Climate-Driven Housing Price Prediction

NAVIGATING ENVIRONMENTAL DISPLACEMENT IN THE U.S.

Quintessa Guengerich, MSPPM-DA

Hannah Nguyen , MSPPM-DA

Jewel Yoko Kentilitisca, MSHCA

OUR HYPOTHESIS

As climate change reshapes U.S. cities, temperature anomalies will allow us to predict housing prices.

QUESTIONS

Regression

- Can we predict real estate purchase price based on climate indicators?

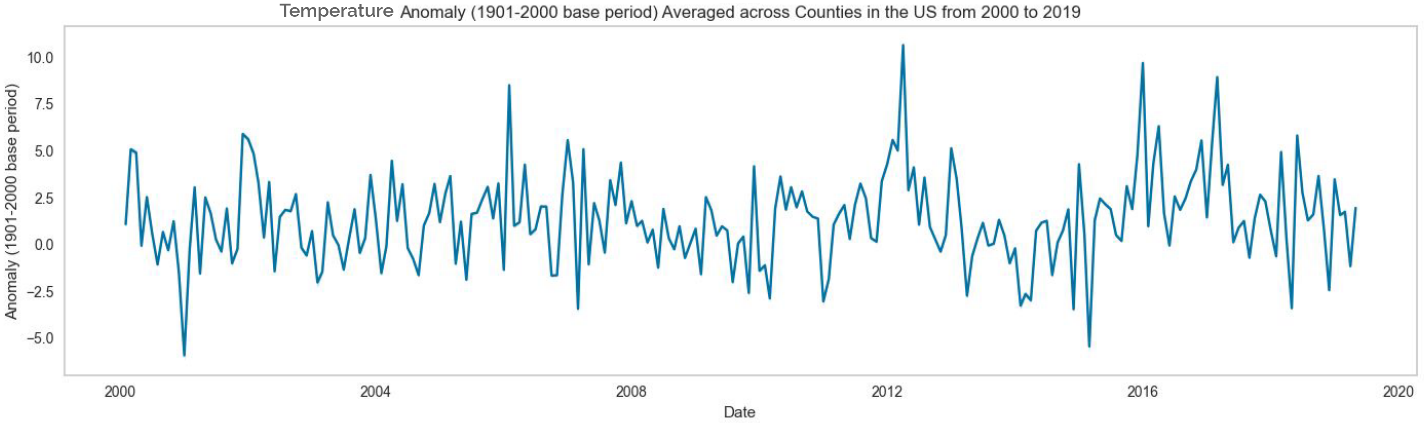
Classification

- Can counties in states with historically different temperature ranges be classified as outliers or not outliers based on real estate purchase price and temperature anomalies?

Unsupervised

- Are there any hidden structures or associations between housing prices and average temperature that is worth exploring?

EXPLORATORY DATA ANALYSIS



METHODOLOGY

Can we predict real estate purchase price based on climate indicators?

Linear Regression

Polynomial Regression

Lasso

Ridge

RANSAC Regression

Can counties in states with historically different temperature ranges be classified as high-risk?

Logistic Regression

Random Forest

Adaboost

XGBoost

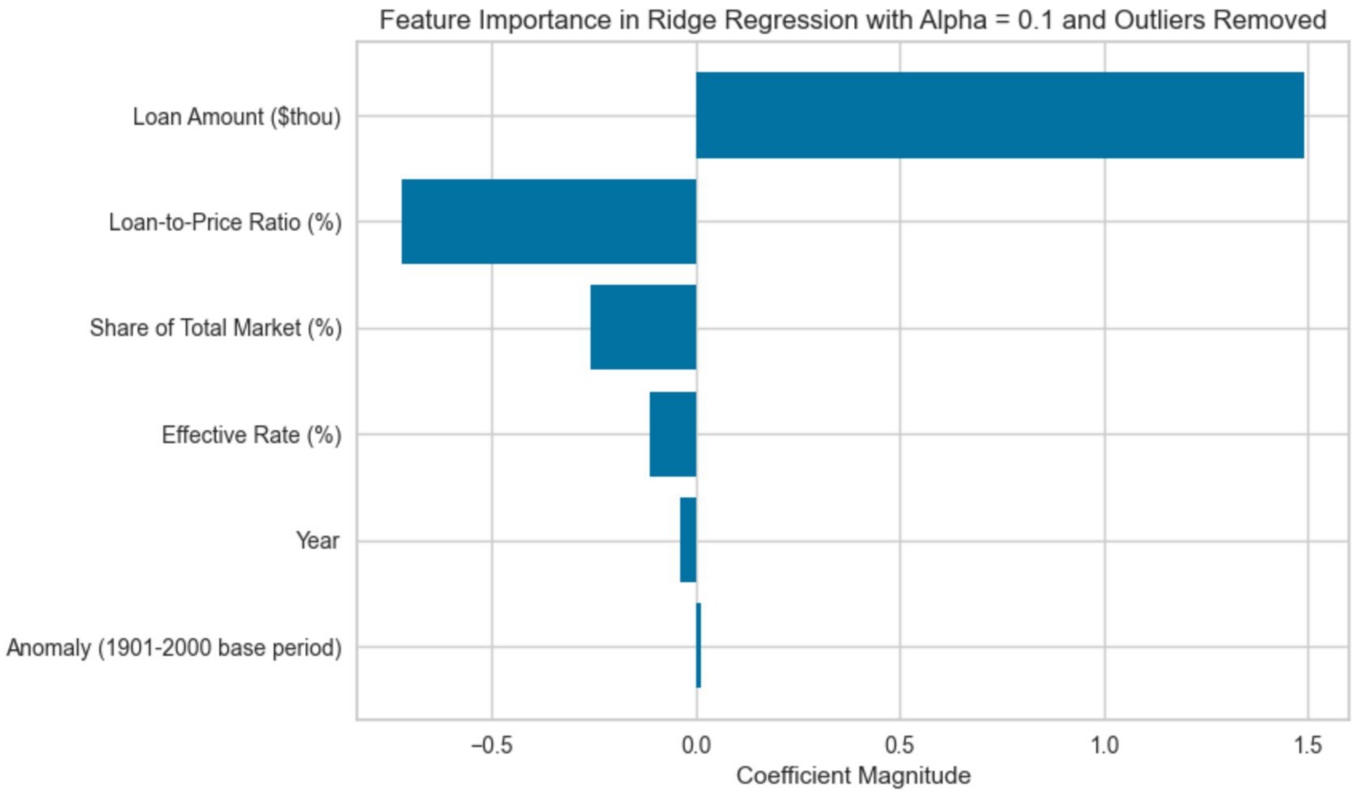
Are there any hidden structures or associations between housing prices and average temperature that is worth exploring?

K-Means Clustering

Hierarchical Clustering

DBSCAN

BEST REGRESSION: RIDGE WITH OUTLIERS REMOVED



R^2 Train:
0.97417

MSE Train:
0.02512

R^2 Test:
0.80338

MSE Test:
0.01435

RANDOM SAMPLING CONSENSUS: FINDING AN INLIER MASK

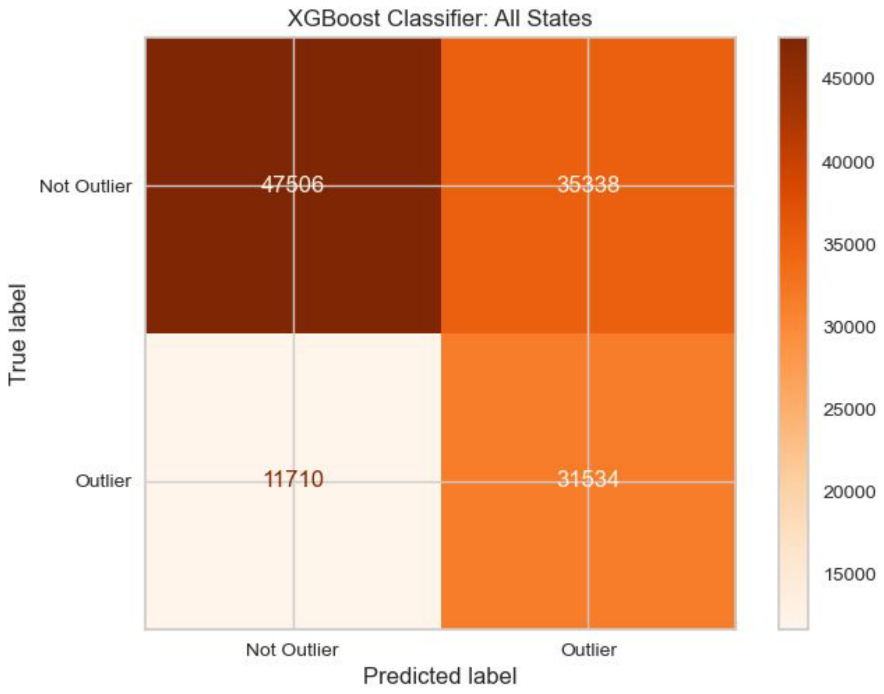
The first time running RANSAC revealed 2008 and 2012 to contain all outliers.

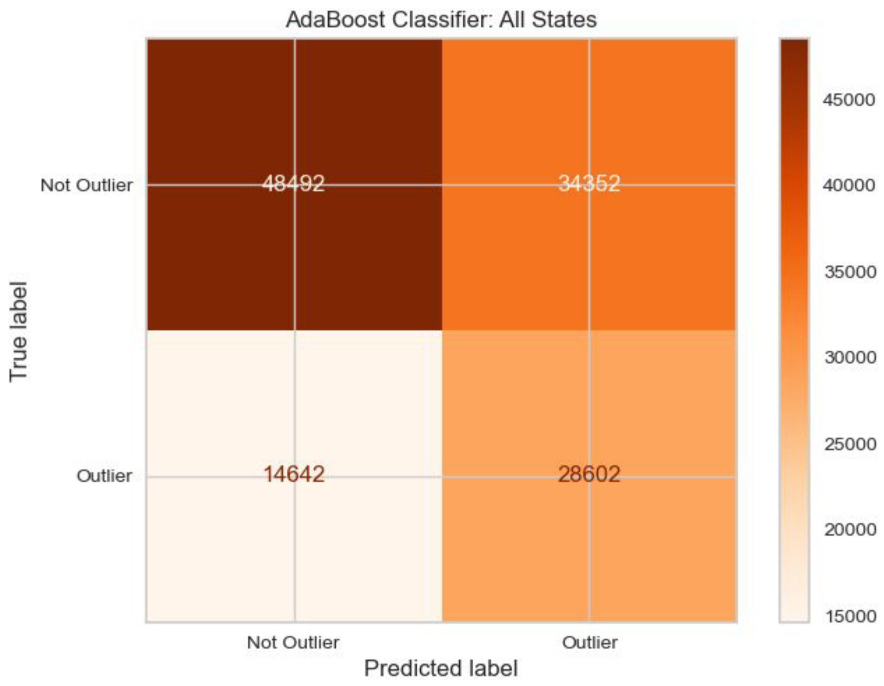
The second time running the algorithm generated a noisier inlier mask which was used for the **classification problems.**

CLASSIFICATION: CAN WE DETECT OUTLIER COUNTIES?

Based on our linear regression (using RANSAC) we determined an **inlier mask.**

Our goal: Classify counties as outliers/inliers based on all other data, with the outlier mask as our target.

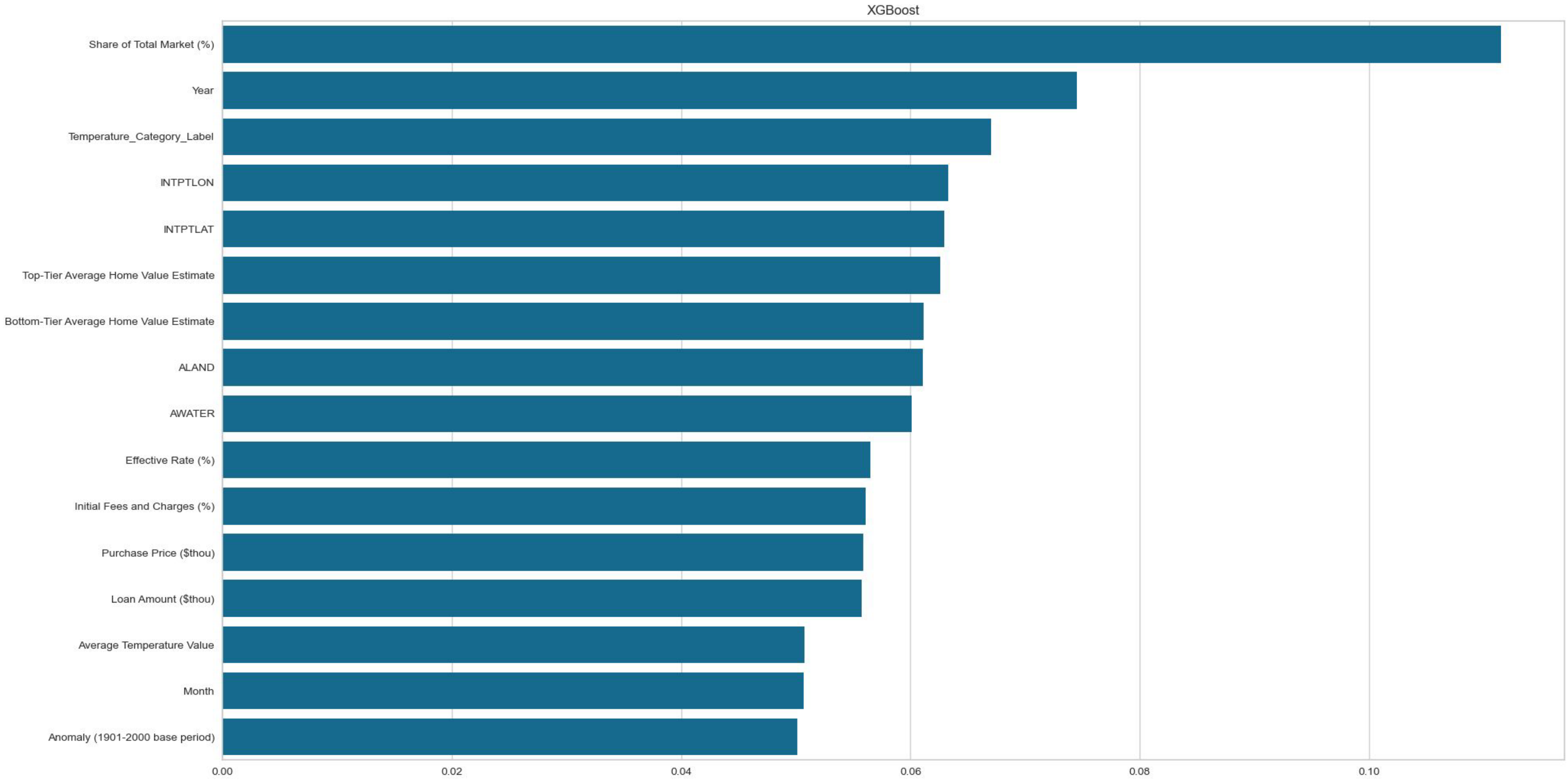




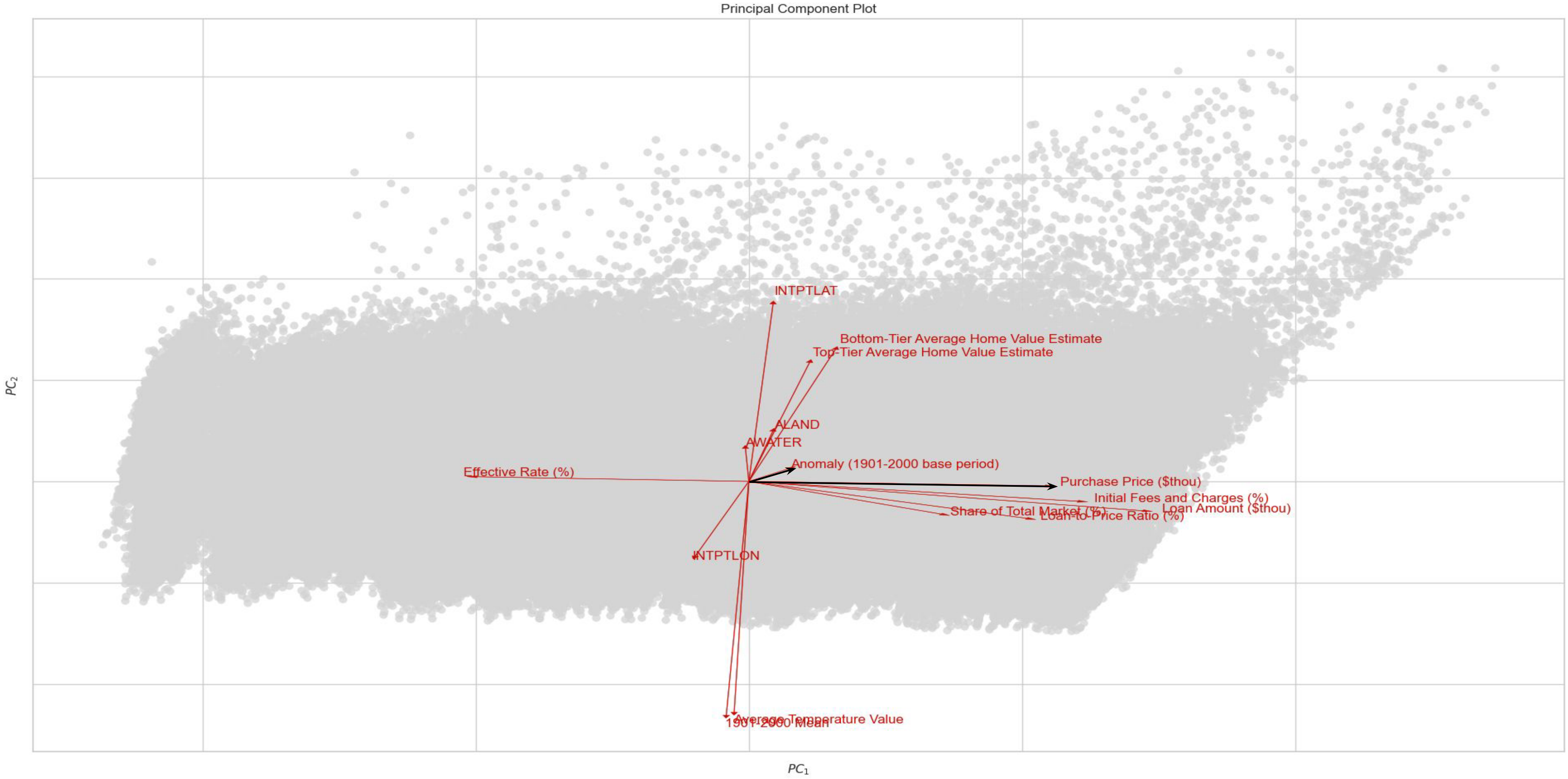
	precision	recall	f1-score	support
Not Outlier	0.80	0.57	0.67	82844
Outlier	0.47	0.73	0.57	43244
accuracy			0.63	126088
macro avg	0.64	0.65	0.62	126088
weighted avg	0.69	0.63	0.64	126088

	precision	recall	f1-score	support
Not Outlier	0.77	0.59	0.66	82844
Outlier	0.45	0.66	0.54	43244
accuracy			0.61	126088
macro avg	0.61	0.62	0.60	126088
weighted avg	0.66	0.61	0.62	126088

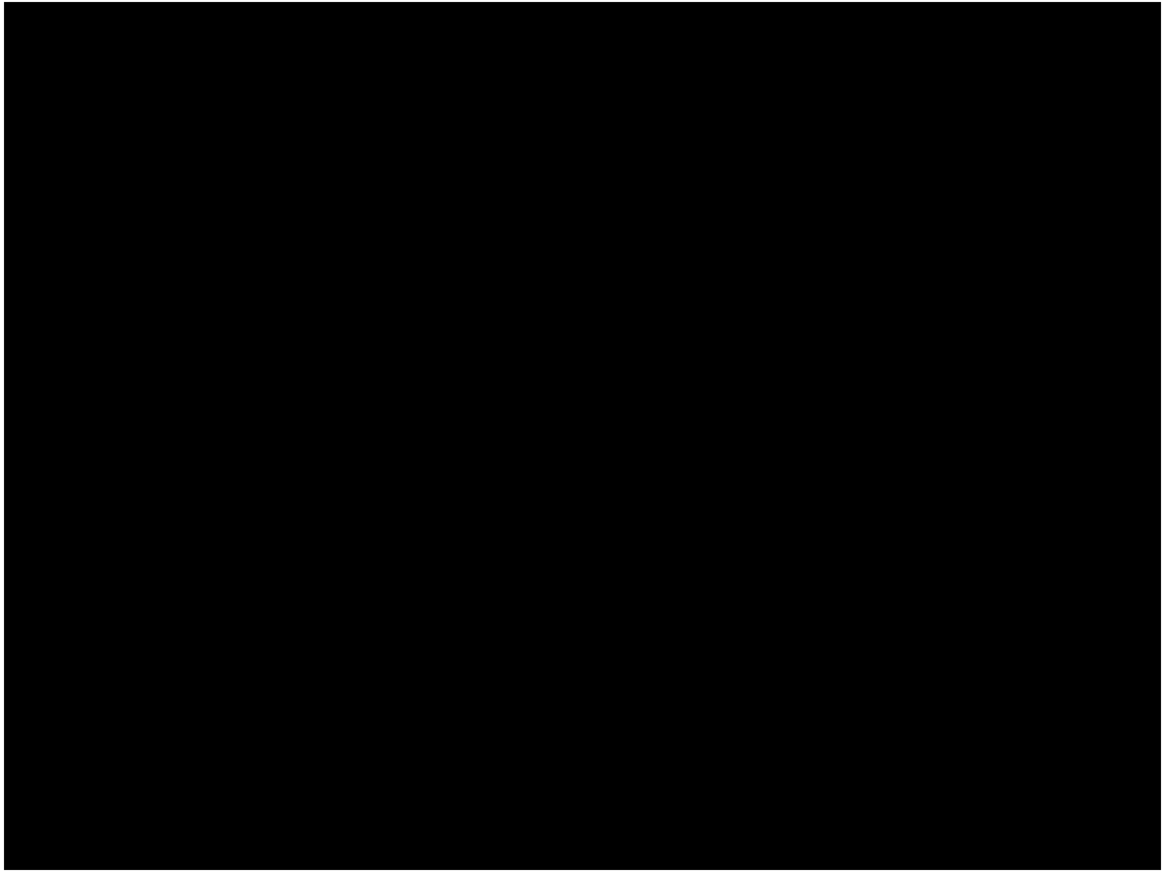
FEATURE IMPORTANCES OF CLASSIFYING OUTLIERS



UNSUPERVISED LEARNING: LATENT TRENDS IN THE DATA



DBSCAN CLUSTERS BY COUNTY OVER TIME



NEXT STEPS

- Rerun clustering algorithms with 2008 and 2012 removed from the dataset
- Gather more data:
  - More climate data: precipitation, air quality, cloud coverage, number of hot days etc.
  - More housing data: rent prices, apartment prices, etc
  - Population data – migration also likely affects housing prices and should be accounted for
- Time series analysis is necessary to more accurately forecast what these weak trends may mean for the future.

CONCLUSIONS

- **Existing trends are not strong enough** to claim true correlation or causality, but they are repeatable.
- **More data is necessary** to attempt to classify counties as high risk.

REFERENCES

Datasets:

1. https://files.zillowstatic.com/research/public_csvs/zhvi/County_zhvi_uc_sfrcondo_tier_0.67_1.0_sm_sa_month.csv?t=1709428647
2. https://files.zillowstatic.com/research/public_csvs/zhvi/County_zhvi_uc_sfrcondo_tier_0.0_0.33_sm_sa_month.csv?t=1709428647
3. Monthly average temperatures by counties in the U.S. from 2000 to 2023 obtained from the National Oceanic and Atmospheric Administration.

Models:

1. https://xgboost.readthedocs.io/en/stable/
2. https://www.mathworks.com/discovery/ransac.html#:~:text=Random%20sample%20consensus%2C%20or%20RANSAC,data%20set%20that%20contains%20outliers.
3. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html

Thank you!

ANY QUESTIONS?