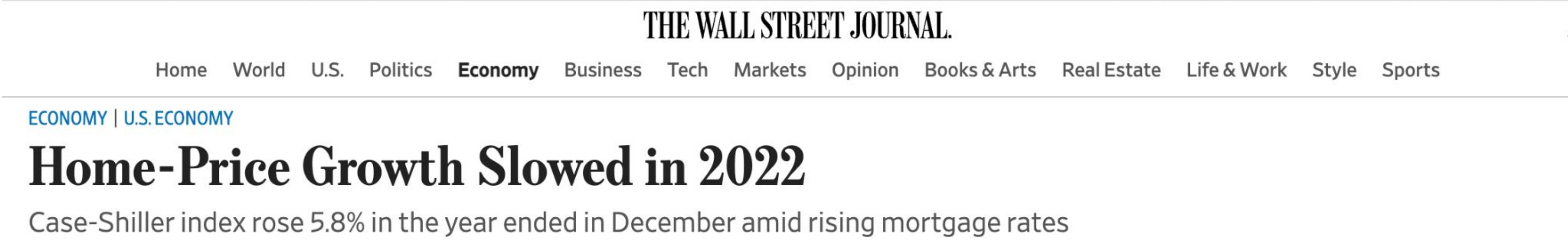
Zillow Listing Optimization

Group 1: Ivy Cheung, David He, Kaitong Hu, Silvia Huang, Alex Mo

Overview

1

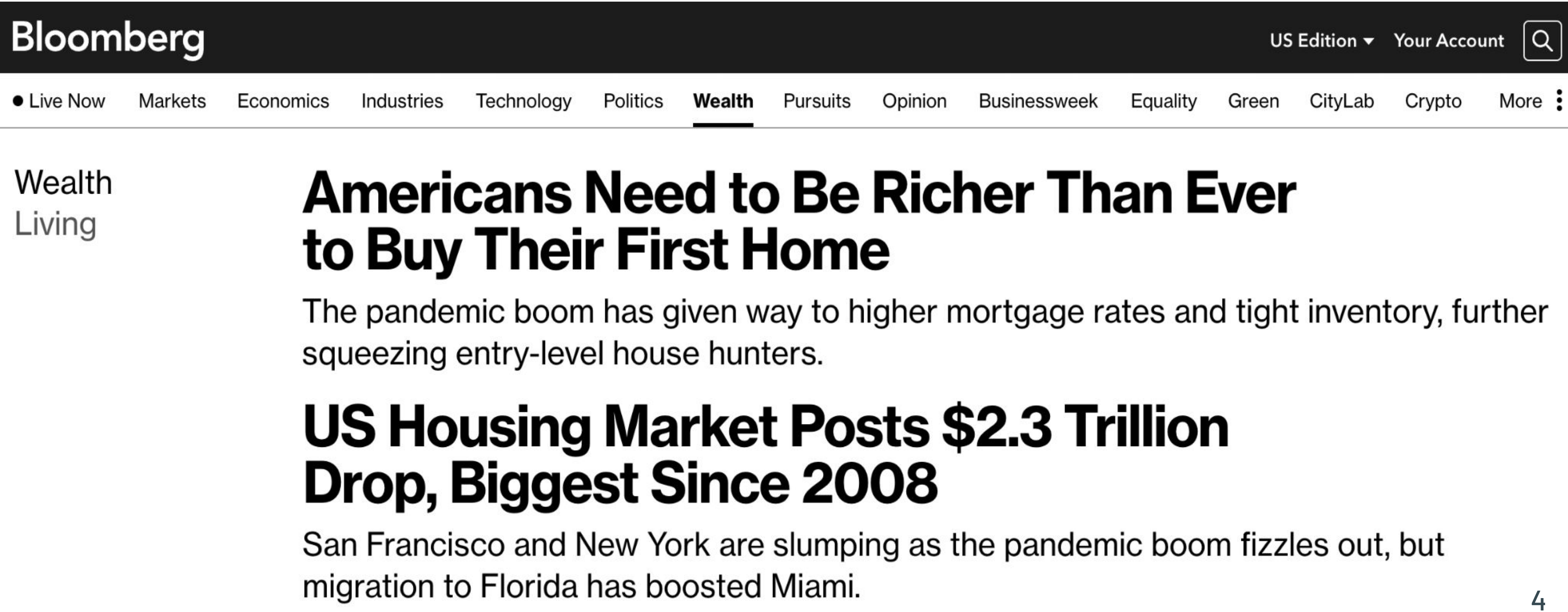
Business Problem



ECONOMY | U.S. ECONOMY

Home-Price Growth Slowed in 2022

Case-Shiller index rose 5.8% in the year ended in December amid rising mortgage rates



Wealth
Living

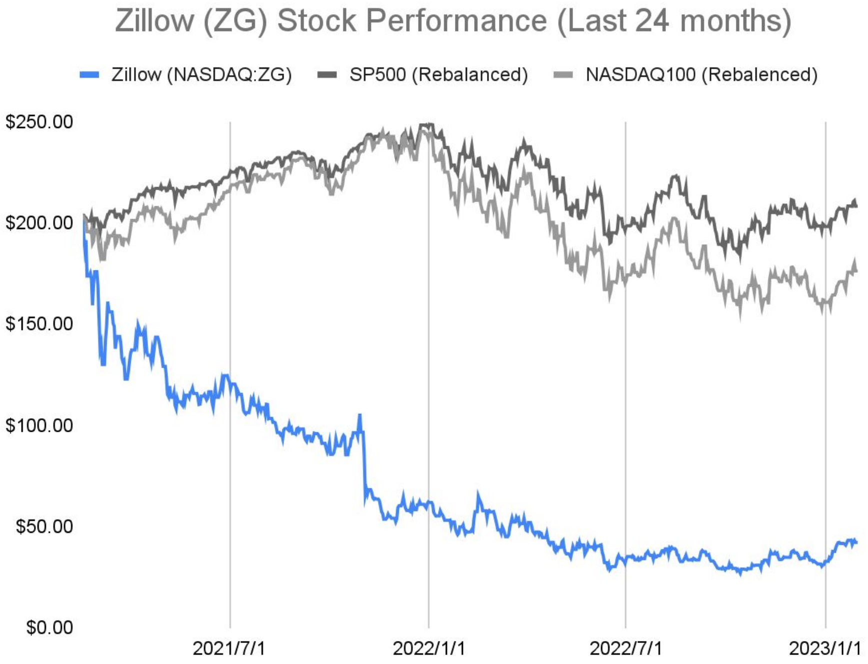
Americans Need to Be Richer Than Ever to Buy Their First Home

The pandemic boom has given way to higher mortgage rates and tight inventory, further squeezing entry-level house hunters.

US Housing Market Posts $2.3 Trillion Drop, Biggest Since 2008

San Francisco and New York are slumping as the pandemic boom fizzles out, but migration to Florida has boosted Miami.

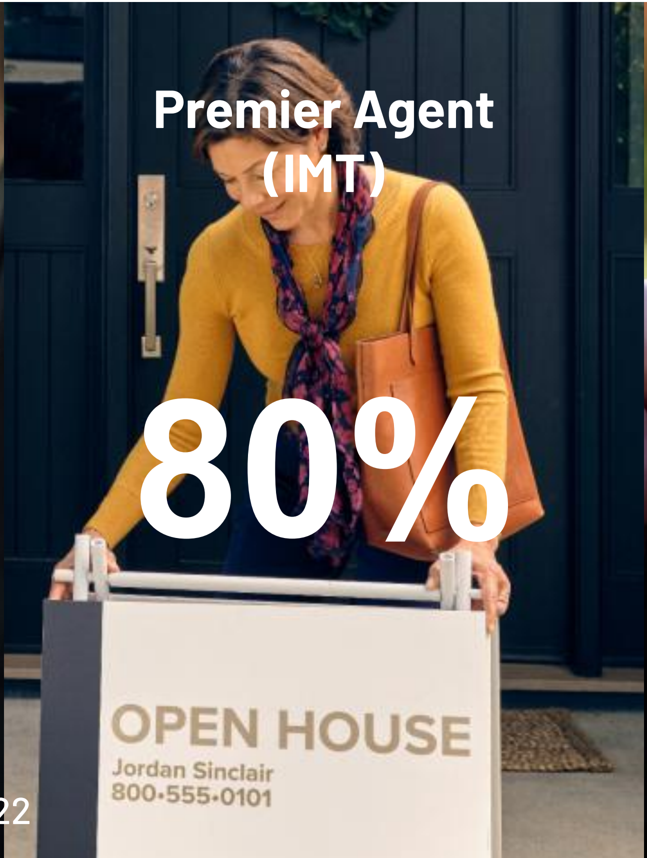
Zillow is in need of a state-of-art data mining system to enhance its revenue


Zillow (ZG) Stock Performance (Last 24 months)

Zillow is an online real-estate marketplace with three main business sectors

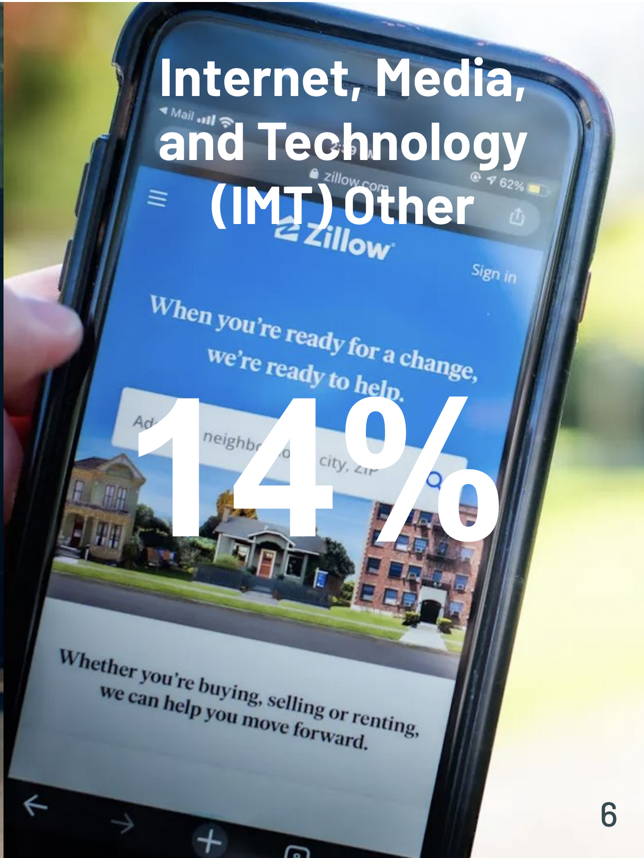
Mortgage

6%



Premier Agent (IMT)

80%


zillow.com
Zillow
Sign in
When you're ready for a change, we're ready to help.
Whether you're buying, selling or renting, we can help you move forward.

Internet, Media, and Technology (IMT) Other

14%

Percentage composition as of FY2022

A two-part methodologies can effectively target Zillow’s business segments and grow revenue

Geo-spatial Clustering

- Improved web search system to offer targeted recommendation on available homes for sale
- Help company identify market patterns and trends overtime

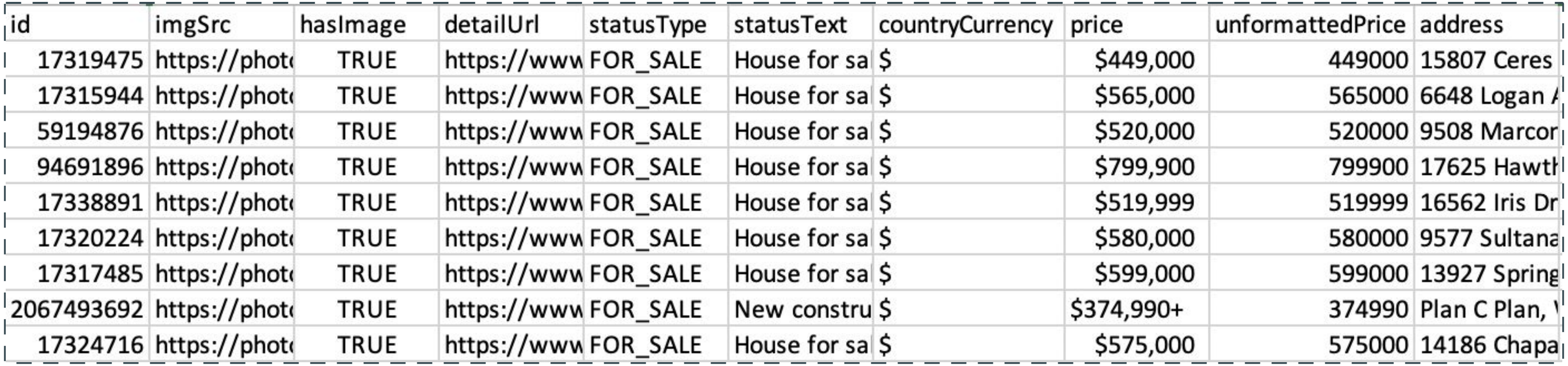
Supervised Learning

- Provide timely, market-based valuations to enhance profitability of Zillow’s home loans business
- Provide data-driven strategy for home realtor to finetune listing prices and facilitate sales

2

EDA

Each row represents a property listed on Zillow as of data extraction date



id	imgSrc	hasImage	detailUrl	statusType	statusText	countryCurrency	price	unformattedPrice	address
17319475	https://phot	TRUE	https://www	FOR_SALE	House for sa	$	$449,000	449000	15807 Ceres
17315944	https://phot	TRUE	https://www	FOR_SALE	House for sa	$	$565,000	565000	6648 Logan A
59194876	https://phot	TRUE	https://www	FOR_SALE	House for sa	$	$520,000	520000	9508 Marcor
94691896	https://phot	TRUE	https://www	FOR_SALE	House for sa	$	$799,900	799900	17625 Hawth
17338891	https://phot	TRUE	https://www	FOR_SALE	House for sa	$	$519,999	519999	16562 Iris Dr
17320224	https://phot	TRUE	https://www	FOR_SALE	House for sa	$	$580,000	580000	9577 Sultana
17317485	https://phot	TRUE	https://www	FOR_SALE	House for sa	$	$599,000	599000	13927 Spring
2067493692	https://phot	TRUE	https://www	FOR_SALE	New constru	$	$374,990+	374990	Plan C Plan, V
17324716	https://phot	TRUE	https://www	FOR_SALE	House for sa	$	$575,000	575000	14186 Chapa

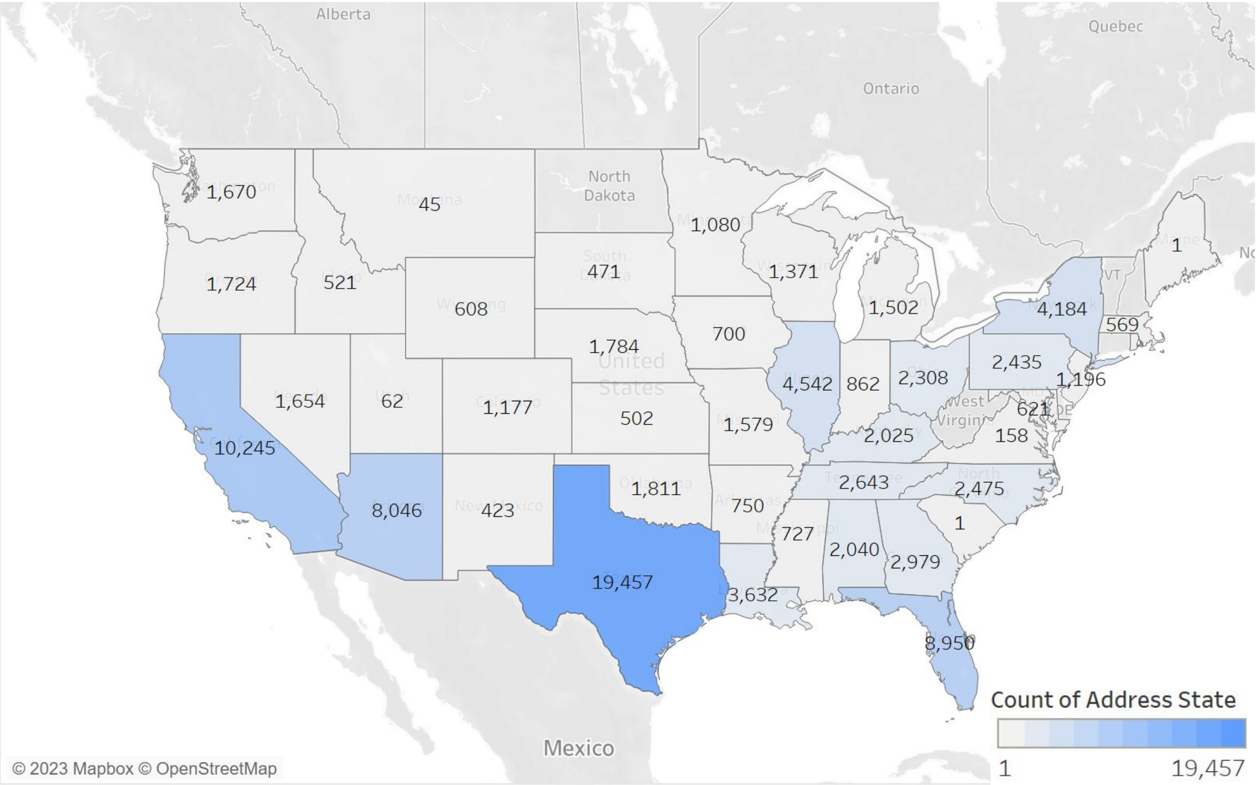
Original data: 127014 rows * 63 cols

Unique identifiers: property listing id

Variables: addressState, addressZipcode, beds, baths, area, PriceReduction, etc.

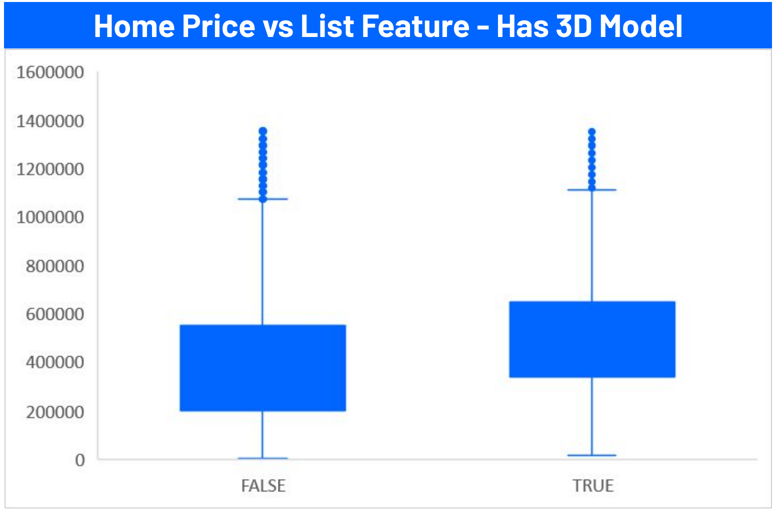
Number of variables with >20% NAs: 16 columns

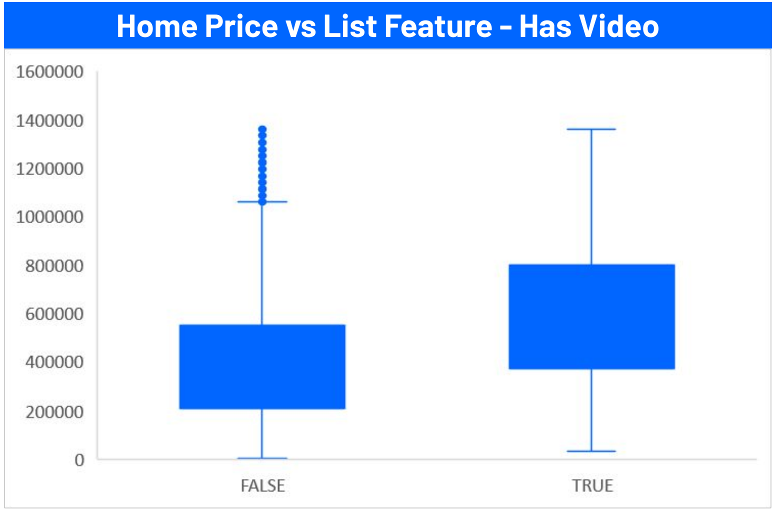
Properties listed are concentrated in a small number of states including TX, CA, and FL



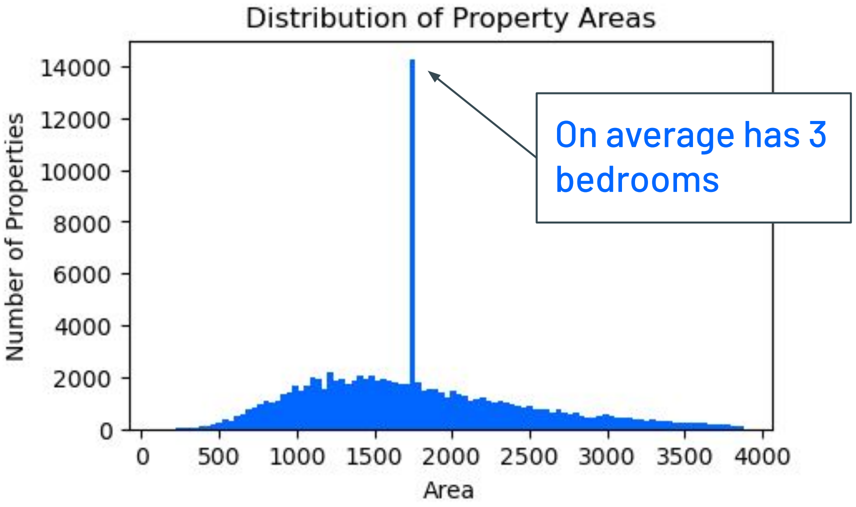
© 2023 Mapbox © OpenStreetMap

Properties with video and 3D models are listed for higher prices compared to others

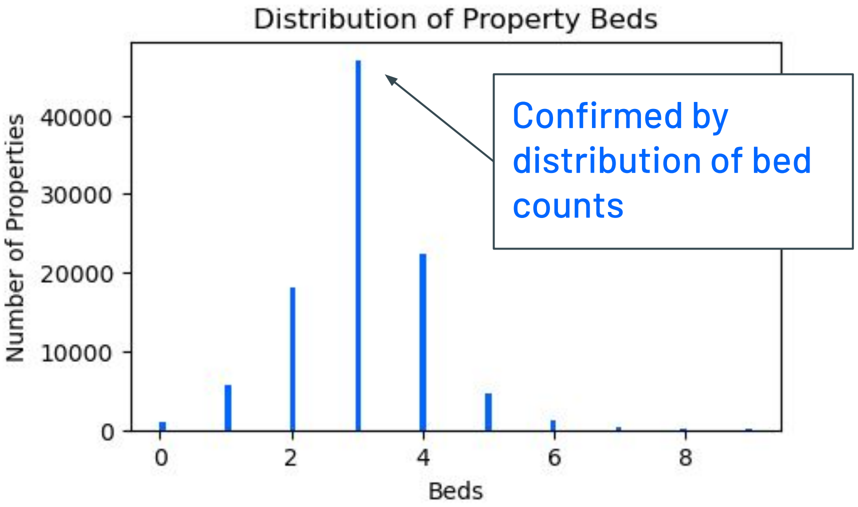




Properties have an abnormally high frequency of being sized at 1737 Sqr. Feet



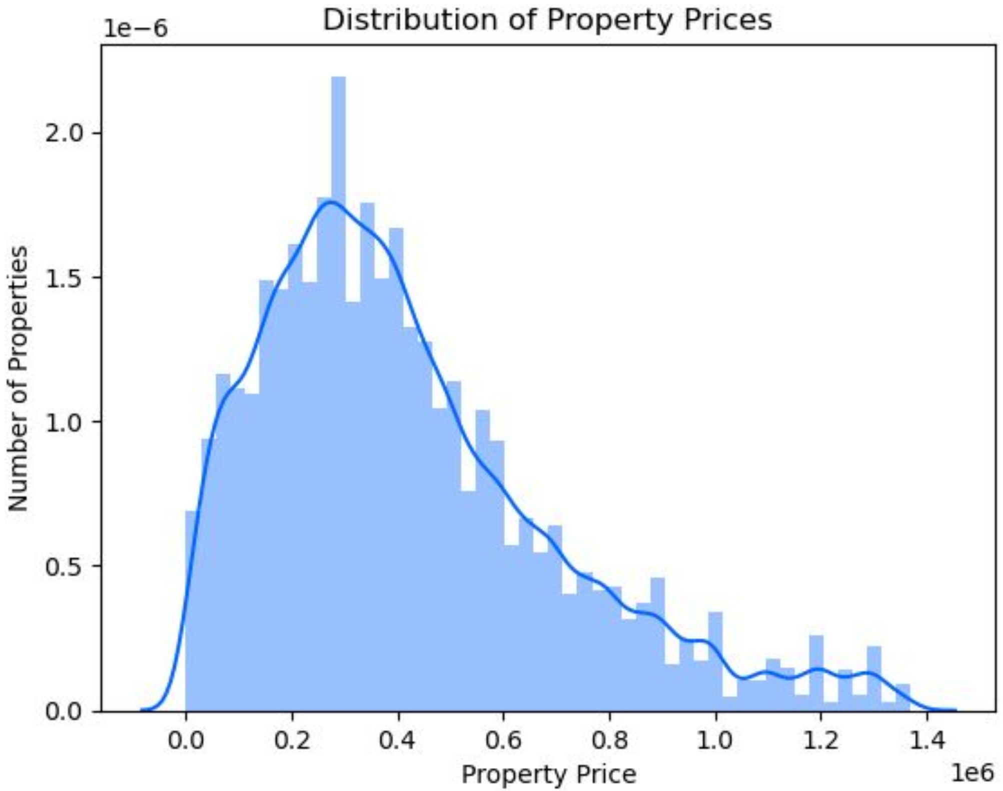
On average has 3 bedrooms



Confirmed by distribution of bed counts

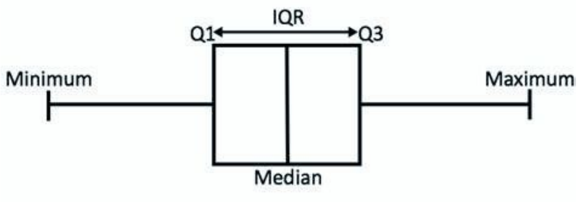
We accepted the anomaly as a natural feature of the dataset after careful analysis

Properties prices are right-skewed with a mean of USD 0.42 Million Dollar

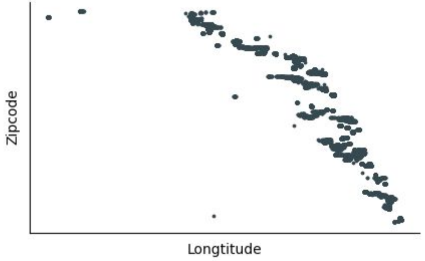


Variables were transformed and engineered according to EDA insights

Outlier Detection



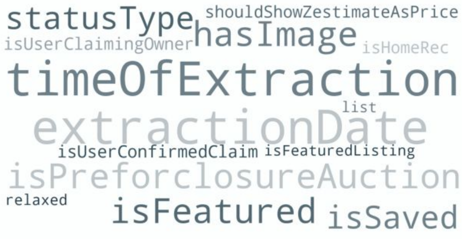
Removed correlated variables



(Zip Code, Longitude)

Corr. = -0.900607

Dropped uninformative variables



Encode categorical variables

<u>Binary:</u>
-has3DModel
-hasVideo
-IsZillowOwned

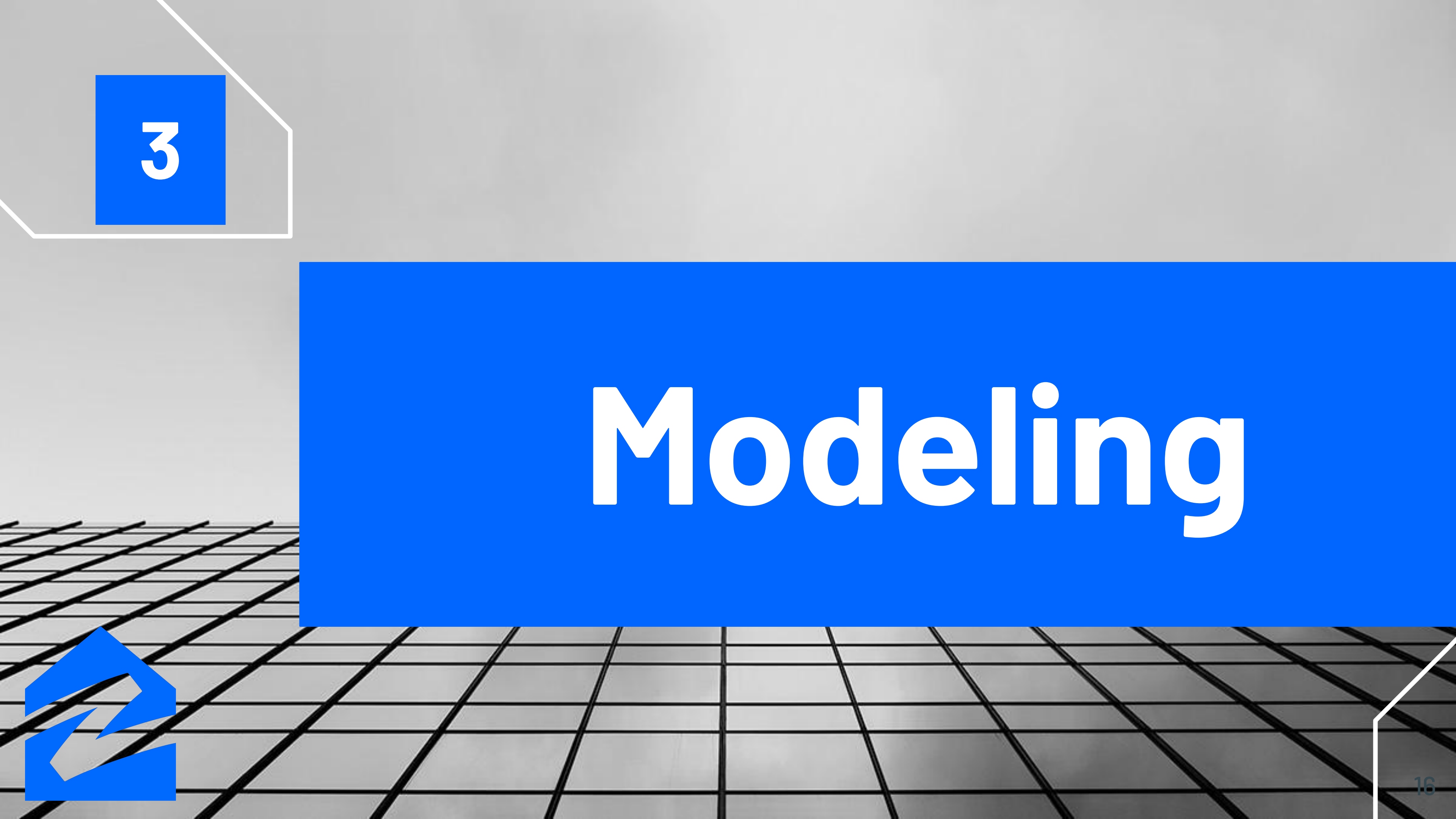
<u>Multi-category:</u>
-Property Status
-Home Type

Number of columns was reduced from 63 to 26

100770 rows * 26 cols

Numerical	price	Listing price of the property
	area	Area of the property
	priceReductionRatio	The percentage change in the price of the property
	lotAreaRaw	The total outdoor areas that is included with a property
Categorical	has3DModel	The property is listed with a 3D model
	hasVideo	The property is listed with a video
	homeType	The style or design of a residential property

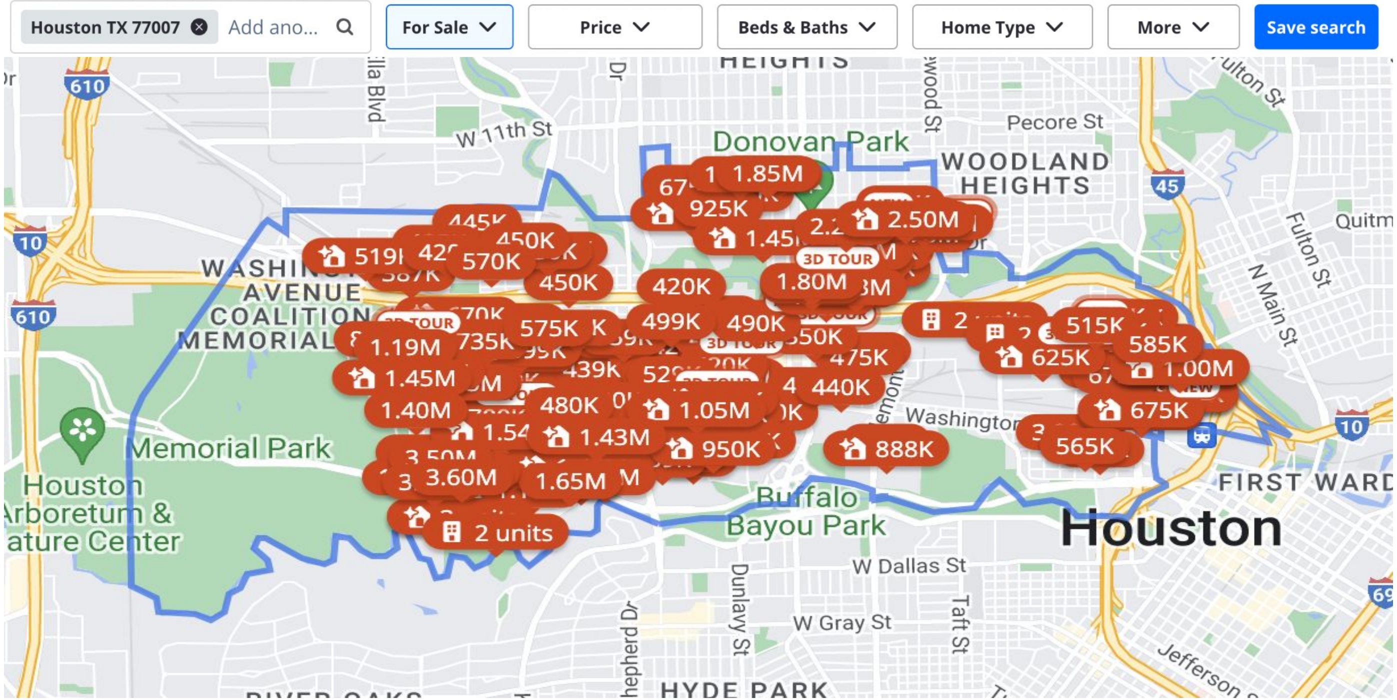
...



3

Modeling

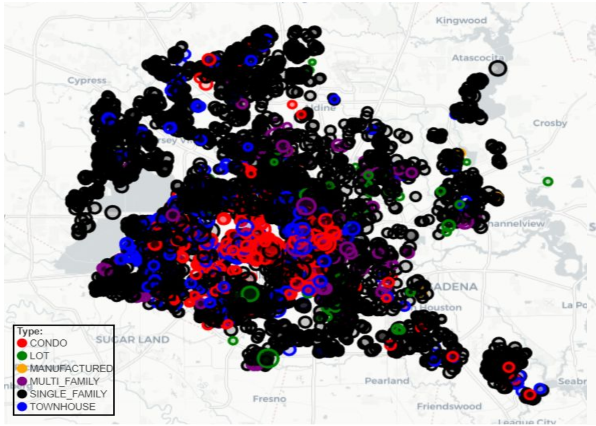
Zillow’s current method of region segmentation is not intuitive and convenient for property buyers



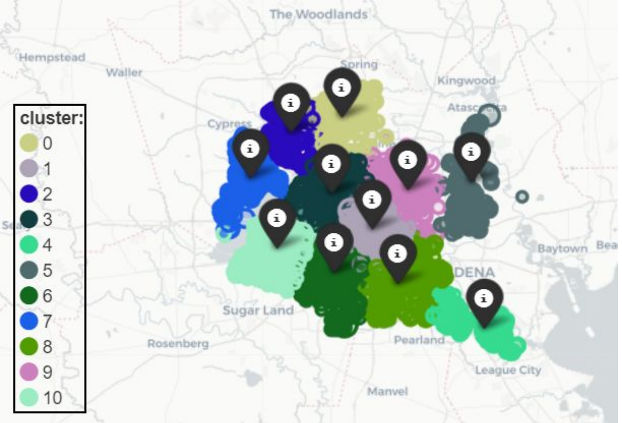
Geo-spatial clustering improves neighborhood identification for decision making

Example: Houston, TX

K-Means - 11 clusters
Silhouette score: 0.44



Original



Post-clustering

Opportunity assessment

Identify hot markets for better resource allocation

Recommendation system

Enhance recommendation system on the Zillow platform

Targeted advertising

Provide intelligence regarding optimal advertising region to realtors

Five models were fitted onto the dataset to find the optimal model being XGBoost

Regression type	Model rationale	Performance MAE, RMSE
Gamma Reg.	Generalized linear model that assumes a gamma distribution of the dependent variable	MAE: 217574.40 RMSE: 278160.97
Basic LR	Statistical method for modeling the linear relationship between a dependent variable and one or more independent variables	MAE: 178361.20 RMSE: 238032.96
Decision Tree	Non-parametric model that partitions features into subsets based on the information gain to make predictions	MAE: 113352.34 RMSE: 182586.19
Random Forest	Ensemble algorithm that combines multiple trees to improve predictive accuracy and reduce overfitting	MAE: 86662.26 RMSE: 133781.59
XGBoost	Gradient boosting algorithm that sequentially builds trees to achieve high predictive performance	MAE: 89624.21 RMSE: 132614.31

K-fold cross validation was used during modeling to refine performance measures and prevent overfitting

0.8-0.2 train-test split and 10-fold cross validation was employed

	MAE	Training RMSE	CV RMSE	R^2
Gamma Reg.	217574.40	278160.97	279219.85	-0.0001
Basic LR	178361.20	238032.96	237524.93	0.2676
Decision Tree	113352.34	182586.19	182319.98	0.5691
Random Forest	86662.26	133781.59	136506.43	0.7687
XGBoost	**89624.21**	**132614.31**	**132691.21**	**0.7727**

Under fitted

Slightly overfitted

Hyperparameter tuning of XGBoost model with grid search cross validation

Max Depth = 8

Maximum depth of ensemble decision trees

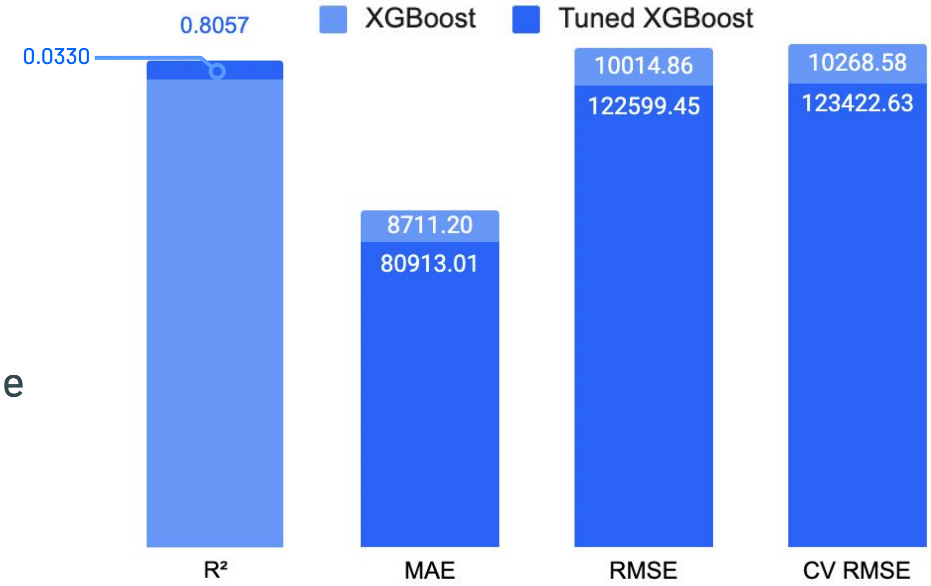
Learning Rate = 0.2

The weight of each tree in the final ensemble

Min Child Weight = 5

Minimum number of samples in each leaf node of decision trees

Model Performance Improvement



Use XGBoost to provide properties’ valuation guidance

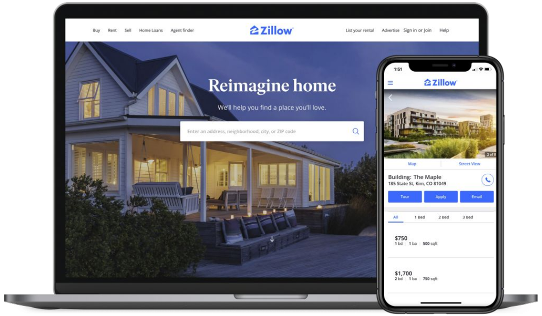
Top 5 Variables

Feature Names	Importance Score
homeType_MANUFACTURED	0.251008
baths	0.147519
homeType_LOT	0.116948
sgapt_Foreclosure	0.074820
homeType_CONDO	0.059859

Conclusions

4

Proposed Future Extensions



Reimagine home

We'll help you find a place you'll love.

Enter an address, neighborhood, city, or ZIP code

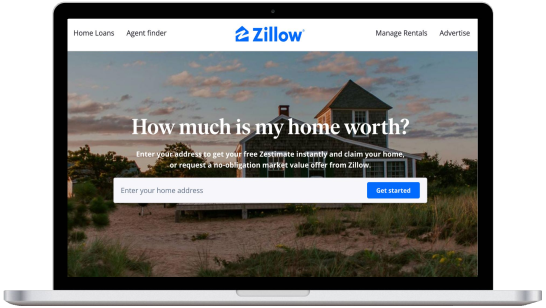
Building: The Maple
185 State St, Kim, CO 81049

$750
1 bd | 1 ba | 500 sqft

$1,700
2 bd | 1 ba | 750 sqft

IMT Searching

Revamp search results rendering through clusters



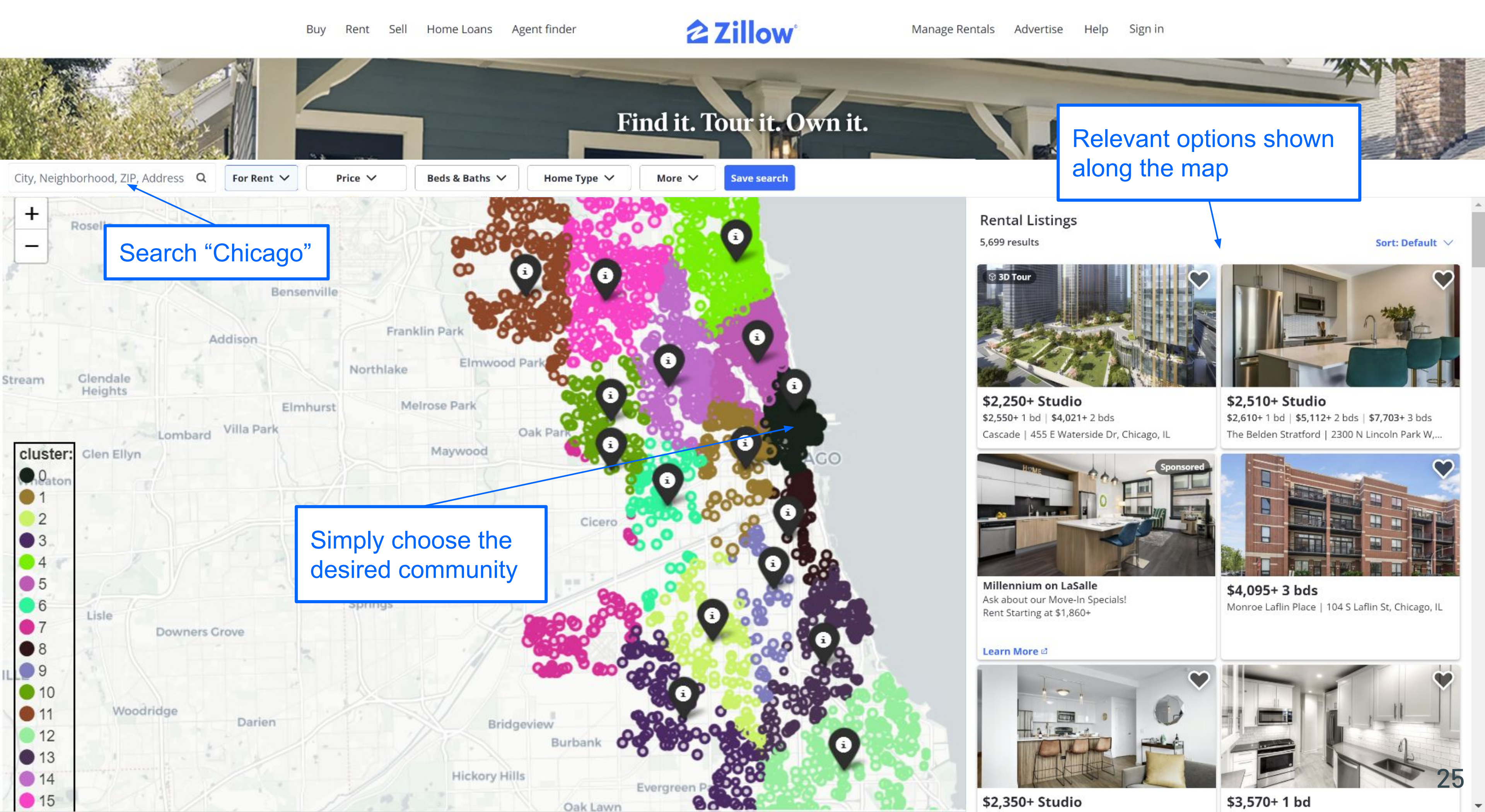
How much is my home worth?

Enter your address to get your free Zestimate instantly and claim your home, or request a no-obligation market value offer from Zillow.

Enter your home address

Zestimate

Provide enhanced price estimation for all listings



Rental Listings

5,699 results

$2,250+ Studio
$2,550+ 1 bd | $4,021+ 2 bds
Cascade | 455 E Waterside Dr, Chicago, IL

$2,510+ Studio
$2,610+ 1 bd | $5,112+ 2 bds | $7,703+ 3 bds
The Belden Stratford | 2300 N Lincoln Park W,...

Millennium on LaSalle
Ask about our Move-In Specials!
Rent Starting at $1,860+
Learn More

$4,095+ 3 bds
Monroe Laflin Place | 104 S Laflin St, Chicago, IL

$2,350+ Studio

$3,570+ 1 bd

Relevant options shown along the map

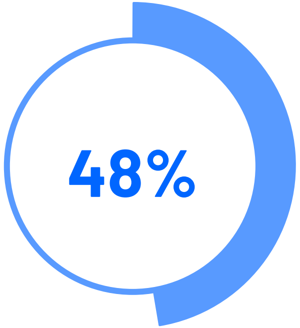
Search “Chicago”

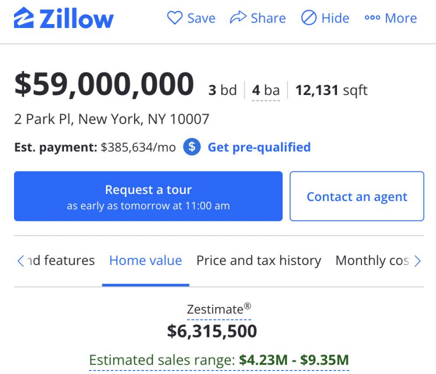
Simply choose the desired community

Enhanced Zestimate with XGBoost

Current

Current Zestimate





$59,000,000 3 bd | 4 ba | 12,131 sqft

2 Park Pl, New York, NY 10007

Est. payment: $385,634/mo **Get pre-qualified**

Zestimate®

$6,315,500

Estimated sales range: **$4.23M - $9.35M**

Data Missing

Far-off Estimates

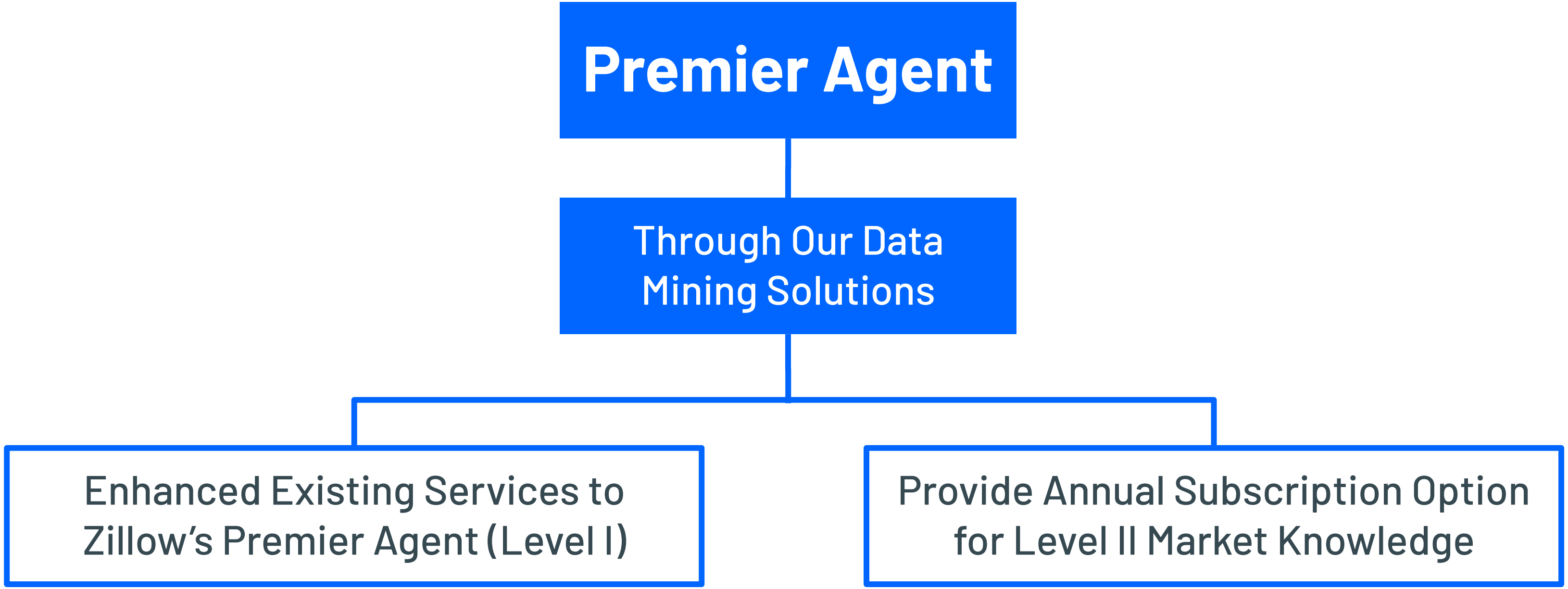
Future

Enhanced Model

- ✓ Lower MAE
- ✓ Lower RMSE
- ✓ No Overfitting

Robust Performance

Generate Extra Revenue though Premier Agent's Service Differentiation



Premier Agent

Through Our Data Mining Solutions

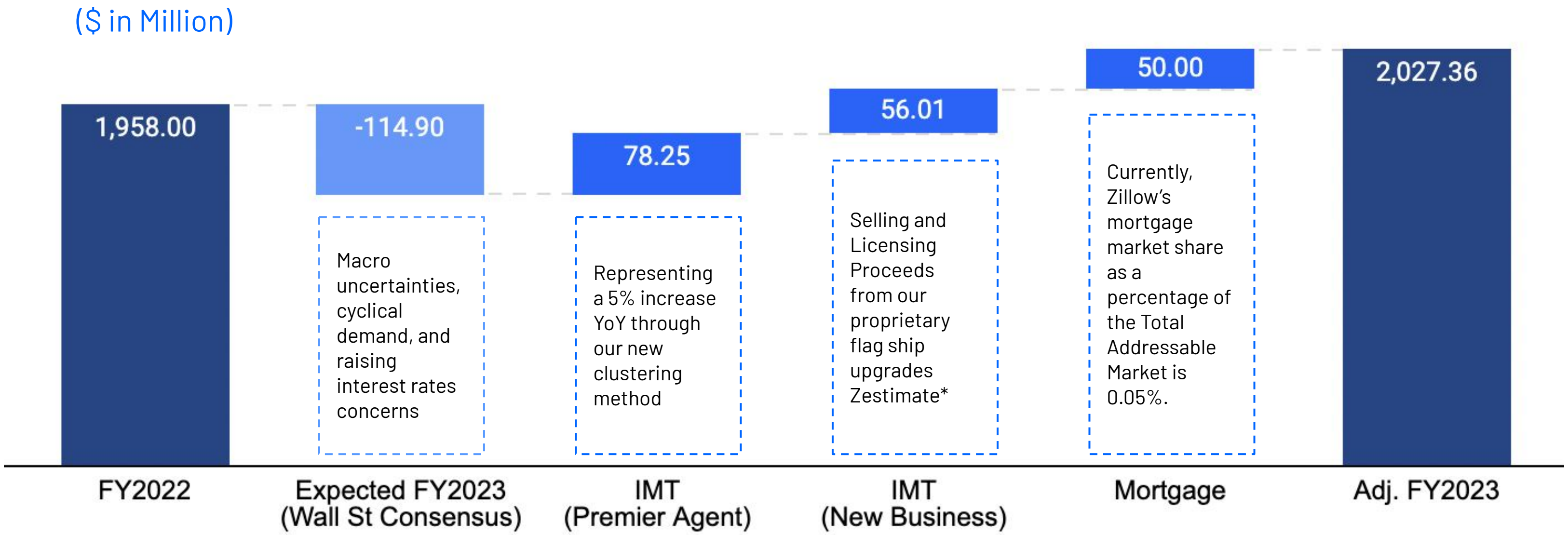
Enhanced Existing Services to Zillow's Premier Agent (Level I)

Provide Annual Subscription Option for Level II Market Knowledge

Improve Zillow's Baseline Service

Develop New Source of Revenue

Implementation synergies can bring $69.36 M in revenue upside despite macroeconomic headwinds



Macro uncertainties, cyclical demand, and raising interest rates concerns

Representing a 5% increase YoY through our new clustering method

Selling and Licensing Proceeds from our proprietary flag ship upgrades Zestimate*

Currently, Zillow's mortgage market share as a percentage of the Total Addressable Market is 0.05%.

From:
Down
114.9 M YoY
(Basecase)

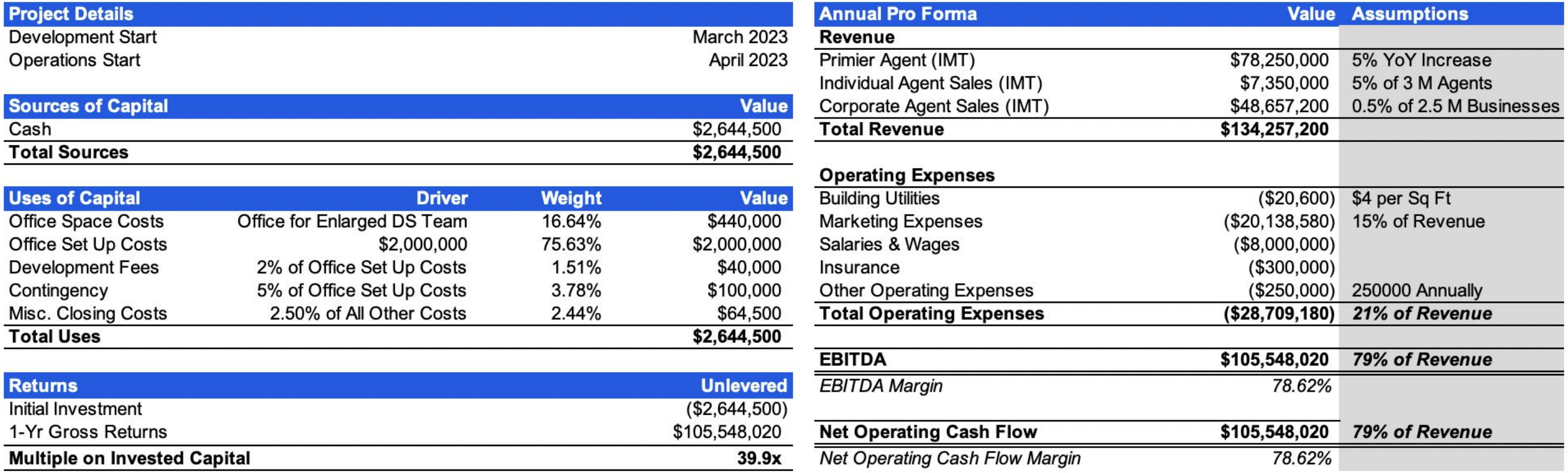
5.86%

State of Art Turnaround

3.54%

To:
Up
69.36 M YoY
(Estimated)

Executive Financial Summary of Project Implementation



Project Details			
Development Start			March 2023
Operations Start			April 2023

Sources of Capital			Value
Cash			$2,644,500
Total Sources			**$2,644,500**

Uses of Capital	Driver	Weight	Value
Office Space Costs	Office for Enlarged DS Team	16.64%	$440,000
Office Set Up Costs	$2,000,000	75.63%	$2,000,000
Development Fees	2% of Office Set Up Costs	1.51%	$40,000
Contingency	5% of Office Set Up Costs	3.78%	$100,000
Misc. Closing Costs	2.50% of All Other Costs	2.44%	$64,500
Total Uses			**$2,644,500**

Returns			Unlevered
Initial Investment			($2,644,500)
1-Yr Gross Returns			$105,548,020
Multiple on Invested Capital			**39.9x**

Annual Pro Forma	Value	Assumptions
Revenue		
Primier Agent (IMT)	$78,250,000	5% YoY Increase
Individual Agent Sales (IMT)	$7,350,000	5% of 3 M Agents
Corporate Agent Sales (IMT)	$48,657,200	0.5% of 2.5 M Businesses
Total Revenue	**$134,257,200**	
Operating Expenses		
Building Utilities	($20,600)	$4 per Sq Ft
Marketing Expenses	($20,138,580)	15% of Revenue
Salaries & Wages	($8,000,000)	
Insurance	($300,000)	
Other Operating Expenses	($250,000)	250000 Annually
Total Operating Expenses	**($28,709,180)**	***21% of Revenue***
EBITDA	**$105,548,020**	***79% of Revenue***
EBITDA Margin	*78.62%*	
Net Operating Cash Flow	**$105,548,020**	***79% of Revenue***
Net Operating Cash Flow Margin	*78.62%*	

39.9x
Capital Returns

~100 M
Operating Income

78.62%
Operating Margin

Thank you for Listening!

References

- Zillow Business Model
- Zillow FY2022 Form 10-K
- Zillow Data Source

Risks Factors

- Zillow's business has and may continue to be impacted by the current and future health and stability of the economy and United States residential real estate industry, including inflationary conditions, interest rates, housing availability and affordability, labor shortages and supply chain issues.

- Zillow may be unable to adequately protect or continue using our intellectual property or prevent others from copying, infringing upon, or developing similar intellectual property.

- Zillow obtains certain real estate data, such as transaction history, property descriptions, tax-assessed value and property taxes paid, under licenses from third-party data providers. Zillow use this data to enable the development, maintenance and improvement of our marketplace and information services, including Zestimates, Rent Zestimates and our living database of homes. Zillow have invested significant time and resources to develop proprietary algorithms, valuation models, software and practices to use and improve on this specific data. Zillow may be unable to access certain of this data from vendors or government agencies if changes in local laws or regulations or other prohibitions on data sharing are implemented or because the quality and quantity of data available to these third parties changes. Zillow may also be unable to renew our licenses with these data providers or enter into new data license agreements, or Zillow may be able to do so only on terms that are less favorable to us, which could harm our ability to continue to develop, maintain and improve these information services and could harm our business, results of operations and financial condition.

Disclaimer

- The Content is for informational purposes only, you should not construe any such information or other material as legal, tax, investment, financial, or other advice. Nothing contained on our presentation constitutes a solicitation, recommendation, endorsement, or offer by any third-party service provider to buy or sell any securities or other financial instruments in this or in any other jurisdiction in which such solicitation or offer would be unlawful under the securities laws of such jurisdiction.

- All Content on this presentation is information of a general nature and does not address the circumstances of any particular individual or entity. Nothing in the presentation constitutes professional and/or financial advice, nor does any information in the presentation constitute a comprehensive or complete statement of the matters discussed or the law relating thereto. You alone assume the sole responsibility of evaluating the merits and risks associated with the use of any information or other Content on the Presentation before making any decisions based on such information or other Content. In exchange for using the presentation, you agree not to hold presenters or any third-party service provider liable for any possible claim for damages arising from any decision you make based on information or other Content made available to you through the presentation.

- We do not represent or guarantee that any content in the presentation is accurate, nor that such content is a complete statement or summary of the marketplace.