PROJECT 3: DATA SPRINT

Group : HDBee

Jimmy Myo Myint Aung
Ying Weng Seng
Nathaniel Tan
Shadni Wardiah

30 OCT 2025

AGENDA

PROJECT OVERVIEW 01

Background & Problem Statement
Objectives
Data Sources

DATA PRESENTATION & ANALYSIS 02

Exploratory Analysis (EDA)
Feature Engineering
Insights from the Data

MODELING & RESULTS 03

Correlation Analysis – What Drives Price?
Model Overview
Model Comparison
What Singaporeans Really Value in an HDB Home?

IMPACT & LESSONS 04

Assumptions & Caveats
Business Impact & Recommendations
Lessons Learned & Next Steps
Retrospective

PROJECT OVERVIEW

BACKGROUND

- Similar HDB flats (same town/size) can differ by S$100k+ (e.g., S$400k vs S$550k).
- Key question: is the gap from floor level, MRT/schools, flat age, or market timing?
- Upfront, explainable pricing helps with faster decisions and fair negotiations.
- We will focus on observable drivers (e.g. flat attributes & location), not opaque or hard-to-interpret features.

Team goal: turn raw transactions into simple, defendable insights for everyday use.

PROBLEM STATEMENT

- HDB resale prices are dynamic and competitive, shifting with buyer preferences, economic conditions, and location demand.
- WOW! agents need to deliver accurate, competitive price estimates across all flat types and neighborhoods.
- Without strong guidance, sellers risk undervaluing and buyers risk overpaying.
- Our goal is to build a transparent, data-driven model to support fair pricing and confident client advice.

OBJECTIVES

IDENTIFY KEY PRICE DRIVERS 🔑

- Analyze how variables such as flat size, age, location, and proximity to amenities (MRT, schools, malls, hawker centres) influence resale prices

COMPARE PRICING PATTERNS 💰

- Examine price variations between neighborhoods and flat models to highlight high-value and emerging areas.

DEVELOP A PREDICTIVE MODEL FOR RESALE PRICE ESTIMATION 🤖

- Build and test data-driven models to estimate resale prices based on key property attributes

VISUALIZE INSIGHTS FOR DECISION-MAKING 💡

- Create interactive dashboards and visual tools to help agents and clients easily interpret price trends and factors

SUPPORT BUSINESS STRATEGY AND CLIENT ADVISORY 🏢

- Use analytical findings to enhance client consultations, improve pricing accuracy, and strengthen WOW!'s market positioning

DATA SOURCES

DATASET

train.csv -- this data contains all of the training data for our model.
test.csv -- this data contains the test data for our model.
sample_sub_reg.csv -- An example of a correctly formatted submission to evaluate the model accuracy on the test data.

DATA PRESENTATION & ANALYSIS

EXPLORATORY DATA ANALYSIS

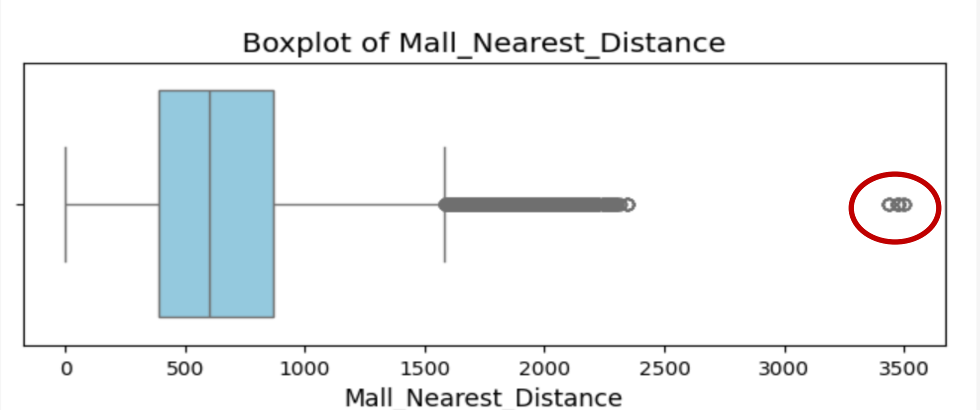
Dataset of **147,976 records** and **49 variables**

DATA CLEANING

- **Identifying duplicates**
- **Identifying nulls** (7 columns)
  - Mall_Nearest_Distance
  - Mall_Within_500m
  - Mall_Within_1km
  - Mall_Within_2km
  - Hawker_Within_500m
  - Hawker_Within_1km
  - Hawker_Within_2km
- Verifying **data types**
- Checking **outliers**

DATA HANDLING

- No duplicates found
- Replace nulls with 0 for 'Mall_Within_XX km' and 'Hawker_Within_XX km'
- Removing rows where by Mall_Nearest_Distance are nulls (829 rows)
- Changing data types (date, Boolean) for columns needed
- Remove outliers (Skewed high values)



Example of identifying skewed outliers

FEATURE ENGINEERING

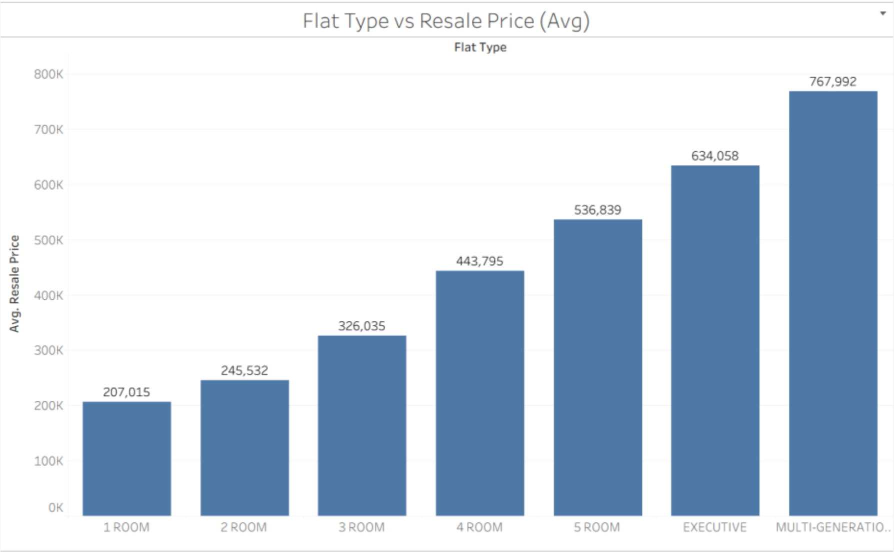
(+) ADDED NEW FEATURES

- floor_level_range
  - Floor categorised based of flat max floor
- years_of_lease_left
  - 99 years – HDB age
- mrt_region
  - Classify MRT names into country regions based to URA region guidelines

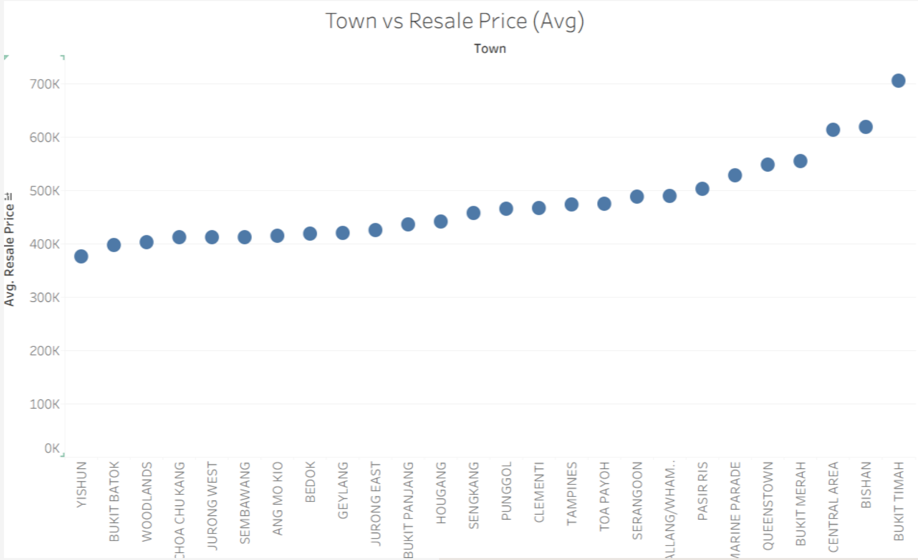
	mid_storey	max_floor_lvl	floor_level_range
0	11	25	Mid Lower Level
1	8	9	Upper Level
2	14	16	Upper Level
3	3	11	Mid Lower Level
4	2	4	Mid Upper Level
...	...	...	...
150629	5	12	Mid Lower Level
150630	5	16	Mid Lower Level
150631	11	15	Mid Upper Level
150632	8	10	Upper Level
150633	8	12	Mid Upper Level

INSIGHTS FROM THE DATA

Size matters

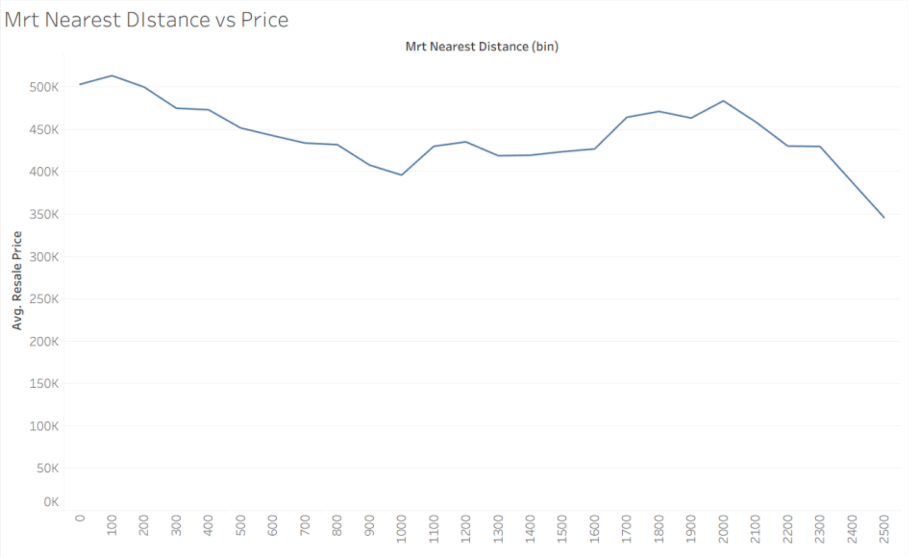


Location/Town

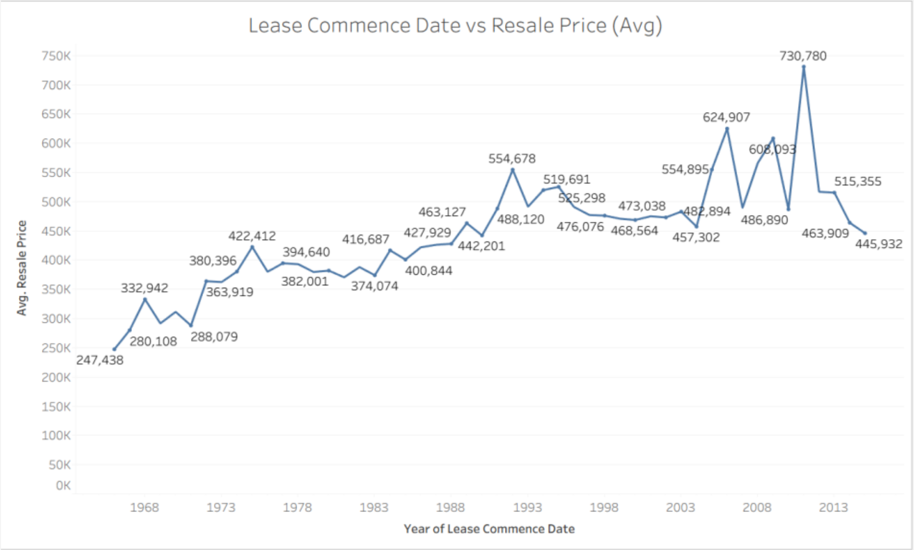


INSIGHTS FROM THE DATA

Amenities distance



Age of HDB



MODELING

& RESULTS

CORRELATION ANALYSIS

	Tranc_YearMon	loor_area_sqf	commercial	narket_hawke	tistorey_carp	ll_Within_50	all_Within_1k	all_Within_2k	ker_Within_5	vker_Within_	vker_Within_2	nearest_dista	rt_interchang	h_nearest_dis	rs_of_lease_l	2 ROOM	3 ROOM	4 ROOM	5 ROOM	EXECUTIVE	lid Lower Leve	lid Upper Leve	Upper Level	east	north	north_east	west	esale_price
Tranc_YearMonth	1	0.02410743	-0.0336211	-0.0001834	0.00051592	0.00882021	0.00318925	0.00583041	-0.0331898	-0.029533	-0.0281385	0.05751384	0.05738941	0.00732766	0.18888874	0.02437752	-0.0537956	0.01947732	0.02678776	-1.63E-06	0.0072679	0.00438462	0.00248918	-0.0232974	-0.0133594	0.05021029	-0.0165774	-0.0264982
floor_area_sqft	0.02410743	1	-0.1364179	-0.0155194	0.0045035	0.06707389	0.04171807	-0.0613842	-0.3128922	-0.3289504	-0.2935995	0.05295092	0.07617868	-0.1113498	0.29095349	-0.2387306	-0.7038948	-0.0501624	0.48486707	0.56921131	0.03157979	-0.0017354	-0.038843	0.03127835	0.06796242	-0.0030577	0.0747583	0.65669984
commercial	-0.0336211	-0.1364179	1	0.0206404	-0.0010953	0.00165326	0.01096087	0.03908522	0.07887984	0.10041891	0.104567	-0.0464804	-0.0622693	0.05052475	-0.2342501	-0.0106744	0.17238597	-0.0270128	-0.0933938	-0.0775395	-0.0161986	0.01153427	0.01668432	0.04288857	-0.042307	-0.0516421	-0.0184644	-0.1408056
market_hawker	-0.0001834	-0.0155194	0.0206404	1	-0.0001697	0.02083456	0.02266437	0.06911835	0.04622444	0.03880034	0.02310052	-0.0153162	0.01722966	0.02998295	-0.0077233	0.04073232	0.00623541	-0.0083048	-0.0056555	-0.0029864	-0.005627	-0.0006853	0.01022182	-0.0048806	-0.0045443	-0.0055632	-0.0057287	-0.0059146
multistorey_carpark	0.00051592	0.0045035	-0.0010953	-0.0001697	1	0.03102551	0.01228525	-9.62E-05	-0.0108302	-0.0139778	-0.01329	0.00490503	0.02883336	-0.0139881	0.00701242	-0.0019157	-0.0036302	0.00408653	-0.0010057	0.00089055	-0.0037626	0.0056976	0.0047515	-0.0081676	-0.0076047	-0.0026737	0.02309036	-0.0077384
Mall_Within_500m	0.00882021	0.06707389	0.00165326	0.02083456	0.03102551	1	0.51316893	0.21211008	0.02508665	0.00412171	-0.0851625	-0.1244223	0.16863877	-0.0551676	0.07721571	-0.0174501	-0.0651636	0.02862559	0.01769384	0.03430772	0.01475459	-0.008605	-0.0094692	-0.063335	-0.0009408	-0.0006174	0.09943885	0.08871682
Mall_Within_1km	0.00318925	0.04171807	0.01096087	0.02266437	0.01228525	0.51316893	1	0.43360695	-0.0023605	0.03298885	-0.0766516	-0.111142	0.26731752	-0.0709866	0.08584193	-0.0154729	-0.0496138	0.01667612	0.03084274	0.0105636	0.01828436	-0.0069949	-0.0019574	0.03158535	-0.054856	-0.0008472	0.068428	0.08580538
Mall_Within_2km	0.00583041	-0.0613842	0.03908522	0.06911835	-9.62E-05	0.21211008	0.43360695	1	0.17833128	0.17285557	0.23202697	-0.1494213	0.10393627	0.04133596	0.02299646	0.05929401	0.00614884	0.00264679	0.00143105	-0.0412209	0.00102173	-0.0021421	0.00681668	-0.0073249	-0.181191	0.00372266	-0.0835204	0.11424498
Hawker_Within_500m	-0.0331898	-0.3128922	0.07887984	0.04622444	-0.0108302	0.02508665	-0.0023605	0.17833128	1	0.75351535	0.63571958	-0.1677552	-0.1874739	0.16508225	-0.3732009	0.09701037	0.31495323	-0.1287296	-0.13136	-0.1113766	-0.0148268	-0.0157885	0.02033359	0.05547472	-0.1358496	-0.1229482	-0.1641747	0.01278653
Hawker_Within_1km	-0.029533	-0.3289504	0.10041891	0.03880034	-0.0139778	0.00412171	0.03298885	0.17285557	0.75351535	1	0.82856603	-0.1779478	-0.2211398	0.18652518	-0.4073639	0.08373418	0.31819183	-0.1057562	-0.1422774	-0.1383116	-0.0091115	-0.0195689	0.01712316	0.05978311	-0.1959309	-0.1379342	-0.2356212	0.09213981
Hawker_Within_2km	-0.0281385	-0.2935995	0.104567	0.02310052	-0.01329	-0.0851625	-0.0766516	0.23202697	0.63571958	0.82856603	1	-0.2135633	-0.2089764	0.19905865	-0.4017409	0.05925987	0.29523061	-0.1057603	-0.1202921	-0.1265629	-0.0052944	-0.0176495	0.00947563	0.05475675	-0.2564383	-0.1397248	-0.296816	0.18514101
mrt_nearest_distance	0.05751384	0.05295092	-0.0464804	-0.0153162	0.00490503	-0.1244223	-0.111142	-0.1494213	-0.1677552	-0.1779478	-0.2135633	1	0.11234382	0.04288405	0.14455805	-0.0002332	-0.0640502	0.02536291	0.02703003	0.01329493	0.01064431	-0.0091727	0.00587993	-0.0399635	-0.0667774	0.23154076	0.03531659	-0.1274085
mrt_interchange	0.05738941	0.07617868	-0.0622693	0.01722966	0.02883336	0.16863877	0.26731752	0.10393627	-0.1874739	-0.2211398	-0.2089764	0.11234382	1	-0.0390023	0.24325724	-0.0042176	-0.1358503	0.05884356	0.06371602	0.01612785	0.01753917	-0.0054051	0.00273587	-0.0572622	-0.1587373	0.21954013	0.04633447	0.06455517
pri_sch_nearest_distance	0.00732766	-0.1113498	0.05052475	0.02998295	-0.0139881	-0.0551676	-0.0709866	0.04133596	0.16508225	0.18652518	0.19905865	0.04288405	-0.0390023	1	-0.1823404	0.0322382	0.1381813	-0.0655255	-0.0531128	-0.0350179	-0.006367	-0.0112069	0.00971033	0.0093509	-0.0860487	-0.1415218	0.08689602	-0.0072261
years_of_lease_left	0.18888874	0.29095349	-0.2342501	-0.0077233	0.00701242	0.07721571	0.08584193	0.02299646	-0.3732009	-0.4073639	-0.4017409	0.14455805	0.24325724	-0.1823404	1	-0.0216057	-0.4684101	0.22648598	0.20430493	0.03895012	0.03635731	0.00492858	-0.0201219	-0.1943497	0.10292602	0.22696069	0.02488497	0.35018899
2 ROOM	0.02437752	-0.2387306	-0.0106744	0.04073232	-0.0019157	-0.0174501	-0.0154729	0.05929401	0.09701037	0.08373418	0.05925987	-0.0002332	-0.0042176	0.0322382	-0.0216057	1	-0.0676083	-0.0937426	-0.063838	-0.0337102	-0.013718	-0.0001206	0.01124563	-0.0047453	-0.023175	0.00580549	-0.0327063	-0.1605935
3 ROOM	-0.0537956	-0.7038948	0.17238597	0.00623541	-0.0036302	-0.0651636	-0.0496138	0.00614884	0.31495323	0.31819183	0.29523061	-0.0640502	-0.1358503	0.1381813	-0.4684101	-0.0676083	1	-0.490474	-0.334009	-0.1763763	-0.0329679	0.009863	0.03208329	0.04090244	-0.0797421	-0.0668728	-0.0294222	-0.5061808
4 ROOM	0.01947732	-0.0501624	-0.0270128	-0.0083048	0.00408653	0.02862559	0.01667612	0.00264679	-0.1287296	-0.1057562	-0.1057603	0.02536291	0.05884356	-0.0655255	0.22648598	-0.0937426	-0.490474	1	-0.4631217	-0.2445554	0.01791734	-0.0112955	-0.0132472	-0.0348711	0.05097718	0.04393119	-0.0096041	-0.0245584
5 ROOM	0.02678776	0.48486707	-0.0933938	-0.0056555	-0.0010057	0.01769384	0.03084274	0.00143105	-0.13136	-0.1422774	-0.1202921	0.02703003	0.06371602	-0.0531128	0.20430493	-0.063838	-0.334009	-0.4631217	1	-0.1665403	0.0080219	-0.0038837	-0.0057766	-0.0210925	0.01432349	0.02586981	0.0248379	0.34785892
EXECUTIVE	-1.63E-06	0.56921131	-0.0775395	-0.0029864	0.00089055	0.03430772	0.0105636	-0.0412209	-0.1113766	-0.1383116	-0.1265629	0.01329493	0.01612785	-0.0350179	0.03895012	-0.0337102	-0.1763763	-0.2445554	-0.1665403	1	0.01386931	0.01097824	-0.0235037	0.03284411	0.02475522	-0.0131101	0.04043622	0.38477496
Mid Lower Level	0.0072679	0.03157979	-0.0161986	-0.005627	-0.0037626	0.01475459	0.01828436	0.00102173	-0.0148268	-0.0091115	-0.0052944	0.01064431	0.01753917	-0.006367	0.03635731	-0.013718	-0.0329679	0.01791734	0.0080219	0.01386931	1	-0.3638615	-0.3470578	-0.0086993	-0.0072355	0.01703321	0.00087356	0.0087973
Mid Upper Level	0.00438462	-0.0017354	0.01153427	-0.0006853	0.0056976	-0.008605	-0.0069949	-0.0021421	-0.0157885	-0.0195689	-0.0176495	-0.0091727	-0.0054051	-0.0112069	0.00492858	-0.0001206	0.009863	-0.0112955	-0.0038837	0.01097824	-0.3638615	1	-0.4044766	-0.0015487	0.00135213	-3.79E-06	0.0093561	0.015815
Upper Level	0.00248918	-0.038843	0.01668432	0.01022182	0.0047515	-0.0094692	-0.0019574	0.00681668	0.02033359	0.01712316	0.00947563	0.00587993	0.00273587	0.00971033	-0.0201219	0.01124563	0.03208329	-0.0132472	-0.0057766	-0.0235037	-0.3470578	-0.4044766	1	0.00132178	0.0021112	0.00671649	-0.0139268	0.03554236
east	-0.0232974	0.03127835	0.04288857	-0.0048806	-0.0081676	-0.063335	0.03158535	-0.0073249	0.05547472	0.05978311	0.05475675	-0.0399635	-0.0572622	0.0093509	-0.1943497	-0.0047453	0.04090244	-0.0348711	-0.0210925	0.03284411	-0.0086993	-0.0015487	0.00132178	1	-0.2186914	-0.2677271	-0.2756926	0.02128902
north	-0.0133594	0.06796242	-0.042307	-0.0045443	-0.0076047	-0.0009408	-0.054856	-0.181191	-0.1358496	-0.1959309	-0.2564383	-0.0667774	-0.1587373	-0.0860487	0.10292602	-0.023175	-0.0797421	0.05097718	0.01432349	0.02475522	-0.0072355	0.00135213	0.0021112	-0.2186914	1	-0.2492755	-0.256692	-0.1741129
north_east	0.05021029	-0.0030577	-0.0516421	-0.0055632	-0.0026737	-0.0006174	-0.0008472	0.00372266	-0.1229482	-0.1379342	-0.1397248	0.23154076	0.21954013	-0.1415218	0.22696069	0.00580549	-0.0668728	0.04393119	0.02586981	-0.0131101	0.01703321	-3.79E-06	0.00671649	-0.2677271	-0.2492755	1	-0.3142482	0.009977
west	-0.0165774	0.0747583	-0.0184644	-0.0057287	0.02309036	0.09943885	0.068428	-0.0835204	-0.1641747	-0.2356212	-0.296816	0.03531659	0.04633447	0.08689602	0.02488497	-0.0327063	-0.0294222	-0.0096041	0.0248379	0.04043622	0.00087356	0.0093561	-0.0139268	-0.2756926	-0.256692	-0.3142482	1	-0.1028206
resale_price	-0.0264982	0.65669984	-0.1408056	-0.0059146	-0.0077384	0.08871682	0.08580538	0.11424498	0.01278653	0.09213981	0.18514101	-0.1274085	0.06455517	-0.0072261	0.35018899	-0.1605935	-0.5061808	-0.0245584	0.34785892	0.38477496	0.0087973	0.015815	0.03554236	0.02128902	-0.1741129	0.009977	-0.1028206	1

(+) CROSS CHECK ON CORRELATIONS BTW FEATURES

(-) CROSS CHECK ON CORRELATIONS BTW TARGET & FEATURES

CORRELATION ANALYSIS

	resale_price
floor_area_sqft	0.65669984
commercial	-0.1408056
market_hawker	-0.0059146
multistorey_carpark	-0.0077384
Mall_Within_500m	0.08871682
Mall_Within_1km	0.08580538
Mall_Within_2km	0.11424498
Hawker_Within_500m	0.01278653
Hawker_Within_1km	0.09213981
Hawker_Within_2km	0.18514101
mrt_nearest_distance	-0.1274085
mrt_interchange	0.06455517
pri_sch_nearest_distance	-0.0072261
years_of_lease_left	0.35018899
2 ROOM	-0.1605935
3 ROOM	-0.5061808
4 ROOM	-0.0245584
5 ROOM	0.34785892
EXECUTIVE	0.38477496
Mid Lower Level	0.0087973
Mid Upper Level	0.015815
Upper Level	0.03554236
east	0.02128902
north	-0.1741129
north_east	0.009977
west	-0.1028206

(1) CORRELATION BETWEEN FEATURES

- Mall/ Hawker within [~~500m | 1km |~~ 2km]

(2) CORRELATION BETWEEN FEATURES & TARGET

- Market Hawker
- Multistorey Carpark
- Pri School Nearest Distance

Almost close to 0 with no correlation

MODEL OVERVIEW

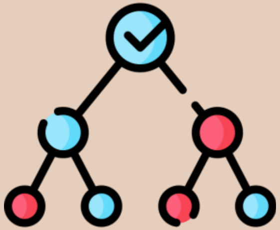
MODELS USED

LINEAR REGRESSION MODEL

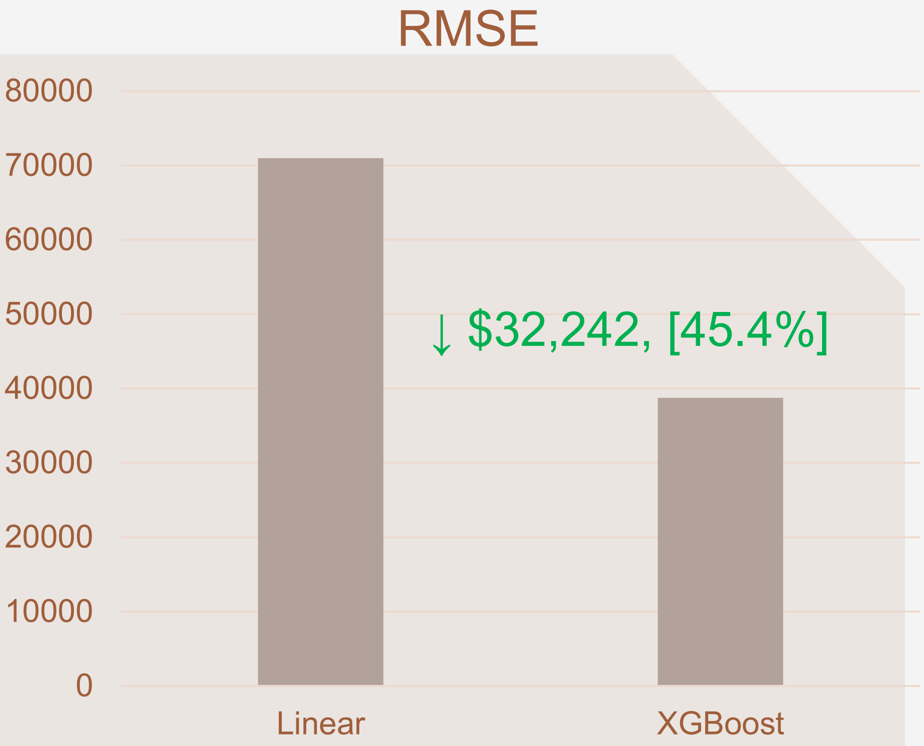
"If one factor increases, how much does the price change?"

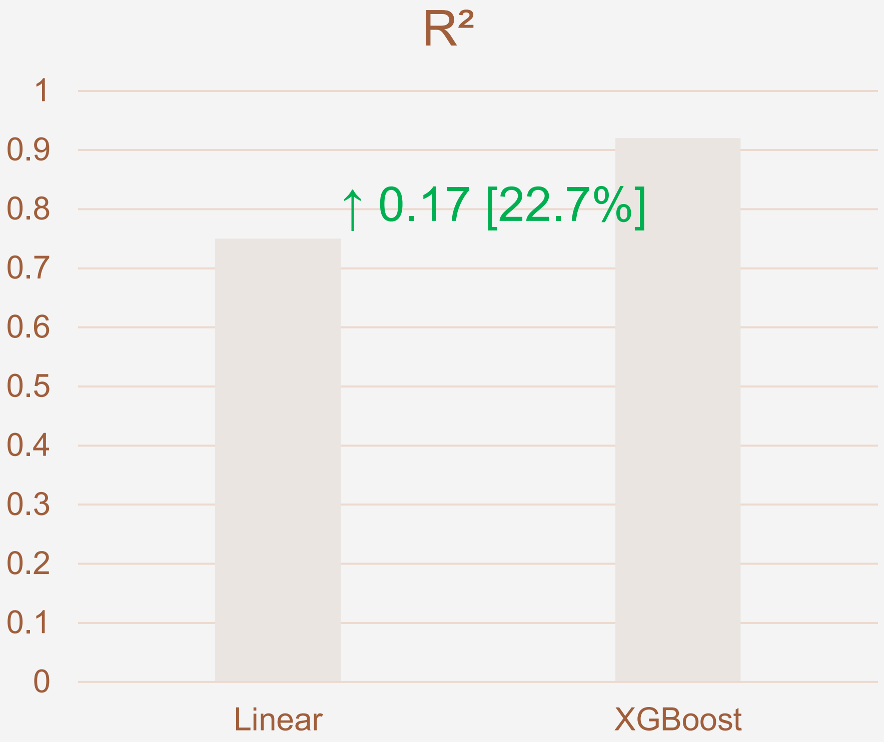
XGBOOST

Instead of one line, it builds *hundreds of small trees* that learn from each other's mistakes.



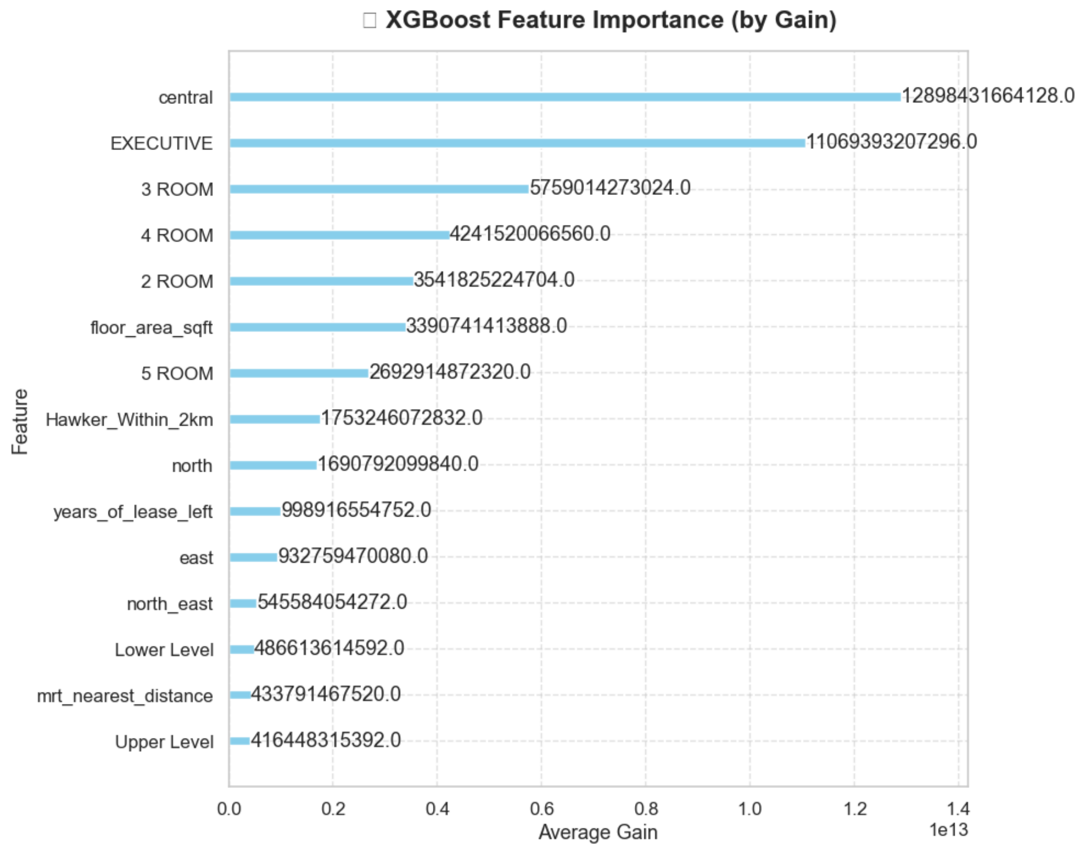
MODEL COMPARISON





What Singaporeans Really Value in an HDB Home?

1. Central is everything
2. **BIGGER** the better
3. Proud Foodies
4. Practical
5. Everyday Conveniences



IMPACT

& LESSONS

ASSUMPTIONS & CAVEATS

🧩 Assumptions

📊 Market behaves rationally – resale prices reflect measurable factors.

🧮 Data sources are reliable and updated.

⏱️ Past trends represent near-term market patterns.

⚠️ Caveats

🏗️ Uncaptured factors – interior condition and renovation not included.

📍 Approximate proximity – map data, not travel time.

📈 Market shifts – sudden policy or sentiment changes may affect predictions.

BUSINESS IMPACT & RECOMMENDATIONS

📊 Data Insights

Enables fair and transparent pricing for both buyers and sellers.

🤖 Model Application

Predictive model allows accurate resale price estimation, reducing guesswork and bias.

🧭 Agent Empowerment

Provides data-backed confidence in negotiations, boosting agent credibility

💼 Business Impact

Strengthens WOW!'s position as a trusted, analytics-driven real estate brand.

💡 RECOMMENDATIONS

- 🧰 Integrate model into agent workflow through a Streamlit-based pricing app.
- 🔄 Continuously update data with new transactions to maintain accuracy.
- 🎯 Use results for targeted marketing (e.g., towns with rising demand).
- 🏘️ Explore rental and BTO market extensions.

LESSONS LEARNED

💡 **Business Relevance**
Insights create measurable impact.

📊 **Interpretability**
- Simple models build trust

💬 **Communication**
Bridges technical and business understanding.

🧹 **Data Quality**
Data drives accuracy

AND STILL GOING STRONG!

🌏 **Market Expansion**
Apply to rental and private housing

📈 **Forecasting**
Predict future resale price trends

💻 **Streamlit App**
Deploy pricing dashboard for agents

🧹 **Data Enhancement**
Add renovation and environmental features

NEXT STEPS

RETROSPECTIVE

From messy data to actionable insights — our journey combined teamwork, technical depth, way forward and storytelling.

TEAMWORK → What works?

Task division and collaboration

TECHNICAL DEPTH → What must be done?

Correlation and feature selection

WAY FORWARD → To move on or put a pin on it?

Exploration to implementation

STORYTELLING → How can we show/explain?

Make complex analytics insights accessible to all audience

STREAMLIT DEMO


QUESTIONS?

THANK YOU

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution.