The Cost of Buying or Renting: Applying Predictive Analytics to Visualization to Support Home Seekers in Los Angeles, California

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 proposes 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. Driven by a Python backend for data preprocessing and analysis, COBRA will utilize D3 to provide a web-based visualization of affordable homes in the Los Angeles area, tailored to an individual's financial information and lifestyle preferences.

2 Current Practice

In the modern era, technology makes it possible to take a "do it yourself" approach to home ownership [2]. People rely less on experts like real estate agents as more digitally-based tools become available to help them decide which house to buy, or whether to buy at all.

Statistical and machine learning models are improving at recognizing patterns in large real estate databases in order to make accurate housing price predictions. Depending on the availability of data from listing services, brokerages, and county and tax assessor records, Zillow's Zestimates use such methods to estimate housing prices within 10 percent of the actual sale price [3]. Although Zillow and similar services can make recommendations about the decision to rent versus buy, their tools lack the geographical component of HomeSeeker, an interactive visual analytics system developed by Mingzhao et al. [4] to help users understand Australian real estate markets. Our team strives to combine the big data analytics of Zestimates with elements of HomeSeeker's visualization design to create a tool that streamlines the process of house-hunting.

3 Our Approach

We will collect data from Zillow's housing prices database and from our users through a website form. Zillow's database provides median housing prices by region and house type. Additionally, we will leverage transactional housing price data for Los Angeles, which Zillow provides through Kaggle. The users of our tool will provide the following information through our website: total cash assets available, yearly salary, projected yearly raise, number of years planned, and accommodation preferences (e.g. number of bed/bathrooms).

COBRA will then analyze Zillow's datasets against the user's preferences using a custom Python program to create a map-based visualization (Figure 1), providing the total cost of buying versus renting each available home within the time period specified

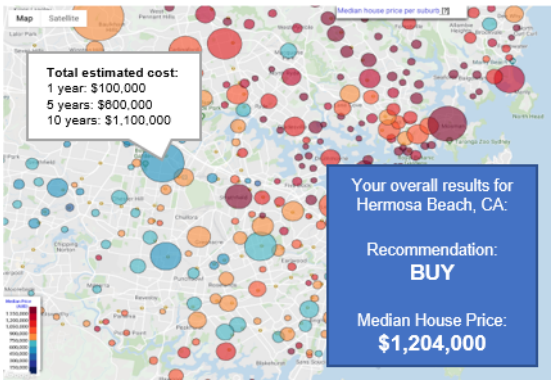

Figure 1: COBRA User Interface Example

by the user. Our tool will also provide the average cost of living in the user's specified neighborhood and an overall recommendation of whether to buy or rent. Our approach is different from the existing rent-or-buy tools in that it provides user-friendly geographic visualizations, total overall costs, and recommendations for a specific time range.

Our approach is inspired by various methods for predicting housing prices. Controlling for heterogeneity in housing data, the hedonic method uses regression techniques to convert the constituent characteristics of a home into a price index [5]. Vineeth et al. [6] experiment with various machine learning algorithms, namely simple linear regression, multiple linear regression, and artificial neural networks. The authors propose the mean square error as a metric for selecting the best performer. Additional algorithms - including k-nearest neighbors, support vector machines, regression tree ensembles - are discussed in the literature [7, 8, 9]. All aforementioned methods require the data to be represented by a discrete set of hand-crafted features, such as construction year, square footage, or postal code. Accordingly, we will commit a significant percentage of our effort toward data curation. Our rent-or-buy logic is inspired by the living cost comparison model [10], which provides a detailed framework for determining whether to make a buy-or-rent decision based on cost savings overtime. Fehérová's framework uses median income and interest rates from Slovakia, whereas COBRA will integrate our user's income and the anticipated mortgage interest rate for Los Angeles.

4 Impact and Measurement of Success

COBRA can help adults who are looking for a new home to determine if it is a better investment to buy or rent. According to research done by Levy and Lee [11], there are 5 stages in the home buying process: (1) problem recognition, (2) product specification, (3) information search, (4) alternative evaluation, and (5) final choice. Their model helps our team better understand consumer incentives so that we can better tailor our tool to help users complete stages 2-4. One important aspect of our tool is the recommendation for certain home seekers to rent rather than buy. Moeller and Wittkowski [12] suggest that the desire for home ownership is influenced by pressures of "possession importance" and "trend orientation" rather than "price consciousness". Our tool helps the customer focus on what is financially practical, rather than what is socially expected by automating the process of determining affordability of available homes. Our impact can be measured through quick pulse surveys, which appear after the user has finished using the tool. The survey would ask the user on a scale of 1-5: "To what extent did this tool help you with your home renting or buying decisions?"

5 Risks and Payoffs

One major risk our team faces is data scarcity. Baldominos et al. [13] experiment with various machine learning techniques to identify real estate investment opportunities in Madrid, Spain. The authors engineer features from data on high-end real estate assets to analyze a snapshot of the market in a six-month period. However, as the authors suggest, prediction performance can be substantially improved using temporal information to model time series over an extended period. The ability to forecast housing prices, then, may be limited by the amount of historical housing data. We will ascertain the availability of such data before investing too much time implementing various regression methods.

Gindelsky et al. [14] use the Zillow Transactional and Assessment Database (ZTRAX) to value housing services on a user cost approach. ZTRAX, however, is not available by application to masters students. Furthermore, it does not contain rent prices and listing data such as time on market. However, Zillow does provide smaller databases for public use that may be downloaded directly or retrieved using an API. Access to datasets such as the Zillow Home Value Index (ZHVI) or Zillow Rent Index (ZRI) using Quandl's free API and Python interface will enable developers to implement an efficient data preprocessing workflow.

The cost analysis guiding the decision to rent or purchase a house is complicated. From a developer's perspective, acquiring a working knowledge of real estate and personal finance necessary to automate this analysis may divert effort from designing the visualization.

Purchasing a house is a significant financial decision; as such, the information conveyed by our visualization tool must be accurate and its origin made transparent. A disclaimer should be included to explain the experimental nature of the tool and recommend the end-user consult an accredited financial advisor before making the decision to purchase a home.

If successful, the visualization tool will distill numerous real estate and rental metrics into a form that is easy for the end-user to digest. For example, Clark and Lomax [15] map national estimates of the rent/price ratio, a proxy for housing market stability, derived from listing data for a sample of English zip codes. Our goal is to create a similar visualization, based on an individual's affordability, that incorporates a temporal

component based on predicted housing prices. Aspiring and existing home owners will be able to monitor regional and temporal housing market trends, which play a major role in long-term financial planning.

6 Project Plan, Progress Measurements, and Costs

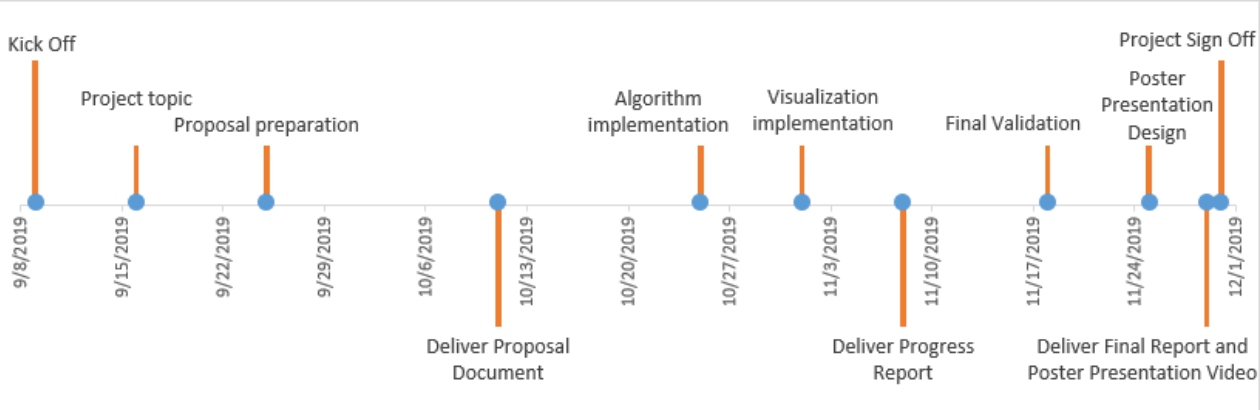

Figure 2: Project Milestones

Our team plans that COBRA will take 3 months to develop and is targeted for completion on November 29, 2019. Costs will be limited to $20 for registering for a website domain.

To ensure progress, our team has created a project plan (Figure 2) that includes project objectives, major milestones and checkpoints. The project plan helps our team organize around a fixed schedule, assign task owners, keep each other accountable, and helps us mitigate risks and employ contingency plans, if needed. The schedule helps us measure progress and determine if we are on-track.

As a precaution, we have planned for testing activities, which will help us verify if the final product meets project objectives prior to submission. Our team will meet weekly to update the project plan, discuss issues, and determine a go-forward plan. Our team has completed Milestones 1-3 (Figure 3) and will move onto Milestones 4-5 (Figure 4).

TASK NAME	START DATE	END DATE	DAYS	1ST MEMBER	2ND MEMBER	PERCENT COMPLETE
Milestones 1: Project topic						
Researching to deciced project topic.	9/9	9/16	8	All		100%
Milestones 2: Proposal preparation						
Every member collect 3 schoolarly articles related to housing	9/17	9/25	9	All		100%
Zillow housing data collection and review	9/17	9/25	9	All		100%
Meeting to select project objectives	9/25	9/25	1	All		100%
Milestones 3: Deliver Proposal Document		10/11				
Q1&2 (9 Heilmeier) contribution	9/26	10/2	7	Anne		100%
Q3 (9 Heilmeier) contribution	9/26	10/2	7	Stephen		100%
Q4&5 (9 Heilmeier) contribution	9/26	10/2	7	Crystal		100%
Q6&7 (9 Heilmeier) contribution	9/26	10/2	7	Matt		100%
Q8&9 (9 Heilmeier) contribution	9/26	10/2	7	Hien		100%
Literature Review Integration	10/2	10/4	3	All		100%
PowerPoint Slide	10/2	10/9	8	Hien	Stephen	100%
Creating Report	10/2	10/9	8	Crystal	Matt	100%
Reference Page	10/2	10/9	8	Anne		100%
Creating presentaion video	10/10	10/10	1	Anne		100%
Final review and submit	10/11	10/11	1	All		100%

Figure 3: Completed Activities

Milestones 4: Algorithm implementation						
Set up development environment (Git, Source Code framework, Web Server)	10/12	10/14	3	Hien	Stephen	0%
Zillow housing data manupilation.	10/12	10/16	5	Matt	Crystal	0%
Prepare a separate dataset for testing	10/16	10/22	7	Crystal	Anne	0%
Implement average cost of buying or living in the desired city	10/12	10/22	11	Anne	Crystal	0%
Implement Buy vs Rent decision algorithm	10/12	10/22	11	Matt	Crystal	0%
Implement calculation of total cost of buying and renting for various time periods	10/12	10/22	11	Hien	Stephen	0%
Algorithm validation	10/22	10/31	10	Stephen	Hien	0%
Milestones 5: Visualization implementatio						
GUI prototype	10/16	10/21	6	Anne	Crystal	0%
3D visualization for average house price, renting price...	10/22	10/31	10	Crystal	Anne	0%
Visualize Buy vs Rent Decision	10/22	10/31	10	Matt	Hien	0%
Visualize Cost of Buying and Renting	10/22	10/31	10	Stephen	Hien	0%
Integration	10/31	11/7	8	All		0%
Milestones 6: Deliver Progress Report		11/8				
Define the format of report	10/28	10/31	4	All		0%
Assignment meeting	10/31	10/31	1	All		0%
Contribute to Progress Report (TBD)	10/31	11/7	8	TBD		0%
Milestones 7: Final Validation						
Integration Testing	11/1	11/18	18	Hien	All	0%
Visualization Testing	11/1	11/18	18	Hien	All	0%
Defect fixing	11/1	11/18	18	All		0%
Milestones 8: Poster Presentation Design						
Design poster presentation	11/9	11/22	14	All		0%
Each member record a single video	11/9	11/22	14	All		0%
Merge presentation videos	11/22	11/25	4	TBD		0%
Milestones 9: Deliver Final Report and Post		11/29				
Assignment meeting	11/18	11/18	1	All		0%
Contribute to Final Report (TBD)	11/18	11/28	11	All		0%

Figure 4: Planned Activities

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).