postal_co	address	closest_m	closest_m	cdb_dist	month	town	flat_type	block	street_na	storey_ra	floor_area	flat_model	lease_con	remaining	resale_pri	year	years_remaining
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2017-04	KALLANG/WH	3 ROOM	1	BEACH RD	07 TO 09	68	Improved	1979	61 years 0	340000	2017	61
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2017-05	KALLANG/WH	3 ROOM	1	BEACH RD	07 TO 09	68	Improved	1979	61 years 0	322000	2017	61
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2017-06	KALLANG/WH	3 ROOM	1	BEACH RD	07 TO 09	68	Improved	1979	61 years 0	370000	2017	61
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2017-11	KALLANG/WH	3 ROOM	1	BEACH RD	07 TO 09	68	Improved	1979	60 years 1	375000	2017	61
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2018-04	KALLANG/WH	3 ROOM	1	BEACH RD	13 TO 15	68	Improved	1979	60 years 0	376000	2018	60
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2018-08	KALLANG/WH	3 ROOM	1	BEACH RD	13 TO 15	68	Improved	1979	60 years 0	342000	2018	60
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2019-05	KALLANG/WH	3 ROOM	1	BEACH RD	07 TO 09	68	Improved	1979	59 years 0	340000	2019	59
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2019-10	KALLANG/WH	3 ROOM	1	BEACH RD	04 TO 06	74	Improved	1979	58 years 1	357000	2019	59
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2020-01	KALLANG/WH	3 ROOM	1	BEACH RD	01 TO 03	68	Improved	1979	58 years 0	320000	2020	58
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2020-03	KALLANG/WH	3 ROOM	1	BEACH RD	04 TO 06	68	Improved	1979	58 years 0	290000	2020	58
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2020-06	KALLANG/WH	3 ROOM	1	BEACH RD	07 TO 09	74	Improved	1979	58 years 0	365000	2020	58
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2020-11	KALLANG/WH	3 ROOM	1	BEACH RD	01 TO 03	61	Improved	1979	57 years 1	320000	2020	58
1.303671	103.8645	190001	1 BEACH R	Nicoll High	441.785	2715.822	2021-02	KALLANG/WH	3 ROOM	1	BEACH RD	07 TO 09	68	Improved	1979	57 years 0	328000	2021	57

Data Source Link -

<https://www.kaggle.com/datasets/tayyeexuan/singapore-hdb-resale-price-dataset2017-jan-2024>

# Data Preprocessing

---

- Dataset checked for missing values, duplicate values, and outliers
- Used ColumnTransformer() for separating different functions to categorical and numerical features
- Encoding Categorical Features: Applied both One-Hot and Labelling Encoding
- Normalization: Scaled numerical features using StandardScaler.
- Tool used : Pandas, Scikit-learn

# Model Selection and Development

---

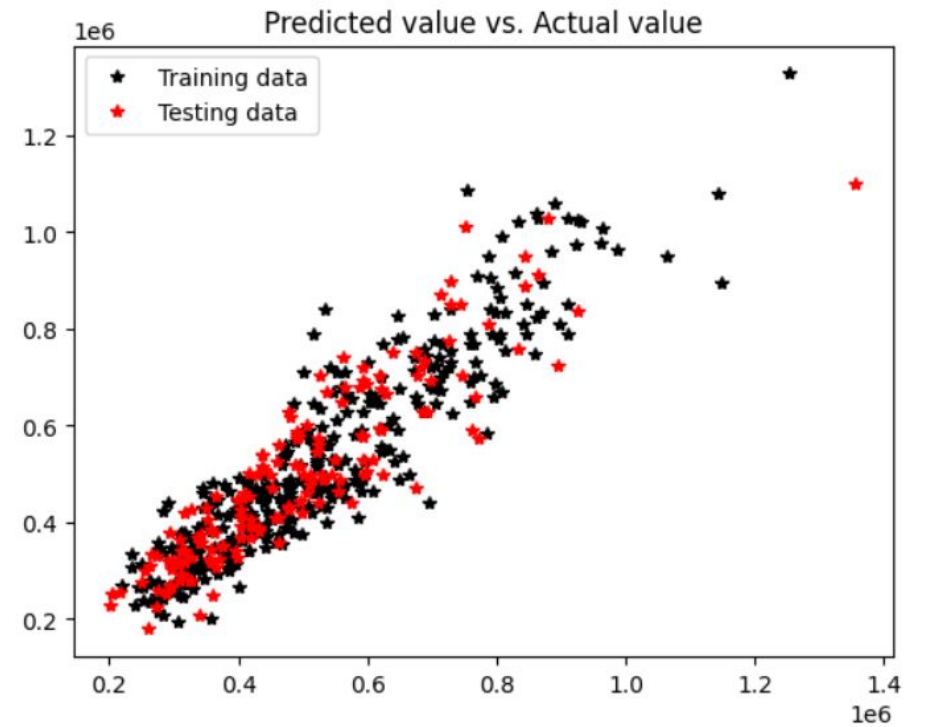
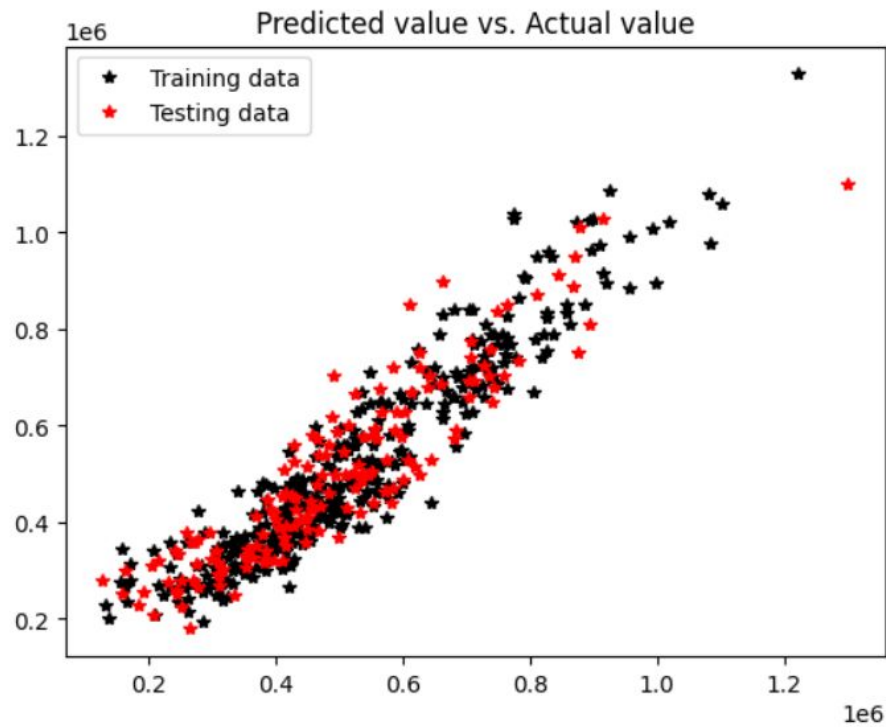
1. Testing Processes
  - Encoding Tested - Both One-Hot & Labelling
  - Modelling Tested - Both Linear & Polynomial regression
  - Features Tested - All Features & Selected Features
2. Cross-validation
  - Hyperparameter tuning using GridSearchCV
  - Polynomial Degree Testing using One-Hot Encoding = 1 to 3
  - Polynomial Degree Testing using Label Encoding = 1 to 9
  - K-folds - up to 10
3. Reproducibility:
  - Detailed steps, metrics comparing pdf files and code snippets provided.

# Comparison with Excel Table for all 7 features

Encoding		One-Hot Encoding	Label Encoding
Optimal Degree		1	2
Training Set	MAE	54289.22881	61234.73486
	MSE	4,850,046,931.58	6,307,269,892.26
	R2	0.883334701	0.848281978
Testing Set	MAE	63138.08388	65002.73915
	MSE	6,148,321,661.83	6,617,114,638.73
	R2	0.821886896	0.808306252



# Visualization with matplotlib



# Conclusion and Future Work



## Summary:

- Best model: Linear Regression with highest  $R^2$  and lowest MAE.
- Importance of accurate price prediction.

## Future Enhancements:

- Incorporate additional features (e.g., economic indicators).
- Continuous model retraining with new data.



**The End**

---