## RENT or BUY?



Introduction Current real estate tools lack crucial property specifics and demographic nuances, oversimplifying decisions. Despite machine learning advancements, challenges like data leakage and market complexity persist. Our advanced valuation system addresses gaps in existing tools, offering comprehensive analysis amid limited properties from increased private equity purchases. It aids users and agents by providing data-driven guidance to navigate the competitive real estate market.

Our Approaches We merged property data categorization by bedroom counts and zip codes with macroeconomic statistics. Featurewiz with SULOV and Recursive XGBoost efficiently selects vital features for model training. Our hybrid models, merging grid search with established machine learning algorithms, ensure a strong predictive base. The Tableau dashboard simplifies complex data into user-friendly visuals for navigation. Leveraging diverse data sources and advanced methods, our approach aims to boost predictive accuracy, addressing real estate analytics challenges.

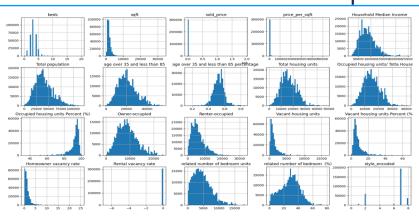
**Data** 

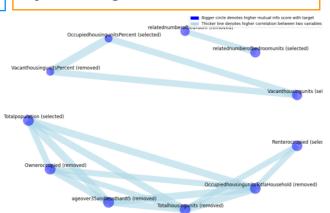
## HomeHarvest Scraped 437k sold records Collected 25 demographic and 107k for-sale and for-rent records in 20 Metro areas

and macro housing features in 2k+ zip codes

## **Featurewiz**

Automated feature selection by setting corr\_limit to 0.5





Experiments and Results Current studies highlight the promise of machine-learning tools combined with socio-economic and geospatial analysis. We built renowned machine learning models, including KNN, Random Forest (RF), Linear Regression (LR), BayesianRidge (BR), XGBoost (XGB), and ANN.

RMSE	MAE	MAD	<b>MAPE</b>	<b>RMSLE</b>	EVS
184,407	30,265	12,025	0.040	0.070	0.943
121,147	3061	609	0.222	0.112	0.972
181,329,225	731,546	55,607	0.013	0.017	0.832
107,404	35,073	20,775	0.060	0.089	0.981
9,766	5,910	3,903	1.127	0.017	0.998
96,195	28,905	21,391	0.039	0.066	0.985
	184,407 121,147 181,329,225 107,404 9,766	184,40730,265121,1473061181,329,225731,546107,40435,0739,7665,910	184,407 30,265 12,025   121,147 3061 609   181,329,225 731,546 55,607   107,404 35,073 20,775   9,766 5,910 3,903	184,407 30,265 12,025 0.040   121,147 3061 609 0.222   181,329,225 731,546 55,607 0.013   107,404 35,073 20,775 0.060   9,766 5,910 3,903 1.127	184,407 30,265 12,025 0.040 0.070   121,147 3061 609 0.222 0.112   181,329,225 731,546 55,607 0.013 0.017   107,404 35,073 20,775 0.060 0.089   9,766 5,910 3,903 1.127 0.017

RF, BR, XGB, and ANN models demonstrate an EVS > 0.95, indicating strong predictive capabilities. The Random Forest (RF) model was chosen for deployment as it outperforms others with superior MAE and MAD scores, indicating strong accuracy and consistency in model prediction.



Visualization Our dashboard delivers intuitive visualization with color-coding, navigational aids, filters, and a variety of charts including maps, plots, heatmaps, bar, and box actionable URL links. Users can swiftly find tailored investment or purchasing choices. At a high level, it spotlights ideal buying or renting locations and the effects of mortgage rate changes on the market. At a detailed level, it shows property counts, price trends based on user filters, and customized buy or rent guidance.

Conclusions and Future Enhancements Our model integrates analytics, a robust machine learning framework, intuitive visualization and an dashboard, offering critical insights into housing trends and opportunities and providing guidance on investment returns and buy-or-rent decisions. Our approach faces limitations like biases in public data and a lack of real-time updates. Future enhancements involve a broader dataset. real-time data integration for accuracy, iterative dashboard testing for improved user experience and tool utility.

