Broker? Or Broker Than Before? How Much A Broker Really Matters

Carey Chang and Niko Ustin

Datasci 112 Winter 2025

Abstract

There is often uncertainty in the process of buying or selling a home. Our research seeks to examine whether the broker affects home sale price through an analysis of the spread of average sale prices of comparable homes by broker. We found that while the broker can have a significant effect on sale price, the level of impact is only semi-predictable based on the features of an individual home.

Research Questions

Does the broker that one chooses for buying or selling a home have a significant impact on the expected sale price?

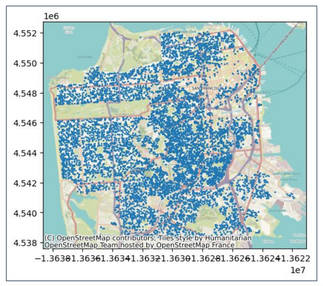
If there really is broker to broker variance, does the variance change depending on a home's characteristics?

Data Collection

We used APIs and web scraping to collect data. The Redfin.com Data API on Rapid API was used to find homes sold in the past five years in San Francisco. It was then used to pull key information on the homes (beds, baths, sale price, etc.). We also web scraped Redfin for home description and broker name for each home in our dataframe using the home's url (gathered from the API). We also downloaded a data set of population and population density across San Francisco from SimpleMaps, joining on the data using zip codes.

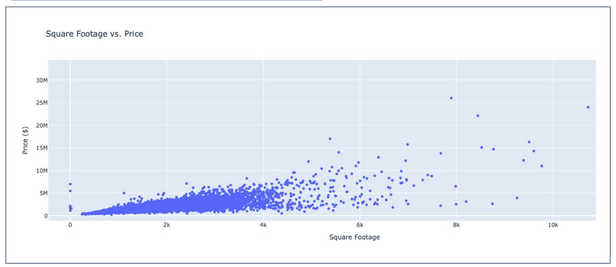
Data Exploration

We created and plotted a GeoDataFrame of houses previously collected to confirm comprehensive coverage within San Francisco. Each home characteristic was also investigated and pruned for irregularities from incorrect listing data.



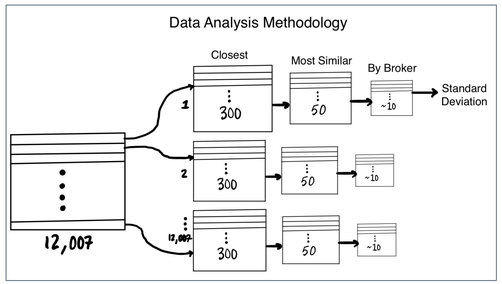
Features:

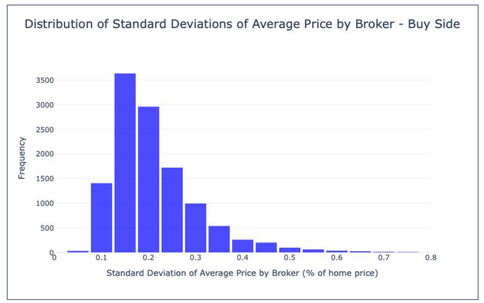
url, lastSoldDate, listingAddedDate, beds, price, lotSize, baths, yearBuilt, sqft, days_on_market, population, density, description, longitude, latitude, seller_agent, seller_broker, buyer_agent, buyer_broker, yearListed, monthListed



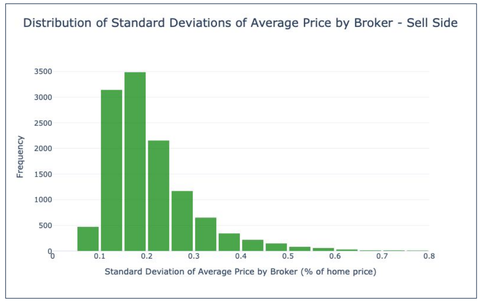
Data Analysis Methodology

In order to assess variance from broker to broker, we first found comparable properties for each home in the DataFrame. For each home, we found the 300 closest homes in the DataFrame using longitude and latitude. Within those 300, we found the 50 most similar homes using pairwise distances, weighting the square footage and description features more heavily. Those 50 homes were grouped by broker, providing a mean sale price by broker for similar, nearby homes. We then took the standard deviation of these average sale prices by broker and stored that as the sell_broker_var or buy_broker_var for each home.





Median = **0.1894**



Median = **0.1821**

Machine Learning

Using data processing, feature engineering, and hyperparameter tuning, we attempted to predict the variance for buyer's broker (buy_broker_var) and seller's broker (sell_broker_var) for given home based on property features. We then evaluated our model through cross validation. Our results for single/stacked regressors is outlined below:

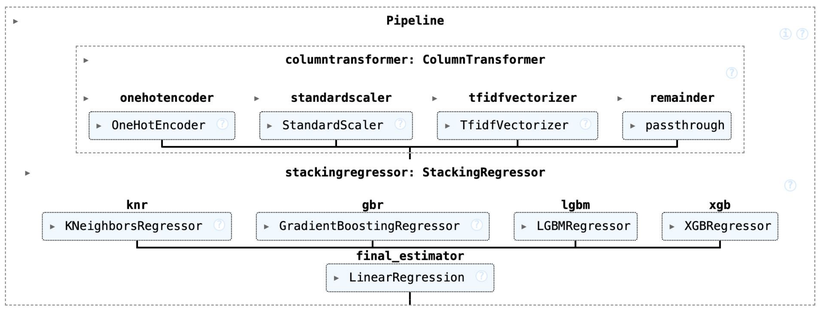
Cross-Validation RMSE

Buyer Broker Variation:

- LightGBM Regressor: 0.0711
- Stacked Regression: 0.0666

Seller Broker Variation:

- LightGBM Regressor: 0.0718
- Stacked Regression: 0.0669



Analysis & Conclusion

Given that the median standard deviation of average sale price from broker to broker across all homes in our DataFrame is 18.94% of home price for the buyer's broker and 18.21% for the seller's broker, broker seems to have significant impact on sale price. If someone randomly selects a broker, they can expect to attain, on average, a sale price around 18% deviated from the median of comparable homes, so choosing a good broker is important. With the best machine learning model achieving a RMSE of 6.7% of home price, the importance of broker from home to home seems semi-predictable.