

CAPSTONE PROJECT: CASHFLOW

+

o

•

Preet Sekhon
01.13.23



AGENDA

Real Estate Discussion
Data Science Workflow
Findings
Recommendations

Problem Statement

+

•

○

A group of large investors are pooling their money together to add real estate to their portfolio. Are you able to design a strategy on how best to allocate capital?

REAL ESTATE

CAPSTONE

4

The 3 Rules of Real Estate

Cash Flow
Cash Flow
Cash Flow



Why Invest for Cash Flow?

Your Cash
Works for You

Growth

Bonuses

Fixed Costs
but Rental
Growth

Principle Pay
Down

Appreciation

CASH FLOW

Revenue

- Rent

Costs

- PITI
 - Principal
 - Interest
 - Taxes(Property)
 - Insurance
 - Maintenance

Profit

- $\text{Revenue} - \text{Costs} = \Profit
- This is your Cash Flow

COSTS

- Initial Cost
 - Referenced as the first “Cash” in Cash on Cash Return
 - 25% Down Payment
- Monthly Costs
 - Mortgage
 - Principle
 - 79% of ‘MLS_Amount’
 - Remaining 75%
 - Closing Costs = 4% of ‘MLS_Amount’
 - 30 Year Fixed Rate Mortgage
 - *6.5% Interest Rate*
 - Property Tax = .71% of ‘MLS_Amount’
 - Insurance = .25% of ‘MLS_Amount’
 - Maintenance = .3% of ‘MLS_Amount’

CASH ON CASH RETURN

Cash Flow
Divided by
Initial Costs

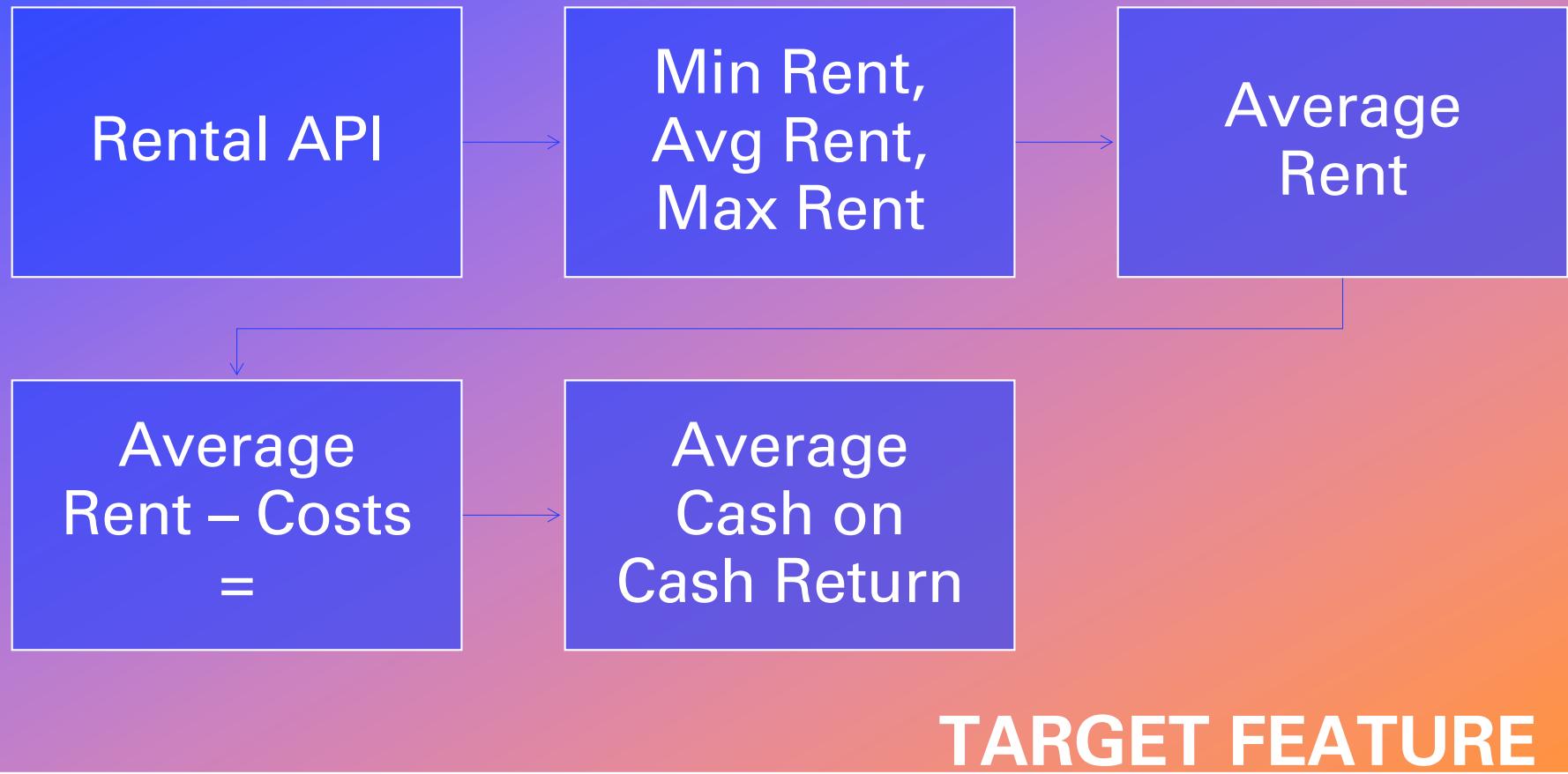


CASH ON CASH RETURN

Q: What is a good cash-on-cash return? •
+ ○

A: It depends on the investor, the local market, and your expectations of future value appreciation. Some real estate *investors are happy with a safe* and predictable *CoC return of 7% - 10%*, while others will only consider a property with a cash-on-cash return of at least 15%.

Stessa, A Roofstock Company



DATA SCIENCE WORKFLOW

Data Science Workflow

Collection

Cleaning and Preprocessing

EDA

Modeling

Conclusions

WORKFLOW: COLLECTION

14

CA.Gov – Voter Data

- 2020-2022

Census Info 2021

- Population
- Income

Real Estate Listings

- Propstream.com

Rental Listings

- API

Source	Rows	Columns
Voter Info	41	28
Census Info	41	147
Real Estate Listings	22,639	23
Rental Listings	1323	25

THE DATA

WORKFLOW: CLEANING AND PREPROCESSING



A large portion of this project was spent on Data Collection and creating one Data Set



Final Data Set (14546 , 261)*



All Nulls Removed

+

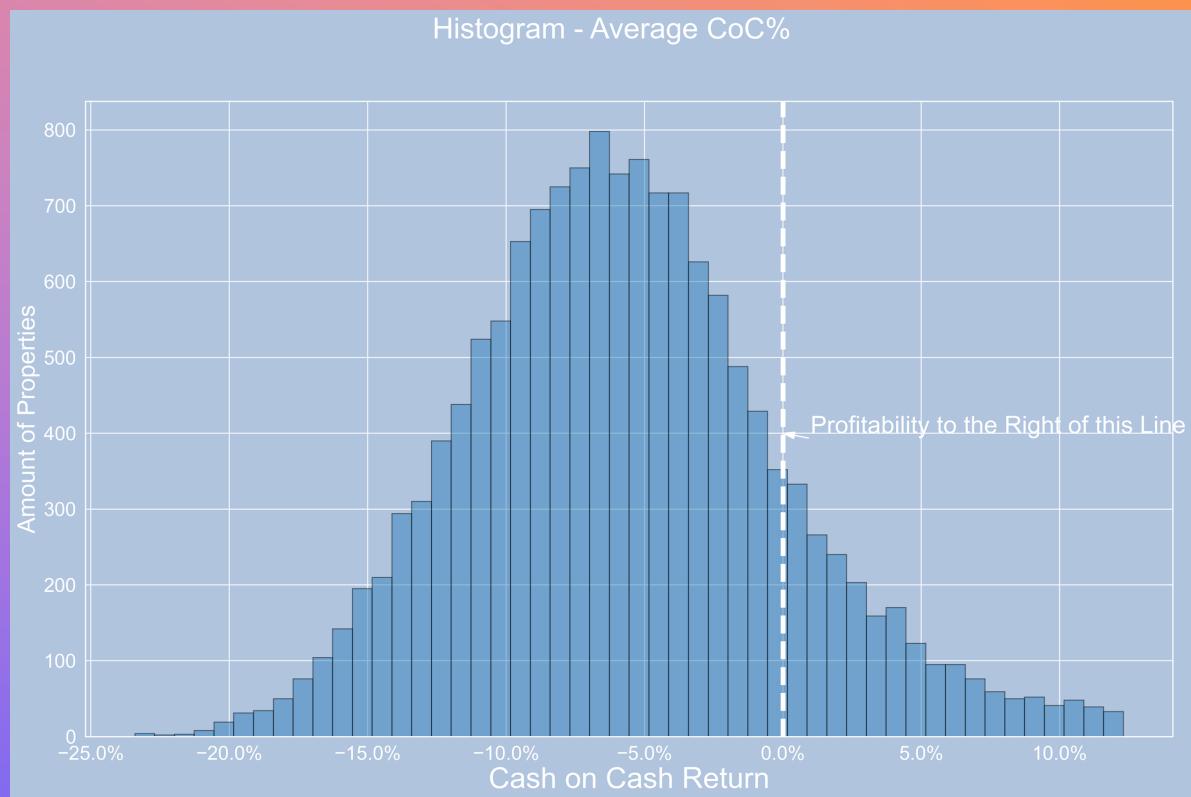
•

◦

EXPLORATORY DATA ANALYSIS

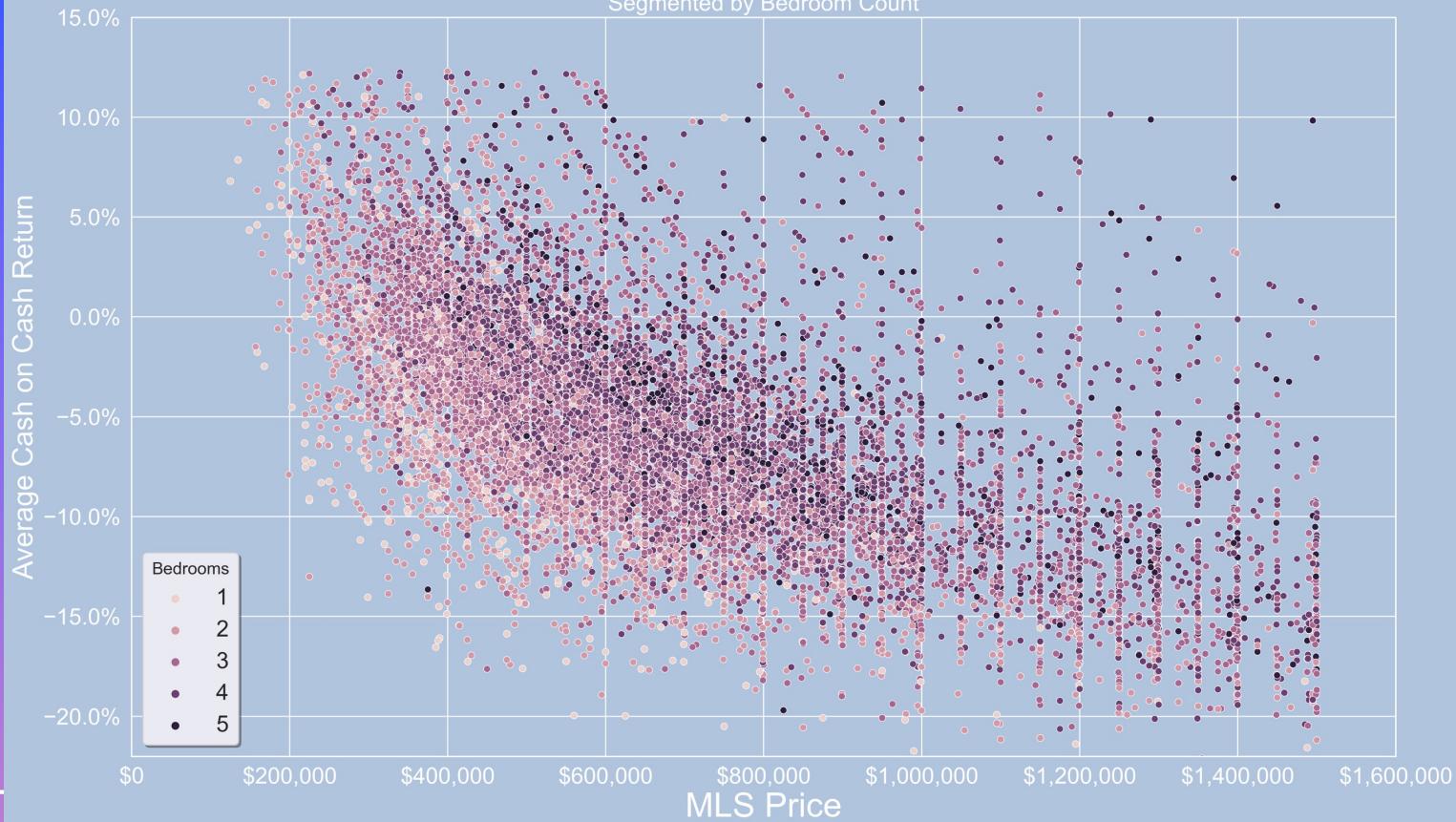
DISTRIBUTION

14.8% of Listings
are Profitable



Average CoC Return by MLS Price

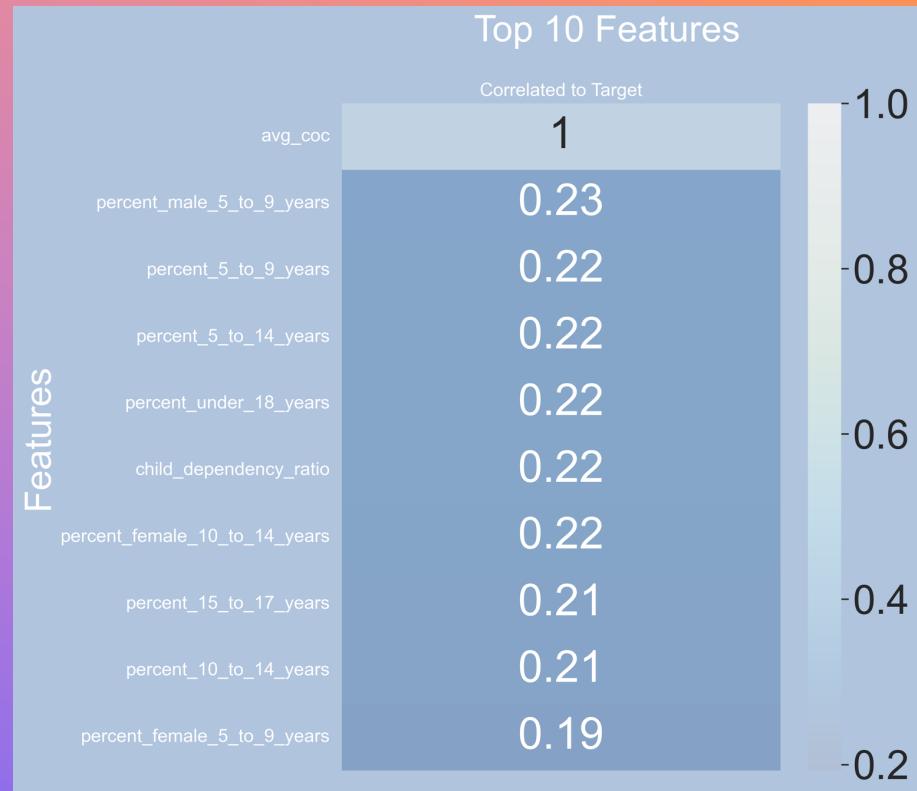
Segmented by Bedroom Count



Less +
Bedrooms
Less
Expensive
Homes

HEAT MAP

- + Highest Correlation to our Target
- Feature is associated with Renters who are Families



- +
-
-
- # MODELING

Linear Regression Models

With GridSearch CV

Regression Models	Train Score	Test Score	RMSE
Linear	.342	.343	.046
Random Forest	.918	.434	.043
Decision Tree	.999	-.098	.059
Support Vector Machines	-320	-316	.548

Pivot: Linear Regression to Classification

After much consideration and Regression Modeling the decision was made to shift to Classification Models. Our new target variable, Target Cash Flow Positivity(TCFP) was created.

Average Cash on Cash Return above or equal to 4% makes TCFP = 1(True), lower than 4% = 0(False).

Distribution of TCFP:

1 - 5.6%

0 - 94.4%

Unbalanced Data

1 - 5.6%
0 - 94.4%

Since our distribution of our target variable was so unbalanced we employed SMOTEENN in our Models.

SMOTEENN is a technique that resamples the data and creates synthetic samples using KNN.

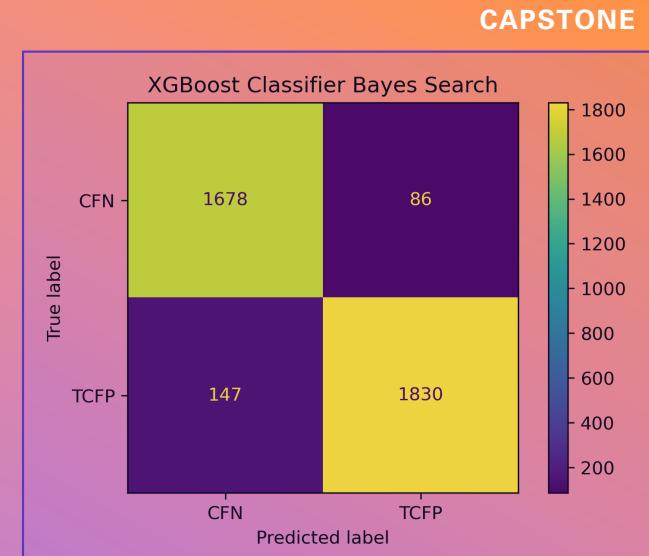
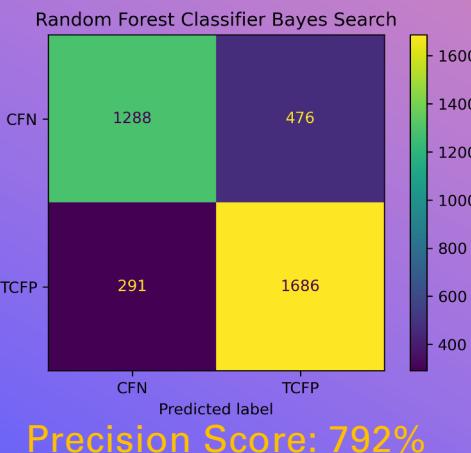
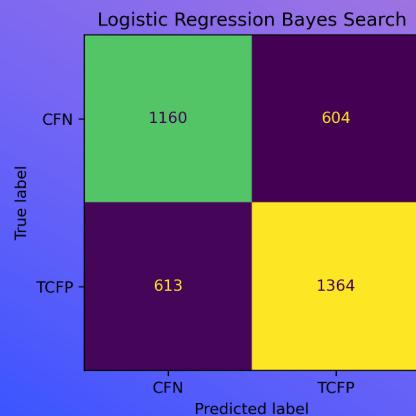
Classification Models

Classification Models	Brier Score Train Set	Brier Score Test Set	Precision Train Set	Precision Test Set
Logistic Regression	.207	.202	.695	.710
Random Forest	.146	.145	.791	.792
XGBoost	.044	.047	.962	.949

Brier Score is better if lower

All models used SMOTEENN and
BayesSearchCV

Classification Models: Confusion Matrix

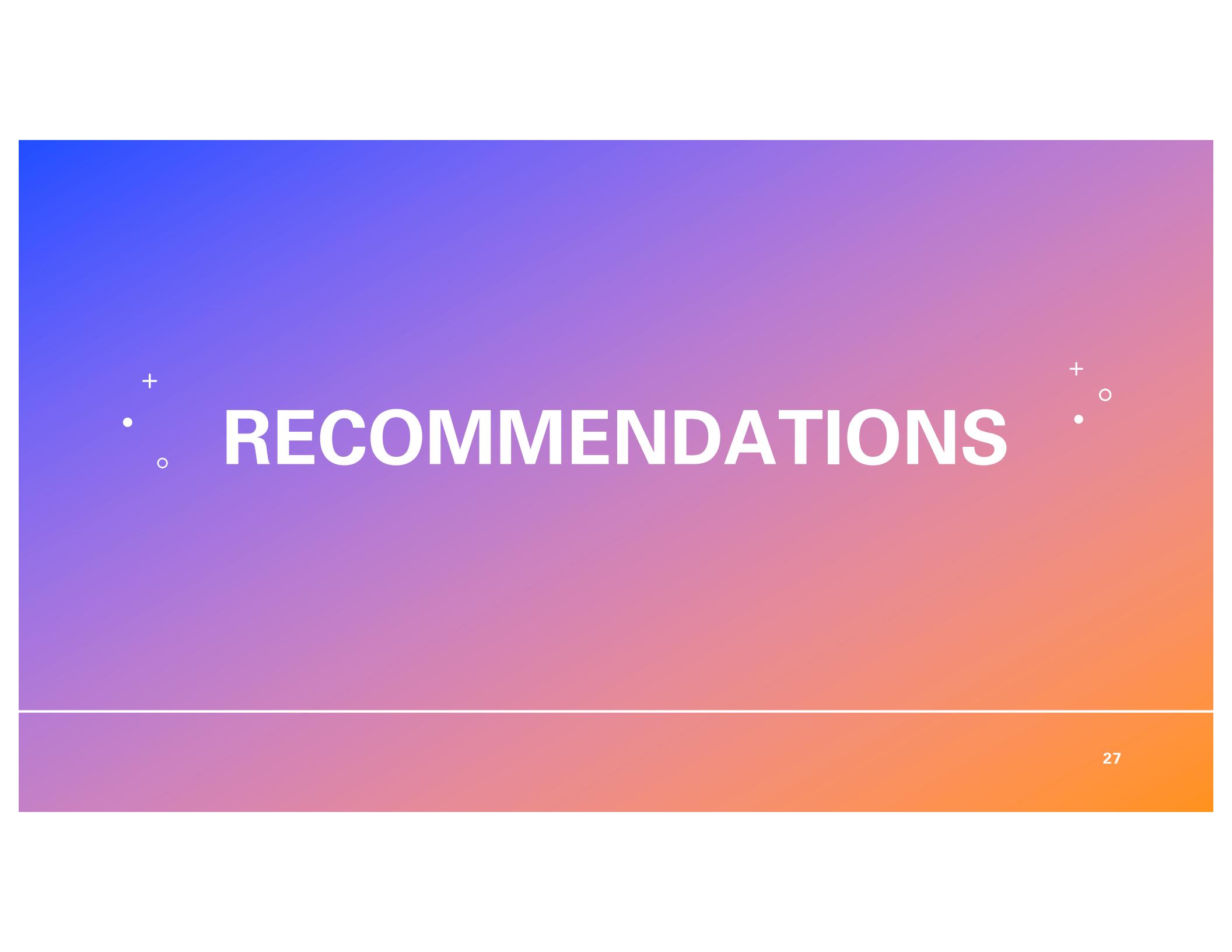


25

XGBoost Feature Importance

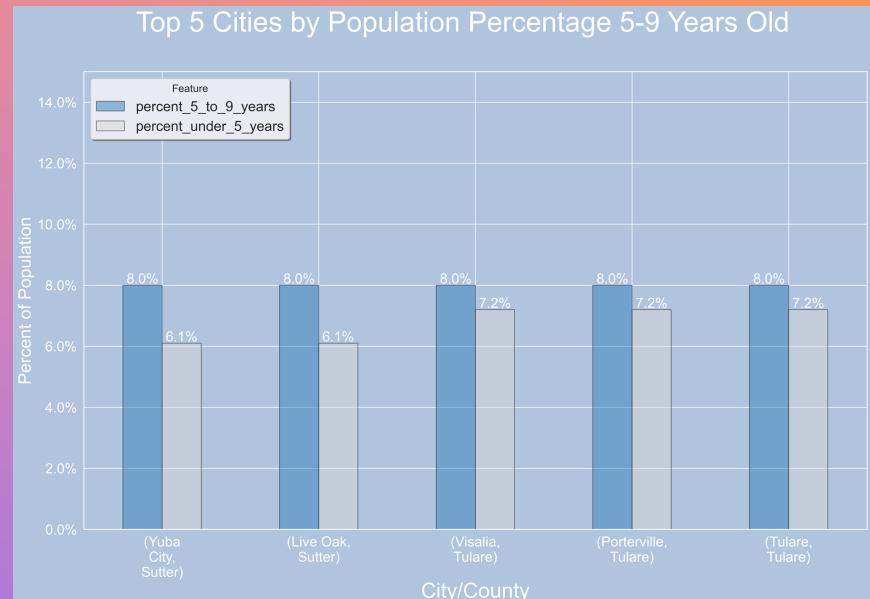
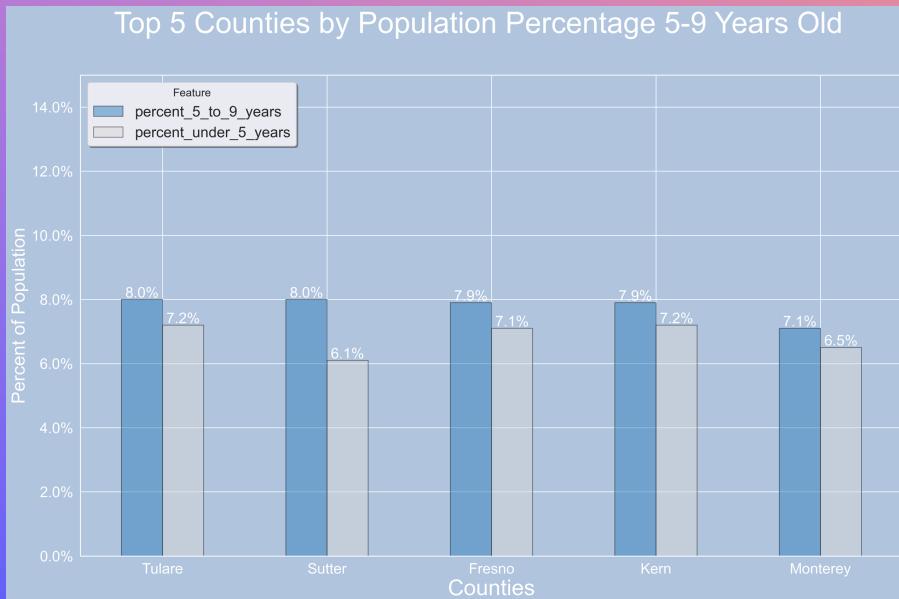
Feature	Importance
Median Household Income - White	.094
3 Bathrooms	.072
Single Family Residence	.054
Median Household Income	.049
2020 Turnout Eligible	.045
Child Dependency Ratio	.042

*Coincide with our Heat Maps from EDA



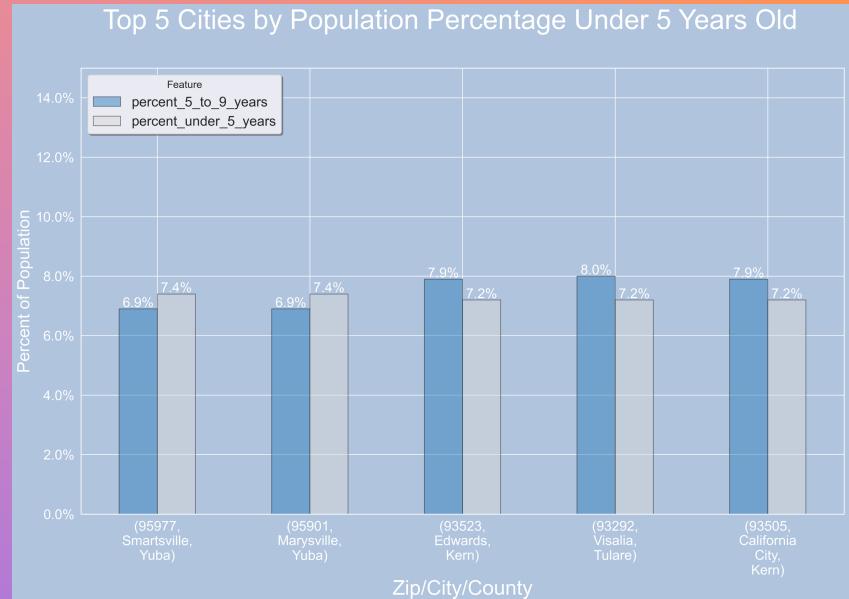
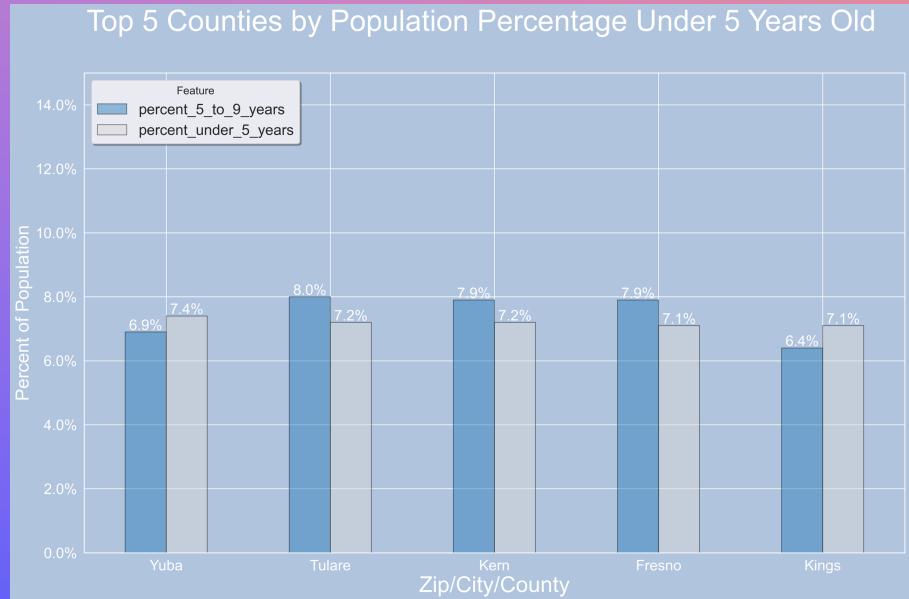
RECOMMENDATIONS

TASK 1A: TOP 5 CITY/COUNTIES FOR OUR TOP CORRELATED FEATURE



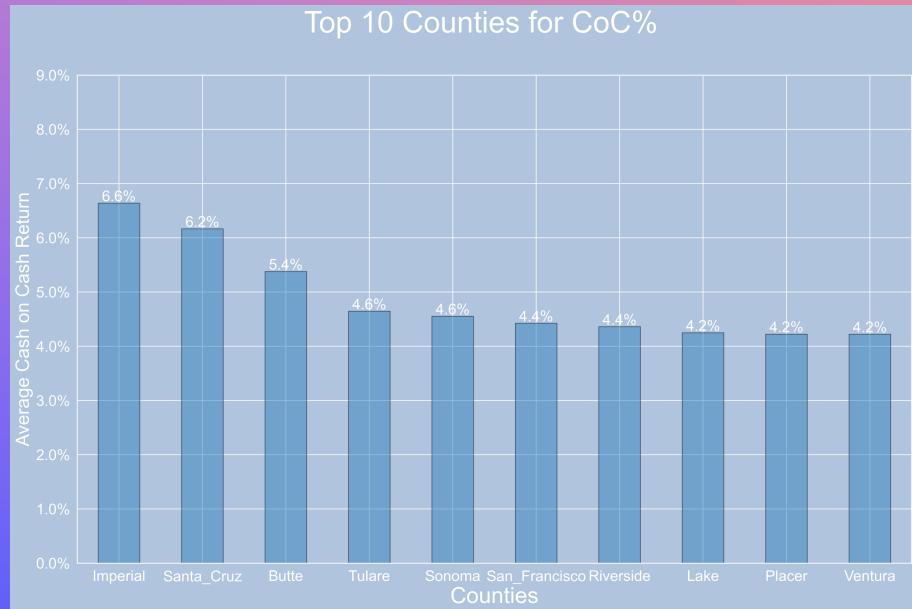
- Population % 5-9 Years
- 6.2% Average per County

TASK 1B: TOP 5 CITIES/COUNTIES FOR FUTURE CORRELATION FEATURE RICH MARKET



- Population % under 5 Years
- 5.7% Average per County

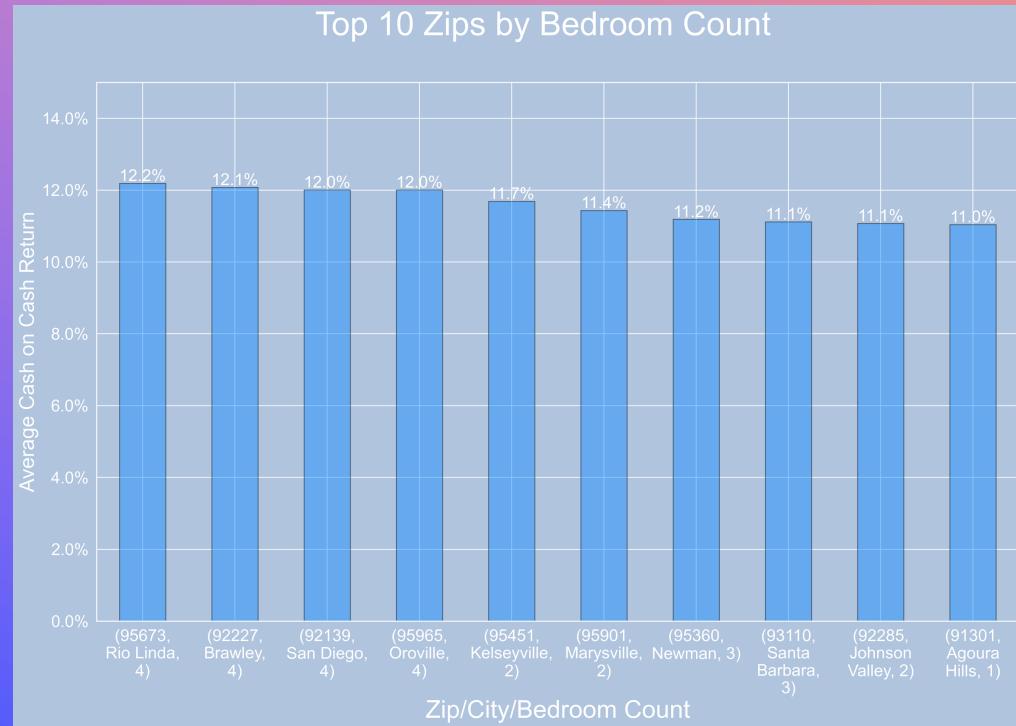
TASK 2: TOP 10 CITIES OR COUNTIES FOR COC%



A more Focused Strategy will drive a Higher CoC Return. On average, over double.

TASK 3: BY ZIP AND BEDROOM COUNT

- +
-
-



For our 3rd part of our Investment Strategy we will be hyper focused and target specific Zip Codes and Bedroom Counts

OVERALL STRATEGY



After further review of Tasks 1-3, look for opportunities to cluster investments. Clustering will provide a reduction of overhead and allow for a certain amount of scalability.

CAPSTONE

+



o



.



THANK YOU

Questions and Comments?

APPENDIX

+

•

○

INVEST FOR THE NOW

CAPSTONE

35



<u>Percent 5 to 9 years</u>	<u>County</u>	<u>Median Rent</u>	<u>Rank</u>
8.4	Merced	\$1,824	34
8	Sutter	\$1,920	28
8	Tulare	\$1,728.	36
7.9	Fresno	\$1,834	33
7.1	Monterey County	\$2,402	20

INVEST FOR THE FUTURE

CAPSTONE

36

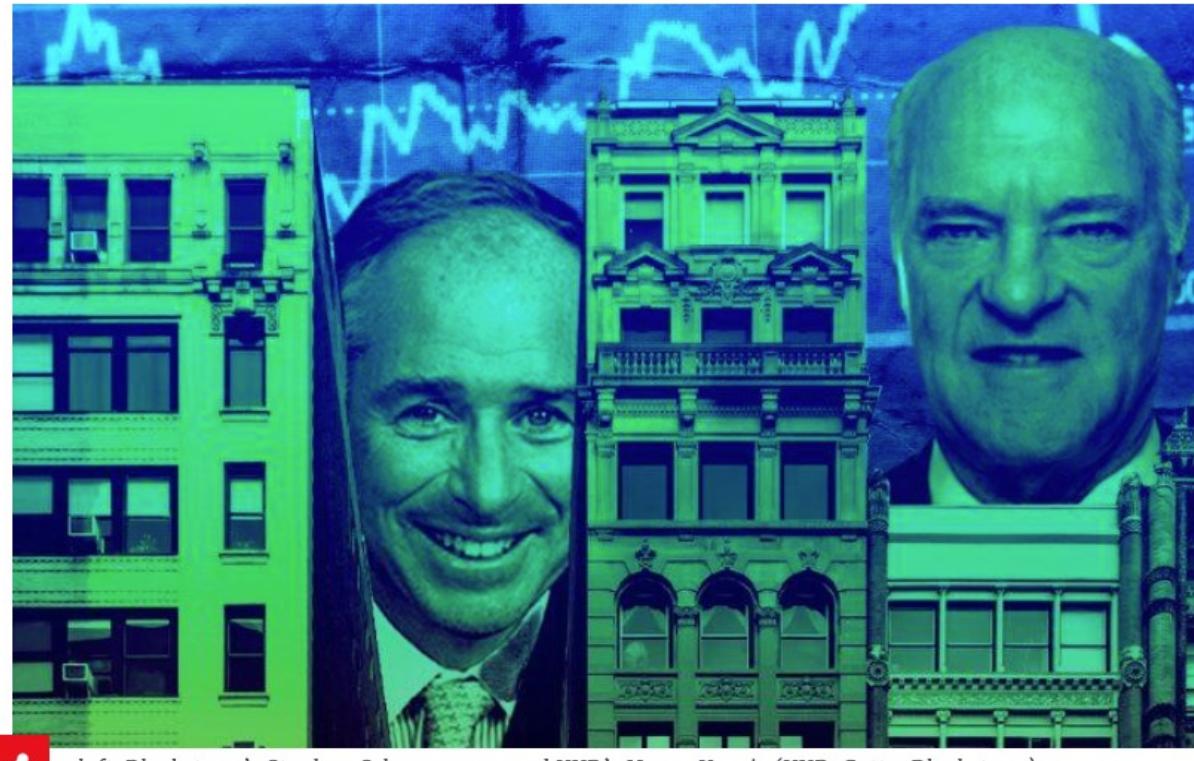


<u>Median Rent</u>	<u>Rank</u>	<u>Percent Under 5 years</u>	<u>County</u>
\$1,893	31	7.4	Yuba
\$1,660	37	7.2	Kern
\$1,729	36	7.2	Merced
\$1,825	34	7.2	Tulare
\$1,835	33	7.1	Fresno
\$1,896	30	6.8	Stanislaus

Wall Street has \$110B for homebuying spree

Blackstone, KKR among firms gearing up to purchase single-family rentals

New York / By TRD Staff



HEAT MAP

Negative

+

•

o

Features	Bottom 10 Features	
	Correlated to Target	Value
building_sqft	-0.27	-0.20
median_hh_income_white	-0.23	-0.21
percent_18_years_and_over	-0.22	-0.22
percent_16_years_and_over	-0.22	-0.23
percent_21_years_and_over	-0.22	-0.24
median_hh_income	-0.2	-0.25
percent_female_55_to_59_years	-0.2	-0.26
percent_85_years_and_over	-0.2	-0.27
2020_turnout_eligible	-0.2	
percent_male_85_years_and_over	-0.2	

