Zillow Analysis

Vishal Aiely, Sehee Hwang, Matt Mohandiss, Marc Muszik, and Selena Xue

I. INTRODUCTION

Home ownership is a beacon of prosperity and a staple of the American Dream. Everyone takes part in the real estate market, whether buyer or seller, renter or landlord. It's an investment that is expected to appreciate with time, and in recent times, that rate is rising faster than ever. That leaves many of us to navigate the most expensive real estate market since before the Great Recession. While we cannot solve the affordable housing crisis on our own, we can mitigate its impact and navigate the market with the power of data.

The objective for this project is to use digital archaeological techniques to retrieve data from Zillow and analyze trends in the housing market. We will be looking to see if there is any correlation between a property's attributes and its price. For example, we will be considering price per square foot in relation to the number of bedrooms and bathrooms. We will also take a look at changes in neighborhood prices if historical data is available.

II. MOTIVATION

The real estate market can be difficult to follow as it is always in flux – housing prices and real estate trends evolve daily. Understanding the market is important as real estate is a large investment. By recognizing correlations in the real estate market, we believe people can more easily navigate the market and make more well-informed decisions. Because our homes are the most expensive item we'll ever own, it is crucial that us, the buyers, understand the true value of a property.

The data we will collect will help us answer important questions that pertain to the decision of whether or not to buy. Questions such as, which neighborhoods have the best value, and which are the most overpriced? What renovation will add the most value to my property? Is now a good time to buy? These are difficult questions to answer, and many attempt to do so without a quantifiable, data-driven approach. Our goal is to bring that data-driven approach to these questions and provide answers that give us real-world insight.

III. DISCUSSION OF DATA

The most detailed, up-to-date and readily available information on the housing market comes from Zillow. Zillow has comprehensive data about all the most important attributes that go into a property, and it makes that data available to us via the internet. We plan to collect any and all attributes that can effect a property's value, such as the location, number of bedrooms, bathrooms, square footage, lot size, and year remodeled, among others. We plan to also collect the Zestimate, or more interestingly, how the list price compares. If time allows, we will collect the same attributes for rental properties as well. The data will be stored in Google Cloud after being read and parsed from a web scraper. The hope is that this data will allow us to see correlations between certain attributes and a property's value, and find opportunity to get an edge in the market.

Another data type we will collect is the historical price change for a given city. This data is made available by Zillow via readily available CSV and TXT files. This data has aggregated home prices from 2017-2021 for many cities throughout the US. This data can be analyzed separately and in conjunction with individual home values.

IV. RESPONSIBILITIES OF MEMBERS

A. *Vishal Aiely*

- Help develop web scraping algorithm
- Integrate Google Cloud data storage
- Assist in visualization of results from data

B. *Sehee Hwang*

- Help develop web scraping algorithm
- Find different data sets to analyze historical trends
- Assist in visualization of results from data

C. *Matt Mohandiss*

- Work with API sources to gather data
- Integrate Google Cloud data storage
- Analyze results, visualize data

D. *Marc Muszik*

- Help develop web scraping algorithm
- Assist in creating dataframe to store data
- Analyze results, find correlations, create model

E. *Selena Xue*

- Help develop web scraping algorithm
- Find different data sets to analyze historical trends
- Analyze results, model visualization

Each team member will take a part in creating the final report and presentation.

V. TIMELINE OF MILESTONES

The milestones for this project are to complete the scrapper and successfully scrape the Zillow web page, to put the web scraper into a working function that accepts a city name or zip code, to collect the data, to integrate Google Cloud for Data, and to clean the data. There are also stretch goals for this project after the milestones are completed. These stretch goals include comparing data between for-sale properties and rental properties as well as adding other comparison attributes such as demographics, school districts, crime rates, and walk-ability.

There are approximately six weeks to work on this project.

Timeline:

Week	Milestones
Week 1-2	• Complete scrapper • Successfully scrape the Zillow web page
Week 3	• Put a web scrapper into a working function • Working function accepts a city name or zip code
Week 4	• Integrate Google Cloud for data storage • Stretch goal: comparing for-sale vs rental properties
Week 5	• Cleaning Data • Stretch goals: New attributes/features: enrich data with • population growth, school district, crime rates, walkability, etc.
Week 6	• Analyzing Data, Visualizing Data, Creating models, Finding Correlations • Writing the final project report

VI. EXPECTED OUTCOME

After gathering and analyzing the data, there are several trends that we hypothesize. The primary outcomes that we expect to discover from our data are two-fold: first, what makes certain properties worth different than others? Second, what causes the overall increase in housing prices across different regions? Beyond that, we hope to understand more regarding the cause of increased price and why some regions may see more of a rise than others. Some of the physical factors of a home that we think will play a role would be more bedrooms, more bathrooms, or whether the home has a pool. Other external factors we expect to see are the quality of nearby schools, crime rates of the area, and trends of houses in the same neighborhood. Some trends we would expect to see from a buyer's standpoint is an increase in non-primary home purchases, as well as an increase in the difference between the closing price and the estimated price prior to bidding. These outcomes may help build an understanding of the market with respect to current economic events such as COVID-19, low-supply in housing, the relocation of individuals from higher priced regions to lower priced regions. Our project hopes to provide people a better understanding of the housing market and how it evolves over time.