September 2024

HDB Price Predictor

Presented By: DJ BAB

Precision Insights for a Smarter HDB Market

- Data Collection & Quality
- Market Volatility & Trends
- Feature Selection & Model Complexity

- Model Accuracy & Reliability
- Time Constraints
- User-Friendly Implementation
- Competitive Market Analysis

Competitive & Volatile HDB Resale Market

Problem Statement

Our agents struggle to close sales due to a time-consuming process that heavily relies on subjective opinions. This reliance leads to inaccurate predictions and inefficient decision-making, hindering their ability to effectively close deals

Solutions

Data-Driven Decision Making

Develop Models to determine true value of HDB & curb speculation

Market Demand and Pricing Trends Analysis

Aim to provide our agents with unparalleled insights and accuracy

Model Accuracy and Reliability

Ensuring accurate and reliable price predictions

150.63K
No. of Transactions

1M
Highest Price

420K
Median Price

Primary School: All

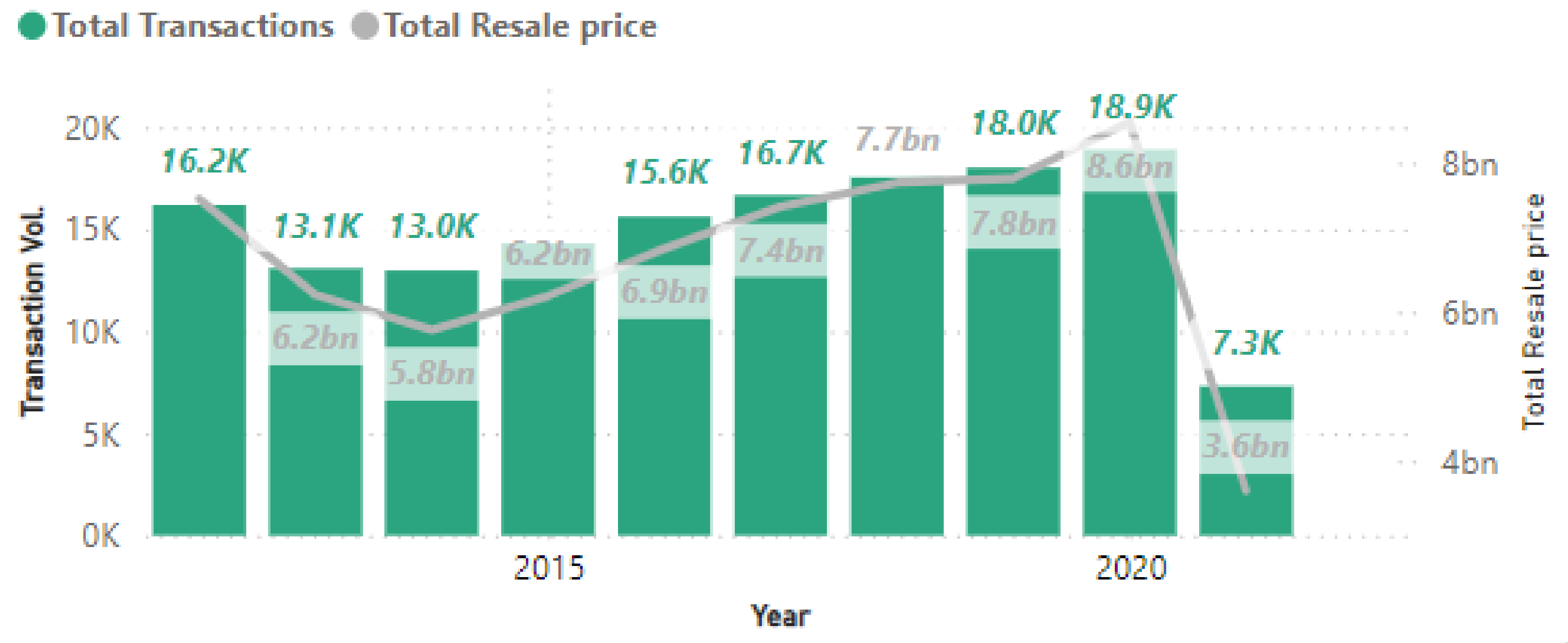
Price Range: 150000 – 1258000

Year: All

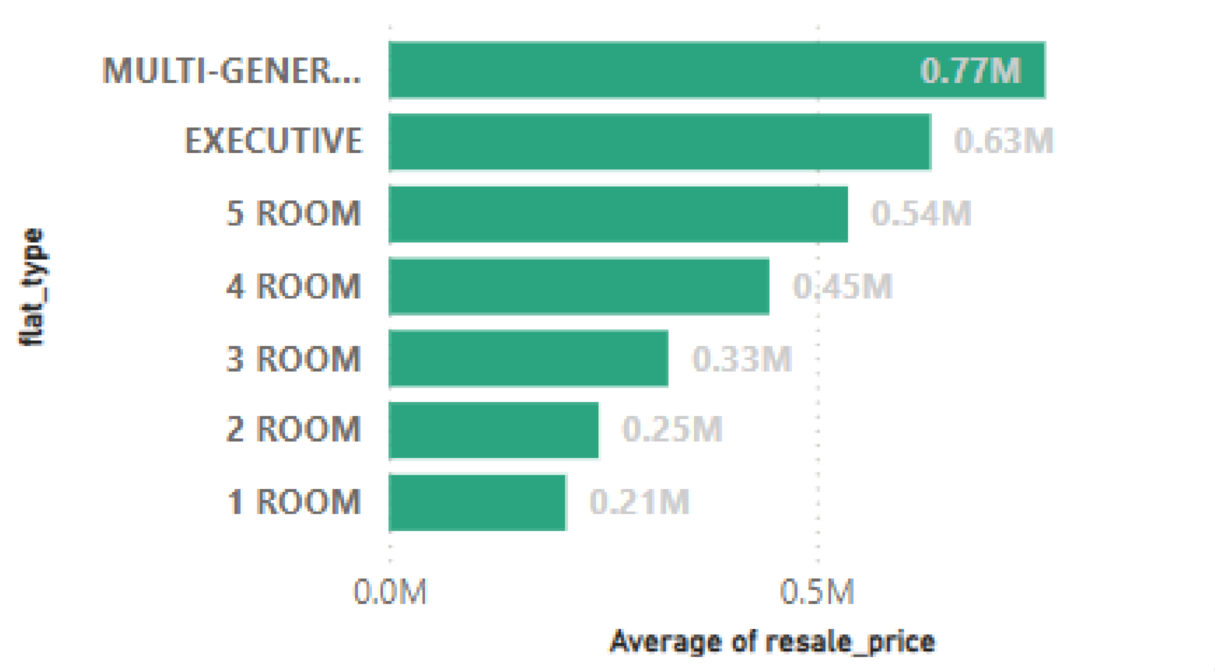
Town: All

town	Total resale_price	No. of Transactions
SENGKANG	5070810154	11069
TAMPINES	4981625953	10506
JURONG WEST	4721029365	11451
Total	**67658993577**	**150634**

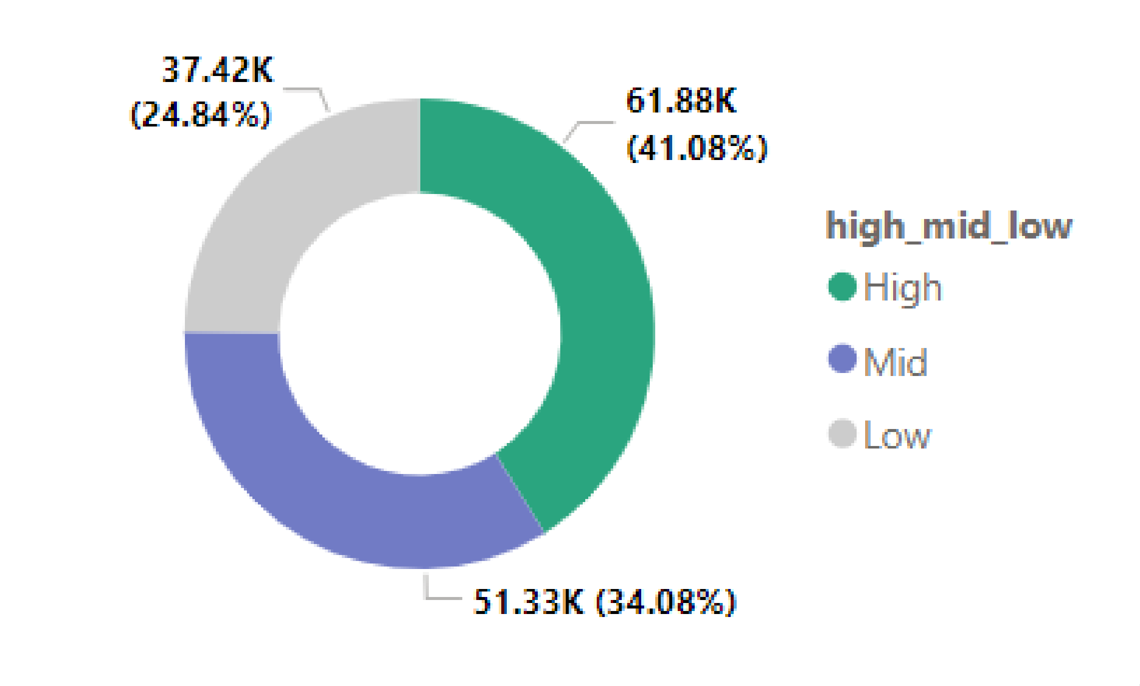
Resale trend by period



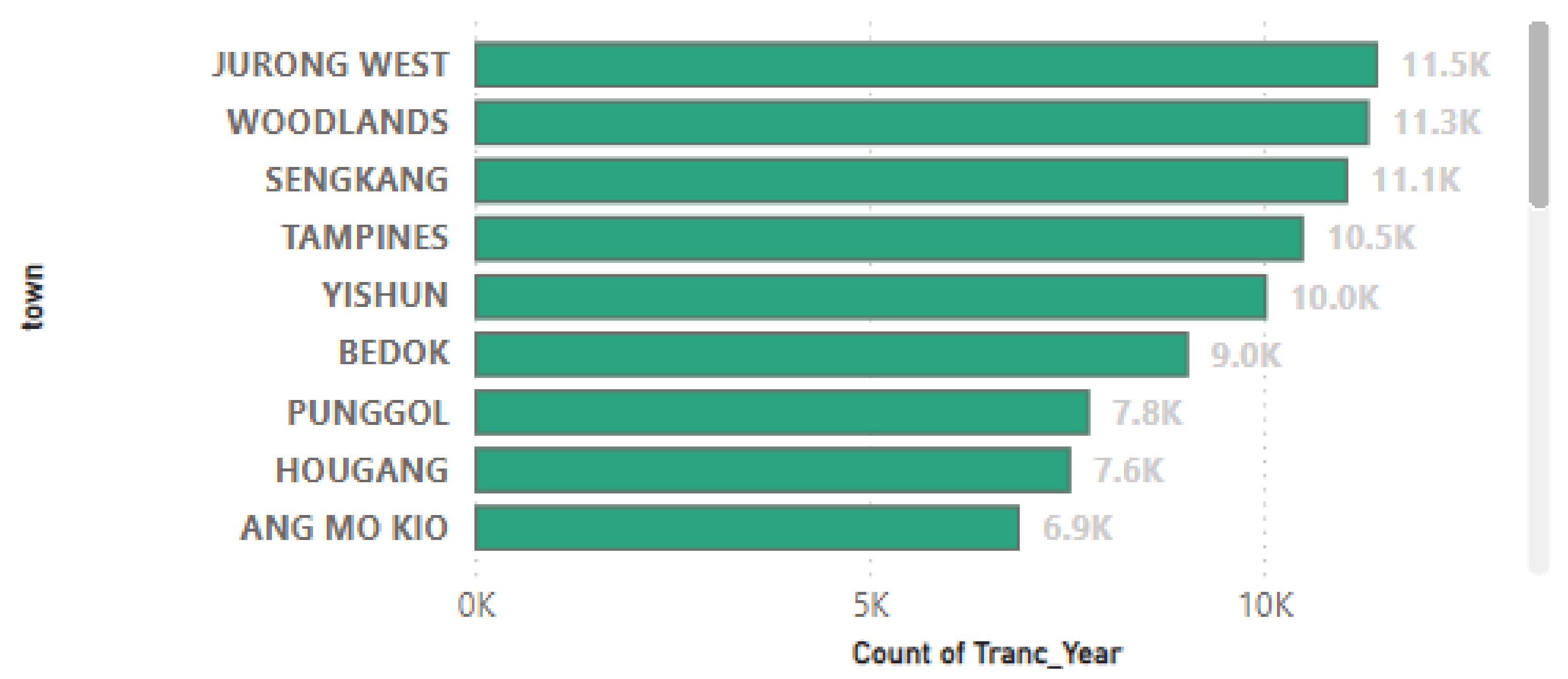
Average Resale Price by flat_type & Story range



No. of Transactions by Story Category

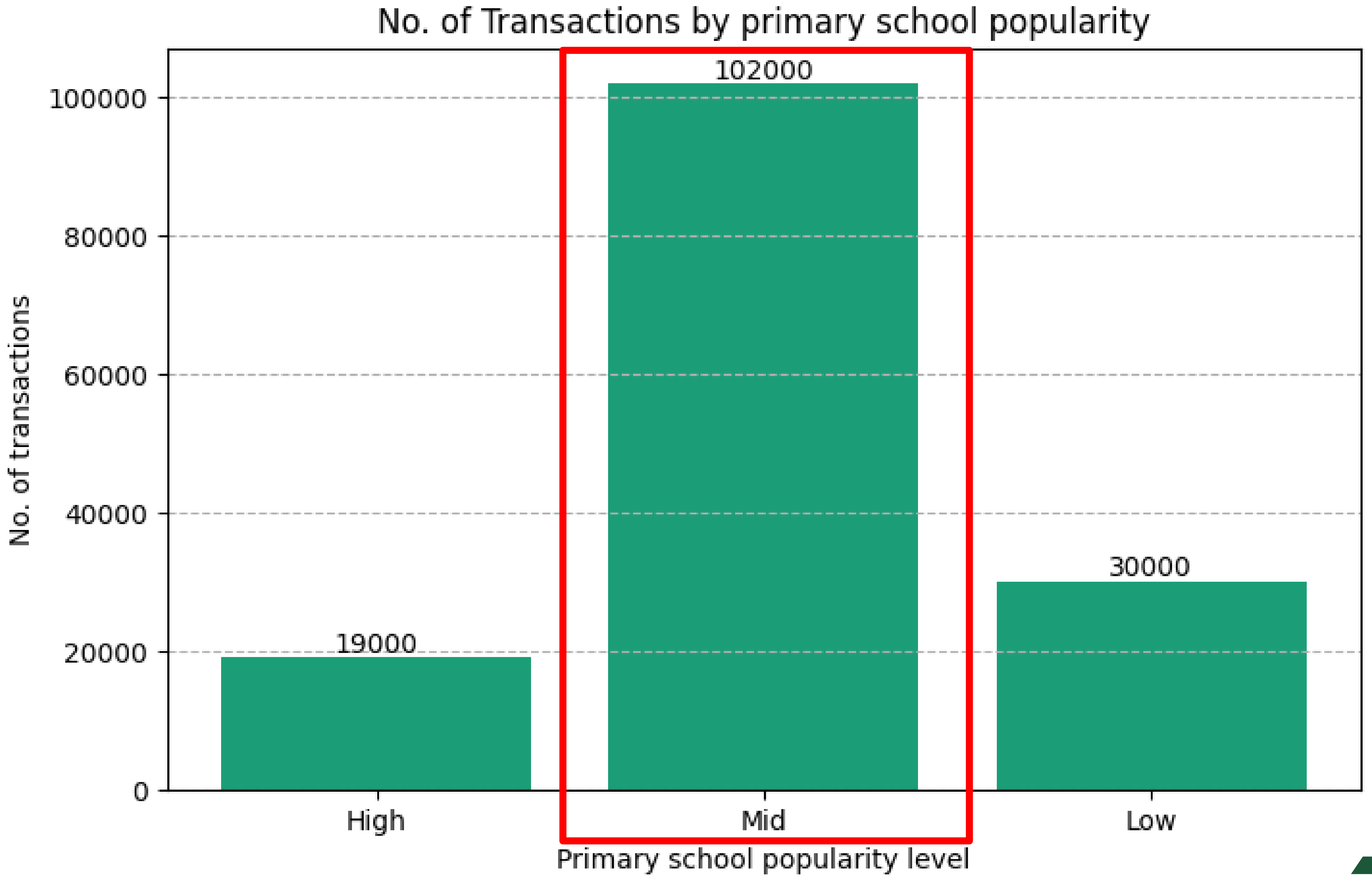


No. of Transactions by Town, Floor lvl & Flat model



Highest number of transactions for primary school with average popularity

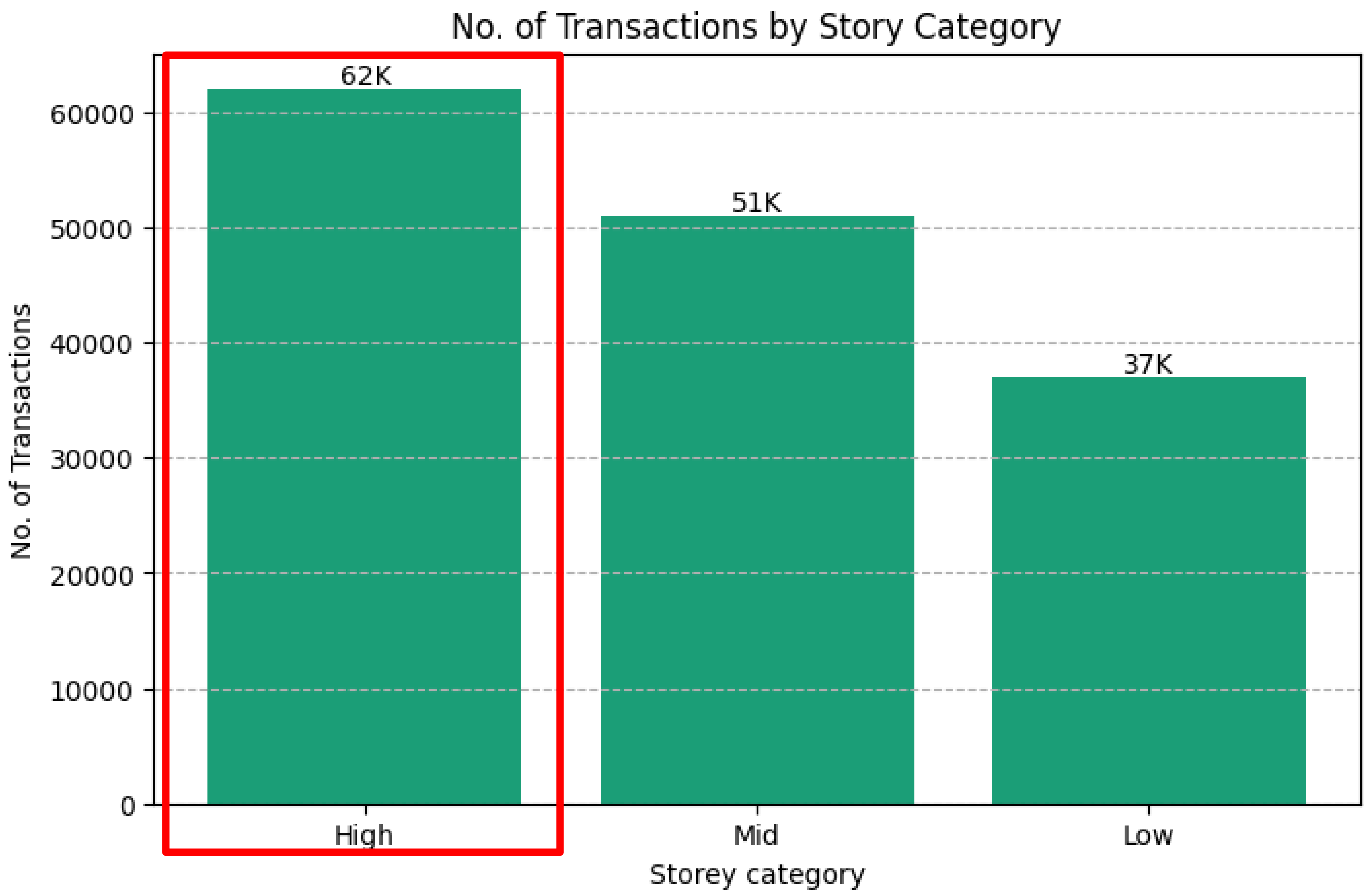
Example: Closing a transaction near to Gan Eng Seng Primary School will fetch you a sales commission of
est. $12, 769!



High Storey has the highest number of transactions

High storey demand average resale price of $453,474.

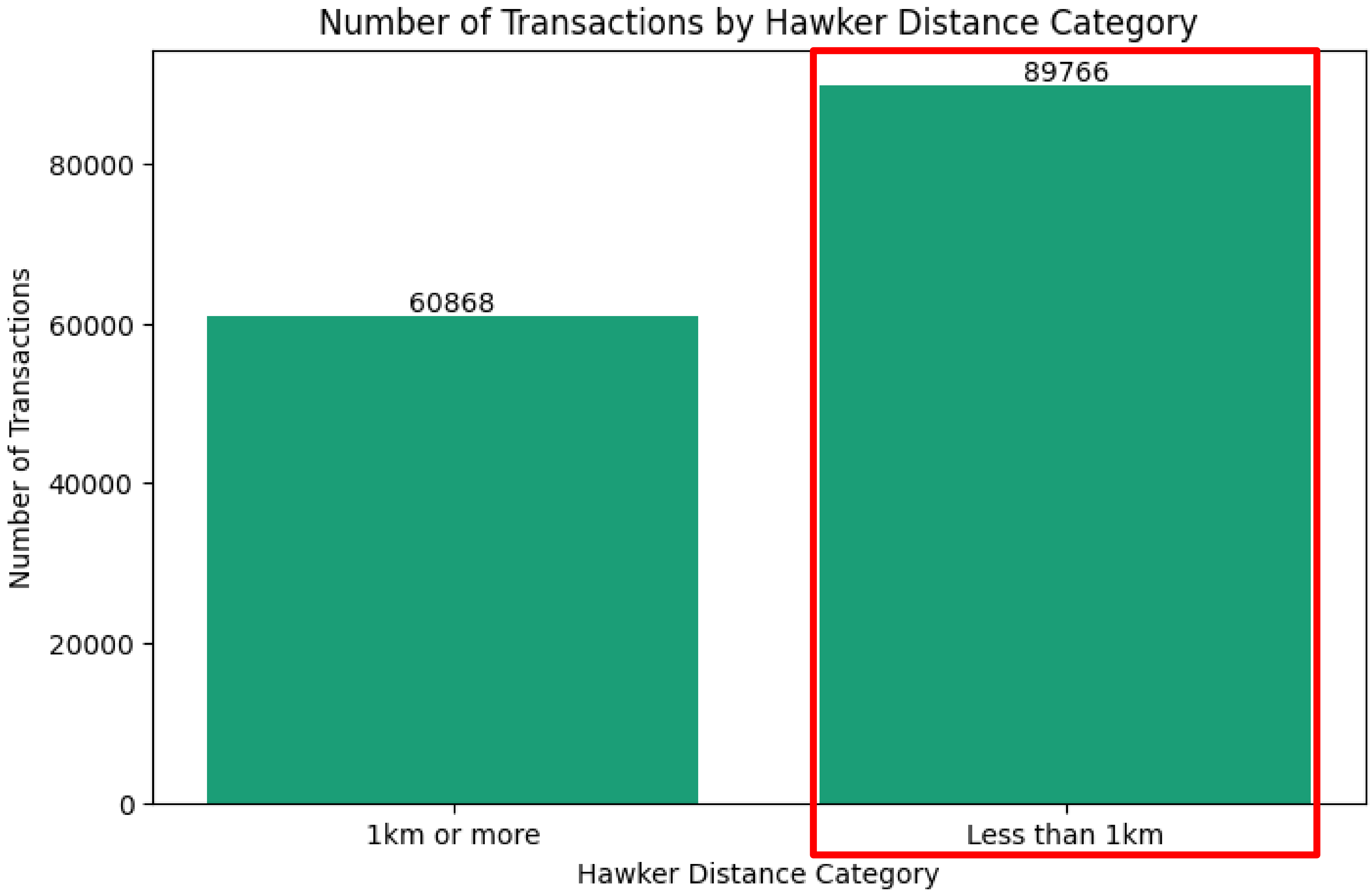
Sale commission of **est. $9,069!**



Flats nearer to hawker has higher number of transactions

Less than 1km demand average resale price of $449,700.

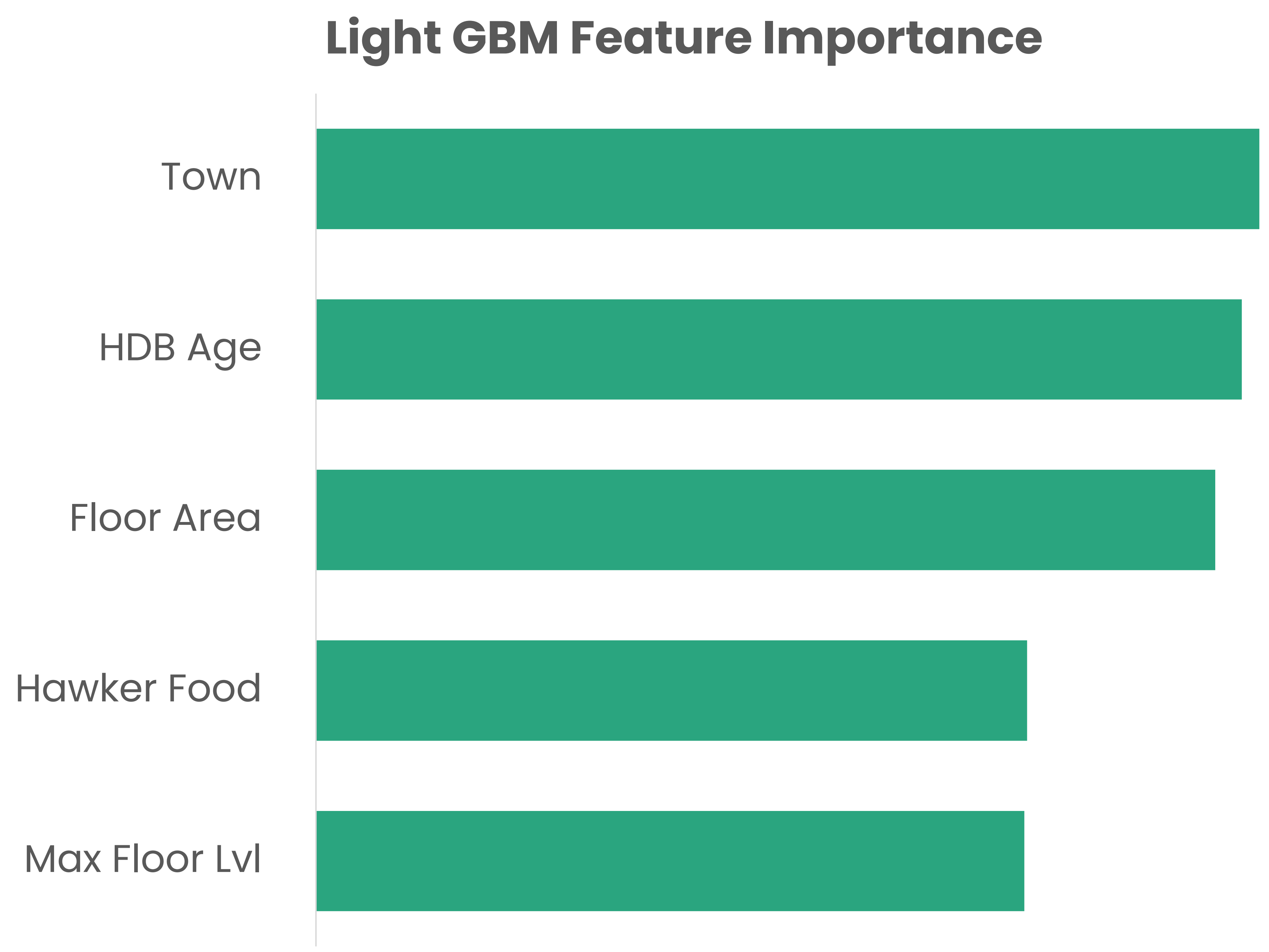
Sale commission of **est. $8,994!**

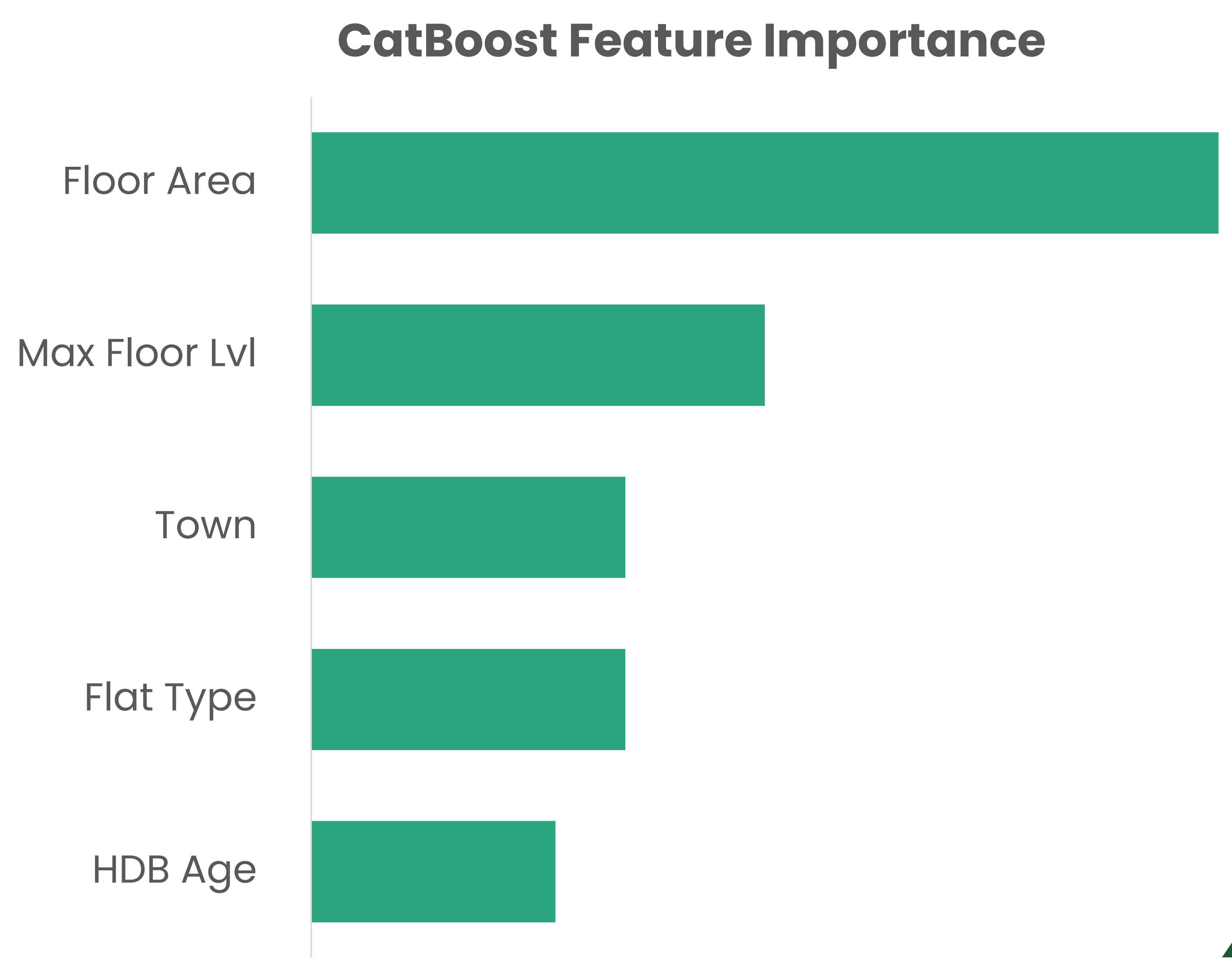


Top 5 models

	Train RMSE	Test RMSE	Train R2	Test R2	Model Run Time (sec)
CatBoost	25,506	25,726	0.9684	0.9677	7.1940
Extra Trees Regressor	2,843	26,685	0.9996	0.9653	14.5840
Random Forest	9,796	25,954	0.9953	0.9672	19.1130
Light GBM	30,903	31,724	0.9535	0.9509	0.783
Decision Tree	2,842	35,901	0.9996	0.9372	1.5770

Light GBM: Not heavily reliant on a particular variable





Model of Choice: LightGBM

95%

Accuracy

RSME: 31,724
Run time: 0.783

Increase Generalization

Avoid Feature bias

Lower Model Run-time

Lower Noise Sensitivity

Streamlit Demo

https://dj-bab-hdb-sales-predictor.streamlit.app/

Revenue per agent dropped by 10%

Increase Market Volatility

- Rise in HDB resale demand
- Increase in market volatility and million-dollars flats “outliers”

Sales Cycle increased by 20%

- Higher price volatility leads to inaccurate price predictions
- Time-consuming process to evaluate trends
- Pricing relies on subjective opinion

Increasingly Competitive Market

- Increasing number of real estate agents in Singapore: ~10% increase between 2022 to 2024
- Some agents are offering 1% commission fee instead of the usual 2%

10%

Dropped Revenue per agent

POC with WOW 50 real estate agents

WOW App has the potential to increase your company's bottom-line by **$3M** per year.

Forecast and Survey

200% **Increase in agents' revenue**

>50% reduction in sales cycle, with the potential to double monthly sales

2x **Buyers' Representative**

With the app, >20% of buyers engaged our agents; up from 10%

3x **Sellers' Representative**

With the app, 60% of sellers engaged our agents; up from 20%

Further Enhancements

Increase Loading Speed

Integration with WOW existing system

Map Features to identify nearby amenities

Any other enhancements required; per user feedback

Extend solution to condominiums

Path to Commercialization

June’25
White label solution to other corporates
Extend solution to condominium and other markets

Feb’25
Go Live
Deployment to all agents

Dec’24
Soft Launch (Beta)
Controlled roll out
Integrate with our existing system

Now
Sept’24
Product Development
Testing

Comprehensive Solution that delivers 6x return

90%

Problem

90% of our agents have trouble closing sales.

Current process is time consuming and rely on subjective opinions, leading to inaccurate predictions.

$0.5M

Solution

Leverage cutting-edge machine learning algorithms, the app analyzes vast amounts of data to identify trends and patterns.

$3M

Impact

Increase agents' productivity by 125%

Increase company’s revenue by 3m per annum.

Thank You

Team’s Reflection

wow!

What went well

- PowerBI and Streamlit were deployed successfully
- Modelling was completed within allocated time frame
- Data engineering improved our model accuracy
- Team Bonding

What didn’t go well

- Some analysis were not relevant to the presentation
- Overlapping work done
- Markdown was insufficient, hence the team needs to further improve it
- Translate technical information into relevant information for stakeholder

Improvements

- Alignment and Communication
- Important to keep our Trello board up to date so that everyone knows their tasks
- Proper allocation of tasks according to team’s capability

Appendix

Streamlit App Demo

Choose Options

Select options from the dropdown menus to display the predictions and data.

Select Town:

BISHAN

Select Flat Type:

5 ROOM

Select Lease Commencement Date:

2019

Select Storey Range:

10 TO 12

🏠 HDB Resale Price Predictor

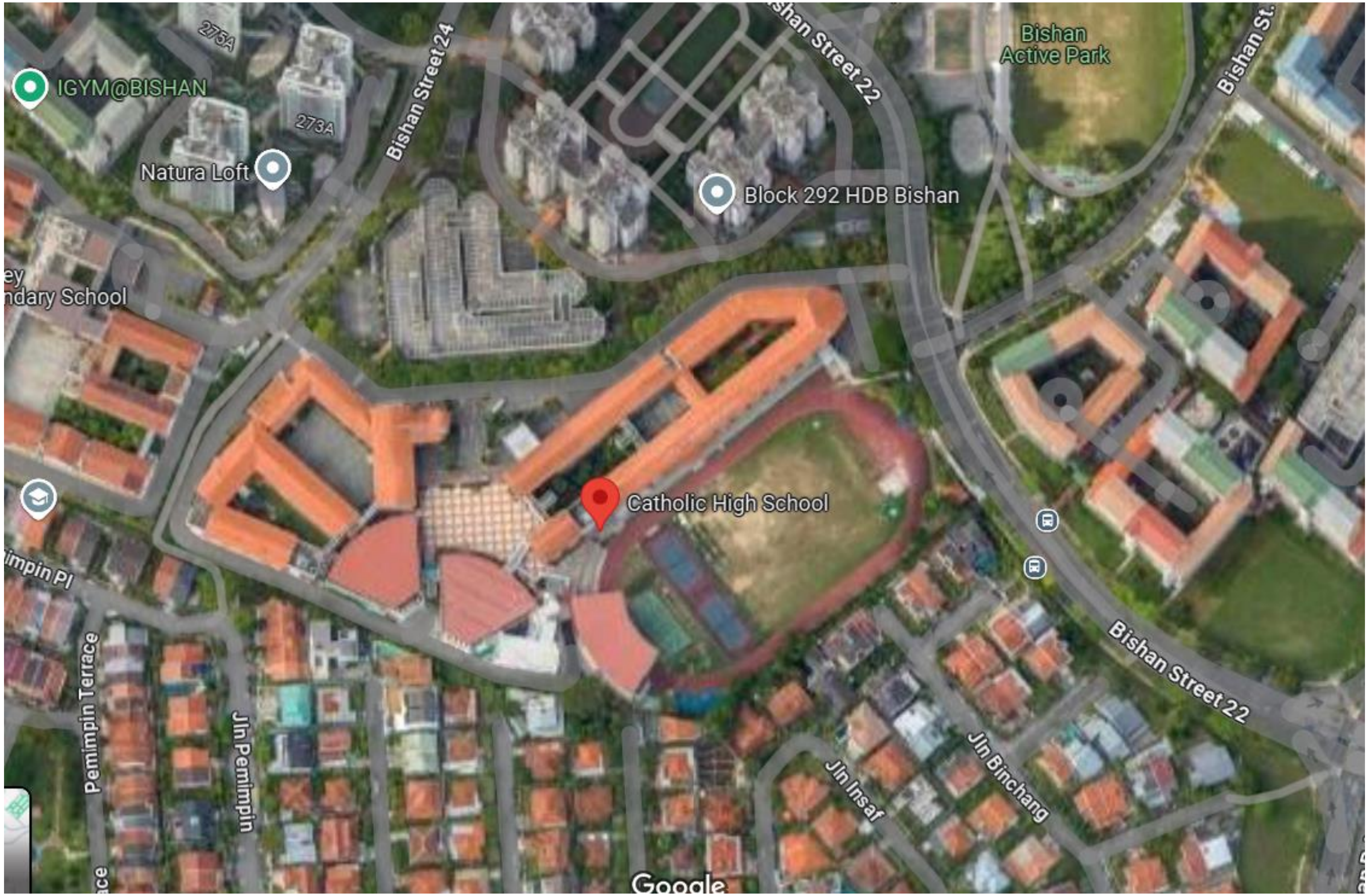
This HDB Resale Price Predictor is created by DJ BAB! Using a LightGBM regression predictive model of history data from 2012-2021

All Predictions Results

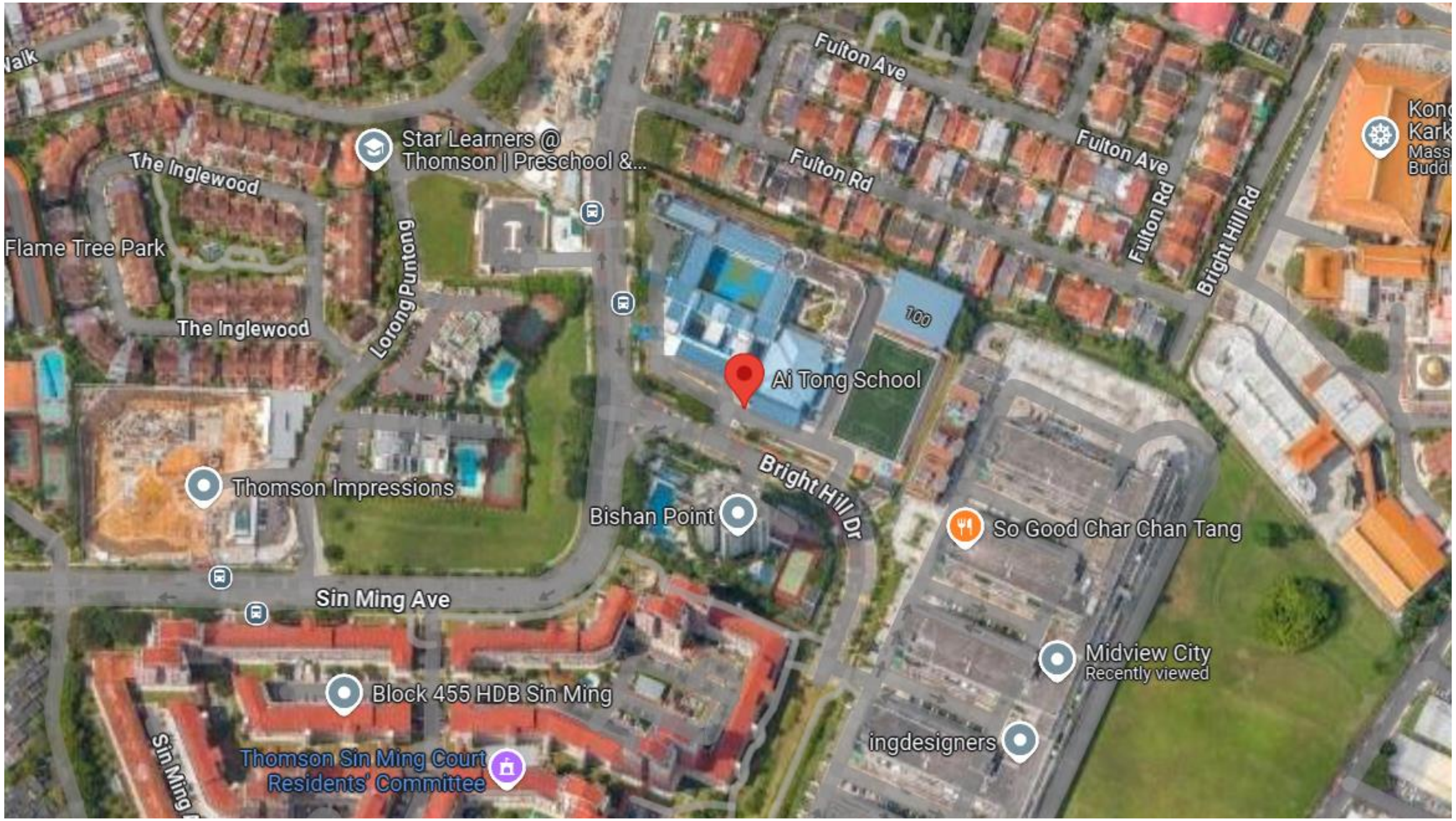
Resale Price: $879,381.23

	Town	Flat Type	Lease Commencement Date	Storey Range	Floor Area (SQ FT)	Resale Price
0	BISHAN	5 ROOM	1975	10 TO 12	1,046	$728,897.28
1	BISHAN	5 ROOM	2019	10 TO 12	1,046	$879,381.23

Catholic High School



Ai Tong School



Mean resale price for transactions near to Primary School with average popularity

pri_sch_name	resale_price
Cantonment Primary School	704068
Kuo Chuan Presbyterian Primary School	644453
Gan Eng Seng Primary School	638478
Queenstown Primary School	620912
Zhangde Primary School	616046
Saint Joseph's Institution Junior	606185
Kong Hwa School	571901
Maris Stella High School	559182
Elias Park Primary School	557036
Changkat Primary School	555870
Alexandra Primary School	544970
Yangzheng Primary School	537439
Poi Ching School	535687
Haig Girls' School	534421
Blangah Rise Primary School	530234
Ngee Ann Primary School	527871

Mean resale price for transactions based on storey categories

high_mid_low	resale_price
High	453474
Mid	450699
Low	439920

Mean resale price for transactions based on hawker_distance categories

	hawker_distance_category	resale_price
0	1km or more	448368
1	Less than 1km	449700

90% of agents demand an app for price recommendations

Advanced Algorithms

Leveraging cutting-edge machine learning algorithms, the app analyzes vast amounts of data to identify trends and patterns.

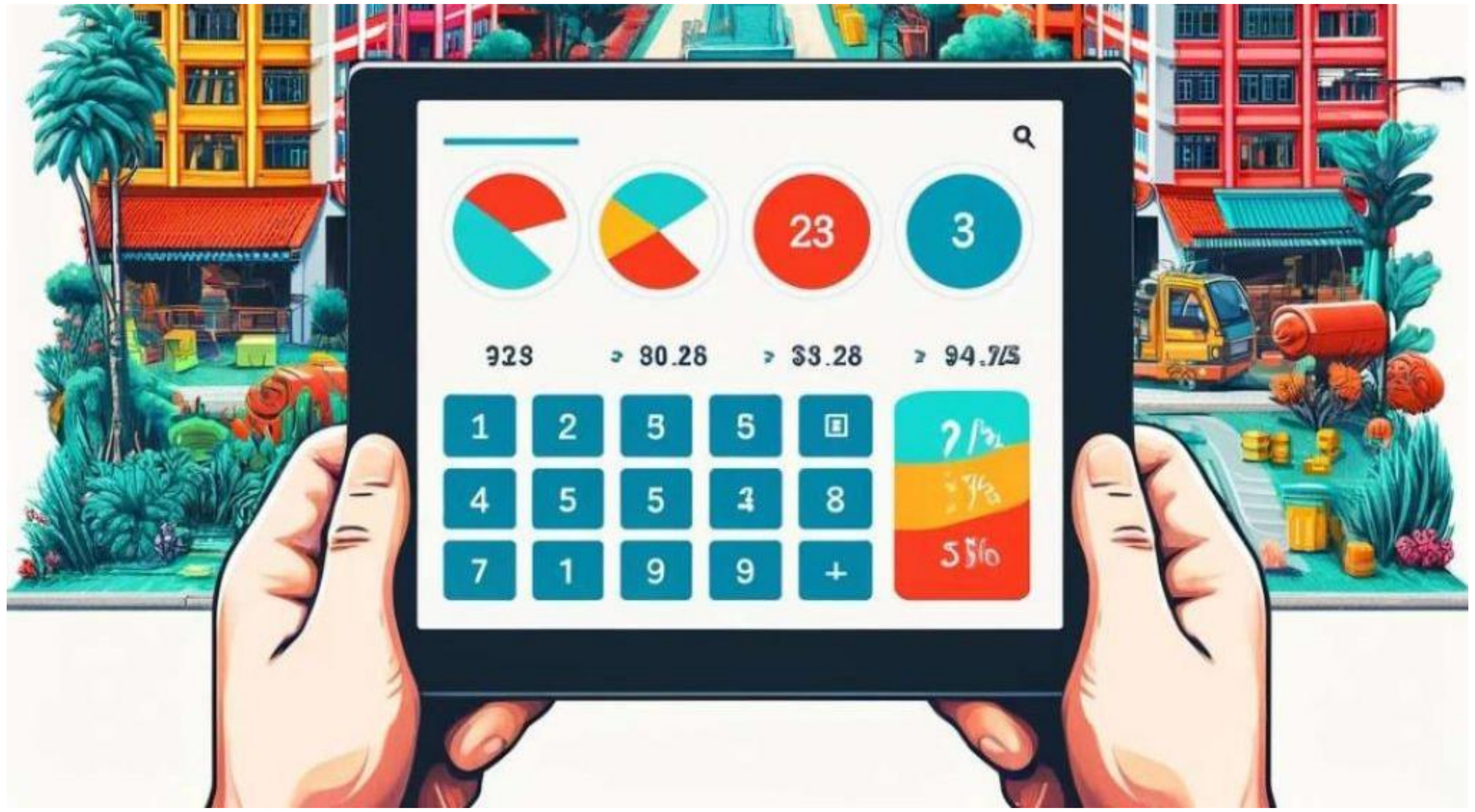
Real-Time Data

The app integrates real-time data feeds from multiple sources, including public records, market trends, and property listings.

Personalized Predictions

Provides personalized price predictions tailored to individual HDB Flats based on their unique characteristics and location.

The App that WOWs!



How the App Works

1

Data Collection

The app gathers data from various sources including HDB resale statistics.

2

Data Processing

Collected data is cleaned, standardized, and transformed into a format suitable for analysis.

3

Machine Learning

Advanced algorithms analyze processed data to identify patterns, trends, and relationships.

4

Price Prediction

The app generates personalized price predictions for specific HDB flat.

Cleaning Train & Test Dataset

Remove column in csv file for simplicity and reduce redundancy
e.g. postal, floor_area_sqm

Remove null value
e.g. replacing the null value for 'Mall_Nearest_Distance' with average distance

Check for duplicate value
e.g. check for duplicate value based on id column

Check for correct format in the dataset
e.g. postal should be int and not object

Data Augmentation & Feature Engineering

Create Boolean Value for Mall/MRT/Hawker within 1km

Create Proximity: Sum of all Boolean value of Mall/MRT/Hawker within 1km (Highest 3 to lowest 0)

Create storey_ratio: low <= 0.33, high > 0.66 and mid [Comparing Mid Storey with Max Floor Level]. Additional column to reflect it as High, Mid and Low (Storey Category).

Create pri_sch_pop: based on primary school vacancy (highp <= 35, lowp > 70, averagep)

Exploratory Data Analysis (Train)

Correlation Analysis	
Variable	**Correlation**
Flat_type (encode)	0.66
Max_floor_lvl	0.50
year_completed	0.35
sec_sch_nearest_dist	0.10
MRT_within_1km_boolean	0.09
Proximity	0.08
Pri_sch_pop	0.02
High_mid_low	-0.01
Mall_Nearest_Distance	-0.09
mrt_nearest_distance	-0.13
hdb_age	-0.35

Exploratory Data Analysis

Expensive Town	
Town	Aver. resale price
BUKIT TIMAH	**S$704, 417**
BISHAN	S$618, 370
CORE CENTRAL REGION	S$604, 930
BUKIT MERAH	S$555, 344
CLEMENTI	S$466, 308
BUKIT PANJANG	S$436, 084
BEDOK	S$419, 066
ANG MO KIO	S$414, 215
CHOA CHU KANG	S$413, 042
BUKIT BATOK	S$397, 436

Popular Flat Type	
Flat_type	count
4 ROOM	**61136**
3 ROOM	39060
5 ROOM	36415
EXECUTIVE	11989
2 ROOM	1896
1 ROOM	82
MULTI-GENERATION	56

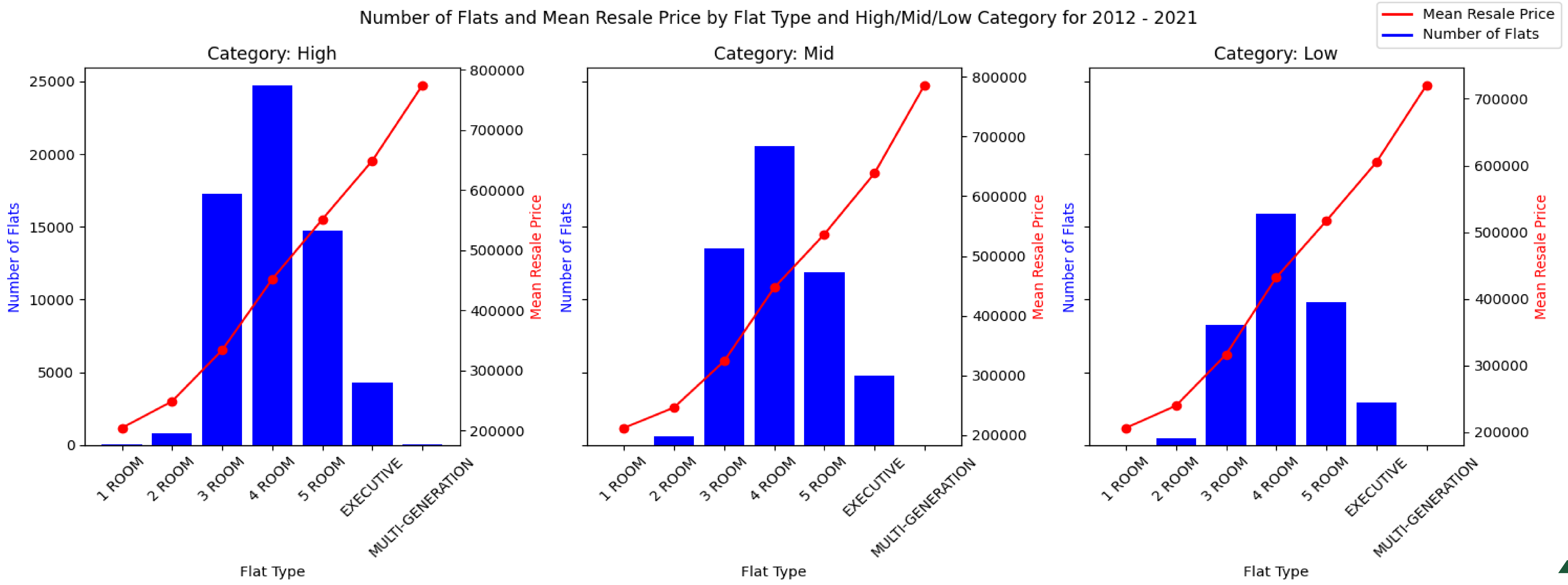
Popular Storey Category		
Flat_type	Aver. resale price	count
High	**453474**	**61882**
Mid	450699	51334
Low	439920	37418

Correlation Analysis	
Variable	Correlation
Flat_type (encode)	0.66
Max_floor_lvl	0.50
year_completed	0.35
sec_sch_nearest_dist	0.10
MRT_within_1km_boolean	0.09
Proximity	0.08
Pri_sch_pop	0.02
High_mid_low	-0.01
Mall_Nearest_Distance	-0.09
mrt_nearest_distance	-0.13
hdb_age	-0.35

Aver. resale price based on pri sch popularity		
Pri sch popularity	Aver. resale price	count
High	S$466, 200	19217
Average	S$448, 170	101888
Low	S$446, 235	29529

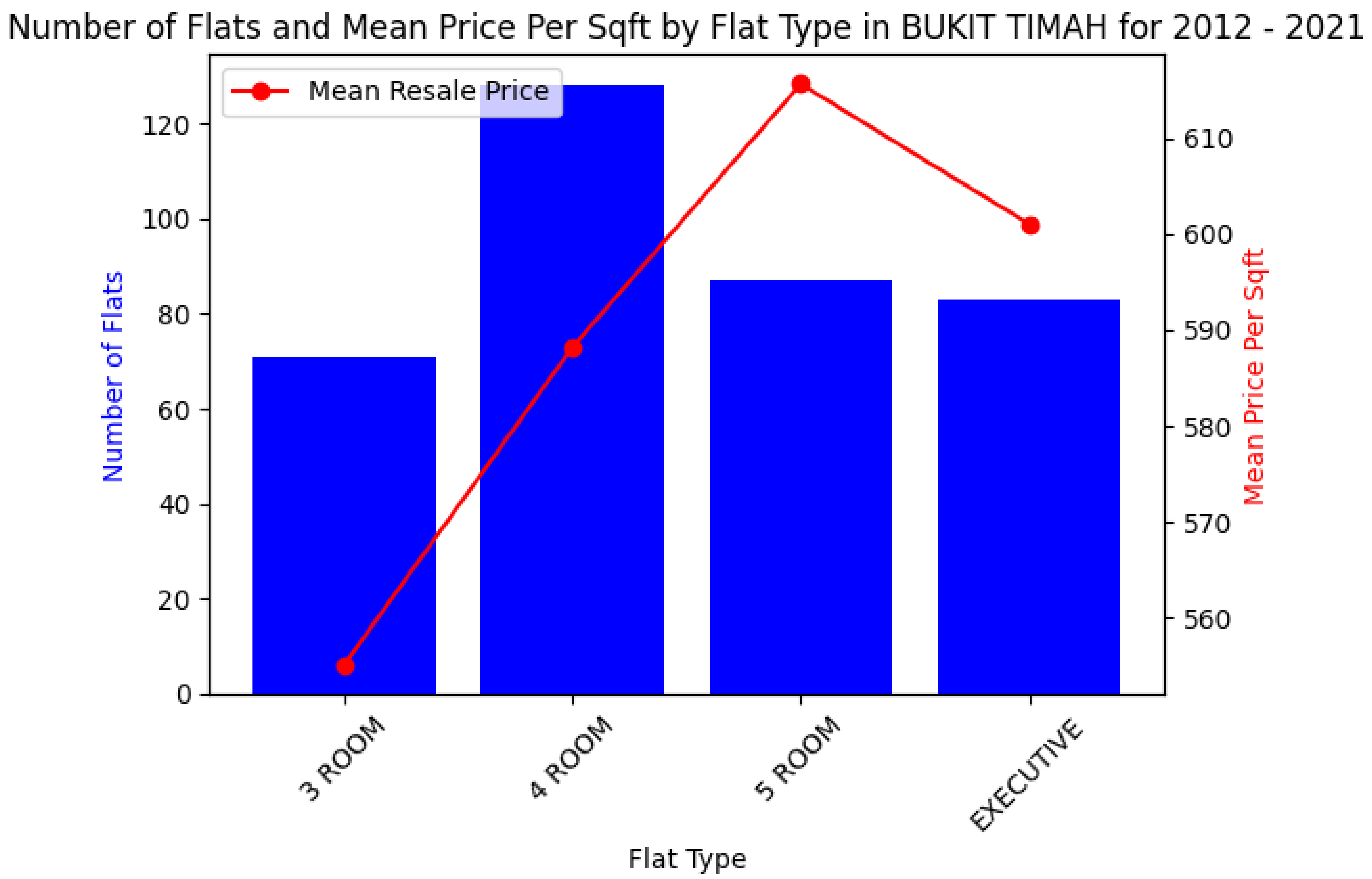
Aver. resale price based on proximity		
Proximity	Aver. resale price	count
3	$S464, 357	59135
0	$S441, 753	1596
2	$S441, 174	63751
1	$S434, 726	26152

4 Room Flat: Most popular flat type among floor category



Bukit Timah: Most popular town with highest resale price

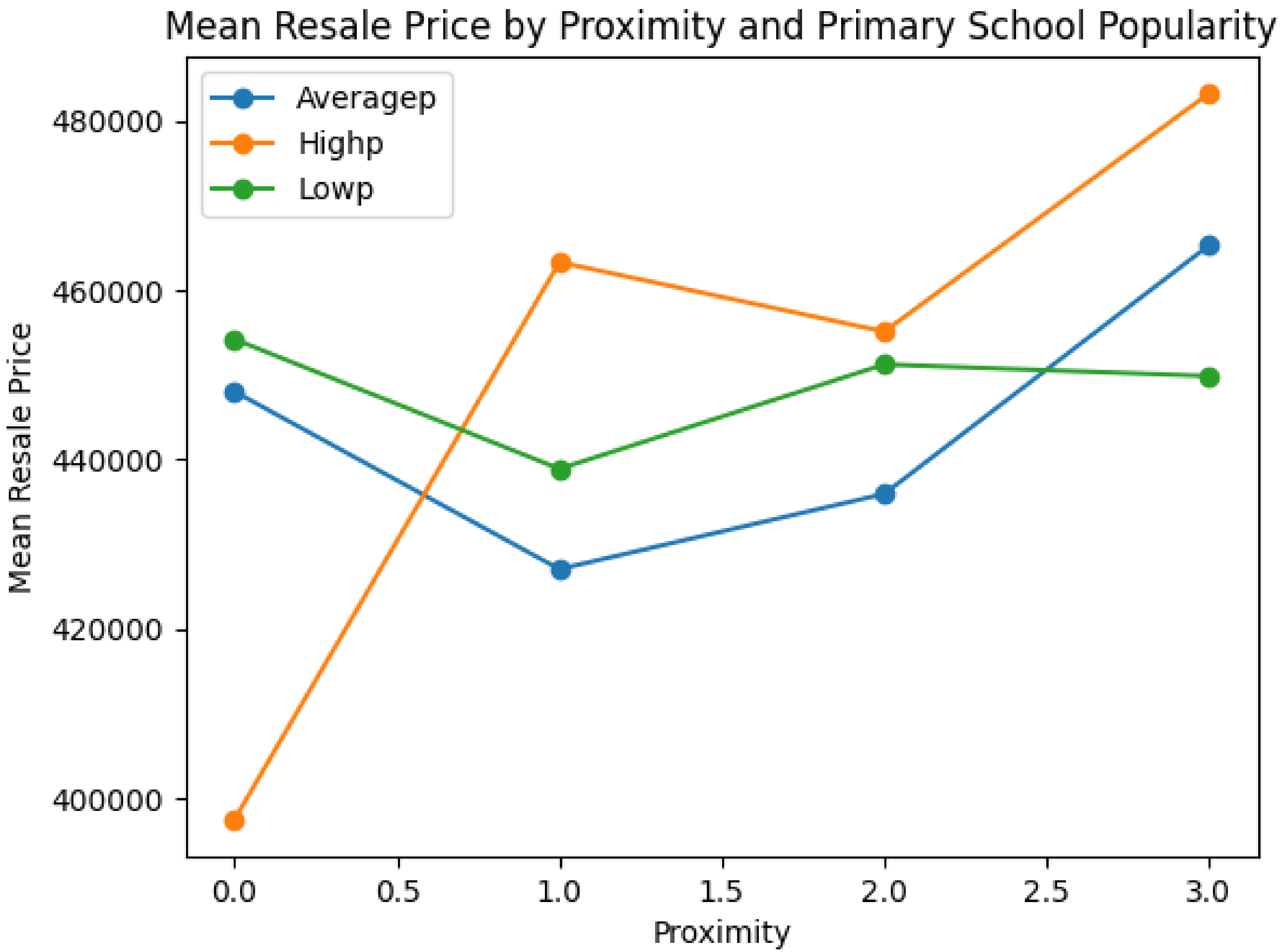
Town	Transaction Year 2012 – 2021 Ave. resale price (S$)
BUKIT TIMAH	704, 417
BISHAN	618, 370
CORE CENTRAL REGION	604, 930
BUKIT MERAH	555, 344
CLEMENTI	466, 308
BUKIT PANJANG	436, 084
BEDOK	419, 066
ANG MO KIO	414, 215
CHOA CHU KANG	413, 042
BUKIT BATOK	397, 436

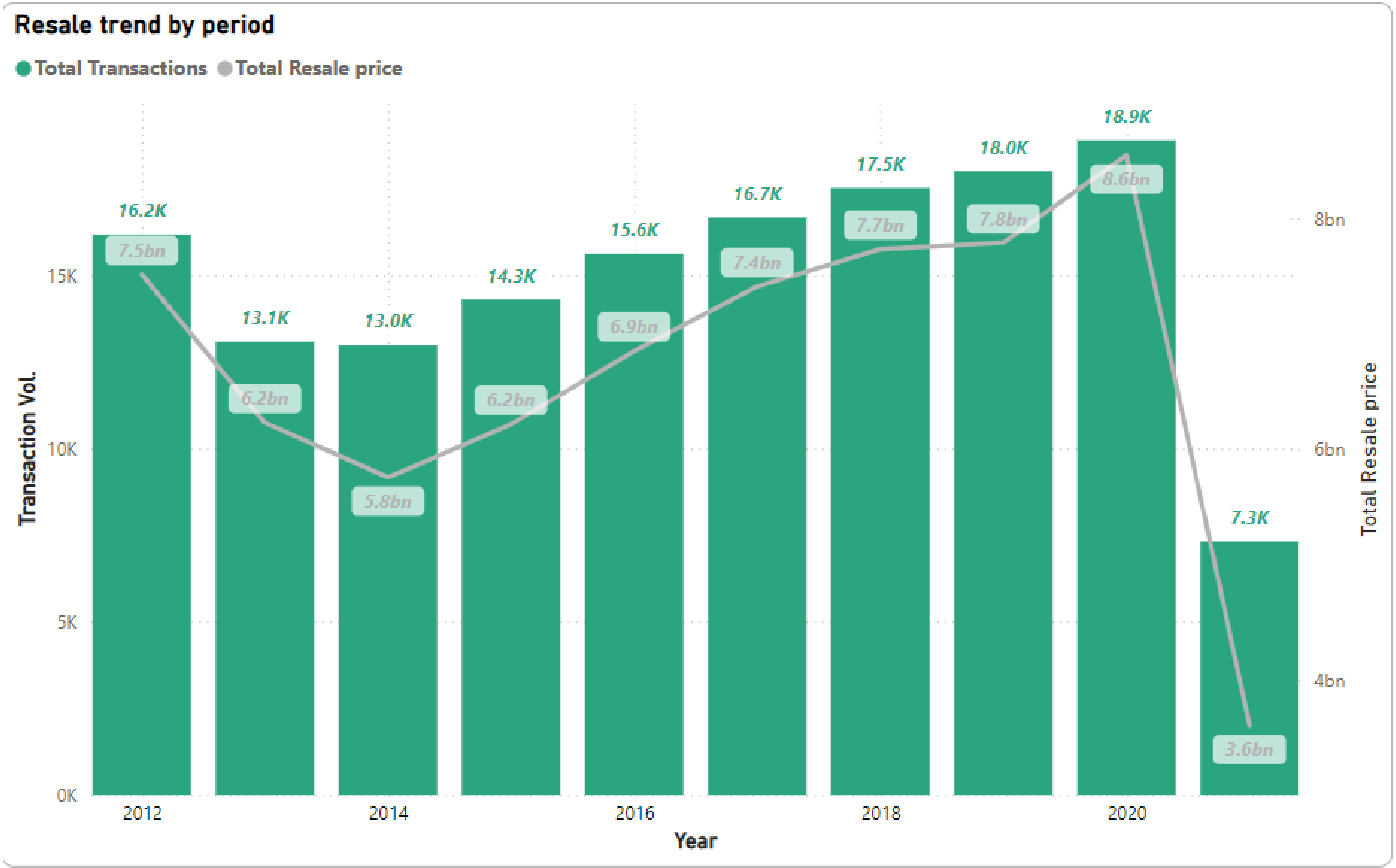


Exploratory data analysis (Train)

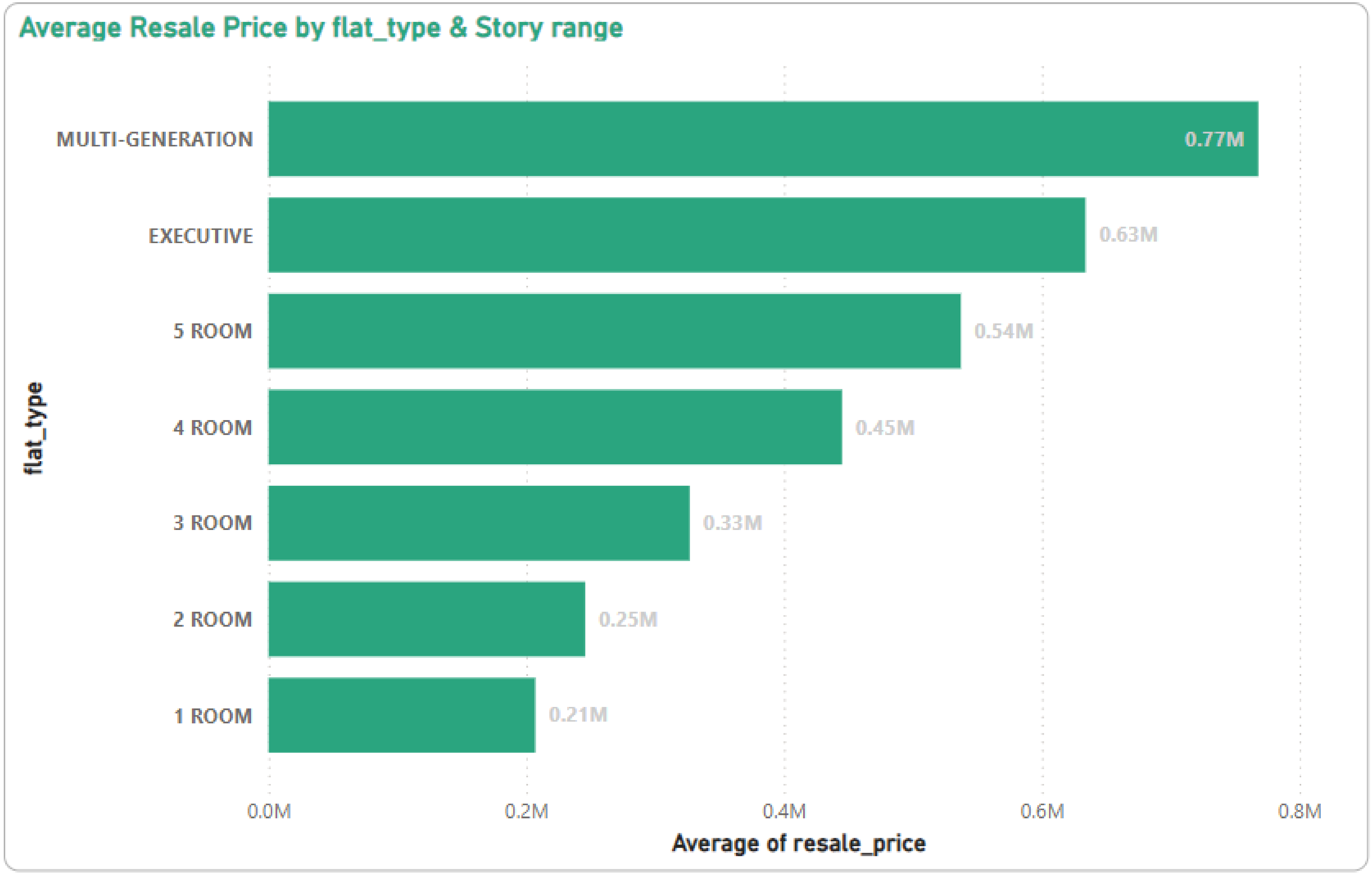
Aver. resale price: Pri sch popularity 2012 – 2021		
Pri sch popularity	**Aver. resale price**	**Count**
High	S$466, 200	19,217
Average	S$448, 170	101,888
Low	S$446, 235	29,529

Aver. resale price based on proximity 2012 - 2021		
Proximity	Aver. resale price	count
3	$S464, 357	59135
0	$S441, 753	1596
2	$S441, 174	63751
1	$S434, 726	26152

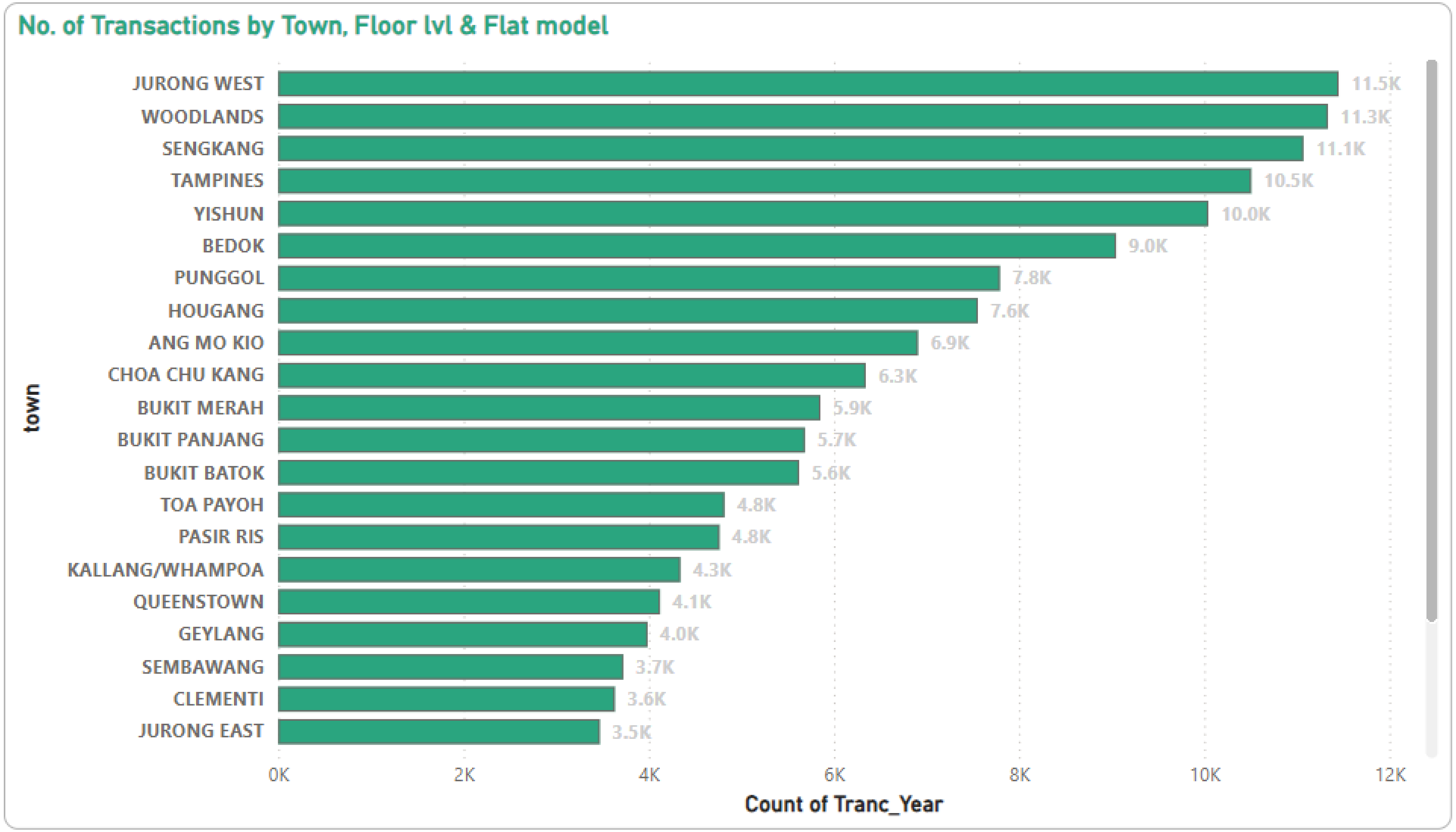




Resale trend by period
Total Transactions
Total Resale price
Transaction Vol.
15K
10K
5K
0K
16.2K
13.1K
13.0K
14.3K
15.6K
16.7K
17.5K
18.0K
18.9K
7.3K
7.5bn
6.2bn
5.8bn
6.2bn
6.9bn
7.4bn
7.7bn
7.8bn
8.6bn
3.6bn
8bn
6bn
4bn
Total Resale price
2012
2014
2016
2018
2020
Year

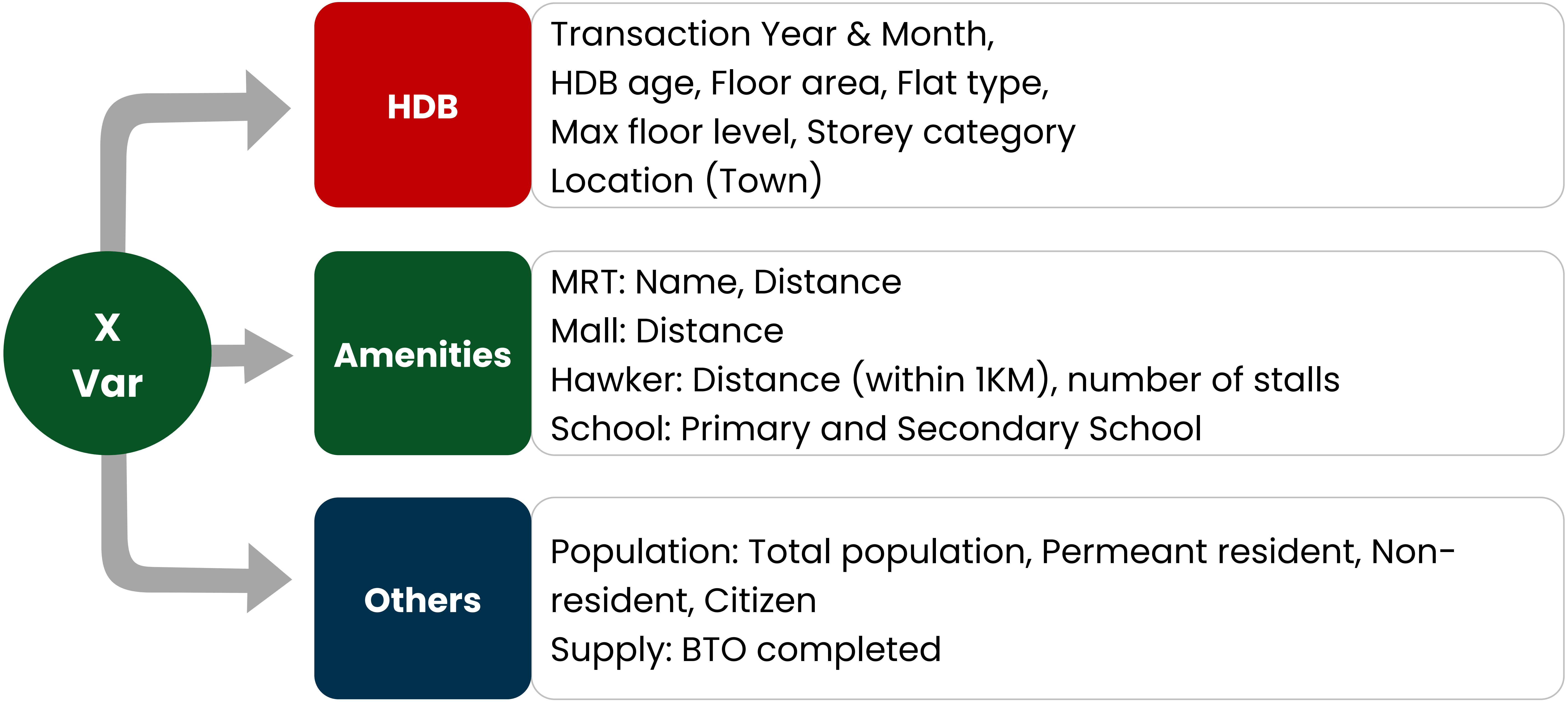


Average Resale Price by flat_type & Story range
MULTI-GENERATION
EXECUTIVE
5 ROOM
4 ROOM
3 ROOM
2 ROOM
1 ROOM
0.77M
0.63M
0.54M
0.45M
0.33M
0.25M
0.21M
flat_type
0.0M
0.2M
0.4M
0.6M
0.8M
Average of resale_price



No. of Transactions by Town, Floor lvl & Flat model
town
JURONG WEST 11.5K
WOODLANDS 11.3K
SENGKANG 11.1K
TAMPINES 10.5K
YISHUN 10.0K
BEDOK 9.0K
PUNGGOL 7.8K
HOUGANG 7.6K
ANG MO KIO 6.9K
CHOA CHU KANG 6.3K
BUKIT MERAH 5.9K
BUKIT PANJANG 5.7K
BUKIT BATOK 5.6K
TOA PAYOH 4.8K
PASIR RIS 4.8K
KALLANG/WHAMPOA 4.3K
QUEENSTOWN 4.1K
GEYLANG 4.0K
SEMBAWANG 3.7K
CLEMENTI 3.6K
JURONG EAST 3.5K
0K
2K
4K
6K
8K
10K
12K
Count of Tranc_Year

Features Selected



25 Features Selected

[10]:

```
# Select the variables used in the model
selected_variables = [
    'Tranc_Year', 'Tranc_Month', 'floor_area_sqft', 'flat_type_Encoded', 'max_floor_lvl',
    'storey_high_mid_low_Encoded', 'hdb_age', 'town_Encoded', 'mrt_name_Encoded',
    'mrt_nearest_distance', 'Mall_Nearest_Distance',
    'Hawker_Nearest_Distance', 'Hawker_Within_1km_boolean', 'hawker_food_stalls',
    'hawker_market_stalls', 'pri_sch_name_Encoded', 'pri_sch_nearest_distance',
    'pri_sch_pop_Encoded', 'sec_sch_name_Encoded', 'sec_sch_nearest_dist',
    'Total Population', 'Permanent Resident Population (Number)',
    'Non-Resident Population (Number)', 'Singapore Citizen Population (Number)','BTO completed'
]

# Count the number of selected variables
num_selected_variables = len(selected_variables)

# Print the count
print(f'The number of selected variables is: {num_selected_variables}')
```

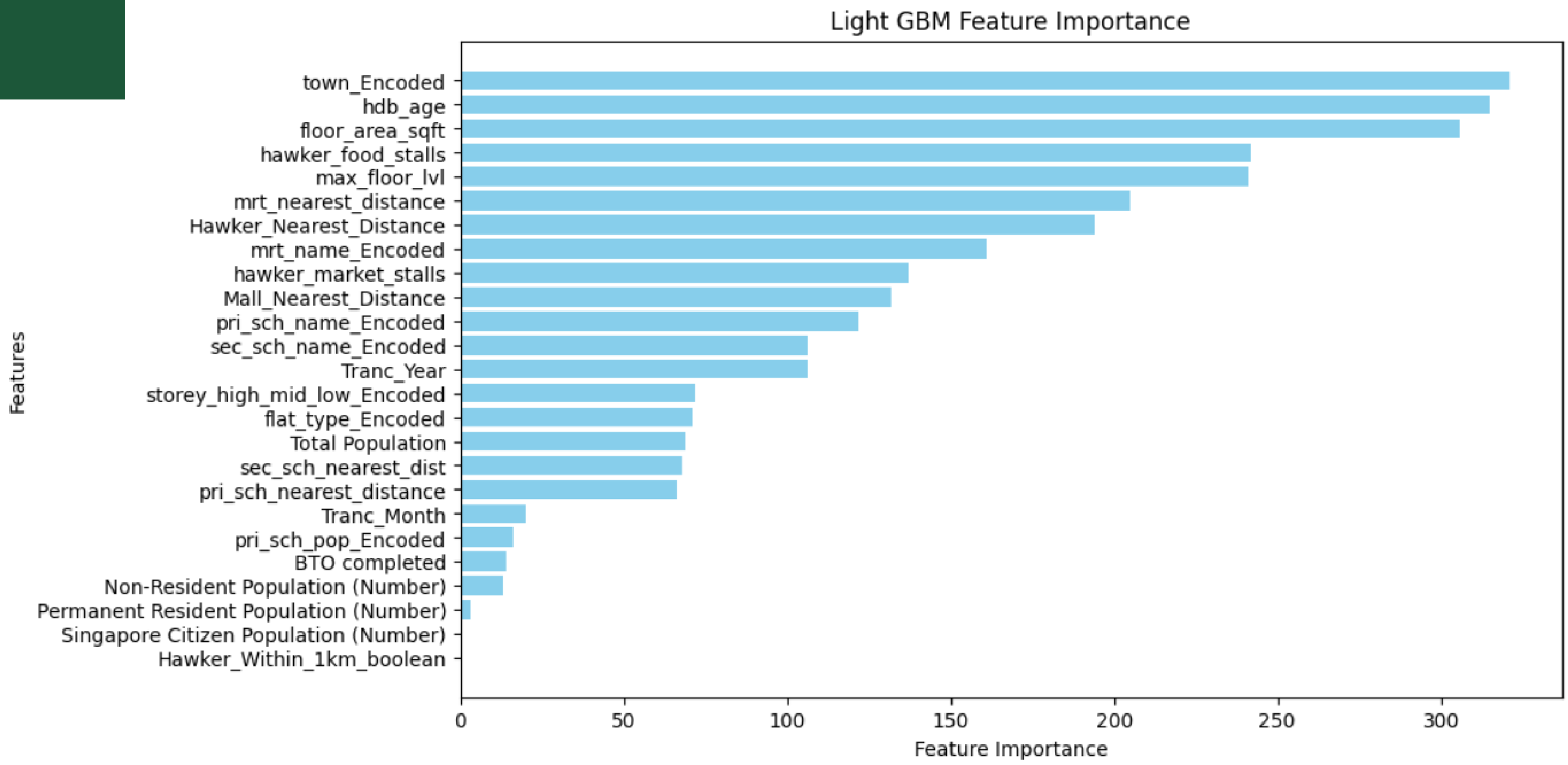
```
The number of selected variables is: 25
```

Pycaret to identify top 5 models with the lowest RMSE

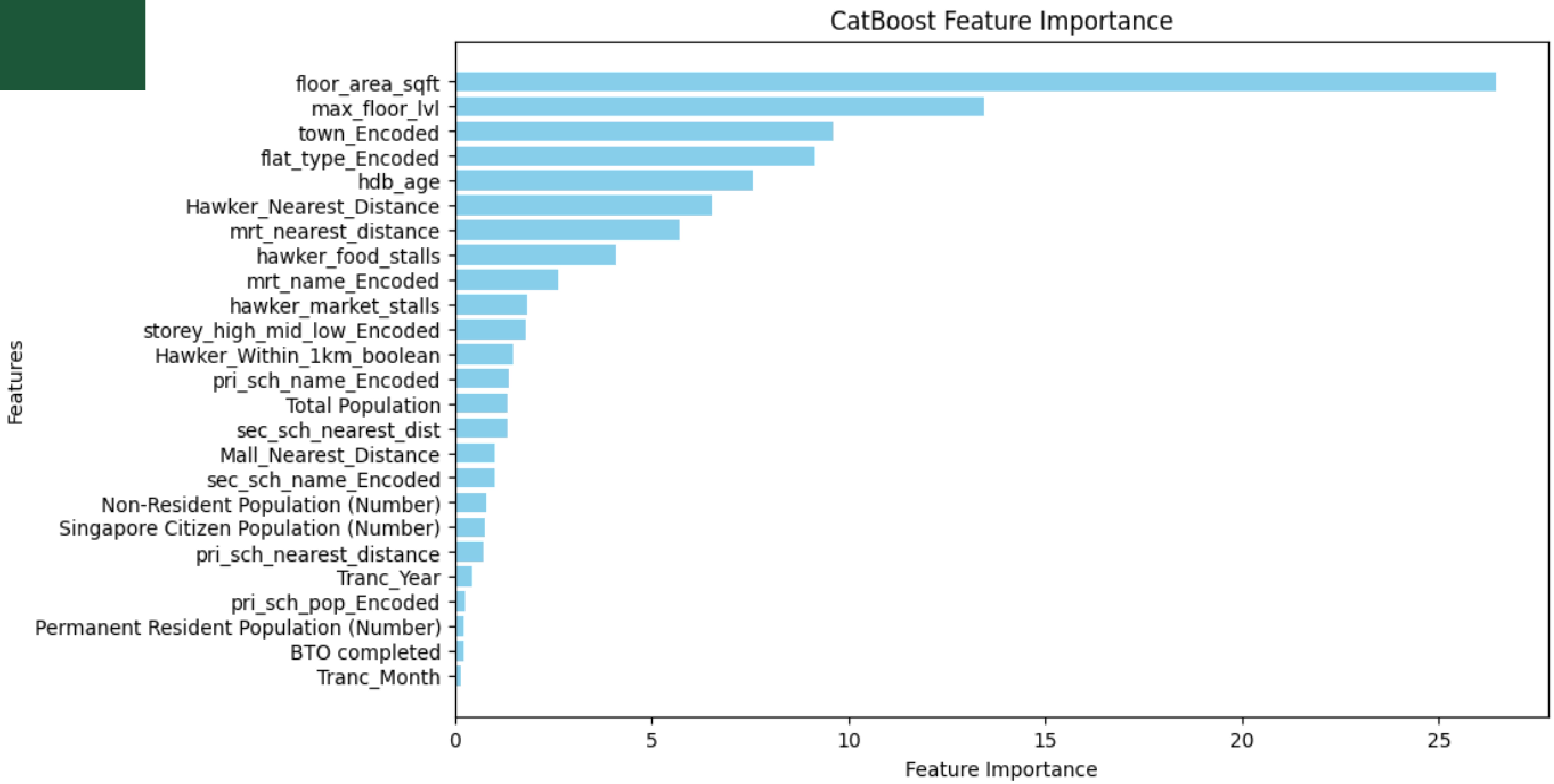
```
# Compare models
best_model = compare_models()
```

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
catboost	CatBoost Regressor	18534.9889	652925838.7059	25549.7432	0.9683	0.0549	0.0418	7.5260
et	Extra Trees Regressor	18878.1415	703613329.7832	26523.1299	0.9658	0.0565	0.0424	19.2110
rf	Random Forest Regressor	19026.8238	716545472.0194	26766.3580	0.9652	0.0570	0.0427	19.6150
lightgbm	Light Gradient Boosting Machine	23089.5720	996113361.0327	31558.0078	0.9516	0.0671	0.0518	1.0810
dt	Decision Tree Regressor	25932.6001	1370561240.3513	37016.2570	0.9335	0.0789	0.0582	0.4730
gbr	Gradient Boosting Regressor	34161.8724	2305212130.0407	48007.2094	0.8881	0.0979	0.0753	7.0390
knn	K Neighbors Regressor	34830.7023	2792316544.0000	52835.9953	0.8644	0.1062	0.0765	1.3750
br	Bayesian Ridge	56154.2299	5628637653.8799	75018.7665	0.7267	0.1600	0.1261	0.2570
ridge	Ridge Regression	56154.0903	5628598939.3947	75018.4994	0.7267	0.1600	0.1261	0.2490
lasso	Lasso Regression	56166.3607	5629707692.9220	75025.9312	0.7267	0.1600	0.1261	3.6990

Light GBM: Not bias on a particular variable



Example: Catboost feature Bias



Increase company’s bottom line by $3M per year

Total number of WOW agents	5,000	
Number of agents focuses on HDB	2,500	50% of agents focuses on HDB resale
Average revenue generated per agent per month	5,714	<= 4000/0.7
Number of HDB agents that uses WOW APP	750	Assumes 30% HDB agents use WOW APP
Rev generated by WOW user (per user, per month)	6,857	20% increase in productivity
Rev generated by WOW users per month	5,142,857	6,857*750
Rev generated if the agents did not use WOW	4,285,714	5,714*750
Increase in revenue per month	857,143	5,142,857 - 857,143
Increase in revenue per year (A)	10,285,714	857,143*12
Revenue share between company and agent (B)	7,200,000	70% commission to agents
Increase in company's bottom line	**3,085,714**	<= (A) – (B)

Trello Board