Predicting Home Values in Los Angeles' South Bay

Springboard | Capstone #1 Proposal

By: Lauren Broussard

Business Problem:

Home prices in Los Angeles County remain consistently high, which can make purchasing a home difficult for those unfamiliar with the market, the neighborhoods, or the popular home features in those neighborhoods (i.e. number of bedrooms, or proximity to a certain school district). Being able to predict home prices within the neighborhoods in the county would be a benefit to many. For example, for a first-time home buyer, the decision to put a bid on a home is a big one, and knowing early on whether or not a particular home is over-(or under-)valued would save them time, money, and stress. Additionally, this kind of information may be useful for a new home builder, as they assess the features that would be important to get the most value for their new build. In this project, I will look at a specific set of neighborhoods in the southwest corner of Los Angeles County, South Bay. I will look at home sales from the past two years in this area.

Potential Client(s):

- *New Home Buyer/Investor:* To know whether or not a home is worth bidding on - i.e. it is undervalued in the market; also for an individual selling their home, what would be a suitable asking price.
- *New Home Builders:* To determine what features may be important to add to a home to get the highest asking price for the neighborhood.
- *Real Estate Agents*: To determine what price they could list a home for a client.

Primary Dataset:

Redfin Home Prices: For this project, I will look primarily at Redfin home sales data for the South Bay neighborhoods. Redfin's website provides current home sale data as well as previous home sale data going back as far as three years. The site includes features like: sale price, sale date, lot size, # of bedrooms, # of bathrooms, zip code, etc. The neighborhoods comprising the South Bay will be determined by the neighborhood's included in LA Metro's South Bay service map.

Proposed Methodology:

Through the data, I will see which attributes in the data affect the price of a home (i.e. # of bedrooms, neighborhood, # of bathrooms, zip code, home type - townhome, single family, condo), and will see if I can train a model to take in a list of home features and predict the price of said home.

Deliverables:

All code for this project will be placed on Github, and deliverables will include code, a written paper, and a slide deck of results and analysis.

Additional Resources:
Other resources that may be helpful in the course of the project are below, although I will focus primarily on the data from Redfin.

Zillow/Kaggle Data: This dataset is provided from Zillow as part of a Kaggle competition, and includes multiple features such as number of bedrooms, bathrooms, pool, air conditioning, etc. from the years 2016 and 2017

LA County Assessor Parcel Data: I will also look at parcel data from the LA County Assessor to add additional features to the original Zillow/Kaggle dataset, where appropriate. This dataset includes 33.3M rows and 51 columns, with home value data from 2006-2019, and including features like number of bedrooms, number of bathrooms, neighborhood, home value, and year built.