CSE 6242 Data and Visual Analytics

Final Report

Rental Recommendation for Commuters

Instructor:

Prof. Duen Horng Chau

November 03, 2018

Team 09

Members:

Bosheng Jian, Lijun Tong, Weixing Tang
Yibo Wang, Zhehan Cao, Zhou Fang

Introduction and Problem Definition

Nowadays, more than 90 percent people use web recommendation systems to buy or sell their real estates. [1,2] There are plenty of platforms which integrate the real estate data and present them in one website, like Google, Yelp, Zillow. [3]

The existing problem is that none of these current systems is designed for meeting all needs, especially for college students and daily commuters in one place. Those people, when they are looking for housing rental, need to search among lots of different websites gaining all the information they need such as commuting time from home to workplace and the safety information. For example, one might need to search for houses on Zillow, then move to Google maps for transit time for every house he or she is interested in and then look up reviews of leasing offices or agencies. Furthermore, one might have additional preferences such as shopping location, favorite store and dog parks.

So we want to build a housing rental recommendation website for those who commutes daily, which can provide renting recommendations based on users' personal references, such as basic information, safety anticipation, and transportation requirement. This recommendation website is designed for off-campus living students and daily commuters who have their special needs when choosing their apartments regarding to pricing, transportation and locations.

Survey Review

Various attributes were found to be common when people search for a house to rent: the property location, property type, rent, number of bedrooms, number of bathrooms, the area of living space, availability of parking space, whether they are furnished or serviced and the daily commuting distance and time. Many researches have been done to analyze the relationship between transportation and other metrics based on renters' choices of renting [4,5,6,7]. Researches also show that the renters' choices are also highly related with their occupations. Those researches show how we may improve our project in the future, which is using the date came from our own website to build model and improve the recommendation system.

At the same time, crime rate is an important factor when it comes to residential decision. It is known that people would prefer to live in safer areas, even if it's more expensive. [8] used a two-stage residential sorting model and the results showed that people are willing to pay more to move to a location with lower violent crime occurrences [8,10]. However, it's hard for us to quantify the extent as we don't have dataset of how much more money people are willing to pay for an apartment with lower violent crime.

Also, it is undeniable that crime rate and house price are innately correlated. [11] used instrumental variable (IV) method to analyze the relationship of crime rate and land price in

Japan and concluded that every 10% decrease in burglaries rate would result an average 1% rise in land price. Given that it was conducted in Japan, we still want to use the same method to analyze the relationship (between crime rate and house price) in Atlanta. [12,13,14] differentiate between the effects of property crime and violent crime, by using hedonic regression, then quantify the ‘‘intangible cost’’ of crime. The results differed from poor, middle and rich neighborhoods and, violent crime is the dominant factor affecting house prices. Thus we aim to only focus on violent crimes in Atlanta.

Our system uses multiple datasets from different sources, including user input data and verified or unverified public datasets. [14] points out that plenty of information input by end-users are inaccurate. The author uses data event analysis and error cluster analysis for investigating errors. Data event analysis involves a review of all the processes that capture or manipulate the data and error cluster analysis centers on a data subset that contains data suspected of being incorrect. We plan to use similar methods to verify inaccurate data and clean inaccurate data.

As for recommendation system, [15,16,17] shows how to construct a recommendation system. [15] build recommendation system with Hadoop and utilizes recommendation algorithms in Mahout Architecture. To solve problems of natural language ambiguity, [17] proposes a classification of semantic approaches into top-down and bottom-up, rely on the integration of external knowledge sources and a lightweight semantic representation based on the hypothesis. To build an semantic recommendation system, [16] proposes that adding a semantic matcher in recommendation system to match semantic documents from web service users. Thus we plan to try similar recommendation system framework as to construct an online semantic content-based recommendation system.

Proposed Methods (45%)

1.Intuition

Comparing to other rental housing website like Zillow, we creatively come up with a new method to analyze data and present data people want to know.

a. The dataset we are using is comprehensive and timely. In terms of collecting data, we are using real time data because people want to know the up-to-date information. Besides, our dataset is comprehensive. For example, we analyze the crime data which is not included in all other websites’ datasets.
b. We also analyze data from users’ perspective. We rank the rental houses in three different ways, which are historical review ranking, convenience ranking and safety ranking. The data we show to users is useful, significant and innovative. These data are not user-accessible in other websites.

2. Approaches

Data Collection and Cleaning

Our data includes downloaded data and scraped data . And there are 5 main sources of our data, including crime data downloaded from Atlanta Police Department, apartment data scraped from Zillow, reviews and rating data scraped from Yelp and via Google Map API , map data via Google Map API, and neighborhood convenience data scraped from Walkable.

a. For crime data: we downloaded 11 types crime data of Atlanta Police Department from 2009 to 2018. The data includes crime type, location, time and neighborhood.

b. For map data: based on the daily commute location, daily commute method and time limit, we calculate daily commute distance and then we get all the zipcodes in the circle with daily commute location as central and daily commute distance as radius. Further when we get candidate apartments, based on users' interest places like gym or theater, we use Google Map API to get the closest users' interest places around each candidate apartment and the distance between them.

c. For Zillow data, we used Python scraping to access apartment basic information by zipcodes we get in map data and users' requirements from frontend. Then we remove all the apartments that don't satisfy the time limit constraint.

d. For review and rating data, we use Python scraping to access each candidate apartment's customer review and rating from Yelp and also get similar information via Google Map API.

e. For neighborhood convenience data, we use Python scraping to access information like walk score, bike score and transition score from Walkable by longitude and latitude information of each candidate apartment.

Report Number	Occur Date	Location	Shift Occurence	Loca	UCR #	IBR Cc	Neighborhoc	Latitude	Longitude
102010746	7/20/18	736 PONCE DE LEON AVE NE	Day Watch	24	312	1201	Grant Park	33.74253	-84.36814
112761812	10/3/18	25 PEACHTREE ST NE	Evening Watch	12	630	2303	Lindridge/Ma	33.82267	-84.35754
161640706	6/13/18	232 FORSYTH ST	Day Watch	34	690	2399	Sylvan Hills	33.70369	-84.40795
173161436	11/12/17	835 MARTIN LUTHER KING JR DR	Evening Watch	12	630	2303		33.75504	-84.41457
173401142	12/6/17	2800 FOREST PARK RD SE	Day Watch	29	730	2434		33.67835	-84.35791
173612019	12/27/17	2348 CASCADE RD SW	Evening Watch	18	640	2305		33.72146	-84.46522
173622049	12/28/17	677 CASCADE AVENUE SW	Evening Watch	13	110	912		33.73709	-84.43589
173641217	12/23/17	950 MARIETTA ST NW	Day Watch	18	710	2404		33.77682	-84.40822
180010043	12/31/17	1341 BEECHER ST SW	Morning Watch	13	640	2305		33.7335	-84.43148
180010146	12/31/17	443 MORELAND AVE SE	Morning Watch	18	710	2404		33.74179	-84.34973
180010177	12/31/17	267 MARIETTA ST NW	Morning Watch	18	640	2305		33.76071	-84.39469
180010190	1/1/18	JOSEPH E LOWERY BLVD SW / OC	Morning Watch	13	440	1399	West End	33.73583	-84.41772
180010270	12/31/17	349 EDGEWOOD AVE SE	Morning Watch	18	640	2305		33.75419	-84.37615

Figure.1 Sample Crime Raw Dataset

1. Method of ranking apartments

In the beginning, users are required to select their daily commute location. Based on this location, 30 candidate apartments will be selected by scraping the data from Zillow. Then we would rank the apartments based on our algorithm.There are three criterions in ranking apartments, including historical review, safety and convenience.

In order to rank the apartments, we calculate the scores for each apartment. For historical review scores, we directly scrape the apartment's score on Yelp. For the convenience scores, the scores are obtained from Zillow by real-time scraping. The convenience scores includes walk score, bike score and transition score.

For the safety score, we calculate the number of last 10 years' accidents happened around each apartment. The equation of calculating the safety score is

$$Score\ of\ Apt\ i = \frac{Num\ i - Min}{Max - Min}$$

where Num i means the number of accidents happened around apartment i. Let Ni denote the number accidents happened around apartment i. Min represents the minimum value of Ni and Max represents the maximum value of Ni.

Based on these three scores, the overall grade of each apartment is calculated by taking the average of these three scores. The recommendation ranking is done based on the overall grade.

2. User Interactive and Visualization

1) Interfaces and their goals

A welcome Interface: This page includes an introduction of our project and a simple tutorial about how to use it. It aims to help the users begin their first try on our recommendation system.
Search-filter Interface: This page has 4 sections. Basic information section contains check-boxes, drop down menus and range sliders which enables users to customize their preference including price range, beds, baths, rating on Google/Yelp, etc. In transportation part, user need to locate their place of work/study and preferred distance/time of commuting. In keyword search section, user can input their preferred places.
Result Interface: A page shows a map with our recommended rental position ordered by rank. In this page, users are able to scroll around the map and see on the map where our recommended apartments are and the keyword search results. Other basic map operations such as zoom in and out are also allowed.
Detailed Information Interface: When users click a marker on in the result page, a detailed information page will pop up, and tell them on why we recommend it.

2) **How to build the interface (Tools)**

HTML5, CSS3 and JavaScipt: In general, our website can be divided into two main functional parts. The first one is obtaining filter options from users and the second is display our result in map. The fundamental tools we use are HTML5 with CSS3 and Javascript. Meanwhile, some powerful library in Javascript like Leaflet.js are used to help us build the website.

React (Search-filter Interface): After visiting the welcome page, a page with result map and user preference options module will be displayed. For our search filter options, we consider Reactivesearch a wonderful Javascript library to realize it. Reactivesearch contains various pre-built module for filter section in map search, which is very efficient and convenient. For the actual display of map, we use Reactivemap library which contains interactive component for map display.

3) **Final effect**

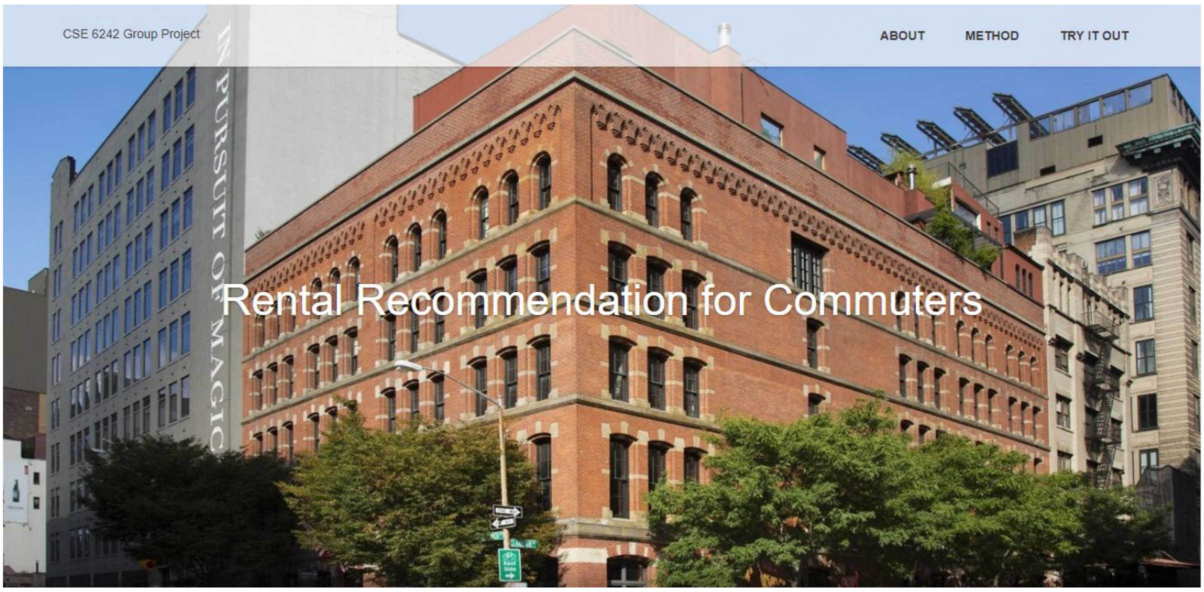


Figure.2 Welcoming Page

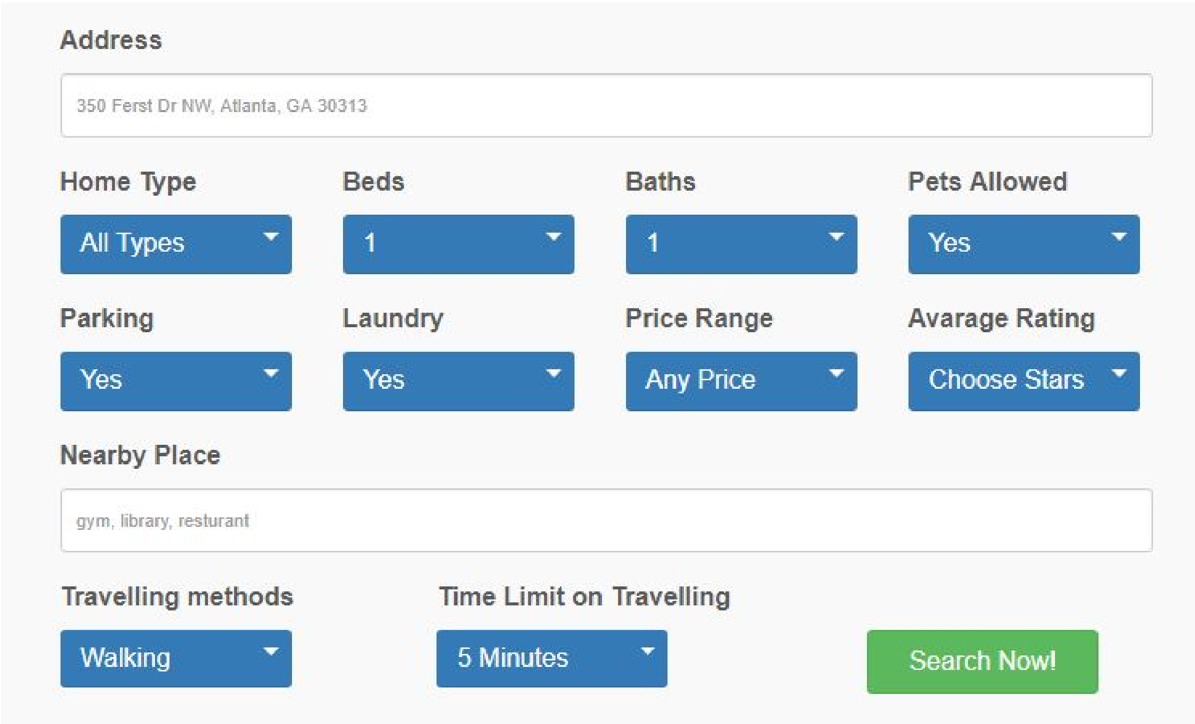


Figure.3 Search Page

Evaluation

(1) What's our result: what question our expriment wants to answer (5%)

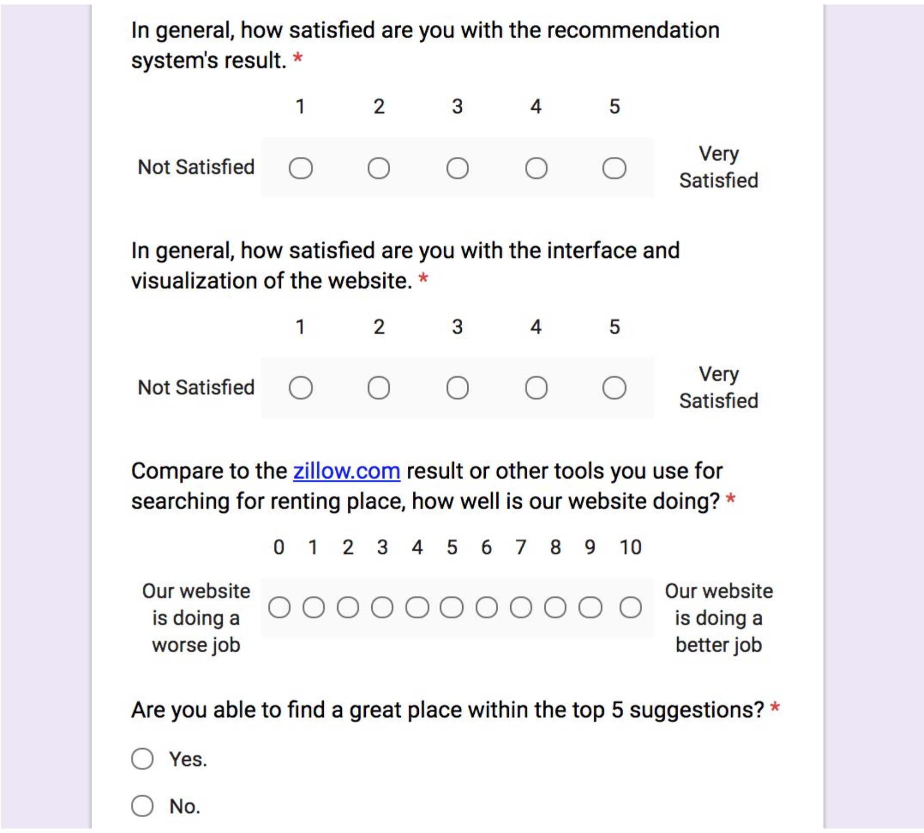
For our recommendation system, we mainly want to add more features compared to existing recommendation service to satisfy users' safety anticipation and transportation requirement. Therefore, in order to evaluate our project, we asked for the feedback from our testing users. We first had the testing users utilize our website for searching for their next potential rentals and then we gave them the survey asking for their evaluation based on their experience with our website.

Also, since user experience is one of the top concern of a recommendation website, our website needs to give back results as quick as possible to limit the user's waiting time. Given that our website needs to handle data from different sources and even with "real-time" scraping, efficiency of our codes are essential. We want to have our website generates searching and ranking results within 2 seconds based on our server's computation resource. This time limit will be our most important evaluation for program efficiency and user experience.

Our questions had two categories. The first category was for asking experience with our website alone and the other category was asking for comparison between our website and other existing solutions such as zillow.com. We not only asked user's for opinions on the results but also asked for user's satisfaction with our interface and visualization. By collecting the responses, we can answer the questions for our project evaluations - how good is our recommendation results; how is our recommendation system compared to others'; how is our website design.

(2) Detail of our evaluation (25%)

Below is a screenshot of our survey given to testing users.



In general, how satisfied are you with the recommendation system's result. *

	1	2	3	4	5	
Not Satisfied	○	○	○	○	○	Very Satisfied

In general, how satisfied are you with the interface and visualization of the website. *

	1	2	3	4	5	
Not Satisfied	○	○	○	○	○	Very Satisfied

Compare to the zillow.com result or other tools you use for searching for renting place, how well is our website doing? *

	0	1	2	3	4	5	6	7	8	9	10	
Our website is doing a worse job	○	○	○	○	○	○	○	○	○	○	○	Our website is doing a better job

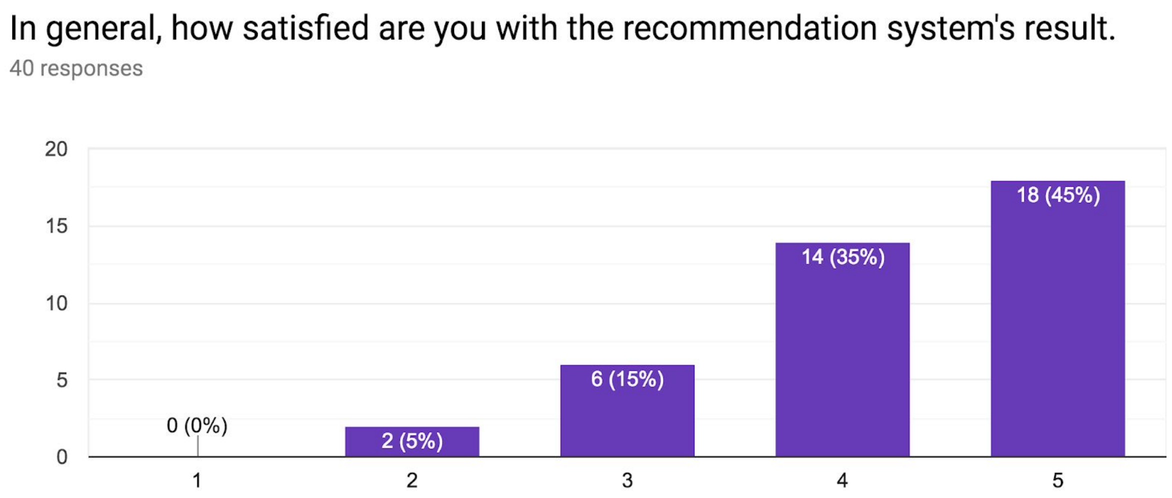
Are you able to find a great place within the top 5 suggestions? *

○ Yes.

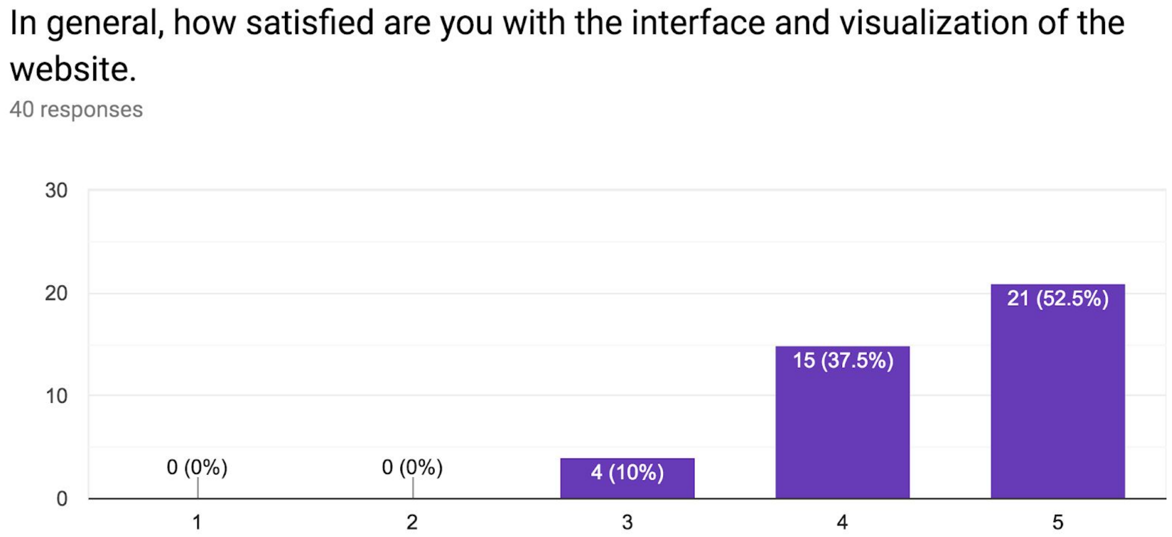
○ No.

We collected response from different groups of users for more accurate results. We had 40 total validate reponses - 3 of them were freshmen or sophomores at Georgia Tech, 9 were juniors or seniors at Georgia Tech, 15 were masters at Georgia Tech, 9 were PhD students at Georgia Tech and 4 were daily commuters working around Georgia Tech campus area.

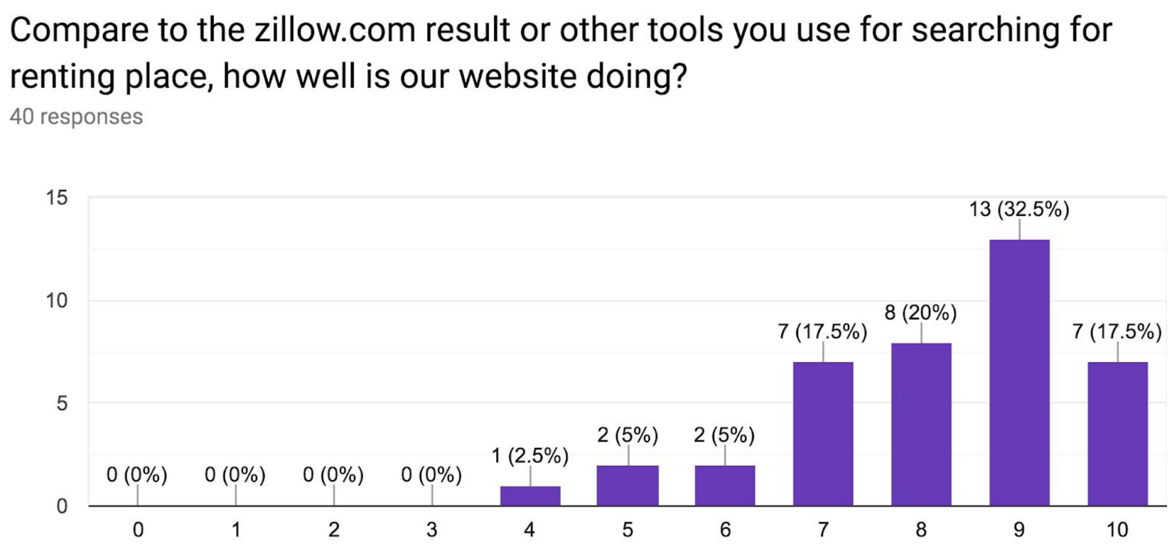
Below are the answers we got from our survey for testing users.



* with 1 unsatisfied and 5 very satisfied



* with 1 unsatisfied and 5 very satisfied



* with 0 our website does a worse job and 10 our website is better

For further quantifying how well our project's recommendations were, we also asked users whether or not they could found a satisfying result within the top 5 suggestions. Only 5 out of 40 users were unable to find one within the top 5 suggestions. The recommendation algorithms and results yielded by the testing user survey met our expectation. However, our website usually took more than 10 seconds for generating the result since we have faced unforseen challenge from scraping web pages. Our project failed meeting the 2-second expected time limit. This was also the main reason why users deducted points from user interface and visualization.

Plan of Activities

Table 1 Distribution of Work

	Done		Doing

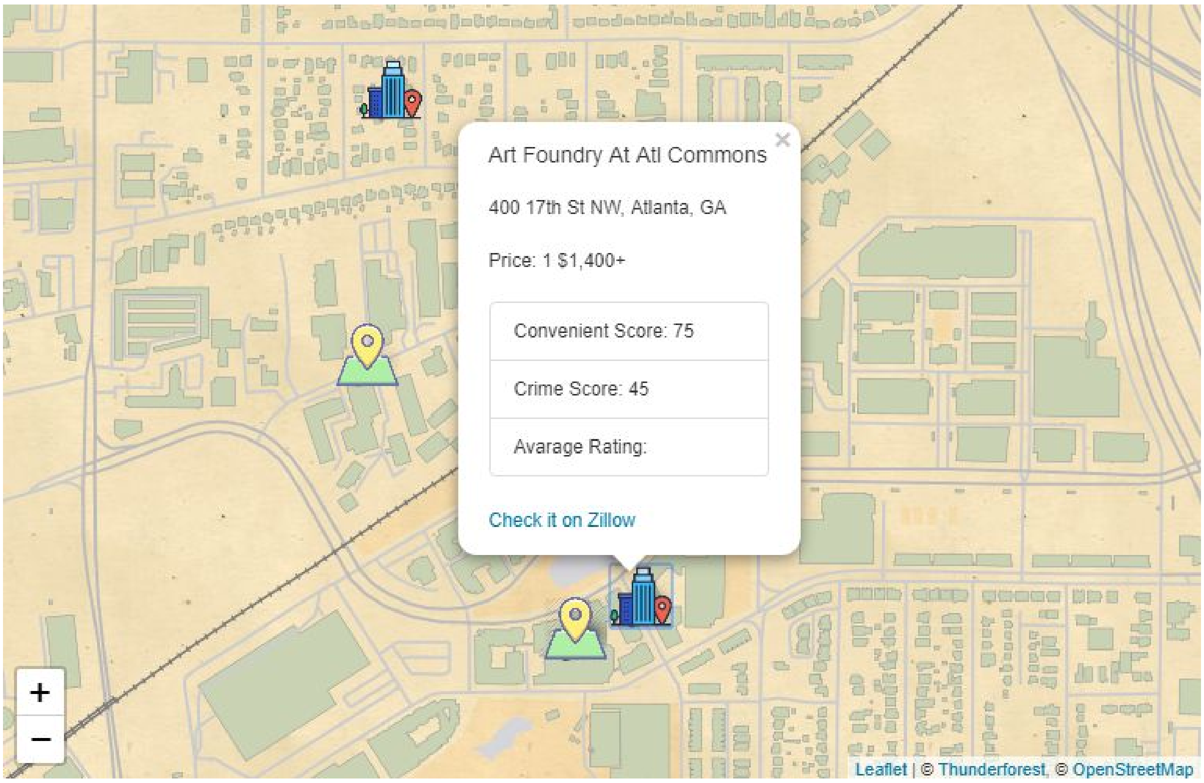
	Team contactor	Proposal presentatio n	Proposal Writing	Progress Report	Implement web interface	Data collection	Data analytics
Time Estimate	N/A	1 week	1 week	1 week	4~5weeks	2 weeks	2 weeks
Zhou Fang	√	√		√	√		
Zhehan Cao			√	√		√	√
Weixing Tang		√		√		√	√
Bosheng Jian			√	√	√		
Yibo Wang			√	√		√	√
Lijun Tong			√	√	√		

Conclusion

In this project, we are going to build a housing rental recommendation website, especially for off-campus living students and daily commuters who have their special needs. There are three main innovation for our project: First, this project innovatively comes up with the idea to design a housing rental recommendation website especially for off-campus students and commuters. Secondly, we used multiple datasets like Yelp, Zillow, Crime datasets to construct multi-feature filtings. Finally, our user interface is interactive and user-oriented. Our project will definitely contribute tremendous value in giving people a easy life.

Reference

[1] National Association of Realtors. *Home Buyer and Seller Generational Trends Report 2017*. National Association of Realtors, 2017.
[2] National Association of Realtors. *Real Estate in a Digital Age Report 2017*. National Association of Realtors, 2017.
[3] Fang, Yao-Min, et al. "An integrated information system for real estate agency-based on service-oriented architecture." *Expert systems with applications* 36.8 (2009): 11039-11044.
[4] Weisbrod, Glen, Steven R. Lerman, and Moshe Ben-Akiva. "Tradeoffs in residential location decisions: Transportation versus other factors." *Transport Policy and Decision Making*1.1 (1980): 13-26.
[5] Bina, Michelle, Valdemar Warburg, and Kara Kockelman. "Location choice vis-à-vis transportation: Apartment dwellers." *Transportation Research Record: Journal of the Transportation Research Board* 1977 (2006): 93-102.
[6] Mathur, Shishir. "Impact of transportation and other jurisdictional-level infrastructure and services on housing prices." *Journal of Urban Planning and Development* 134.1 (2008): 32-41.
[7] Frenkel, Amnon, Edward Bendit, and Sigal Kaplan. "Residential location choice of knowledge-workers: The role of amenities, workplace and lifestyle." *Cities* 35 (2013): 33-41.
[8] Kutsuzawa, Ryuji, et al. *The impact of crime on housing land prices and the effects of police boxes and voluntary groups on crime prevention in Japan*. No. 60. 2010.
[9] Zhang, Zhaohua, and Diane Hite. "House Value, Crime and Residential Location Choice." *2015 Annual Meeting, January 31-February 3, 2015, Atlanta, Georgia*. No. 196826. Southern Agricultural Economics Association, 2015.
[10] Bayer, Patrick, Robert McMillan, and Kim Rueben. *An Equilibrium Model of Sorting in an Urban Housing Market: The Causes and Consequences of Residential Segregation*. No. 860. Center Discussion Paper, 2003.
[11] Tita, George E., Tricia L. Petras, and Robert T. Greenbaum. "Crime and residential choice: a neighborhood level analysis of the impact of crime on housing prices." *Journal of quantitative criminology* 22.4 (2006): 299.
[12] Lynch, Allen K., and David W. Rasmussen. "Measuring the impact of crime on house prices." *Applied Economics* 33.15 (2001): 1981-1989.
[13] Hellman, Daryl A., and Joel L. Naroff. "The impact of crime on urban residential property values." *Urban Studies* 16.1 (1979): 105-112.
[14] Gelman, Irit Askira, and Ningning Wu. "Combining structured and unstructured information sources for a study of data quality: a case study of Zillow. com." *hicss*. IEEE, 1899.
[15] Verma, Jai Prakash, Bankim Patel, and Atul Patel. "Big data analysis: recommendation system with Hadoop framework." *Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on*. IEEE, 2015.
[16] Manikrao, Umardand Shripad, and T. V. Prabhakar. "Dynamic selection of web services with recommendation system." *Next Generation Web Services Practices, 2005. NWeSP 2005. International Conference on*. IEEE, 2005.
[17] de Gemmis, Marco, et al. "Semantics-aware content-based recommender systems." *Recommender Systems Handbook*. Springer, Boston, MA, 2015. 119-159.
[18] Hurley, Gareth, and David C. Wilson. "DubLet: An Online CBR System for Rental Property Recommendation." *International Conference on Case-Based Reasoning*. Springer, Berlin, Heidelberg, 2001.
[19] Guy, Nicolas Nahimana. *A Recommender system for rental properties*. Diss. Strathmore University, 2017.
[20] Kasamani, Bernard Shibwabo, and Guy Nahimana. "A System for Recommending Rental Properties." *Journal of Systems Integration* 8.3 (2017): 10-18.


Art Foundry At Atl Commons
400 17th St NW, Atlanta, GA
Price: 1 $1,400+
Convenient Score: 75
Crime Score: 45
Avarage Rating:
Check it on Zillow
+
−
Leaflet | © Thunderforest, © OpenStreetMap