

Results of SnapHomz Trials

1. Methodology Summary

The analysis focused on real estate datasets, with steps including data cleaning, exploratory data analysis, plotting data distributions, and predictive modeling using Random Forest and Linear Regression.

2. Key Metrics/Statistics

- Highest number of listings in the city of Frisco; lowest in Sunnyvale.
- CA has more listings compared to TX overall.
- Mean prices are higher in TX while median prices are lower, indicating potential outliers.
- RandomForest outperformed Linear Regression (R^2 : 0.502 vs 0.445).

3. Graphs/Plots

Data distribution plots, price vs. square footage visualizations, and listing count charts were generated in the analysis notebook. These visualizations illustrated trends such as higher prices with increased square footage and city-wise price variations.

4. Recommendations/Interpretation

- Focus on TX properties for higher mean returns but account for variability.
- Prioritize Frisco for listing volume and market activity.
- Use RandomForest for more accurate price predictions.
- Further exploration of features like agent performance and listing dates may enhance predictions.

