Beyond The Flat:

A Holistic HDB Price Predictor

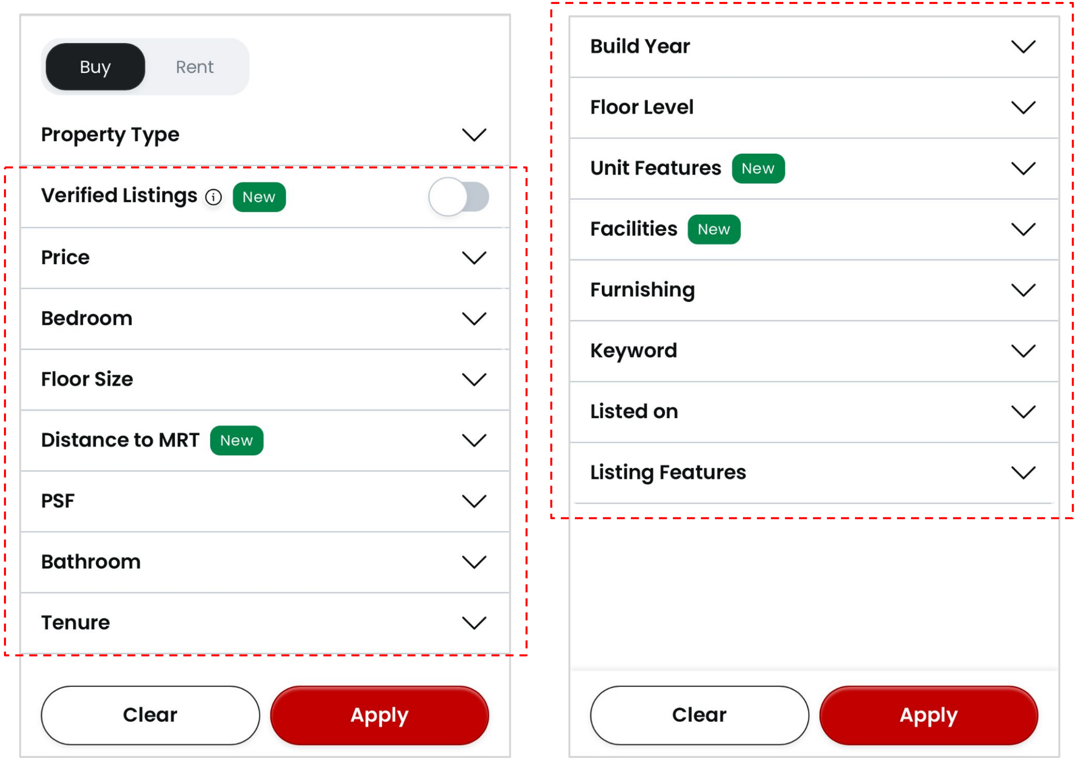
Alina, Daniel, Leon, Nigel, Rei, Tiffany
4 September 2025

PINNACLE VENTURES

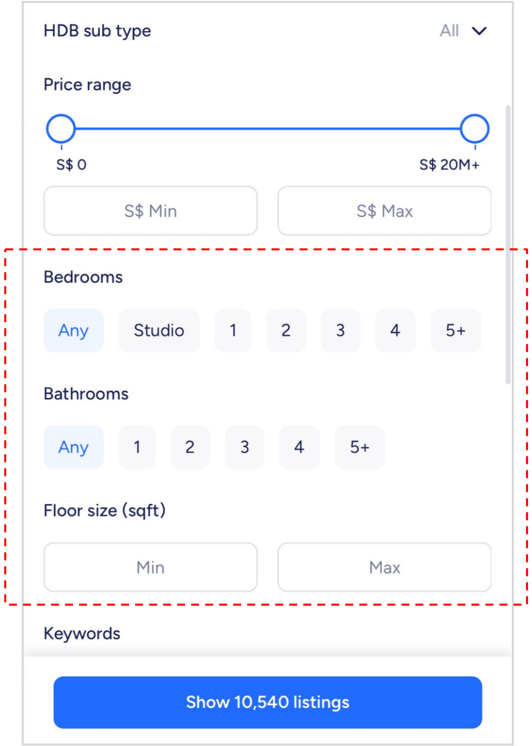
A **data-driven** tool designed to empower agents and homebuyers

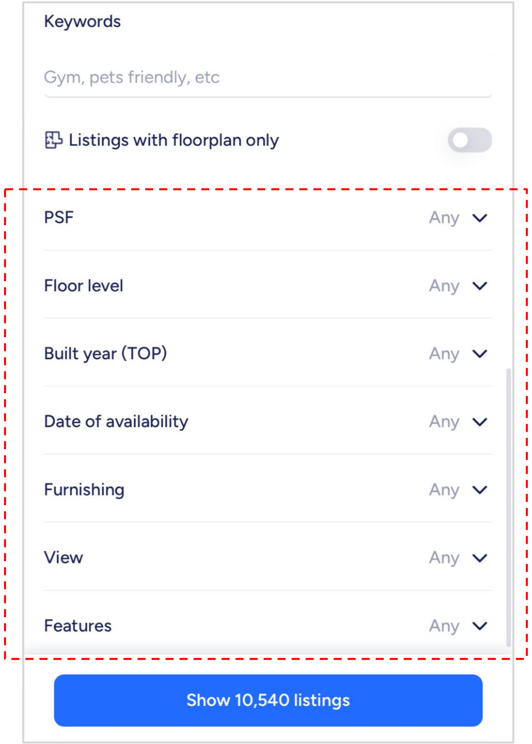
What's on the market **today**?

PropertyGuru



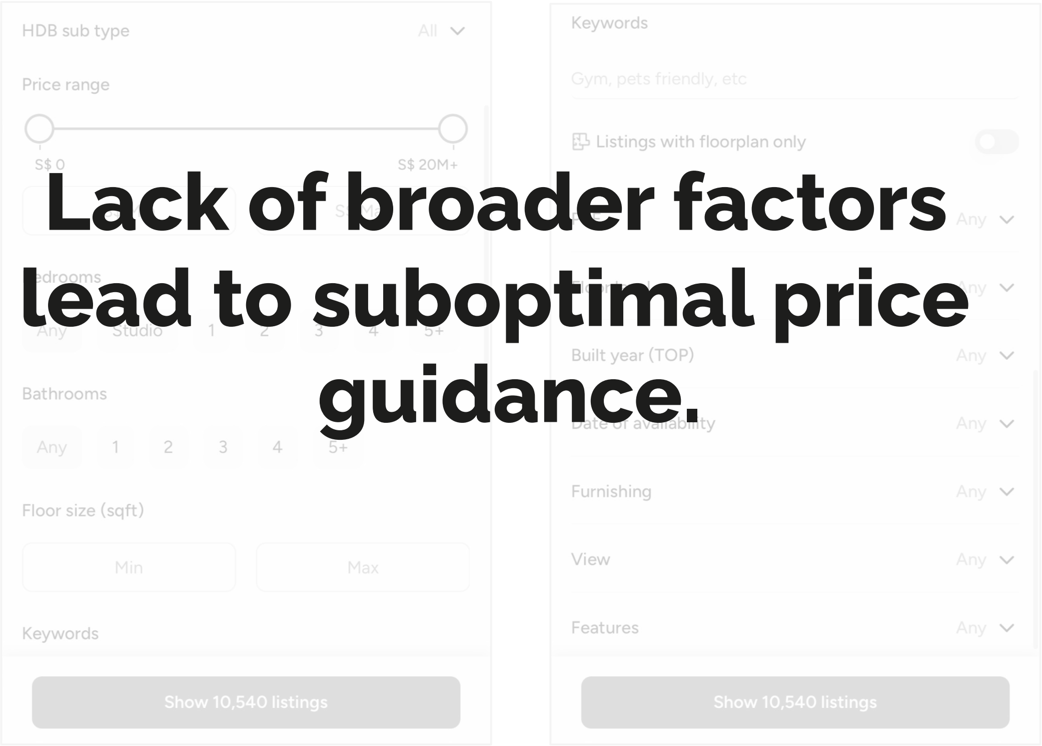
Existing platforms rely on traditional factors





Existing platforms rely on traditional factors

Lack of broader factors lead to suboptimal price guidance.



How can we offer more accurate price guidance using factors that are currently not considered?

The Current Approach

Basic
Micro

Details at the flat or neighborhood level that affect value

Enhancing the Factor Set

Macro

Broader economic conditions that influence prices city-wide

Basic Micro

Details at the flat or neighborhood level that affect value

Engineered Micro

Combined metrics from multiple basic micro factors to better reflect a flat's value.

Enhancing the Factor Set

Macro NEW	Basic Micro	Engineered Micro NEW
Gross Domestic Product	Max Floor Level	Affluence Index
Consumer Price Index	HDB Age	Distance to Nearest Top School
Median Household Income	Lease Commencement Date	Amenities Proximity Score
	Transaction Month	Age at Sale
	Transaction Year	Nearby Amenities (1km)
	Floor Area	Demand
		Transport Proximity Score
		Number of Top Schools
		MRT Developments
		Average School Subscription Rate

Enhancing the Factor Set

Macro NEW	Basic Micro	Engineered Micro NEW
Gross Domestic Product	Max Floor Level	**Affluence Index**
Consumer Price Index	HDB Age	**Distance to Nearest Top School**
Median Household Income	Lease Commencement Date	**Amenities Proximity Score**
	Transaction Month	Age at Sale
	Transaction Year	Nearby Amenities (1km)
	Floor Area	Demand
		Transport Proximity Score
		Number of Top Schools
		MRT Developments
		Average School Subscription Rate

What are the Macro Drivers of Property Value?

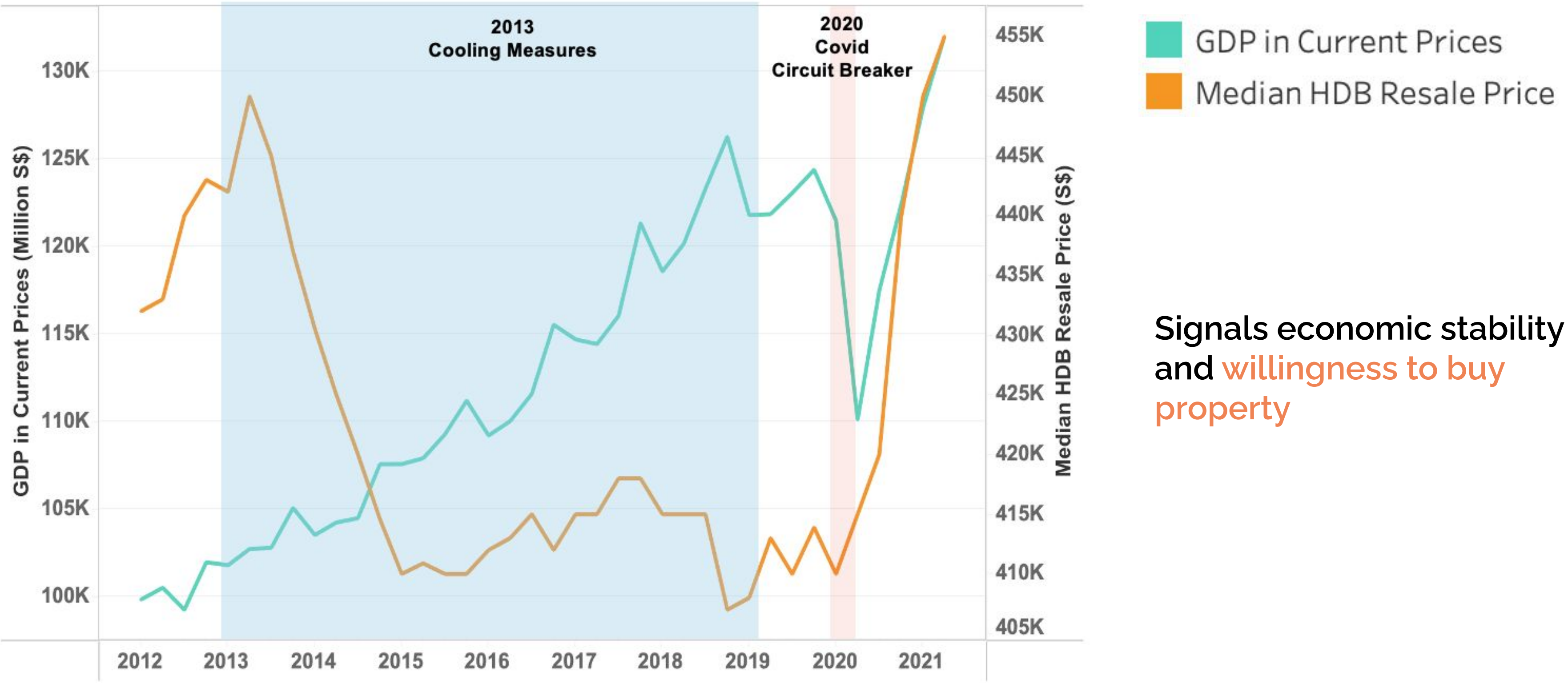
GDP (Gross Domestic Product)

Signals economic stability and willingness to buy property

CPI & MHI

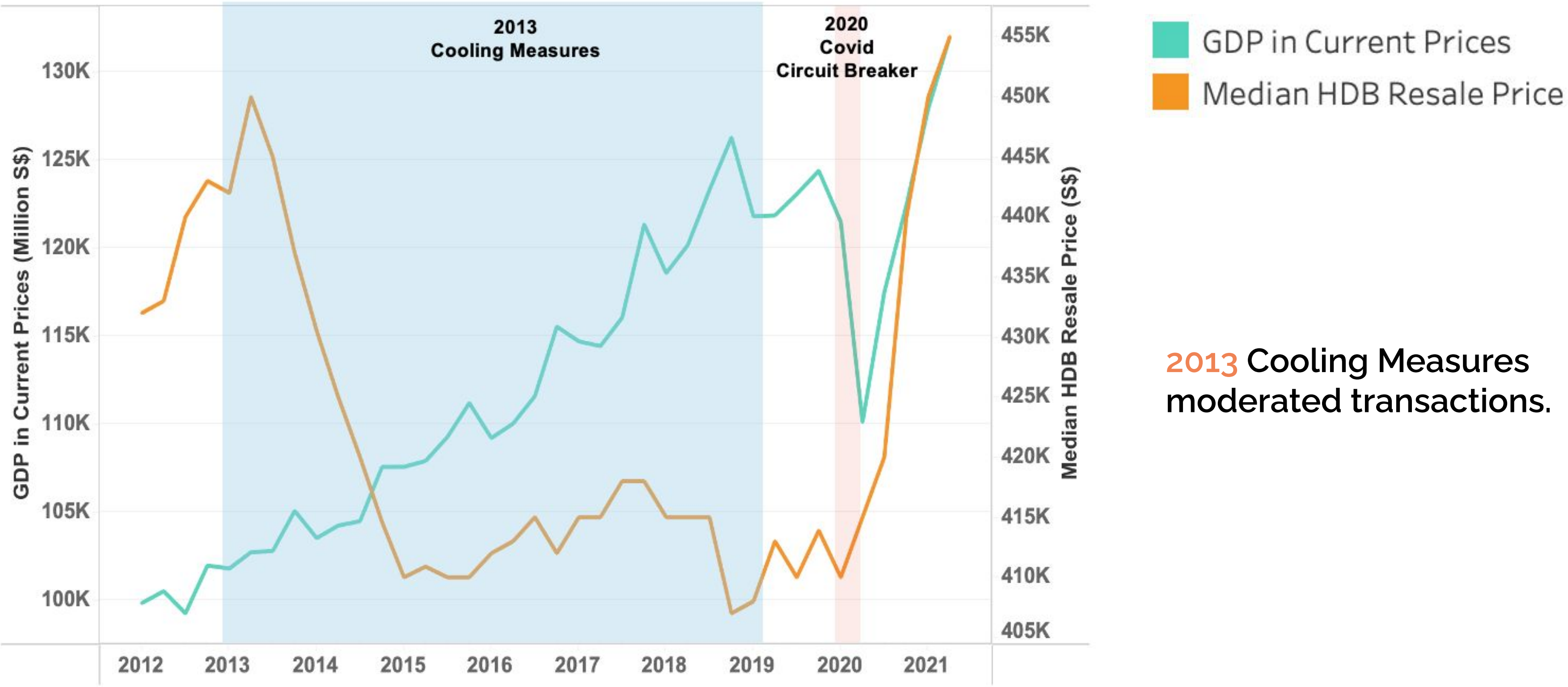
Consumer Price Index & Median Household Income

GDP signals Economic Strength



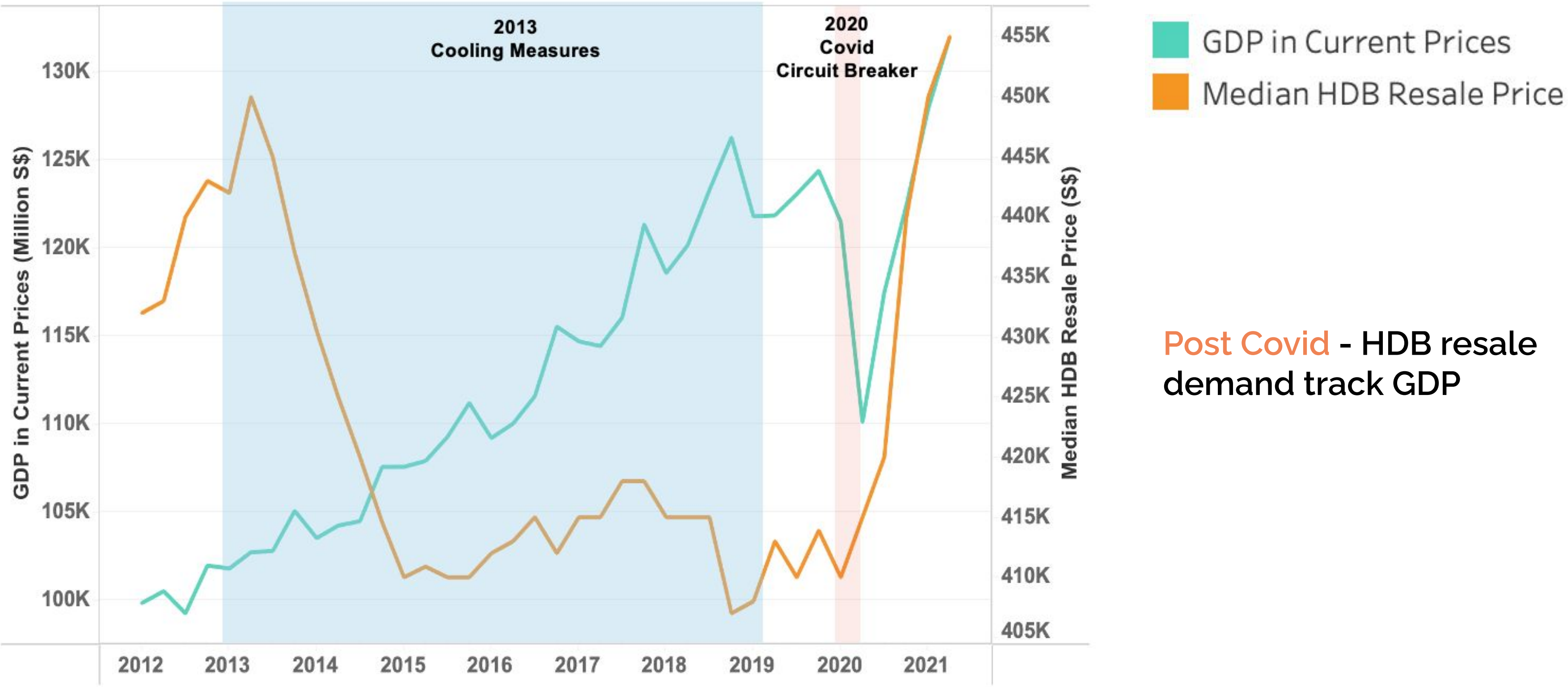
Signals economic stability and willingness to buy property

GDP signals Economic Strength



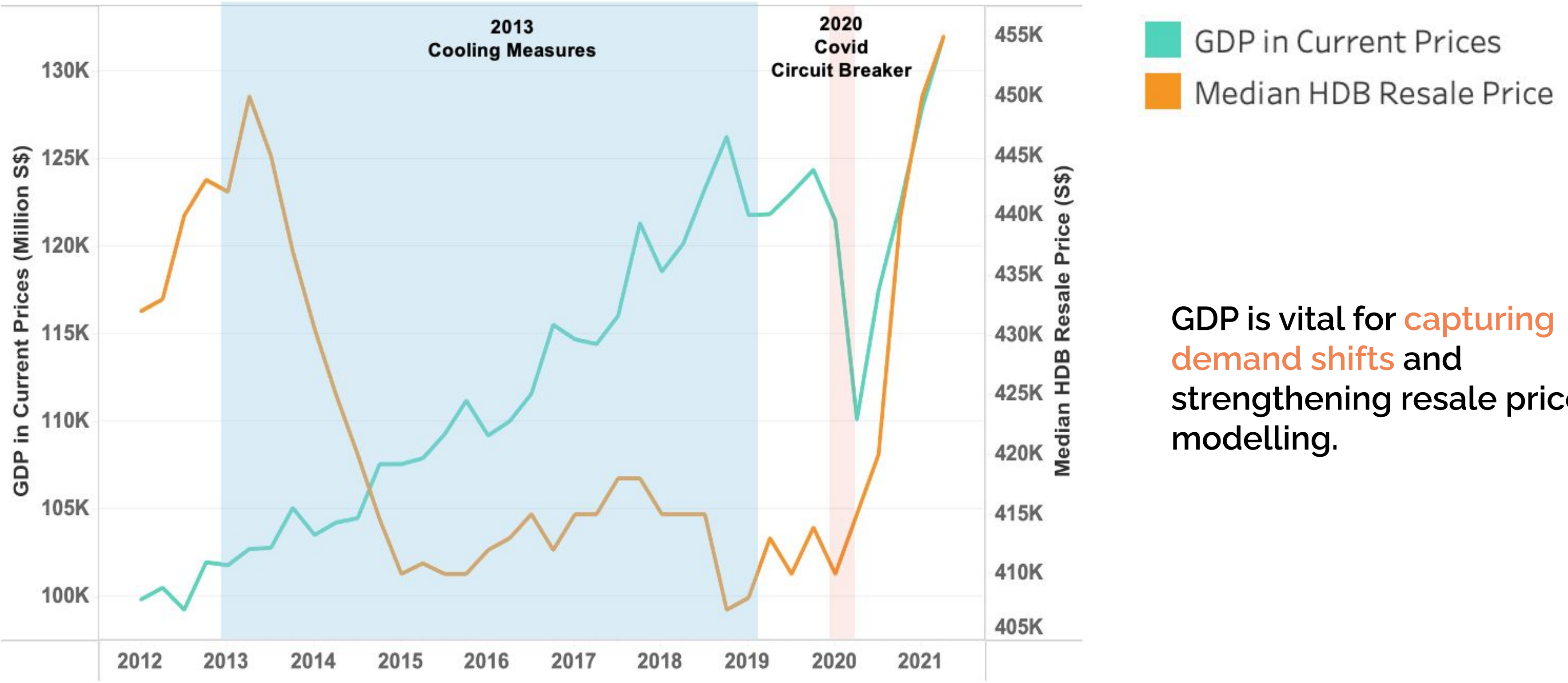
2013 Cooling Measures moderated transactions.

GDP signals Economic Strength



Post Covid - HDB resale demand track GDP

GDP signals Economic Strength



GDP is vital for capturing demand shifts and strengthening resale price modelling.

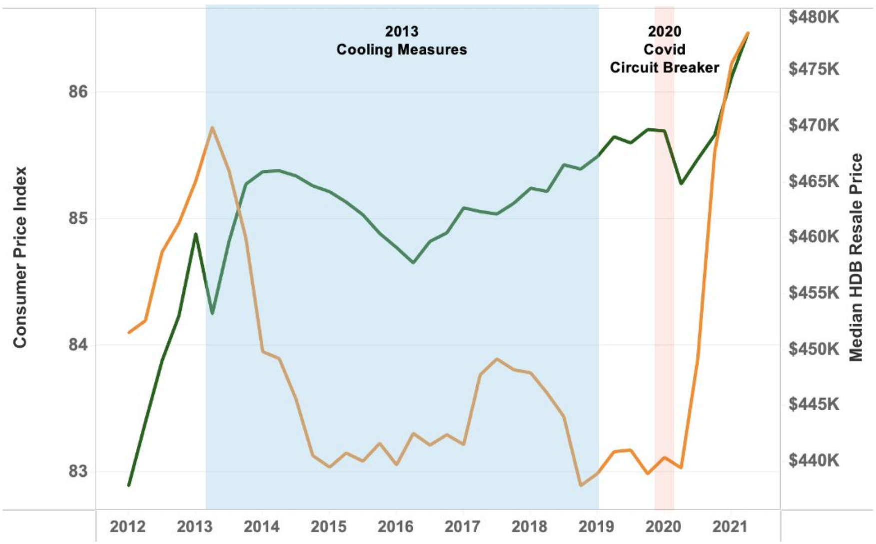
GDP

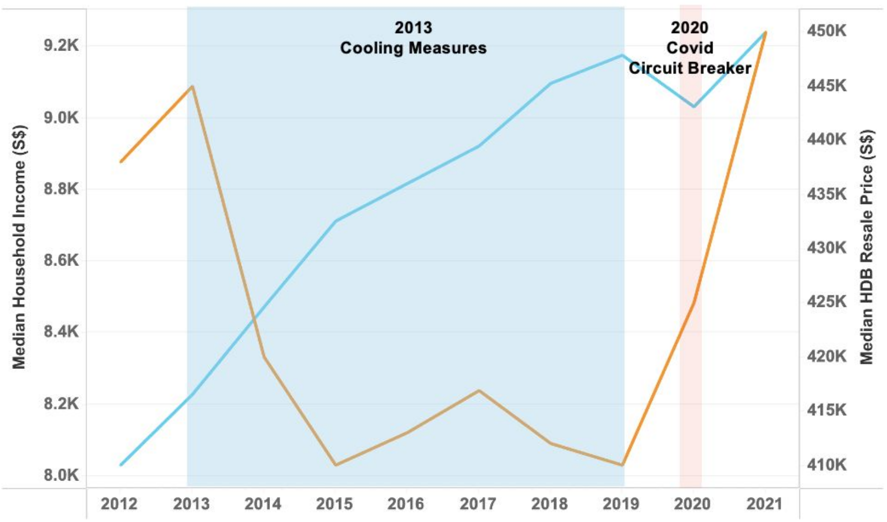
Gross Domestic Products

CPI (Consumer Price Index) & MHI (Median Household Income)

Key factor influencing housing affordability and resale price movements

Resale prices move with CPI and MHI except during Cooling Measure





The Macro Drivers of Property Value

Higher **GDP** drives income growth and property demand

CPI reflects inflation's impact on housing prices

MHI reflects income power for resale price movement

What are the Micro Drivers of Property Value?

Affluence Index

Town's rank x Flat Type x Floor Cat

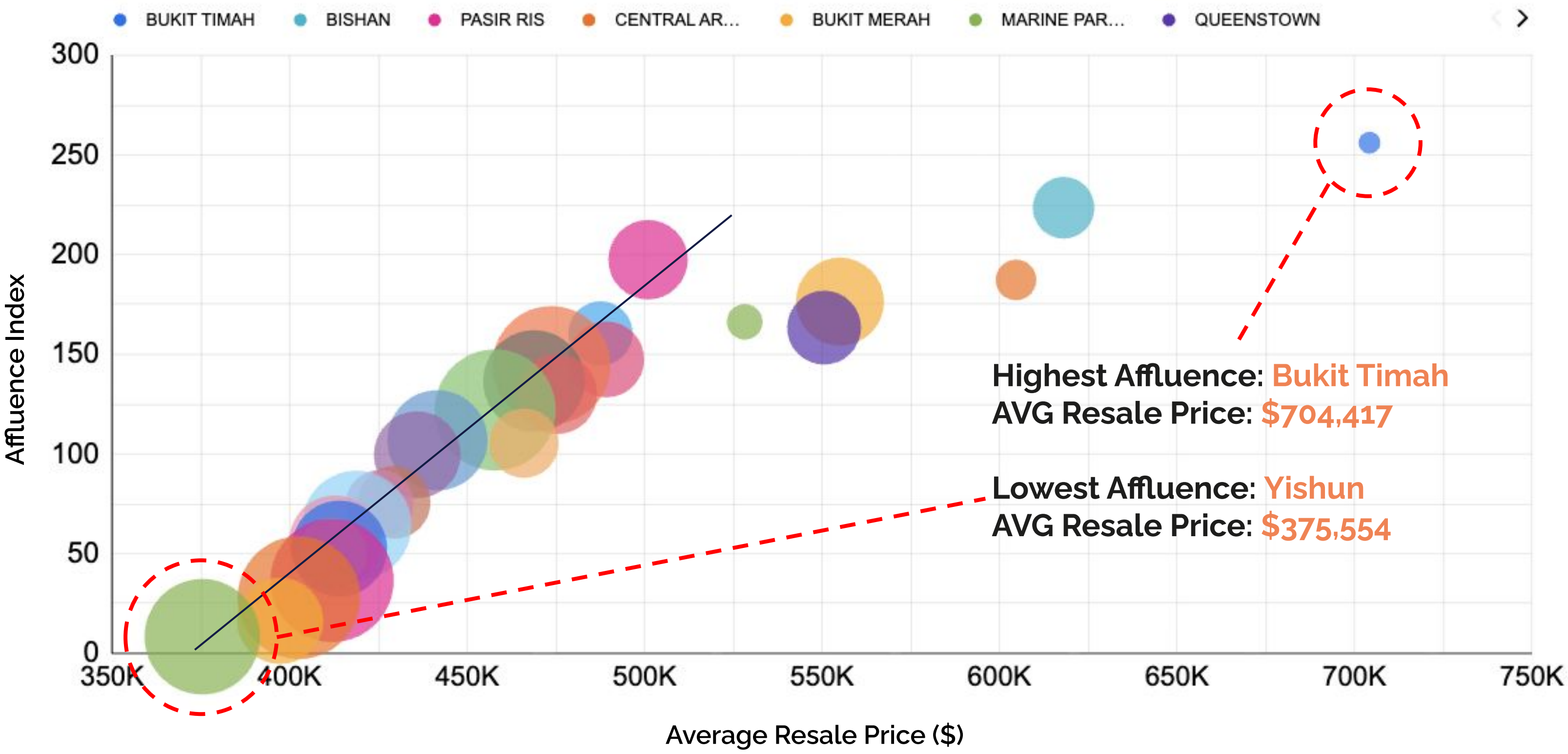
Distance Nearest To Top School

Top School defined as one with very high demand,
where the number of applicants is more than twice the number of available vacancies.

Amenities Proximity Score

It's simply the average distance to nearby key amenities such as hawkers and malls.

Resale Price Increases with Affluence Index



Highest Affluence: Bukit Timah
AVG Resale Price: $704,417

Lowest Affluence: Yishun
AVG Resale Price: $375,554

Affluence Index = Town Rank x Flat Type x Floor Cat

Affluence Index

Town's rank x Flat Type x Floor Cat

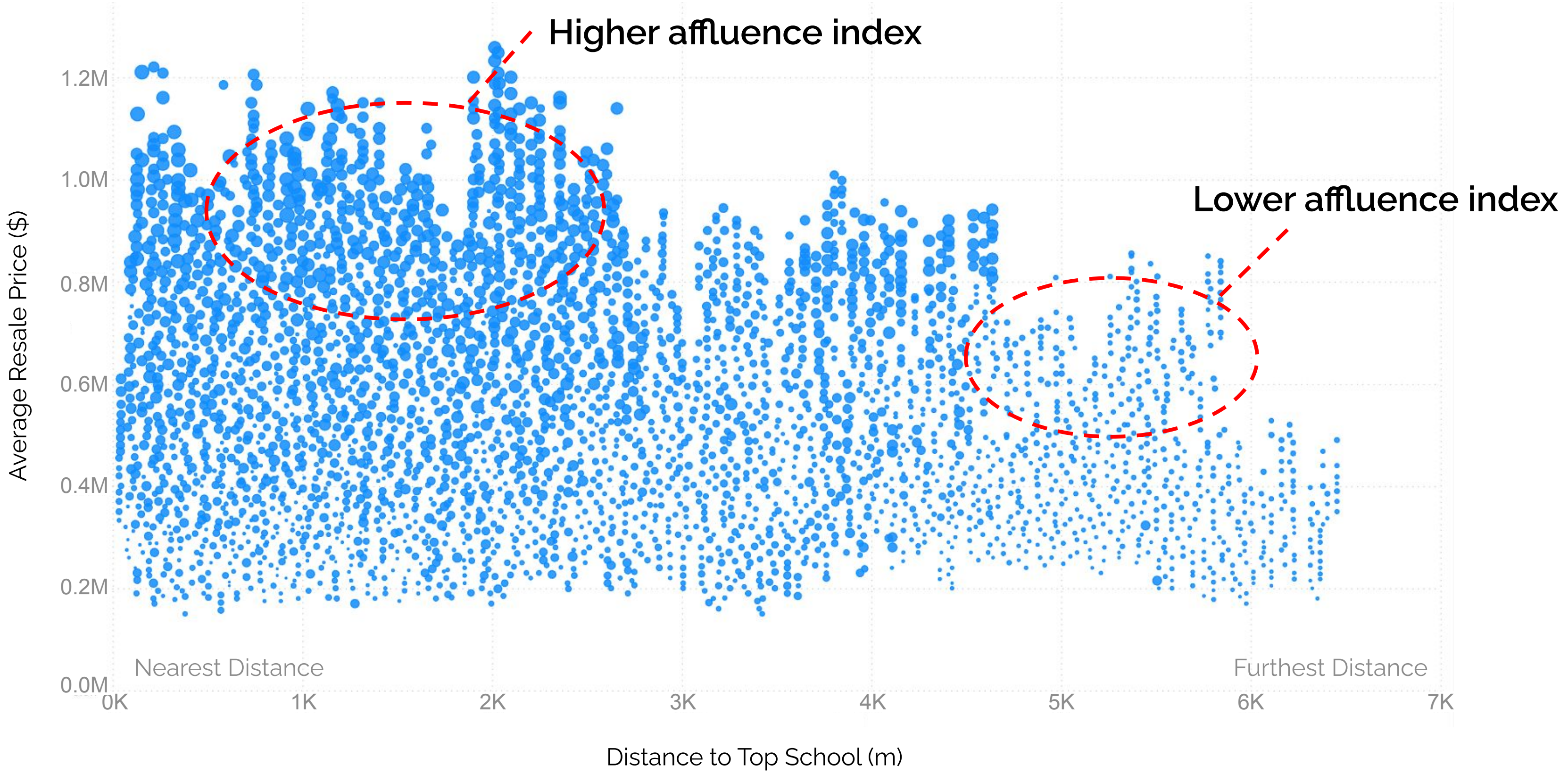
Distance Nearest To Top School

Top School defined as one with very high demand,
where the number of applicants is more than twice the number of available vacancies.

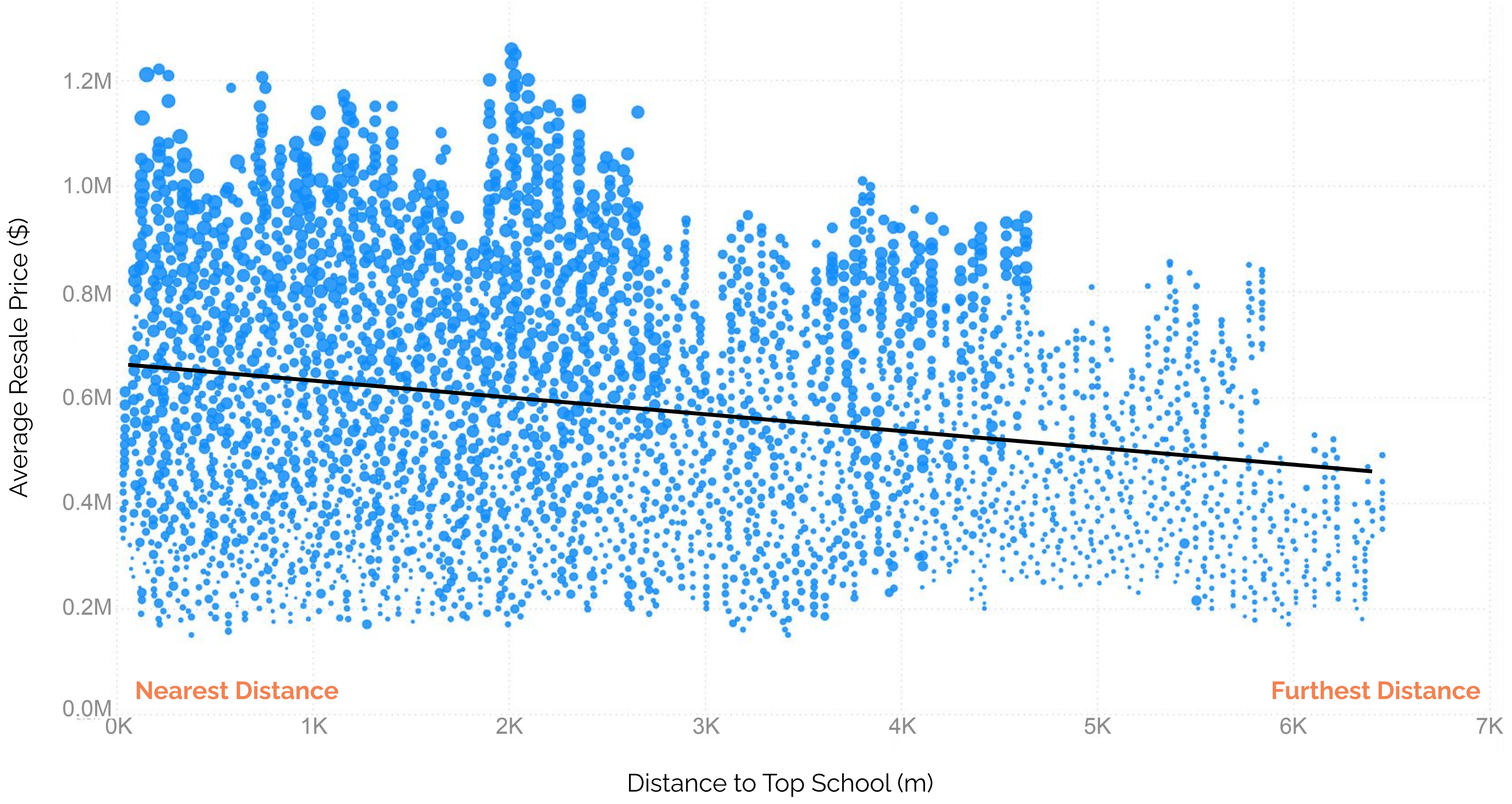
Amenities Proximity Score

It's simply the average distance to nearby key amenities such as hawkers and malls.

Affluence links to preference for top schools



Proximity to Top School drives up the price



Affluence Index

Town's rank x Flat Type x Floor Cat

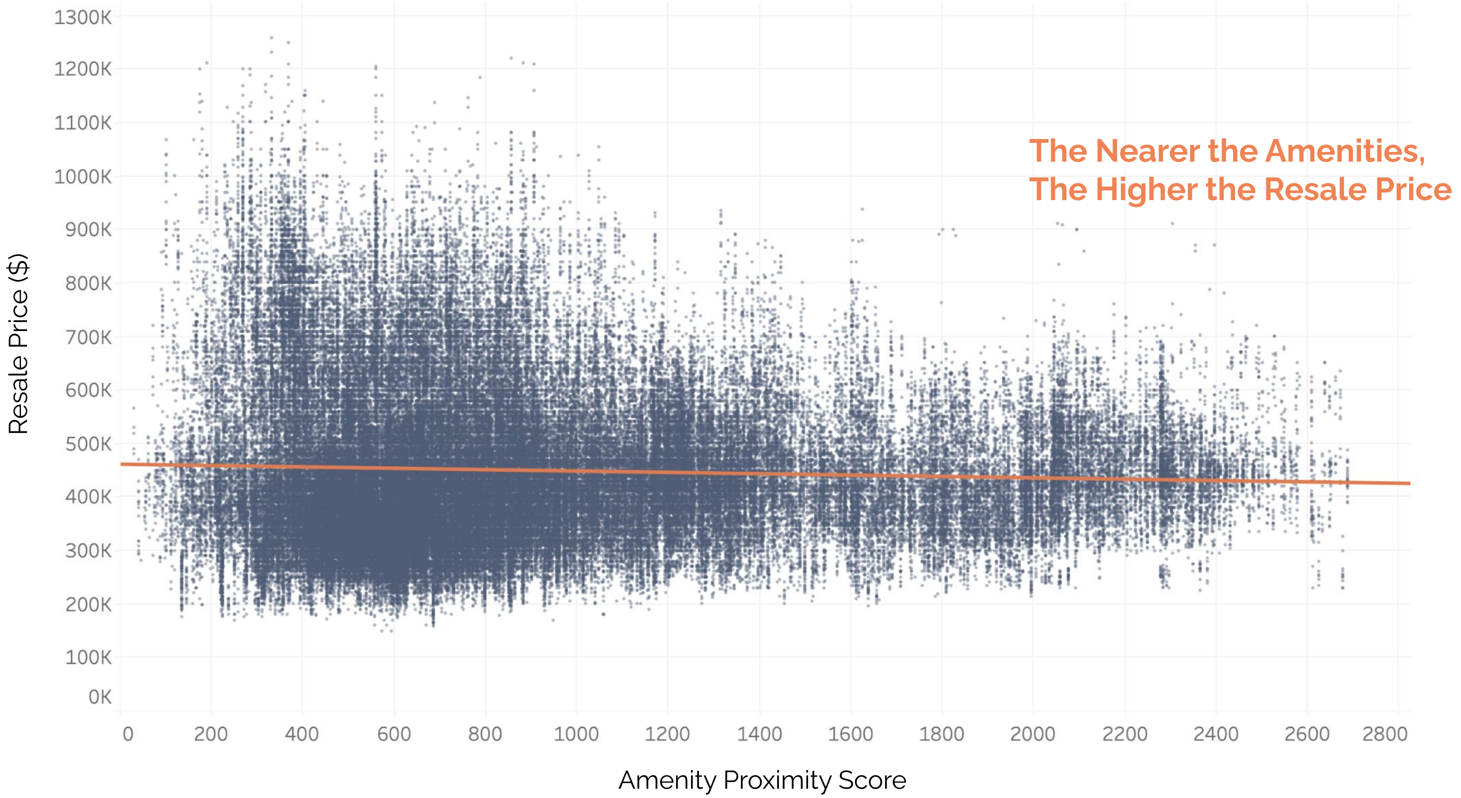
Distance Nearest To Top School

Top School defined as one with very high demand,
where the number of applicants is more than twice the number of available vacancies.

Amenities Proximity Score

It's simply the average distance to nearby key amenities such as hawkers and malls.

Access to Amenities Raises Property Value



Amenities Proximity Score = average distance to nearby key amenities

The Micro Drivers of Property Value

Higher **affluence index** pushes up resale prices.

Homes near **top schools** suggest a higher price premium.

Being close to **amenities** also adds property value.

Modelling on Python

Initial Model Evaluation

We ran an initial experiment to **establish a performance baseline** for improvement.

Model	RMSE	Time
Random Forest Regressor	$31,000	10.04 sec
Extra Trees Regressor	$32,443	6.49 sec
CatBoost Regressor	$33,139	3.71 sec
Light Gradient Boosting Machine	$37,748	0.73 sec

RMSE: Root Mean Square Error
Measures **how far** our predictions are from **actual values** (smaller means more accurate)

Initial Model Evaluation

We ran an initial experiment to **establish a performance baseline** for improvement.

Model	RMSE	Time	
Random Forest Regressor	$31,000	10.04 sec	
Extra Trees Regressor	$32,443	6.49 sec	
CatBoost Regressor	$33,139	3.71 sec	
Light Gradient Boosting Machine	**$37,748**	**0.73 sec**	(baseline)

RMSE: Root Mean Square Error
Measures **how far** our predictions are from **actual values** (smaller means more accurate)

Models Tested

Boosting Model

A model that starts with a baseline that continuously improves from the previous results to achieve a reliable final result.

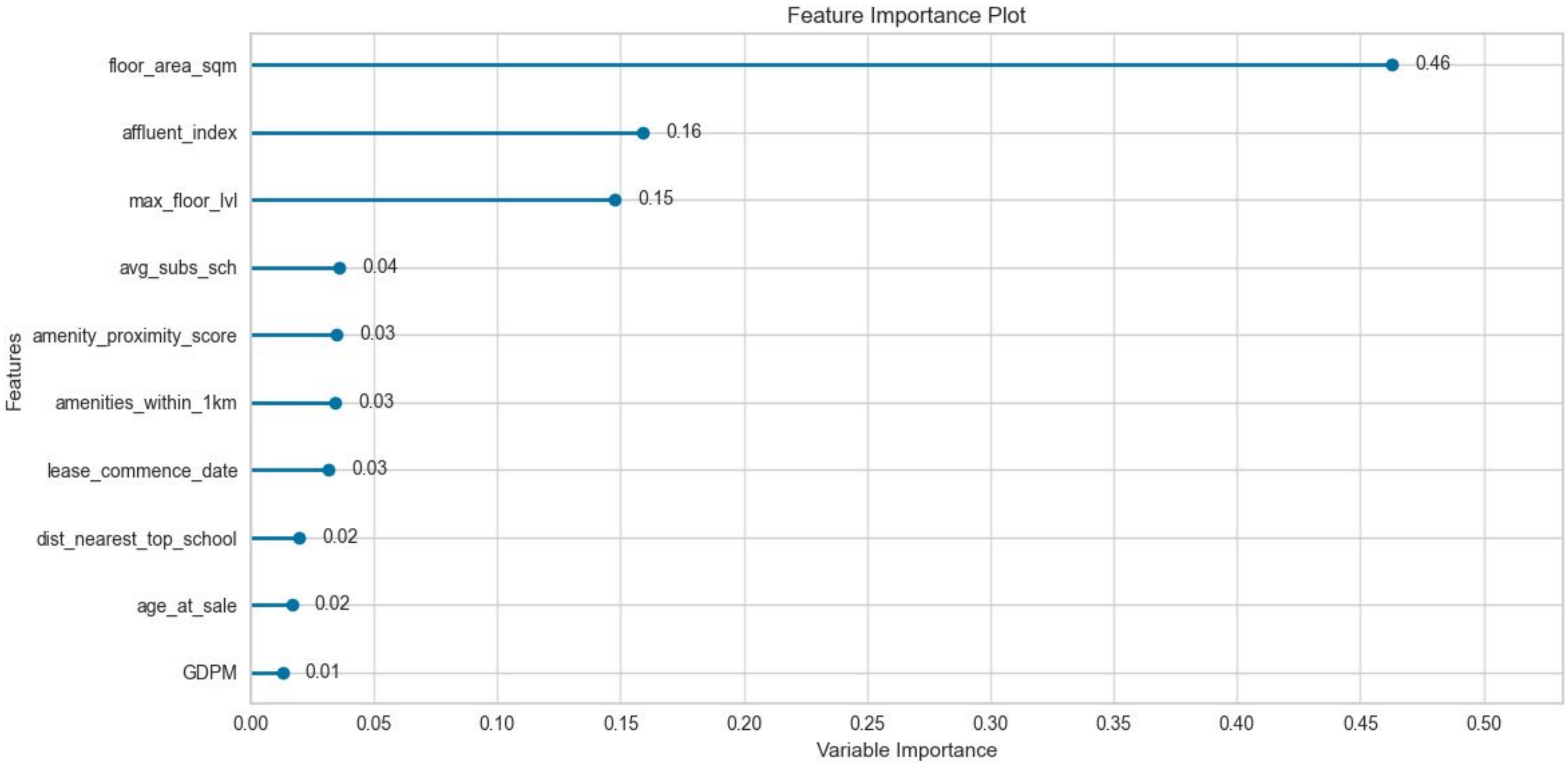
CatBoost
LightGBM

Bagging Model

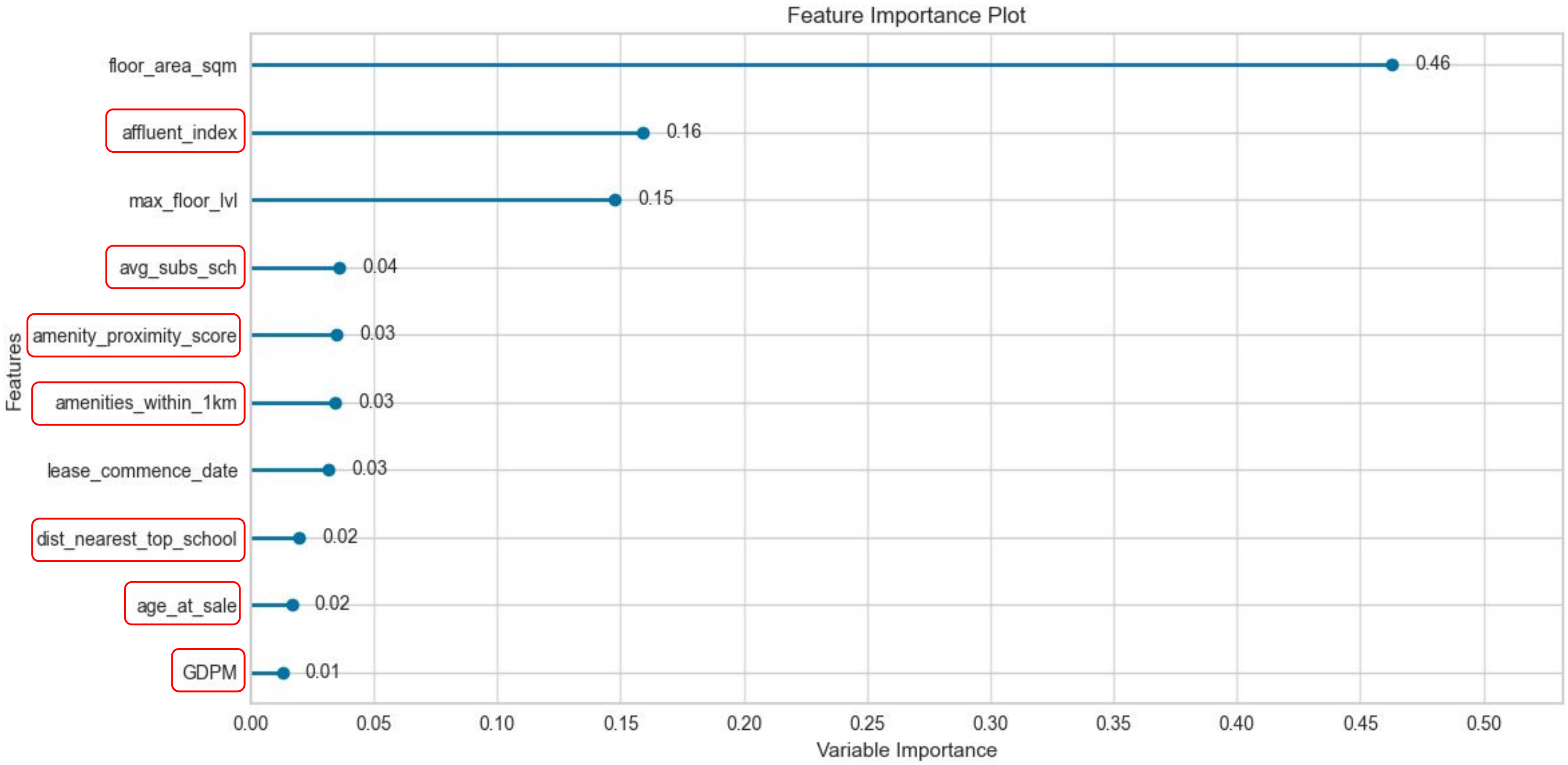
A model that trains many independent versions on different random samples and combines their answers to produce a stable, reliable final result.

Random Forest
Extra Trees

Feature Importance on LightGBM



Feature Importance on LightGBM



Final Modelling Results

Model	TR RMSE	TE RMSE	RMSE Diff	Run Time
Random Forest Regressor	$9,963	$26,372	$16,409	22.9 sec
Extra Trees Regressor	$65	$26,047	$25,982	4.8 sec
CatBoost Regressor	$24,621	$26,371	$1,750	52.3 sec
Light Gradient Boosting Machine	$8,913	$24,622	$15,709	12.5 sec

Overall Score

Model	TR RMSE (Ranking)	TE RMSE (Ranking)	RMSE Diff (Ranking)	Run Time (Ranking)	Score
Random Forest Regressor	\$9,963 (3)	\$26,372 (4)	\$16,409 (3)	22.9 sec (3)	13
Extra Trees Regressor	\$65 (1)	\$26,047 (2)	\$25,982 (4)	4.8 sec (1)	8
CatBoost Regressor	\$24.621 (4)	\$26,371 (3)	\$1,750 (1)	52.3 sec (4)	12
Light Gradient Boosting Machine	\$8,913 (2)	\$24,622 (1)	\$15,709 (2)	12.5 sec (2)	7

Ranking derived based on the total score, the lower points are ranked higher.

Why LightGBM

Accuracy

97%

Model with the highest accuracy

Run time

12.5s

Second fastest run time

Drop in prediction error

35%

Largest drop in prediction error

Demo on Streamlit

What makes our product different?

Exclusive Features

Unique features not found on existing platforms

Resale Connections

Visuals showing how these features connect to resale prices

Benefits

Higher Accuracy

Incorporate engineered micro and macro factors alongside flat details to generate more precise resale price estimates.

More Holistic

Integrate broader economic and neighborhood data into pricing tools for a fuller picture of value.

Recommendations

Deploy Our Model

Integrate the model into existing pricing tools to enhance their accuracy and depth.

Sell to Platforms

Partner with property platforms to commercialize the solution and create revenue opportunities.

Conclusion

Problem

Current platforms mainly rely on **basic micro factors**, which do not capture the full picture of resale price influences.

Solution

Incorporating **engineered micro** factors and **macro** factors improves prediction accuracy.

Impact

Our model shows a nearly **35%** reduction in prediction error, helping users make smarter, more informed decisions in the HDB resale market.

Thank You

PINNACLE
VENTURES

Reflections

1. **Collaborative Problem-Solving**:
   Working together helped us tackle challenges more effectively.

2. **Communication**:
   Staying aligned and sharing progress ensured smooth teamwork.

3. **Iterative Learning**:
   Testing and refining ideas as a team improved our solutions.

4. **The Value of Hypotheses:**
   Starting with a hypothesis gave us focus and direction to our analysis.

5. **Value of Domain Knowledge:**
   By understanding macro factors, aided the team into developing specific engineered features.