Cost of Buying or Renting Algorithm (COBRA) Final Report

Anne Benolkin abenolkin3@gatech.edu
Hien Le hvan6@gatech.edu
Matthew Molinare mmolinare@gatech.edu
Crystal Nguyen cnguyen312@gatech.edu
Stephen Wang stephen.wang@gatech.edu

1 Introduction

In the search for a home, potential renters and buyers can become inundated with endless information found online about homes and real estate opportunities. It is believed that the average human has the capacity to hold approximately seven data points in short-term memory [1]. While this exact number is disputed, it is eclipsed by the number of criteria that factor into identifying the optimal home and deciding whether to rent or to buy. To help people alleviate the stress of looking for the perfect home, our team has implemented the Cost of Buying/Renting Algorithm (COBRA) to help home-seekers process the overwhelming amount of data points involved in finding the perfect home by providing an automated rent or buy recommendation based on the home seeker's financial and personal home preferences.

2 Problem Definition

Housing decisions are complex. There are many variables involved in purchasing or renting a home that it can be difficult making the financially sound choice. Factors such as the economy, market value, change in income, and liquidity can be difficult to reconcile into a single buy or rent decision. In designing COBRA, our team focused on building a useful and accurate tool that is simple enough to appeal to a wide range of users, but complex enough to incorporate all pertinent data points and assumptions.

According to Levy and Lee, every family goes through 4 stages in the process of finding a home: problem recognition, search, evaluation of alternatives, and final choice [11]. Our primary goal is to decrease the difficulty associated with the last 3 stages of the home seeking process. Up until the modern age of technology, each household relied on a combination of financial or real estate advisors, online repositories of available homes, and word of mouth to make a decision on which home to buy or rent. Even with advanced technology, access to vast amounts of information have made finding the perfect home, at the right price, increasingly overwhelming. COBRA attempts to alleviate this problem by helping the user automate the decision making process.

3 Innovation and Differentiation

Companies such as Zillow, Money Under 30, and HomeSeeker have attempted to create tools that help home seekers make a decision on whether to rent or buy in a specific city or neighborhood [4]. However, these tools are typically one-dimensional (i.e., only show one neighborhood at a time) and lack visualizations that enable the user to better digest the information.

Our team's implementation of COBRA, exhibits the following innovations, which have yet to be provided on existing platforms:

- A graphical user interface that provides *geographic visualizations* of rent-or-buy decisions across all zip codes
- A definitive rent-or-buy recommendation for *all zip codes* in the user's selected county
- An *estimated net value* for renting or buying over the desired time period for all zip codes
- An *estimated monthly* cost for renting or buying over the desired time period for all zip codes
- A *recommendation against either renting or buying* in a specific zip code if the user's budget does not meet the estimated monthly cost

The net value of *buying* is defined as the market value of a home over time if the cash assets are used to buy the home. The net value of *renting* is defined as the investment return on cash assets over time if that amount is invested rather than used for a down payment. Other platforms do not use net value as a parameter, which is an additional differentiating factor for COBRA.

4 Our Methodology

COBRA is designed to address the shortcomings of current tools by going beyond providing a simple buy or rent recommendation. Using the following inputs from our users, COBRA recommends whether to rent or buy a home in a specific zip code, estimates the total net value and monthly cost of either renting or buying over various periods of time, and recommends zip codes that the user may be able to afford:

- Maximum monthly budget for housing
- Expected yearly raise in salary
- Years expected to live in home
- Current available cash assets for down payment or investment
- Desired city or zip code of home
- Number of bedrooms and bathrooms
- Desired square footage of home
- Desired range of years that home was built in

On top of user data provided through COBRA's graphical user interface, COBRA is also driven by three of Zillow's housing datasets. The primary dataset used in our algorithm is a record of approximately 2.8 million individual homes for sale or rent in the Los Angeles, California area, which includes information on home features (e.g. number of bedrooms, square footage, etc.), geo-location, and pricing. Figure 1 below illustrates initial exploratory analysis conducted by our group to investigate potential trends within the data, segmented by geography and number of bedrooms.

Price per Square Feet by County

County/Bedrooms	1	2	3	4	5	6
Los Angeles County	$539	$442	$377	$353	$388	$348
Orange County	$504	$428	$385	$351	$375	$374
Other Counties		$869	$612	$436	$362	$271
Ventura County	$427	$348	$330	$305	$305	$294

Figure 1: Exploratory Analysis of Zillow's Average Price per Square Feet

The second dataset is a list of all zip codes reported by Zillow and the corresponding city, neighborhood, and county. The third Zillow dataset includes median rental price by square footage by zip code and number of bedrooms (used to predict rental prices). COBRA combines the user's input data with the 3 Zillow datasets to filter which homes the user may be able to rent or buy and provide a prediction on how much the home will cost within the desired timeframe of occupation. Figure 2 provides a snapshot of various neighborhoods Zillow provides housing data on.

Our team has transformed the first dataset for the purpose of this project by removing irrelevant columns, removing rows with null values for pricing, removing outliers, defining home values based on Zillow's provided tax value estimate and California's average property tax of 0.77%, and integrating estimated home rental prices into the overall dataset. The other two datasets were exported from Zillow's website and filtered for records tagged as "California" in the "State" column. We chose Zillow's database as the source for COBRA as Zillow is generally a trusted source for estimating home values among both users and academics [3].

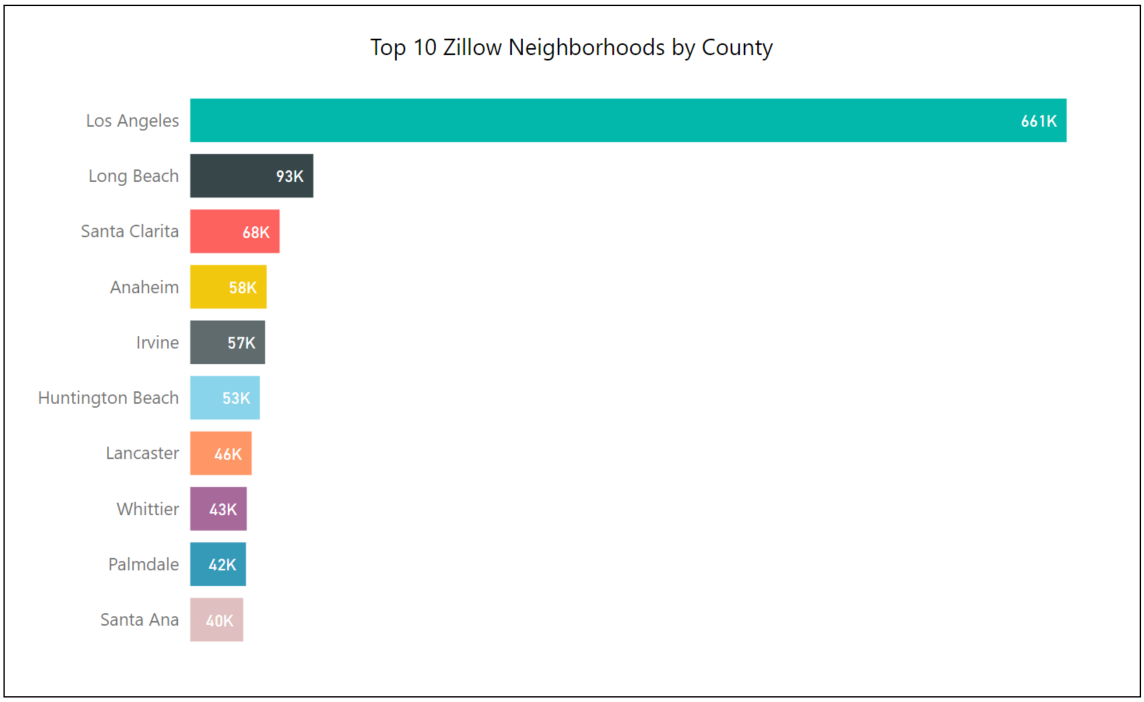

Figure 2: Exploratory Analysis of Zillow's Top 10 Neighborhoods by Count of Homes

The user first interacts with COBRA through a graphical user interface (powered by JavaScript and HTML/CSS) hosted on our website where he or she is required to fill out a form with the aforementioned information (as seen in Figure 3 on the next page). Once the user submits the form, the data provided in the form is converted into input for a Python-based algorithm, which will process and visualize the data in the following steps:

1. Filters Zillow's database of available home to match the user's home preferences (e.g. zip code, number of bed/bathrooms, year built, and desired square footage).
2. Initiates Python algorithm - defines the user's financial inputs as variables: numbers of years to occupy the home (num_years), available cash assets for down payment (down_payment), maximum monthly budget (initial_monthly_budget), and expected annual raise (annual_raise).
3. Computes the user's expected monthly budget over time based on the initial monthly budget, number of years to occupy the home, and expected annual raise.
4. Retrieves the sale and rental value of each home and creates an instance of a Renter object and Buyer object.
5. Calculates the average monthly cost of buying and renting each home based on the user's available cash assets and default values for rental appreciation rate, home value appreciation rate, property tax rate, mortgage rate, and mortgage term.
6. Calculates the renter's return on investment (ROI) at year t as

$$\mathrm{ROI}_t = \Big(1 + r_{\mathrm{return}}\Big)\mathrm{ROI}_{t-1} + \Big(1 + \frac{r_{\mathrm{return}}}{2}\Big)S_{t-1} \tag{1}$$

   where r_{return} is the investment return rate and S is the surplus, or difference in annual cost between renting and buying. The surplus is compounded at half the rate to reflect the fact that it's not available at the beginning of the year.
7. Calculates the net value of buying as the current home value less closing costs and outstanding mortgage balance.
8. Calculates the net value of renting ($\mathrm{NV}_{\mathrm{rent}}$) as,

$$\mathrm{NV}_{\mathrm{rent},\,t} = \mathrm{ROI}_t - r_{\mathrm{tax}} \left(\mathrm{ROI}_t - \mathrm{ROI}_1 - \sum_{i=1}^{t-1} S_i^{+} \right) + \sum_{i=1}^{t} S_i^{-} \tag{2}$$

where r_{tax} is the investment tax rate.

9. Creates a line plot of the average net value for buying and renting a home over the user's desired time range (in years) for each zip code.

10. Creates a line plot of the average monthly cost for buying and renting a home over the user's desired time range (in years) for each zip code.

11. Classifies each zip code as "buy", "rent", or "exclude" (not affordable) based on average net value and average monthly costs. If the net value of buying is greater or equal to that of renting, the recommendation is for the user to "buy". Otherwise, the user should "rent". If the monthly costs of renting and buying both exceed the monthly budget for any year, than we recommend the user neither rent nor buy ("exclude"). This concludes the Python algorithm.

12. Returns a visualization to the GUI (using Google Maps API, JQuery, and D3.JS) with the "rent or buy" classification of all zip codes in the user's selected county, using blue for "buy", green for "rent", and gray for "exclude". When the user hovers over a zip code, a tooltip visualizing the average net value and monthly cost line plots appears. See Figure 4 below.

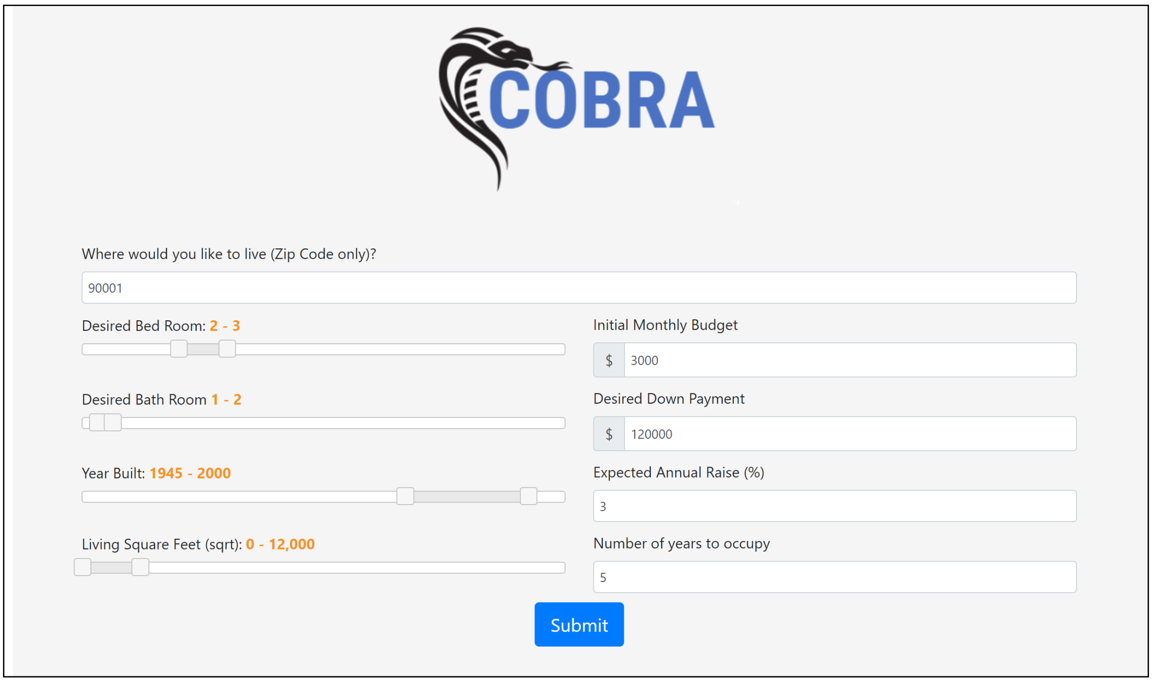

Figure 3: COBRA Graphical User Interface: Questionnaire

The algorithm is influenced by Clark and Lomax's rent-price ratio, which helps us determine the rent multiplier based on home value [15], Gindelsky, Moulton, and Wentland's user cost approach for predicting home prices, which uses Zillow's database of housing prices [14], Liang, Phillips, and Yu's regression model for determining real estate prices based on home attributes [8], and the principal accumulation value of compound interest, which helps us determine change of monthly budget over time based on expected annual raise [16].

COBRA is a front-end web client application framework and is built on the Pyramid framework. On the front-end, COBRA is designed using the Google Map API, JavaScript libraries such as JQuery and D3.JS, and HTML/CSS. On the back-end, COBRA is supported by a Python algorithm which employ various libraries such as Pandas, NumPy, etc. COBRA can be deployed using AWS on the Ubuntu Server 18.04 LTS using web server uWSGI+ nginx.

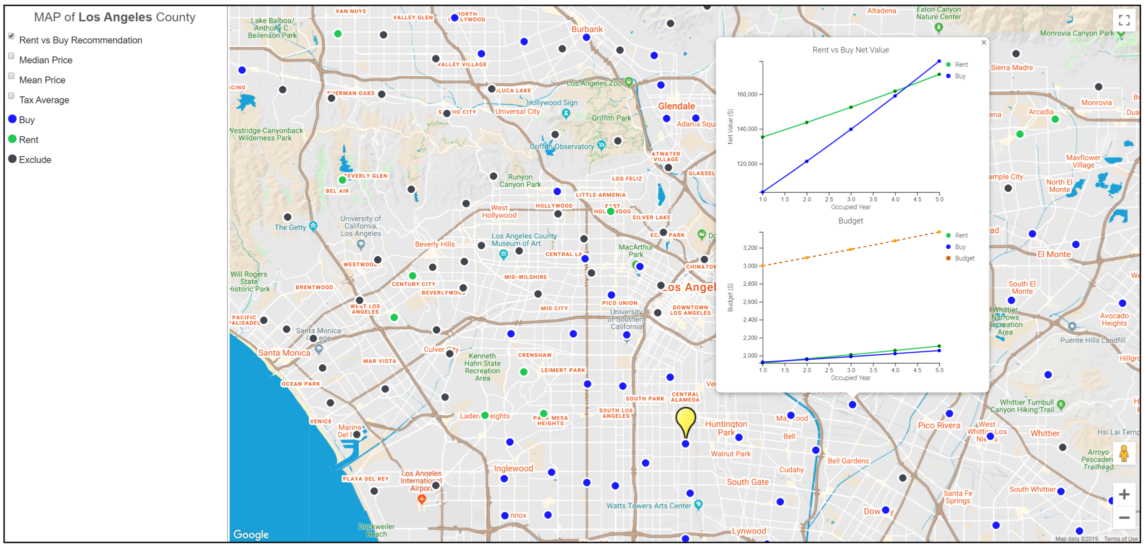

Figure 4: COBRA Results: Visualization of Rent of Buy Recommendations by Zip Code.

5 Experiments and Evaluation

To evaluate the effectiveness of COBRA's algorithm, our team sampled 12 random homes from the dataset and asked 5 participants to make a traditional assessment on whether to rent or buy (considering Fehérová's key factors affecting buy or rent decisions) using provided assumptions on monthly budget, available cash assets, expected annual raise, and years to occupy the home [10]. Our team then compared the results of the traditional, human-based assessment to COBRA's, Zillow's, and Money Under 30's rent or buy assessment to answer the following questions:

- How does COBRA's assessment results compare to traditional, human-based assessments?
- How does COBRA's assessment results compare to other rent or buy tools?
- How does COBRA's alogorithm differ from other rent or buy tools?

Note that COBRA's accuracy cannot be definitively measured as there is no data available on whether an individual decided to rent or buy a specific home nor is there data available on if the rent or buy decision was the optimal decision for the user. Therefore, our evaluation of COBRA's algorithm is meant to compare differences between various assessment methods rather than to measure true accuracy or precision.

Scenario	Neighborhood	Bedrooms	Sq Ft	Est. Cost	Participant Results	COBRA Results	Zillow Results	Money Under 30 Results
A	Orange	2	949	$409,875	Buy	Buy	Buy	Buy
B	Lancaster	3	2395	$440,784	Buy	Buy	Buy	Buy
C	Buena Park	4	2307	$633,257	Buy	Exclude	Buy	Buy
D	Palos Verdes Estates	3	2308	$2,777,070	Exclude	Exclude	Rent	Rent
E	Santa Ana	4	1235	$344,592	Buy	Buy	Buy	Buy
F	Los Angeles	1	1019	$391,683	Buy	Buy	Buy	Buy
G	Whittier	3	1694	$759,723	Rent	Exclude	Buy	Buy
H	Los Angeles	3	1304	$655,796	Rent	Exclude	Buy	Buy
I	Los Angeles	2	874	$628,045	Rent	Rent	Buy	Buy
J	Long Beach	1	689	$427,125	Buy	Buy	Buy	Buy
K	Orange	3	2620	$1,098,681	Exclude	Exclude	Rent	Rent
L	Los Angeles	1	1084	$488,334	Buy	Buy	Buy	Buy

Figure 5: Results from COBRA Evaluation

For each scenario, we assumed the user had $200,000 in cash assets available for a down payment or for investing, a $3,000 monthly budget for rental or mortgage payment, an expected annual raise of 3%, and a desire to buy or rent for 5 years. For many cases, COBRA provides the recommendation to "exclude" the home, which means that the home is not affordable for the user in the time frame specified. As seen in Zillow's and Money Under 30's results, this option is not available, which can be viewed as a shortcoming. Often times, the monthly cost of renting exceeds the user's ability to pay that amount; COBRA considers this factor in its recommendation whereas Zillow and Money Under 30 do not.

In Scenario I, COBRA recommends "rent", which is aligned with the user, but misaligned with Zillow and Money Under 30. In this case, the difference may be due to COBRA's consideration of net value, which suggests that investing the cash assets and paying the expected monthly rent would result in higher returns for the user; Zillow and Money Under 30 only take total costs into account.

Lastly, in Scenario C, G, and H, COBRA recommends "exclude", whereas participants chose either "buy" or "rent" - this is likely due to COBRA's consideration of approximately 2.8 million data points and default assumptions on home appreciation and investment returns, which the user likely has not accounted for. Overall, COBRA's algorithm appears to add additional depth to existing tools, by providing an option to "exclude", incorporating net value, and considering additional assumptions such as home appreciation and expected annual raise.

6 Conclusions

Our team is confident in COBRA's ability to support home seekers in making better financial decisions on whether or buy or rent a home based on personal home preferences and Zillow's extensive database of home value information. COBRA enhances the user experience by employing various data visualizations, such as interactive maps, line plots, and color-coded rent or buy recommendations across multiple zip code to add depth and robustness to its assessment. Additionally, COBRA's algorithm differentiates itself from current tools by comparing investment returns to home value appreciation to predict net value. In its essence, COBRA is a tool that can successfully help many people optimize and alleviate one of the most stressful decisions in life - finding the perfect home. All team members have contributed similar amounts of effort.

References

[1] Miller, George A. *The magical number seven, plus or minus two: Some limits on our capacity for processing information.* Psychological review 63.2 (1956): 81.

[2] DeSimone, Brendon. *Next Generation Real Estate: New Rules for Smarter Home Buying & Faster Selling.* Changing Lives Press (2014).

[3] Corcoran, Charles and Fei, Liu *Accuracy of Zillow's Home Value Estimates.* Real Estate Issues 39.1 (2014): 45-49.

[4] Li, Mingzhao, et al. *HomeSeeker: A visual analytics system of real estate data.* Journal of Visual Languages & Computing 45 (2018): 1-16.

[5] Case, K. E., and R. Shiller. *Prices of single-family homes since 1970: New indexes for four cities.* New England Economic Review (1987): 45-56

[6] Vineeth, N., Ayyappa, M., and Bharathi, B. *House Price Prediction Using Machine Learning Algorithms.* International Conference on Soft Computing Systems (2018): 425-433.

[7] Limsombunchai, Visit. *House price prediction: hedonic price model vs. artificial neural network.* New Zealand Agricultural and Resource Economics Society Conference (2004).

[8] Liang, Jiang, Phillips, Peter C.B., and Yu, Jun. *A New Hedonic Regression for Real Estate Prices Applied to the Singapore Residential Market.* Research Collection School of Economics (2014): 1-22.

[9] Bourassa, Steven, Eva, Cantoni, and Martin, Hoesli. *Predicting house prices with spatial dependence: a comparison of alternative methods.* Journal of Real Estate Research 32.2 (2010): 139-159.

[10] Fehérová, Martina *The Factors Affecting Rent versus Buy Decisions and Implication for the Housing Market in Slovakia.* Almanach: Aktuálne Otázky Svetovej Ekonomiky a Politiky (2018): 19-29.

[11] Levy, Deborah S., and Christina Kwai-Choi Lee. *The influence of family members on housing purchase decisions.* Journal of Property Investment & Finance 22.4 (2004): 320-338.

[12] Moeller, Sabine, and Kristina Wittkowski. *The burdens of ownership: reasons for preferring renting.* Managing Service Quality: An International Journal 20.2 (2010): 176-191.

[13] Baldominos, Alejandro, et al. *Identifying Real Estate Opportunities Using Machine Learning.* Applied Sciences 8.11 (2018): 2321.

[14] Gindelsky, Marina, Jeremy Moulton, and Scott A. Wentland. *Valuing Housing Services in the Era of Big Data: A User Cost Approach Leveraging Zillow Microdata.* Big Data for 21st Century Economic Statistics. University of Chicago Press (2019).

[15] Clark, Stephen, and Nik Lomax. *Rent/price ratio for English housing sub-markets using matched sales and rental data.* Area (2019).

[16] Li, Xudong, and He, Mingxing. *The principal accumulation value of simple and compound interest .* BioTechnology: An Indian Journal (2014).