Personalized Short Term Home Rental Suggestions

VacayAway - Team 22
Stephen Leonard
Deepti Madathany
Jun Xiang
Sudhanshu Manda
Hoon Na

Introduction - Motivation

Short-term home rental websites such as AirBnb, VRBO and HomeAway have become popular over the past decade. However, it is difficult to find a neighbourhood that meets an individual's specific criteria in an unfamiliar locale. We want to improve upon existing short-term home rental websites by creating an integrated service to evaluate listings and their neighbourhoods based on consumer-specific criteria by aggregating multiple data sources. Such a service can be used by both vacationers and apps like AirBnB/hotels to allow their customers to create a one-stop shop for listings. For this MVP, we are evaluating HomeAway listings in Atlanta [7] using local crime data [3] and the quality of nearby businesses/attractions using Yelp [17].

Problem Definition

We provide users with an interface to provide custom weights to filter listings in Atlanta along with weights for the filter criteria. We use this weighted criteria to score all Homeaway listings in the city of Atlanta and return the top 5 listings. The custom criteria includes price range, neighbourhood crime and Yelp business categories.

Survey

There are a few tools that currently try to help users choose the right neighborhood for them. AreaVibes and AARP provide neighbourhood "livability scores". However, these scores heavily weight education, employment and infrastructure, [8] which are irrelevant when searching for a short-term home rental. The results in the study by De Nadai and Lepri on home values used these factors, and will be irrelevant to our goal [4].

There have been many studies on economic values for neighborhoods. According to a study on price estimate methodology, most of the primary factors are categorized into two divisions: 1) tangible elements such as proximity to amenities and walkability [9] and 2) intangible elements such as perception of security and safety and cultural liveliness [4].

Many studies have also shown a relationship between neighborhood quality and crime [18]. In Goncalves' study, crime has been shown to decrease property value by up to $8,841 [2]. Naik et al. [13] provide a tool that calculates the perceived safety level of a neighborhood using an image from Google Maps' street view. This can be useful when scoring a neighborhood based on crime and safety.

According to Okulicz-Kozaryn's study, livability scores and residents' perceptions do not always show a strong correlation [1]. Analogously, using rental ratings may not be the most accurate in finding the most ideal choice for a consumer. Zervas et al. shows that ratings on AirBnb are predominantly above 3.5 stars unlike TripAdvisor, which saw much more varied ratings [6]. Therefore, ratings distributions can be highly variable depending on latent variables. Filtering results by rating would not appreciably reduce the sample set of possible properties. HomeAway allows users to filter on nearby attractions. However, it can quickly filter away all listings if the user gets too specific. Instead, we want to combine HomeAway's custom filters with the livability scores of AreaVibes and AARP to rank available properties on how they fit our user's requirements.

Glaesar et al showed a correlation between new businesses entering an area and increased home prices using Yelp data[5]. This inspired us to look at aggregating Yelp ratings to evaluate neighbourhoods. Josang & Haller proposed a time-based multinomial reputation system based on Dirichlet probability using a longevity factor to develop a dynamic community base-rate reputation model [10]. Shaalan & Zhang added the quality of the rating based on the reviewer's overall reputation to build an exponential model, TQRank, to score products [15]. Tan et al conducted a survey of sentiment detection methods by evaluating the content of reviews which could help on improving the ratings [16].

We looked into multi-criteria geographic visualization studies to tie all of these together. Rinner [14] uses geographical visualization to evaluate urban quality of life in Toronto by asking participants to blend two different evaluation models on a map. The participants loved the instant visual feedback but wanted more variables. Hamilton et al [12] evaluated a web-based geospatial multi criteria decision analysis tool using open-source software. The tool provides an end-to-end workflow to create a criteria model using user-provided decision rules in an iterative manner.

Proposed Method

Innovations

Today, vacation rental websites allow users to narrow down listings on generic criteria and force them to research the actual neighbourhoods on their own. We plan to fill in this gap with the following innovations:

1. We will allow users to search neighbourhoods on multiple custom criteria such as Mediterranean restaurants, etc. to allow them to narrow down listings with minimum effort. We will also let them prioritize the different criteria by adding weights to them.
2. We score the different listings on the above-defined criteria to properly evaluate the recommendations. Today, websites provide livability scores that emphasize factors such as taxes and education that are important to long term residents but unimportant to travellers.
3. Unlike the status quo tools that just measures the distance of businesses, we also measure the quality of businesses nearby when evaluating a listing. In the same vein, we weight the type of crime when coming up with the crime score.

Data Sources

HomeAway

Every rental listing for the city of Atlanta was scraped from HomeAway. This provides us with latitude/longitude, price, rating, review count, and number of bedrooms/bathrooms.

Crime Data

Atlanta Police Department provides crime data over Atlanta through the recent ten years with 315,000 entries. Its major fields include date of occurrence, latitude/longitude and type of crime. Community environment changes with time, so we take crime data in recent five years.

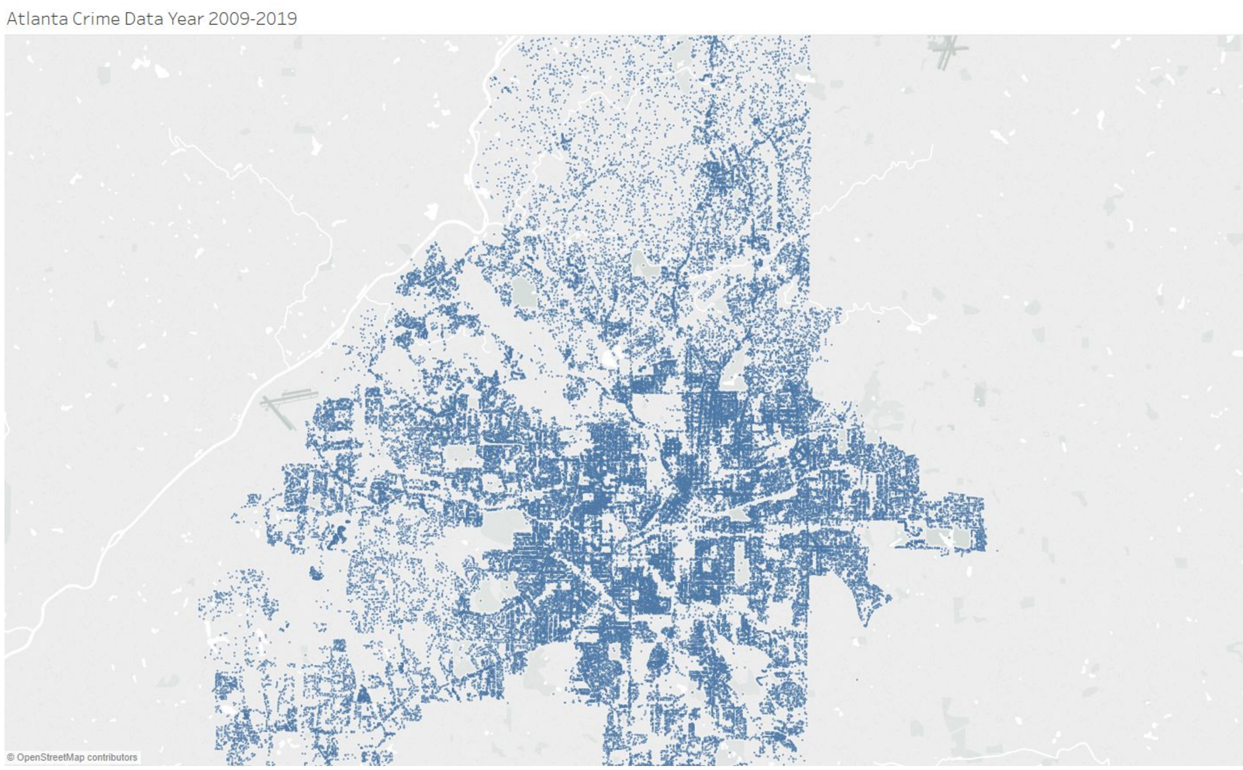


Figure 1. Map of Atlanta with crime distribution.

Yelp

The Yelp open business dataset actually does not contain any of the business information in Atlanta, Thus, we scraped the Atlanta business data from Yelp API. Due to the constraints of the number of rows for each call, we called all the business data for each of the zipcode in Atlanta. Thus, there were some datasets that were not called due to the Yelp's 1,000 row limitation. Our final dataset contains over 22,000 just for Atlanta. The primary features for each business are name, rating, number of reviews, categories, price, latitude and longitude.

A user will select three User Categories (UCs), comprised of Primary/Secondary Category (PC and SC) pairs, and weight their importance. For example, a user might be interested in being around a multitude of bars, so they would pick a PC/SC pair of “Night Life” and “Bars”, which makes up 1 UC.We then find the 5 closest businesses that match this UC and calculate a Distance Score for that UC as follows:

$\boldsymbol{D_x}$ = Distance Score for HomeAway listing

$\boldsymbol{d_i}$ = Value indicating distance of i^{th} closest business listing to HomeAway listing

$$\boldsymbol{D_x} = \sum_{i=1}^{5} \frac{i}{\boldsymbol{d_i}}$$

This Distance Score formula is effective at discriminating between listings with farther away UC businesses from closer businesses. A listing with UC businesses closer to it will have a higher Distance Score than a listing with UC businesses farther away.

Crime Rating

Crime rating is a comparison between crime density over certain radius at listing location and overall crime density in Atlanta. This represents relative severity of crimes around a location. We also assign weights to various types of crime, using the method from the Cambridge Crime Harm Index [19]. Then, we divided the city of Atlanta into 245 neighborhoods. Using the historical crime data with the weights on different types of the data, each neighborhood has a weighted-sum of the crimes as below:

$$Crime\_Score = W_1C_1 + W_2C_2 + \ldots + W_NC_N$$

For each neighborhood, we utilized the ARIMA time-series model, one of the most well-known algorithms in Time-Series to forecast a crime score for a given month. Figure below is the representation of a crime-forecasting for a neighborhood in Atlanta. Based on the plot, the model captures the crime trends of the location in the past and forecasts the future crime rates.

Each listing is assigned the Crime Score of the neighborhood it is located in.

	Crime Type character varying	Crime Harm Index integer
1	LARCENY-NON VEHI...	2
2	BURGLARY-NONRES	5
3	AUTO THEFT	4
4	LARCENY-FROM VE...	2
5	BURGLARY-RESIDEN...	5
6	MANSLAUGHTER	4
7	ROBBERY-RESIDENCE	8
8	AGG ASSAULT	6
9	ROBBERY-PEDESTRI...	8
10	HOMICIDE	24
11	ROBBERY-COMMER...	8

Figure 2. Crime Harm Index for each type of crime

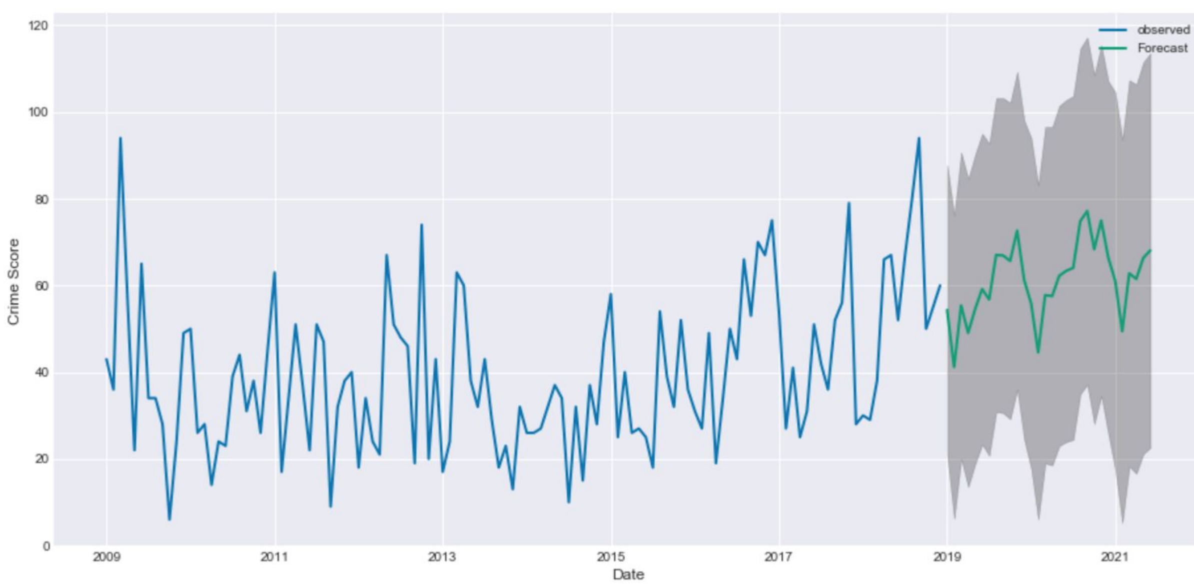


Figure 3. Crime forecasting of a neighborhood in Atlanta

Feature Representation

A feature vector for every rental listing is constructed by appending the crime rating for the listing as well as distance scores from the listing to each different Yelp category. The distance score is calculated by finding the minimum distance from the listing to a business of that category.

Matching Score

The goal of our system will be to output a Matching Score [MS] for each listing that best matches the user's weighted criteria.

Based off the final feature representation, each feature can be represented in the set $F = \{F_1, F_2,...,F_N\}$ for N features. Each feature will have a weight - supplied by the user - represented in the set $W = \{W_1, W_2,...,W_N\}$. At this point in time, it seems reasonable to constrain the list of possible algorithms to encompass permutations of linear regression. Therefore, the algorithm will take the following form:

$$MS = W_1F_1 + W_2F_2 + \ldots + W_NF_N$$

The output will be a ranked list of listings with the highest Matching Scores.

Interactive UI

We built an interactive UI using Tableau to get the user input and return the top 100 listings on an Openstreet Map of Atlanta. (Fig 2). We allow users to weight the importance of different criteria from 0 to 4 with 4 being the most important. We then refresh to pull the listings and scores.The scoring algorithm is deployed to a local tabpy server as a REST endpoint. It receives the user inputs and scores the listings. Tableau ranks the updated listings and scores and displays the top 100 listings along with a hover tooltip that shows the listing id, rank and the match score. The refresh is currently a manual step on Tableau Desktop but can be automated on Tableau

Server.

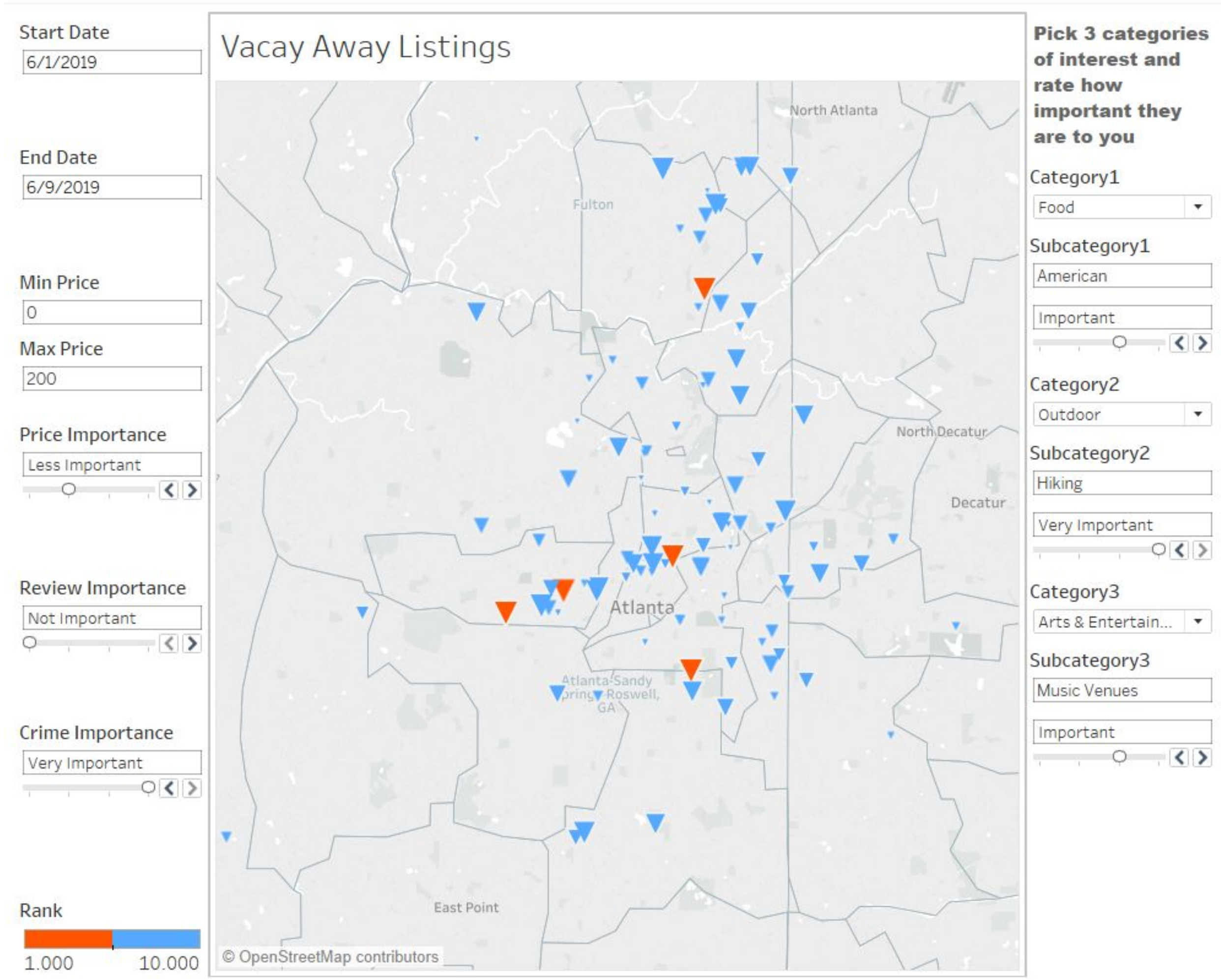


Figure 4. VacayAway UI

In addition, we built a UI to allow users to see the distribution of types of businesses across neighborhoods in Atlanta. (Fig 3) For this, we downloaded the neighborhood GIS data of Atlanta.

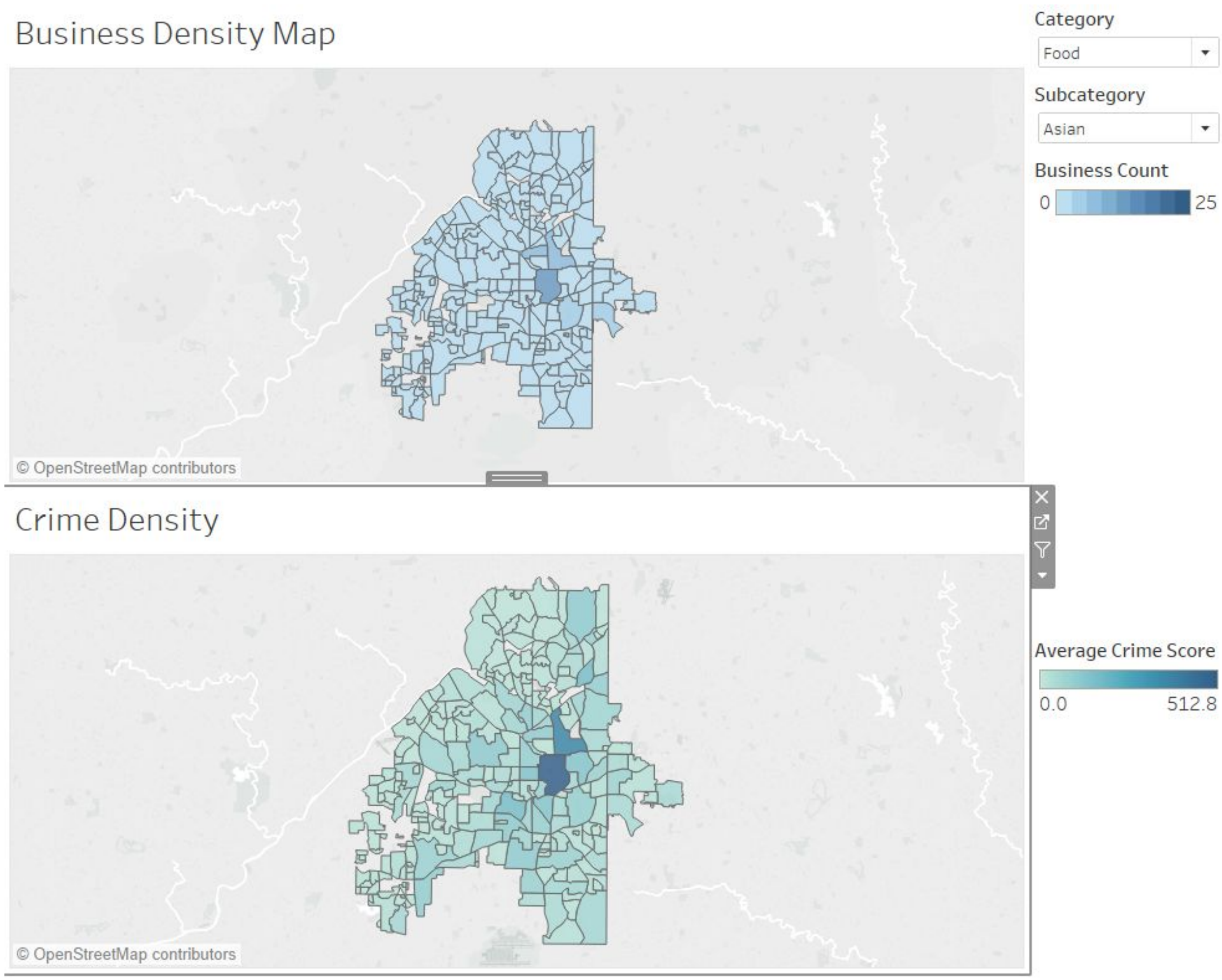
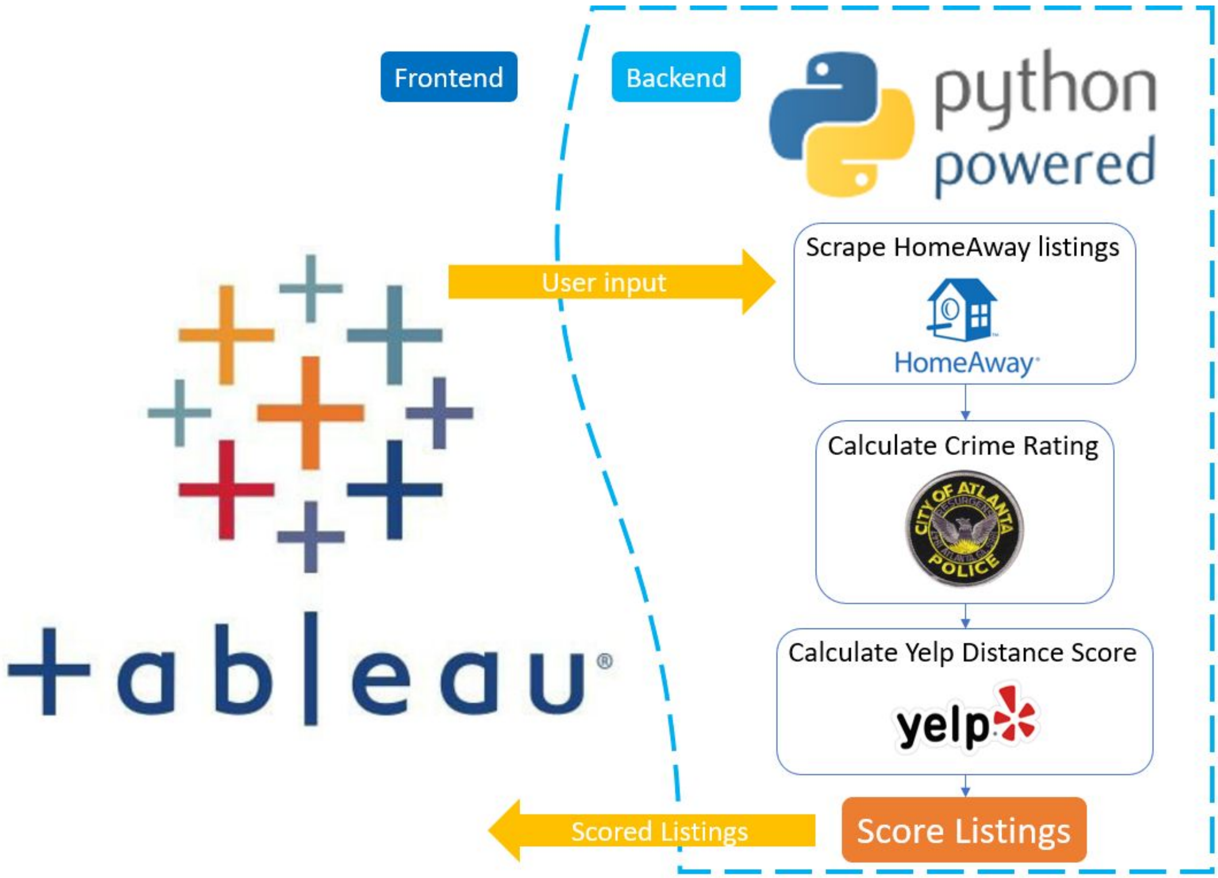


Figure 5. Business and Crime Density in Atlanta Neighborhoods

Pipeline



The application utilizes the following pipeline, graphically represented above:

1. From the aforementioned VacayAway UI, the user selects important vacation rental information, including start and end dates. The user also selects 3 PC/SC that are of interest to them to create UCs. Finally, the user moves the Importance sliders designating how important factor is to their rental selection. For example, a user might think that rental cost is very important, and may therefore move the ‘Price Importance’ slider to the right. Importance sliders are also present for User Reviews, Crime Rates, and for the three designated UCs.The user inputs are sent through a Python Backend, where the Importance sliders’ values are represented as ‘weights’ on the different Factors. The following Factors’ weights are parsed:
    - HomeAway listing price
    - Average review of HomeAway listing
    - Crime rate score for listing
    - The three UCs
2. First, RESTful API calls are made to obtain HomeAway listings that meet the user’s search criteria and start/end dates.

3. The Crime Score for each listing is calculated as described above and appended to each listing.
4. The Distance Scores for each UC are calculated using business entries from the Yelp database and appended to each listing.
5. The aforementioned Factors' values are normalized to a range of [0,1].
6. The normalized Factor values are used to calculate each HomeAway listing's Matching Score
7. The scored HomeAway listings' information is sent back to Tableau - the VacayAway UI then presents and visualizes the listings ranked by Matching Score.

Experiments / Evaluation

We currently have three main experiments planned for our project, asking 20 participants to use our system:

1. ***Determine relationship between user specified importance and rental scores.*** We evaluated this factor by asking users to change the importance of their criteria and asked if the new recommended locations were more in line with their expectations. Over half liked the current design; 18% people need more choices to describe their criteria; 26% people found it hard to find the categories they wanted.

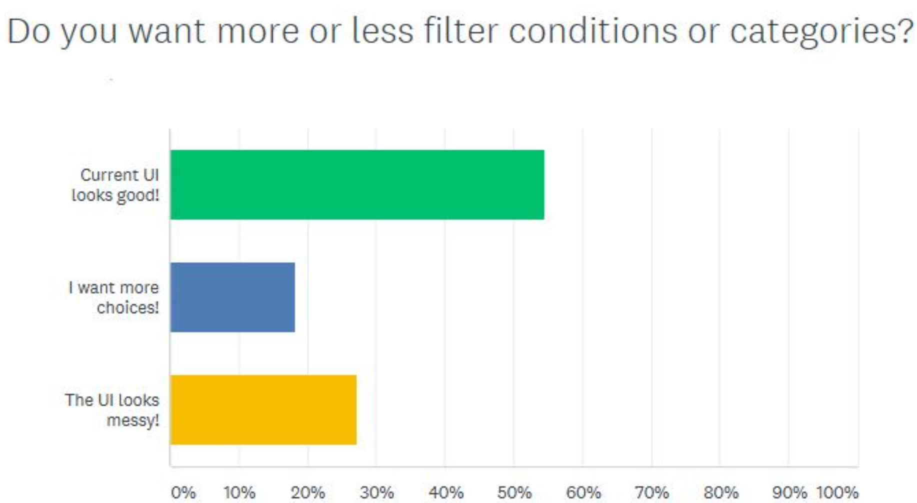


Figure 6. Category survey result.

2. ***Determine usability of the user interface***. We gave 20 participants a task to find a place to stay in Atlanta that fits their needs if they wanted to come here on a weekend vacation. After they completed this task we asked them to fill out a survey rating their experience and detailing what they had trouble with and what they would like to see added or removed. The results of the survey are shown below. One of the common feedbacks we heard from users is that it was hard to see on the map what part of Atlanta a listing is in

because everything looks gray. Another common complaint is not being able to see why it gave certain listings a high score. It currently doesn't show the nearby businesses that fit the criteria they gave, and that is something they wanted. This gives us a good starting point for changes to make going forward.

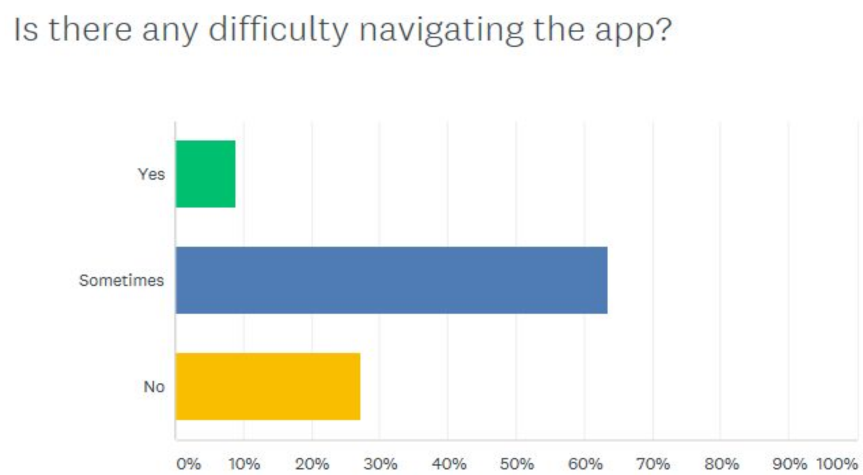


Figure 7. Application UI survey result.

3. ***Determine quality of recommendations to real people.*** After our participants completed their search we reviewed the system-recommended listings. For each location that was recommended to them we asked them how accurately it fit their specified criteria. Of 20 people, 72% satisfied with the results; 18% found the recommendations a bit off their expectation; 9% found the recommendations misleading, due to criteria not correctly reflected on categories.

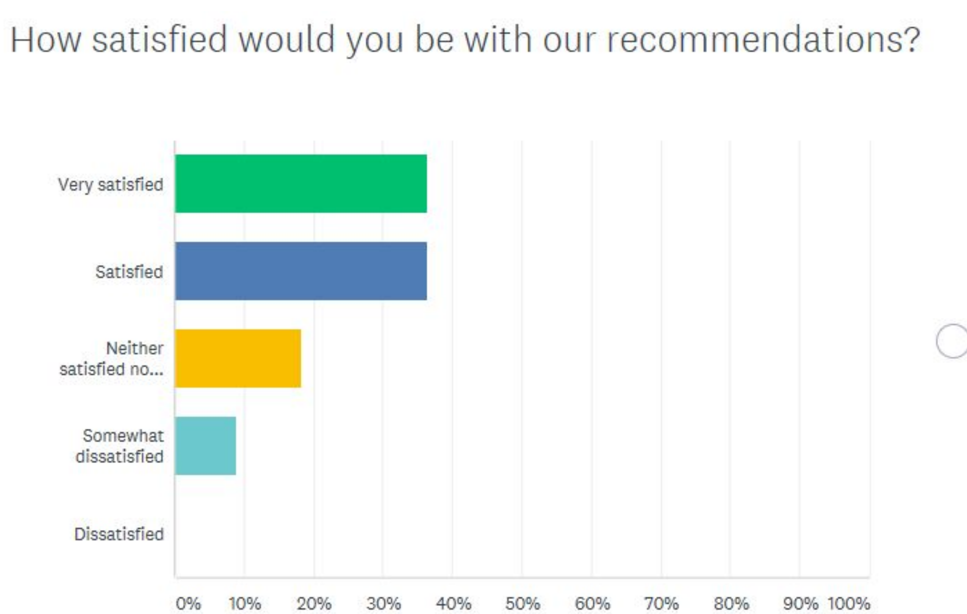


Figure 8. Recommendation satisfaction survey result.

4. ***Determine accuracy of models.*** Along with the survey, for the validation of the performance of our time-series model, we have performed the one-step ahead forecast to forecast the crime scores from 2016 to 2019 using the past data and checks how accurately it performs by comparing with the actual crime data. Below is one of the plots of our one-step ahead forecast and the model generally shows a good accuracy. Numerically, most of the models show the Root Mean Squared Error of 25 or less. However, due to the limited datasets of crimes in some of the neighborhoods, our model showed higher errors in some of the neighborhoods.

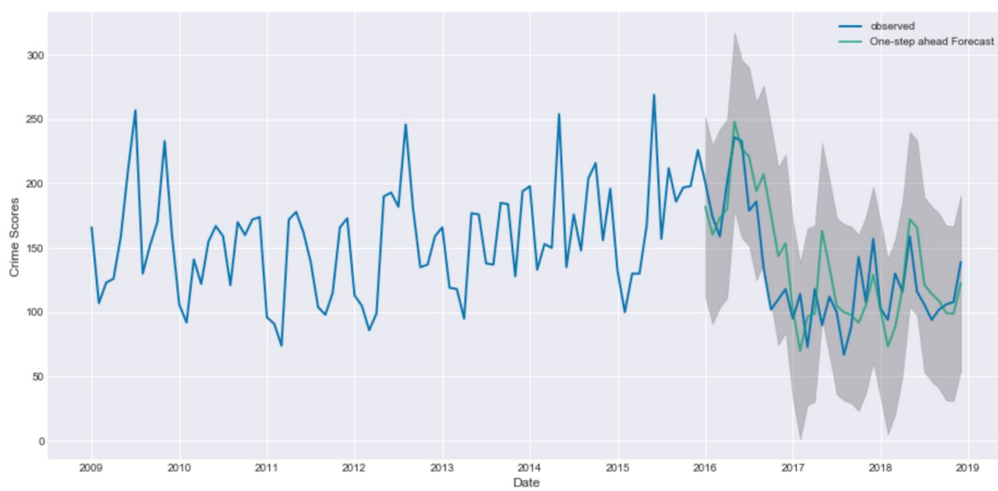


Figure 9. Validation of the model with the one-step ahead forecasting

Conclusions and Discussions

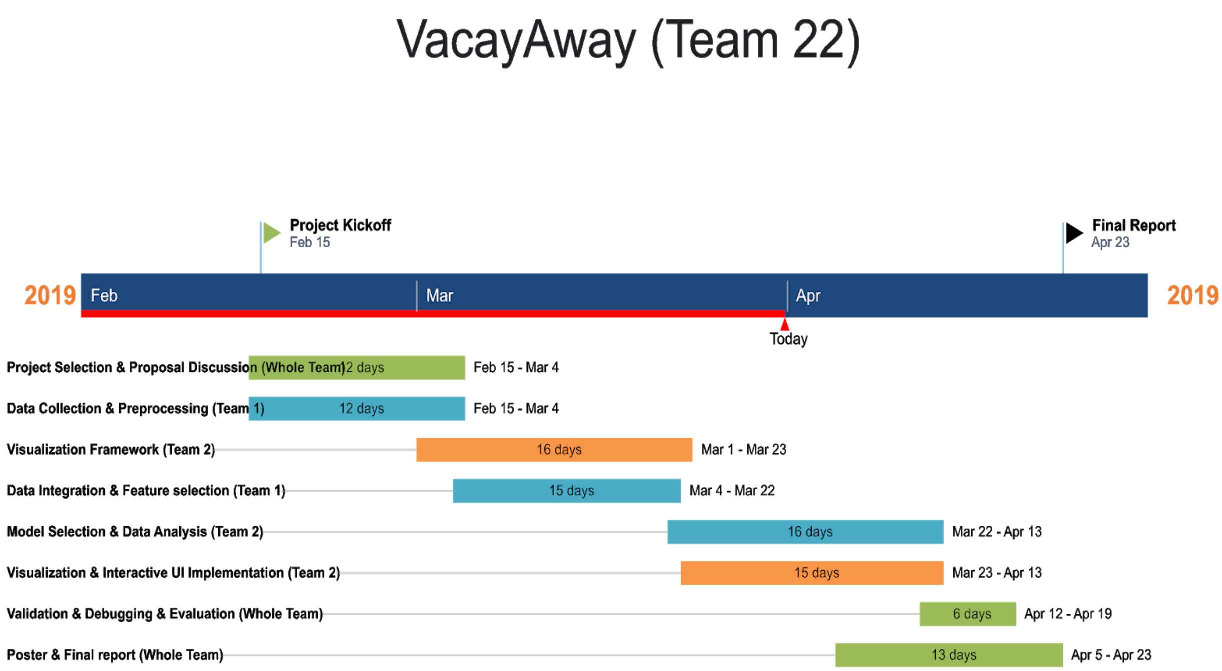
We have successfully built a prototype that does everything we initially set out to do, which was to give rental recommendations to users based on user specified criteria with customized importance weights. There are many limitations with the application that will need to be addressed going forward. One of the primary concerns is the performance of the application, which caused us to only be able to use three categories, but this can be greatly improved by switching from Tableau desktop to Tableau server as this allows our pipeline to integrate seamlessly. Alternatively, we may recommend switching away from Tableau in favor of other visualization tools as it will make the UI more flexible and intuitive to use. The other main concern is expanding the recommendations to areas besides Atlanta. This will require working with the data providers to receive more business and home listing data than we could scrape through their API for free. Another big point of improvement going forward will come from further analyzing the category list to add or remove ones that aren't needed. It may also be helpful to analyze the effects of hyper-weights on listing suggestions by performing A/B testing while varying "hyper-weights" that weight the weighted sum algorithm to determine the Matching Score differently. This would be done by having users go through the application process while varying the hyper-weights used on the backend, and gauging user satisfaction to determine the best set of hyper-weights to use.

Overall, this initial application shows potential as an innovative and helpful tool to augment consumes' experiences with vacation rental platform. There are numerous changes that can be made going forward to improve speed and recommendations to users.

Plan of Action

Team 1: Jun, Sudhanshu, Hoon

Team 2: Stephen, Deepti



***All team members have contributed similar amount of effort.**

References:

[1] Adam Okulicz-Kozaryn. City Life: Rankings (Livability) Versus Perceptions (Satisfaction)

[2] Aliyu, Aliyu & Muhammad, Maryam & Girgiri Bukar, Mohammed & Musa Singhry, Ibrahim. (2016). Impact of Crime On Property Values: Literature Survey and Research Gap Identification. African Scholar Journal of Humanities and Social Science (AJHSS). 4. 68-99.

[3] "Atlanta Police Department." *Atlanta Police Department : Community Policing Programs*, www.atlantapd.org/community/community-policing-programs/youth-programs.

[4] De Nadai, Marco & Lepri, Bruno. (2018). The Economic Value of Neighborhoods: Predicting Real Estate Prices from the Urban Environment. 10.1109/DSAA.2018.00043.

[5] Edward L. Glaeser, Michael Luca, Hyunjin Kim. Nowcasting Gentrification: Using Yelp Data to Quantify Neighborhood Change

[6] Georgios Zervas, Davide Proserpio, and John Byers. A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average

[7] HomeAway.com, Inc. "Homeaway API." *Waikiki Beach Vacation Rentals: Condos & More | HomeAway*, Homeaway, www.homeaway.com/platform/documentation.

[8] Jana Lynott, Rodney Harrell, Shannon Guzman and Brad Gudzinas. The Livability Index 2018: Transforming Communities for All Ages

[9] Jon P. Nelson. Valuing Rural Recreation Amenities: Hedonic Prices for Vacation Rental Houses at Deep Creek Lake, Maryland

[10] Josang, A., & Haller, J. (2007). Dirichlet Reputation Systems. Second International Conference on Availability, Reliability and Security (ARES'07), 112-119.

[11] Leo Breiman. 2001. Random Forests.

[12] Michelle C. Hamilton, John A. Nedza, Patrick Doody, Matthew E. Bates, Nicole L. Bauer, Demetra E. Voyadgis & Cate Fox-Lent (2016) Web-based geospatial multiple criteria decision analysis using open software and standards, International Journal of Geographical Information Science, 30:8, 1667-1686

[13] Nikhil Naik, Jade Philipoom, Ramesh Raskar and César A. Hidalgo.
Streetscore - Predicting the Perceived Safety of One Million Streetscapes. *CVPR Workshop on Web-scale Vision and Social Media* (2014)

[14] Rinner, C. (2007). A geographic visualization approach to multi-criteria evaluation of urban quality of life. International Journal of Geographical Information Science, 21(8), 907-919.

[15] Shaalan, Y., & Zhang, X. (2016). *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9877*, 269-282.

[16] Tang, Tan, & Cheng. (2009). A survey on sentiment detection of reviews. Expert Systems With Applications, 36(7), 10760-10773.

[17] "Yelp Dataset Challenge." *Yelp*, Yelp, www.yelp.com/dataset/challenge.

[18] Yu-Hua Xu, Jin-won Kim, Lori Pennington-Gray. Explore the Spatial Relationship between Airbnb Rental and Crime.

[19] Lawrence Sherman, Peter William Neyroud, Eleanor Neyroud, The Cambridge Crime Harm Index: Measuring Total Harm from Crime Based on Sentencing Guidelines, Policing: A Journal of Policy and Practice, Volume 10, Issue 3, September 2016, Pages 171–183.

[20] “Atlanta Department of City Planning GIS Data” https://dcp-coaplangis.opendata.arcgis.com/datasets/neighborhoods