Zillow kaggle codeup

REGRESSION PROJECT: ZILLOW 2017 KAGGLE PROPERTY TAX VALUE CHALLENGE

BY: ANNIE CARTER

SOURCED FROM: CODEUP, INC AND KAGGLE'S 2017 ZILLOW DATASOURCES PROJECT LOCATED ON GITHUB

AGENDA

Executive Summary

Goals, Take Aways

Findings

Recommendations, Takeaways

Recommendations

Takeaways, Next Steps

Conclusion

EXECUTIVE SUMMARY

1

GOALS

Develop ML Regression model to forecast 2017 assessed values of single-family properties, identify crucial factors, present findings

2

FINDINGS

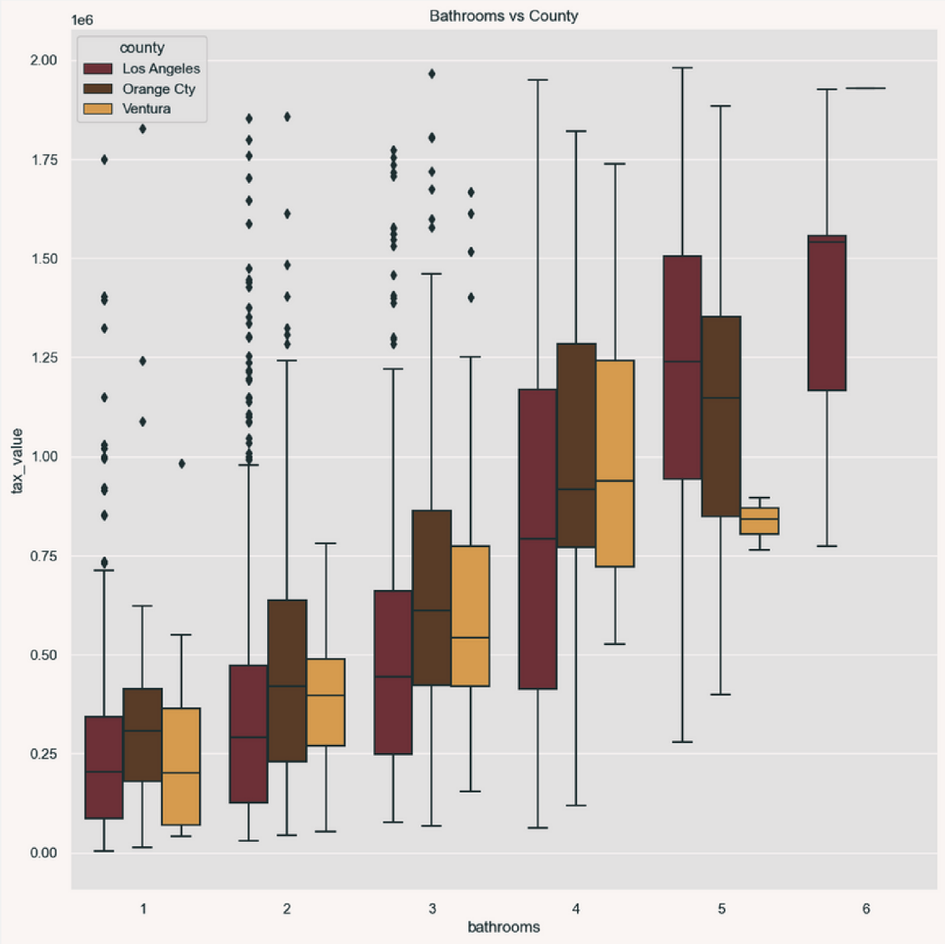
OLS and Lasso+Lars models showed lowest RMS surpassing baseline by 22%; square footage had strongest correlation with tax values, while bathrooms had more impact than bedrooms

3

RECOMMENDATIONS

To enhance prediction accuracy, aggregate at least three years of past data, use the IQR method for data cleaning, incorporate different features, and exclude high null value features

Our Data



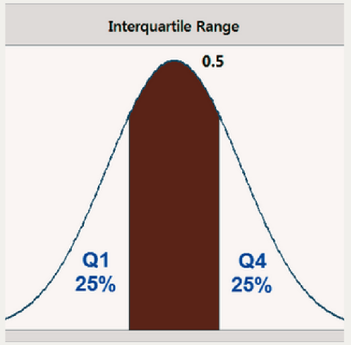
Ex. Bathrooms show visual correlation to tax value in all counties. Statistical testing confirmed significant relationship

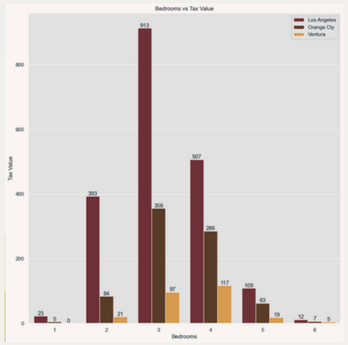
Findings

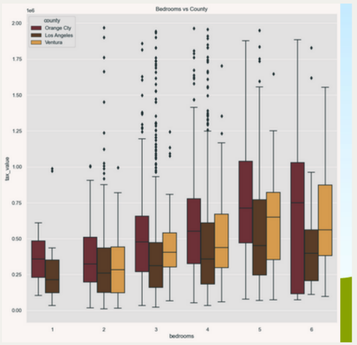
- Statistical testing found Correlation/relationship between all features and property tax value across three counties.
- Square footage and bathrooms demonstrates the strongest regression line.
-
- Ordinary Least Squares regression (OLS) and Lasso+Lars models achieved lowest RMSE values of 275,079 and 278,281, outperforming the baseline by 22%.

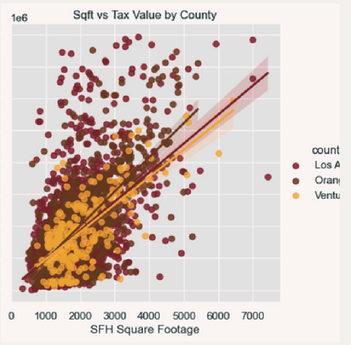
Recommendations

- Aggregate at least three years of past data to enhance prediction accuracy.
- Utilize the Interquartile Range (IQR) method for data cleaning to prevent overfitting and handle data skewness.
- Enhance the prediction model by incorporating the "bathbdcnt" column from the original zillow.csv dataset, eliminating redundancy and improving accuracy.
- Due to a high number of null values, variables such as "fireplace" and "basement" should be excluded from the analysis, despite their significant correlations with tax value prediction.



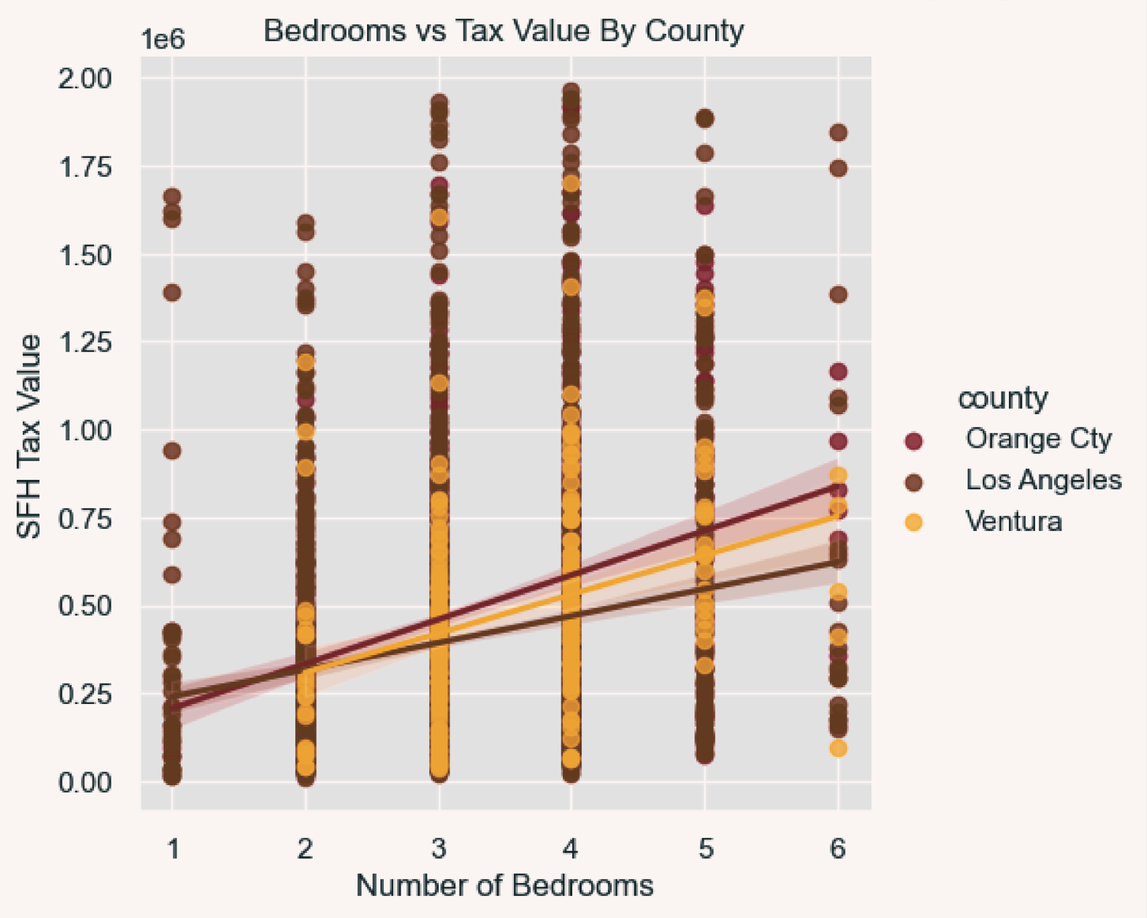






Conclusion

This project developed a machine learning regression model to forecast property values in Los Angeles, Ventura, and Orange counties based on property attributes. Square footage exhibited the strongest correlation with property tax value, and the chosen model, Lasso+Lars (alpha = 0.03), outperformed the baseline with an RMSE of 278,281. Significant relationships were found for bedrooms, square footage, bathrooms, and lot size with property tax values.



Thank You

GITHUB.COM/ANNIE-CARTER

ANNIE.CARTER831@GMAIL.COM

REGRESSION PROJECT