Cost of Buying or Renting Algorithm (COBRA) Implementation Progress 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 information found online about a multitude of 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.

Our team plans to implement the Cost of Buying/Renting Algorithm (COBRA) to help home-seekers process the overwhelming amount of data points involved in buying or renting a home to make an optimal decision.

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, gentrification, and liquidity can be difficult to reconcile into a single buy or rent decision. In designing COBRA, we must consider how we can create an accurate and useful tool that is user friendly enough to be used by anyone but complex enough to incorporate all of the pertinent data.

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.

Companies such as Zillow or projects such as 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 do not estimate the total cost of buying or renting a specific home over a specified time period nor do they recommend which home to buy in a chosen city or neighborhood.

3 Proposed Method

COBRA is designed to address the shortcomings of current tools by going beyond providing a simple buy or rent recommendations. Using the following inputs from our users, COBRA recommends whether to rent or buy a home, estimates the total cost of either renting or buying over multiple periods of time, and recommends homes that the user may be able to afford:

- Yearly salary
- Expected yearly raise
- Years expected to live in home
- Current available assets (e.g. cash)
- Desired city or zip code of home

- Number of bedrooms and bathrooms
- Desired square footage of home

COBRA tool is driven by Zillow's extensive database of available homes in Los Angeles, California and surrounding area. The dataset consists of almost 3 million records of available homes, with information on home features (e.g. number of bedrooms, square footage, etc.), geolocation, and pricing.

Our team has transformed the dataset for the purpose of this project by removing irrelevant columns, removing rows with null values for pricing, removing outliers (e.g. homes with more than 10 bedrooms, non-residential units, etc.), 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. 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]. All data and corresponding reference tables have been transferred to csv files for use in the COBRA interface and algorithm.

The user first interacts with COBRA through a graphical user interface hosted on our website where he or she is required to fill out a form with the aforementioned information.

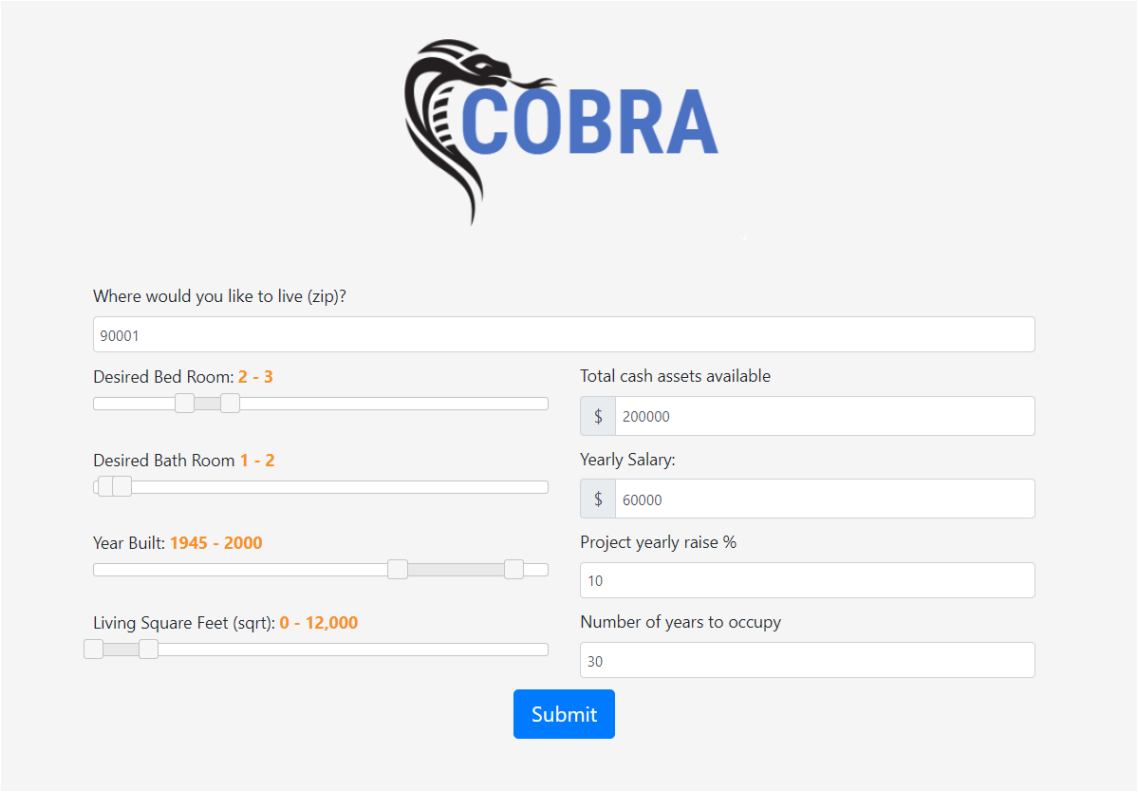

Figure 1: COBRA questionnaire

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 our database of available home to match the user's home preferences (e.g. city or zip code, number of bed/bathrooms, and desired square footage
2. Calculates the mean/median average price and tax amount for buying and renting the user's desired home in the chosen city or neighborhood
3. Estimates the maximum contribution a user can make towards buying or renting a home based on the user's projected salary and lump sum of assets over the chosen time frame
4. Provides a general recommendation on whether to buy or rent in the chosen city or neighborhood
5. Estimates the total cost of buying and renting each remaining home based on estimated interest rates and Homeowners Association Fees over the course of the chosen time frame
6. Filters our database for homes where the user's maximum contribution exceeds the total cost of buying or renting

7. Classifies each remaining home “blue” for rent and “red” for buy
8. Visualizes all available homes for renting or buying on the GUI along with an overall recommendation for renting or buying in the chosen city or neighborhood

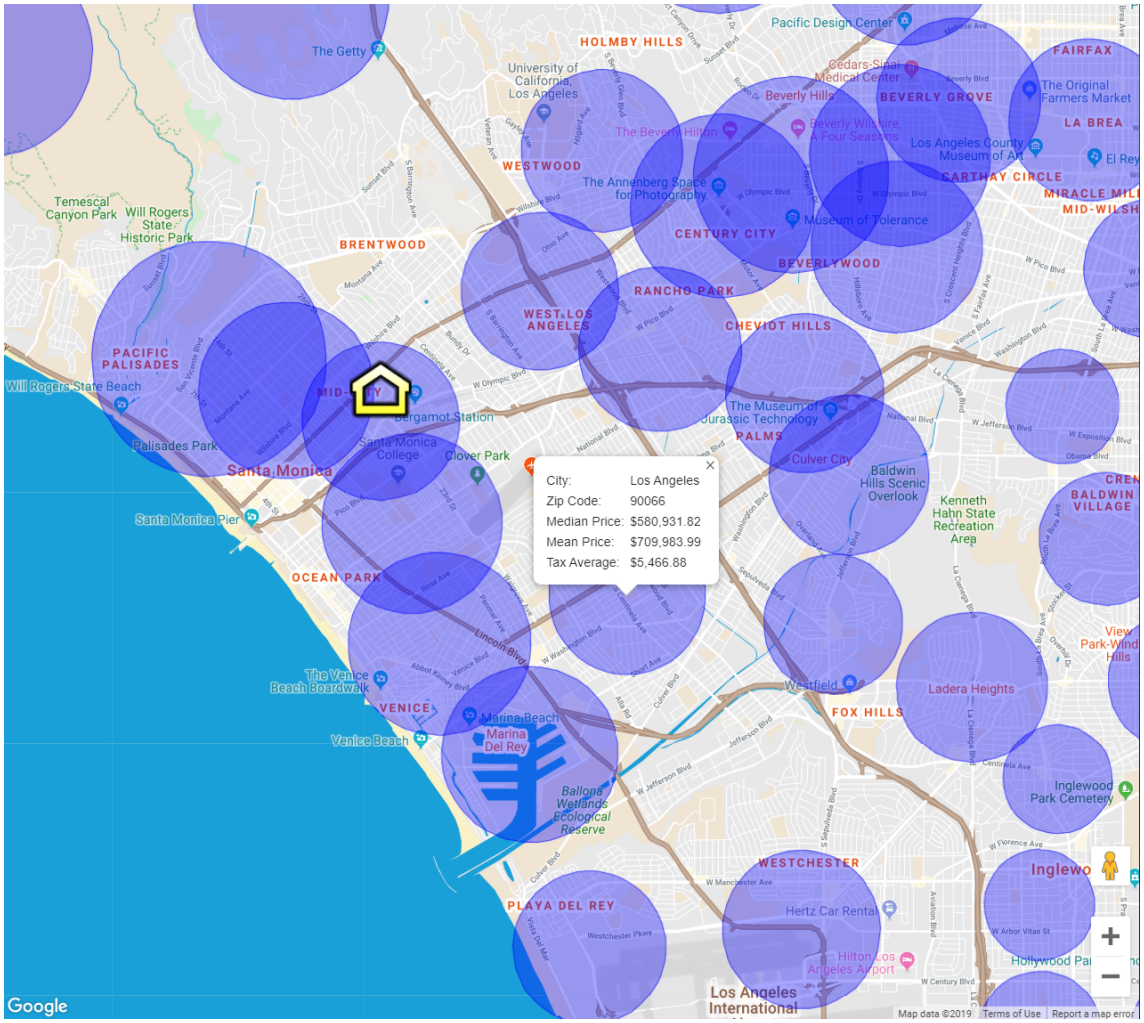

Figure 2: COBRA visualization of mean housing prices by zip code.

The algorithm is influenced by Clark and Lomax’s rent/price ratio, which helps us determine the cost of renting over a range of time [15], Gindelsky, Moulton, and Wentland’s user cost approach for predicting home prices, which uses Zillow’s database of housing prices [14], and Liang, Phillips, and Yu’s regression model for determining real estate prices based on home attributes [8]. Our team is still in the process of determining which components of the aforementioned approaches will live in the final version of COBRA’s algorithm.

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.

Once published, COBRA will exhibit each of the following innovations, which have yet to be provided on existing platforms such as Zillow or Money Under 30:

- A graphical user interface that provide visualizations at the indivual home level for rent or buy decisions
- A definitive rent or buy recommendation for each available home
- An estimated total cost or renting or buying over the desired time period

4 Experiments and Evaluation

Our team has completed the data collection, transformation, and integration phase of our project plan and have initiated the development of our GUI and Python algorithm. COBRA is designed to answer the following questions:

- Should I rent or buy in my neighborhood of choice?
- What will it cost me to rent or buy in my neighborhood of choice?
- What types of home can I afford?

Because we have not implemented our algorithm end-to-end, we do not have an initial set of test results. Once we are able to obtain our initial test results, we plan to test COBRA on the effectiveness of answering the three questions above by sampling approximately 10 homes from the dataset to make a traditional assessment on whether to rent or buy (considering Fehérová's key factors affecting buy or rent decisions) and the total cost of renting or buying and compare it to COBRA recommendations [10]. If our assessment aligns with COBRA's recommendation, then we will proceed to publishing COBRA to a web browser.

Our initial observations suggest that we should focus on optimizing computational time of COBRA's algorithm. Because the dataset contains about 3 million records, inefficient computation of recommendation will cost the user time and result in poor user experience.

Additional information that may be helpful for our team to obtain is: estimated mortgage interest rates for the next 2-10 years, projected inflation rate for the next 2-10 years, and estimated Homeowner Association Fees for various neighborhoods in Los Angeles, CA.

Our revised high-level milestones are as follows:

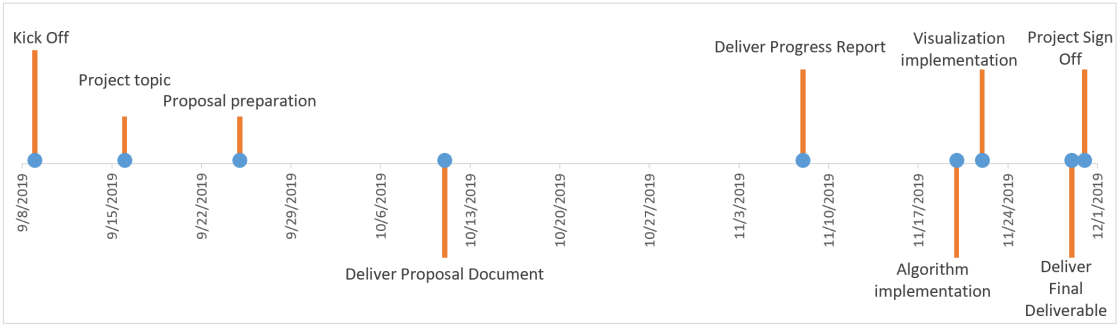

Figure 3: COBRA High-Level Milestones

Our revised detailed project plan is as follows:

TASK NAME	START DATE	END DATE	DAYS	1ST MEMBER	2ND MEMBER	PERCENT COMPLETE
Milestones 1: Project topic						100%
Milestones 2: Proposal preparation						100%
Milestones 3: Deliver Proposal Document						100%
Milestones 4: Algorithm implementation						100%
Set up development environment (Git, Source Code framework, Web Server)	10/12	10/14	3	Hien	Stephen	100%
Zillow housing data manupilation.	10/12	10/16	5	Matt	Crystal	100%
Prepare a separate dataset for testing	10/16	11/15	31	Crystal	Anne	50%
Implement average cost of buying or living in the desired city	10/12	11/15	35	Anne	Crystal	50%
Implement Buy vs Rent decision algorithm	10/12	11/15	35	Matt	Crystal	50%
Implement calculation of total cost of buying and renting for various time periods	10/12	11/15	35	Hien	Stephen	50%
Algorithm validation	10/22	11/18	28	Stephen	Hien	0%
Milestones 5: Visualization implementation						
GUI prototype	10/16	10/21	6	Anne	Crystal	100%
3D visualization for average house price, renting price...	10/22	11/20	30	Crystal	Anne	75%
Visualize Buy vs Rent Decision	10/22	11/20	30	Matt	Hien	75%
Visualize Cost of Buying and Renting	10/22	11/20	30	Stephen	Hien	75%
Integration	10/31	11/20	21	All		0%
Milestones 6: Deliver Progress Report		**11/8**				
Define the format of report	10/28	10/31	4	All		100%
Assignment meeting	10/31	10/31	1	All		100%
Contribute to Progress Report (TBD)	10/31	11/7	8	TBD		100%
Milestones 7: Final Validation						
Integration Testing	11/1	11/22	22	Hien	All	0%
Visualization Testing	11/1	11/22	22	Hien	All	0%
Defect fixing	11/1	11/22	22	All		0%
Milestones 8: Poster Presentation Design						
Design poster presentation	11/9	11/29	21	All		0%
Each member record a single video	11/9	11/29	21	All		0%
Merge presentation videos	11/22	11/29	8	TBD		0%
Milestones 9: Deliver Final Report and Poster F		**11/29**				
Assignment meeting	11/18	11/18	1	All		0%
Contribute to Final Report (TBD)	11/18	11/29	12	All		0%

Figure 4: COBRA Detailed Project Plan

All team members have contributed similar amounts of effort up until this point.

5 Conclusions and Discussion

Currently, we cannot make conclusions about the results of COBRA or gauge the extent of its precision or success. However, we feel confident the the data quality, the proposed Python algorithm, and the current design for the GUI will enable us to innovate a new tool that will help home seekers make better financial decisions on whether or buy or rent a home based on their current life situation.

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).