Cost of Buying or Renting Algorithm (COBRA)

CSE 6242 Data & Visual Analytics

Team 25: Anne Benolkin, Hien Le, Matthew Molinare, Crystal Nguyen, Stephen Wang

Introduction and Problem Definition:

In the search for a home, home seekers can become inundated with information found online about a multitude of real estate opportunities. The number of criteria that factor into identifying the optimal home and deciding whether to rent or to buy is vast and overwhelming. Our team implemented the Cost of Buying/Renting Algorithm (COBRA) to help home-seekers process millions of data points to make an optimal decision on whether to rent or buy their future home. COBRA differentiates itself from other tools by combining user-friendly visualizations with statistically advanced back-end algorithms.

Understanding the Data

Data Acquisition

- Kaggle (Zillow competition) download
  - ~2.8 million data points
  - Home features
  - Home location
  - Home price
- Zillow (published data) download
  - List of zip codes
    - By city
    - By neighborhood
    - By county
  - Median rental price
    - By square footage
    - By zip code
    - By number of bedrooms

Data Cleaning

- Removed irrelevant columns
- Removed outliers
- Removed rows with missing values
- Used property tax to impute home price

parcelid	bathroomcnt	bedroomcnt	squarefeet	latitude	longitude	regionidzip	yearbuilt	numberofstories	est_cost	taxamount	zip	Neighborhood	County	County 2	price_sqft
14599876	2	2	916	33659000	-117673000	96962	1989		463,873	3571.82	92630	Lake Forest	Orange County	Orange County	$506
14601697	2	2	959	33637398	-117588492	96993	1989		562,709	4332.86	92688	Rancho Santa Margarita	Orange County	Orange County	$587
14619750	2	2	943	33716200	-118049000	96967	1989		468,506	3607.5	92649	Huntington Beach	Orange County	Orange County	$497
14632519	2	2	1102	33647093	-117590099	96993	1989		234,704	1807.22	92688	Rancho Santa Margarita	Orange County	Orange County	$213
14632619	2	2	876	33646043	-117591185	96993	1989		370,125	2849.96	92688	Rancho Santa Margarita	Orange County	Orange County	$423
14632719	2	2	888	33645436	-117590995	96993	1989		353,525	2722.14	92688	Rancho Santa Margarita	Orange County	Orange County	$398
14652668	2	2	1124	33554300	-117691000	96987	1989		448,291	3451.84	92677	Laguna Niguel	Orange County	Orange County	$399
14715626	2	2	1211	33631016	-117706199	96962	1989		420,384	3236.96	92630	Lake Forest	Orange County	Orange County	$347
14741636	2	2	1226	33590798	-117720089	96974	1989		343,270	2643.18	92656	Aliso Viejo	Orange County	Orange County	$280
14742464	2	2	1113	33557919	-117688323	96987	1989		282,891	2178.26	92677	Laguna Niguel	Orange County	Orange County	$254
14744189	2	2	1045	33681799	-117712233	96961	1989		664,184	5114.22	92629	Dana Point	Orange County	Orange County	$636
14746806	2	2	965	33663600	-118010000	96966	1989		501,504	3861.58	92648	Huntington Beach	Orange County	Orange County	$520
14746906	2	2	1023	33659500	-117997000	96966	1989		703,288	5415.32	92648	Huntington Beach	Orange County	Orange County	$687
14617146	2	2	864	33760343	-117908842	97050	1989		285,306	2196.86	92843	Garden Grove	Orange County	Orange County	$330
14597547	2	2	1048	33655491	-117685431	96962	1989		359,356	2767.04	92630	Lake Forest	Orange County	Orange County	$343

COBRA employs 2,865,789 records (212 MB) of home value and home feature data from the Los Angeles area.

Exploratory Analysis

Our initial analysis indicates that there is great variability in home values across the Los Angeles area.

Price per Square Feet by County

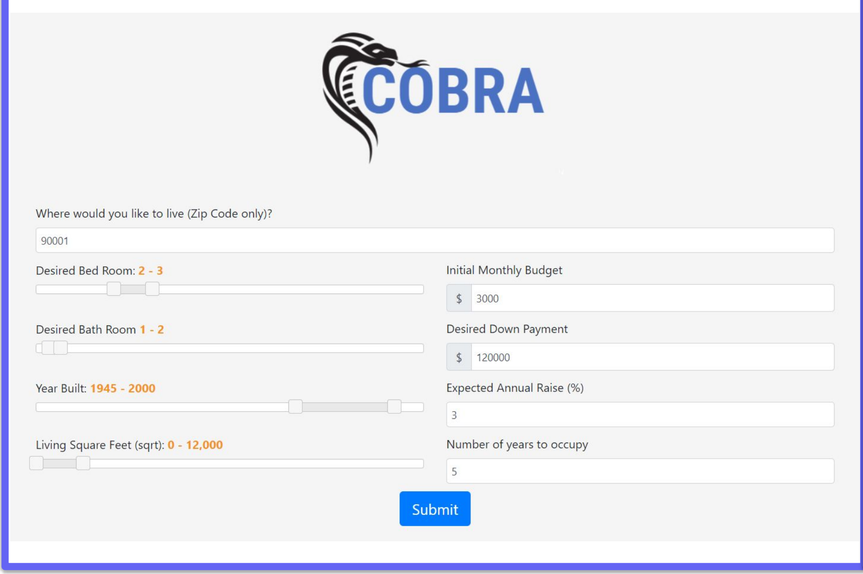
County/Bedrooms	1	2	3	4	5	6
Los Angeles County	$539	$442	$377	$353	$388	$348
Orange County	$504	$428	$385	$351	$375	$374
Other Counties		$869	$612	$436	$362	$271
Ventura County	$427	$348	$330	$305	$305	$294

Approach and Algorithm

The user first interacts with COBRA through a graphical user interface (powered by JavaScript and HTML/CSS) hosted on our website.

Users input

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



Approach and Algorithm (Cont'd)

The data provided in the form by the user is converted into input for a Python-based algorithm and is used to filter Zillow's database of available home to match the user's home preferences. The algorithm is summarized as follows:

- Initiates Python algorithm
- Defines the user's inputs as variables
- Computes the user's expected monthly budget
- Retrieves the sale and rental value of homes
- Calculates the average monthly cost of buying and renting in each zip code
- Calculates the ROI of investing cash assets

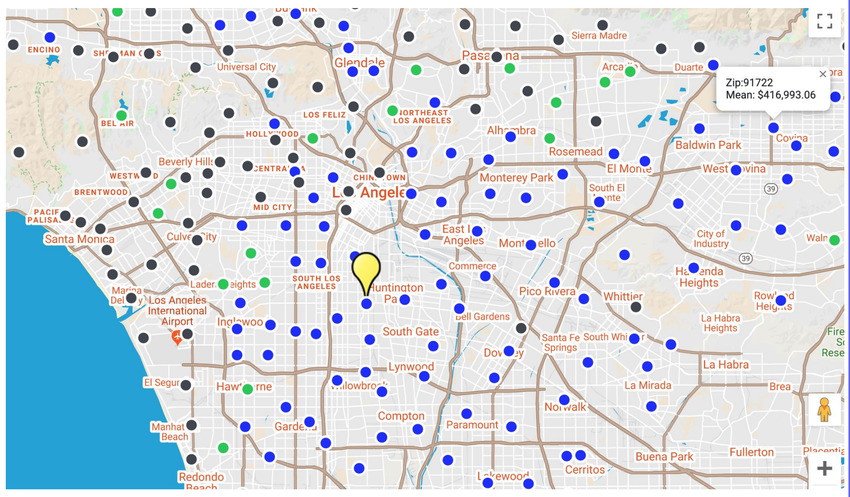
$$\mathrm{ROI}_t = \left(1 + r_{\mathrm{return}}\right)\mathrm{ROI}_{t-1} + \left(1 + \frac{r_{\mathrm{return}}}{2}\right)S_{t-1}$$

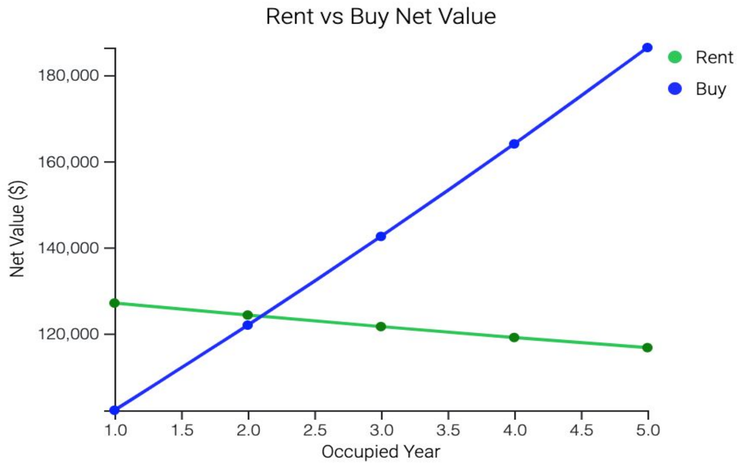
- Calculates the net value of buying and renting

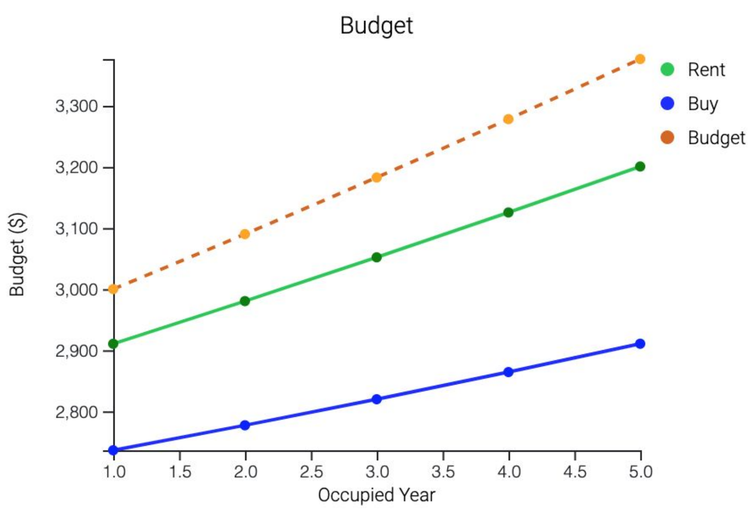
$$\mathrm{NV}_{\mathrm{rent},\,t} = \mathrm{ROI}_t - r_{\mathrm{tax}}\left(\mathrm{ROI}_t - \mathrm{ROI}_1 - \sum_{i=1}^{t-1} S_i^{+}\right) + \sum_{i=1}^{t} S_i^{-}$$

COBRA Output:

- Classifies each zip code as "buy" (blue), "rent" (green), or "exclude" (grey - not affordable)
- Net value of buying/renting line plot for each zip code
- Average monthly cost of buying/renting a home line plot







Results and Evaluation

This tool provides a recommendation, therefore it is difficult to assess the accuracy of the tool definitively. There is no "right" answer to accompany the rent or buy question.

However, models and simulations, like COBRA, are well suited for addressing problems that do no have one right answer.

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.

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

We found that while there was some disagreement across sources, our COBRA results aligned closely with the desires of the participants and the results of similar tools.

Conclusion

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 preferences and Zillow's extensive database of home value information.

COBRA enhances the user experience by employing various data visualizations, such as interactive maps, net value and monthly cost line plots, and color-coded rent or buy recommendations across multiple zip code to add depth and robustness to its assessment.

COBRA's algorithm differentiates itself from current tools by comparing investment returns to home value appreciation to predict net value and displays rent or buy recommendations across multiple zip codes.

COBRA successfully helps many people optimize and alleviate one of the most stressful decisions in life.

Contribution Breakdown:

Proposal

Lit review - all

Final write up - Crystal and Matt

PowerPoint - Hien and Stephen

Video - Anne

Progress Report

Write up- Crystal lead, all

Final Project

User interface brainstorm - all

Data exploration - Crystal lead, all

Algorithm - Matt lead, all

Python code and GUI - Hien and Stephen lead, all

Final report- Crystal lead, all

Poster - Anne lead, all

Video - all

All members contributed equally