

COMPETITIONS BETWEEN AIRBNB ACCOMMODATIONS AND HOTELS IN LOS ANGELES

PROPOSAL OF CIT 550 FINAL PROJECT

Team: Charging Bulls

Team Members: Zhongyin Zhang (johnianz@seas.upenn.edu, Github: Johnianzhang),
Xiaofei Huang (xh2342@seas.upenn.edu, Github: xh2342),
Dian Gu (diangu@seas.upenn.edu, Github: DianGu86),
Liang Tang (Email: tangreb@seas.upenn.edu, Github: LiangTang888)



The picture is adapted from <https://www.vecteezy.com/vector-art/20004043-cartoon-angry-charging-bull-mascot>

Motivation of the application/website

Sharing economy as a shift from ownership of goods to temporary rental of them has emerged since 2009 due to global economic recession, cumulative trust of world wide web, and development of online payment system (Dillahun & Malone, 2015). The lodging industry is probably one of the sectors most impacted by the meteoric development of sharing economy (Johnson & Neuhofer, 2017). Airbnb has risen as a dominant player of P2P accommodate with the net worth of \$67.6B in 2023, which is higher than almost any of the top hotel corporations, including Marriott (53.54B), Hilton (45.03B), and Hyatt (13.43B). Accordingly, it is crucial to investigate the competitions between Airbnb accommodations and hotel providers in a district. Thus, our team is motivated to develop a website for both Airbnb hosts and hoteliers who are interested to improve their competitiveness in the local market. Los Angeles is selected as an investigation site for the website. Specifically, with filters, users could check the list of P2P accommodations considering pricing (i.e., price), functional (e.g., room type, size, etc.) and hedonic (e.g., service, value, review scores, etc.) features as well as spatial dependency (i.e., geographic locations of Airbnb offerings and hotels in a specific area).

Data Source

We adopt the combination of secondary dataset and scraping data for the website. The secondary dataset is gained from <http://insideairbnb.com/>. The website collects data from primary cities on the Airbnb website on a regular basis and make data available to the public. The purpose of the website is to collect evidence to support the website owners to win in the debate against Airbnb. The website owners believe that the majority of Airbnb listings in most cities are entire homes, many of which are rented all year around, disrupting housing and communities. We use the dataset of Los Angeles scraped on Mar 7, 2023, which is available on <http://insideairbnb.com/get-the-data/>. The dataset includes four tables titled with listings, calendar, reviews, and neighborhoods. We plan to use three of them (i.e., listings, reviews, neighborhoods). The table of listings is our primary information source with 113,788KB. This table has 42452 rows and 75 attributes. Please note many attributes are not related to our website purpose, which will not be considered in the present project. We also use the table of price for individual Airbnb listings in continuous days. The table has 1,048,576 rows and 7 attributes. The two tables are connected with Airbnb listing id. Besides this insdeairbnb datasets, we plan to scrape the latitude and longitude information of hotels in Los Angeles on www.hotels.com. It is anticipated 300-500 hotels exist in the city. The sample summary statistics is included in the appendix.

Query #1. Property type/size and price

On this webpage, users could select a combination of functional attributes and for each of the functional attributes they can identify the value range they are interested in. The functional attributes include room type (i.e., entire home, private room, etc.), bedroom number, bed

number, and accommodation size (i.e., number of guests that can be accommodated). Price is also incorporated into the combination of functional attributes on the webpage, since price is the crucial consideration for Airbnb listings. The incorporation of price increases the applicability of the search on the webpage for Airbnb hosts.

Query #2. Information of hosts and price

On this webpage, users could use a customized benchmark of evaluating hosts to filter the Airbnb offerings. For each of the host-related attributes they can identify the value range they are interested in. The host-related attributes include host response time (e.g., within an hour), host response rate, host acceptance rate, whether the host is superhost or not, number of listings the host has. Any host has his/her unique host_id. Similar with query #1, price is also incorporated into the combination of host-related attributes on the webpage.

Query #3 Reviews and price

Airbnb provides review score for six categories, including accuracy, cleanliness, check-in, communication, location, and value as well as the overall review score. On this webpage, users could customize their filters to identify the Airbnb listings which satisfy their needs. Similar with query #1, price is also incorporated into the combination of review scores on the webpage.

Query #4 Spatial analysis of Airbnb offerings and hotels in a district

We combine the geographic information (i.e., latitude and longitude) of the insiderairbnb dataset and that scraped from www.hotels.com. We will further use GeoDa or other similar apps to prepare the GIS map (in the form of picture) for Los Angeles. On the map Airbnb accommodations and hotels are marked in different colors (i.e., red vs. green). We will separate Los Angeles into 4-5 districts. Through clicking each of the districts, users could view enlarged map for that specific area.

Query #5 Top Airbnb offerings upon overall review score in different districts

This query incorporates the information from query #3 and query #4. In each of the different districts, users could gain the top 10 Airbnb offerings upon the overall review score. One query is needed corresponding to each of the districts.

Reference

- Dillahunt, T.R., Malone, A.R., 2015. The promise of the sharing economy among disadvantaged communities. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, New York, NY, pp. 2285-2294.
- Johnson, A., Neuhofer, B., 2017. Airbnb - an exploration of value co-creation experiences in Jamaica. International Journal of Contemporary Hospitality Management 29(90), 2361-2376.

Sample Summary Statistics

Attribute	Mean	Deviation	Max	Min
id	N/A	N/A	N/A	N/A
host_id	N/A	N/A	N/A	N/A
host_response_time	0.31	0.6697	3	0
host_response_rate	0.96	0.1210	1	0
host_acceptance_rate	0.96	0.2034	1	0
host_is_superhost	0.34	0.4734	1	0
host_total_listings_count	169.61	760.3025	8316	1
room_type	0.34	0.5717	3	0
accommodates	3.91	2.7312	16	0
bedrooms	1.79	1.1331	24	1
beds	2.16	1.5839	50	1
price	283.33	1081.1467	99999	0
review_scores_rating	4.70	0.5325	5	0
review_scores_accuracy	4.77	0.3945	5	0
review_scores_cleanliness	4.71	0.4203	5	0
review_scores_checkin	4.84	0.3518	5	0
review_scores_communication	4.83	0.3690	5	0
review_scores_location	4.79	0.3442	5	0
review_scores_value	4.67	0.4227	5	0