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 (Dillahunt & 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).

Obligatory Features of Website

We will include the following obligatory features:

1) search (i.e., property features webpage, property features webpage, host information webpage, top listings by reviews webpage)
2) illustration of tables (i.e., top listings by reviews webpage)
3) illustration of graphs (i.e., property features webpage, host information webpage)
4) illustration of maps (i.e., Airbnb vs. hotels webpage)
5) illustration of text and pictures (i.e., home page)

Optional Features of Website

We plan to use @media queries to create a responsive website. Media queries can be used to check many things, such as width and height of the viewport, width and height of the device, orientation (is the tablet/phone in landscape or portrait mode?), and resolution. The purpose of using media queries is to deliver a tailored style sheet (responsive web design) to desktops, laptops, tablets, and mobile phones.

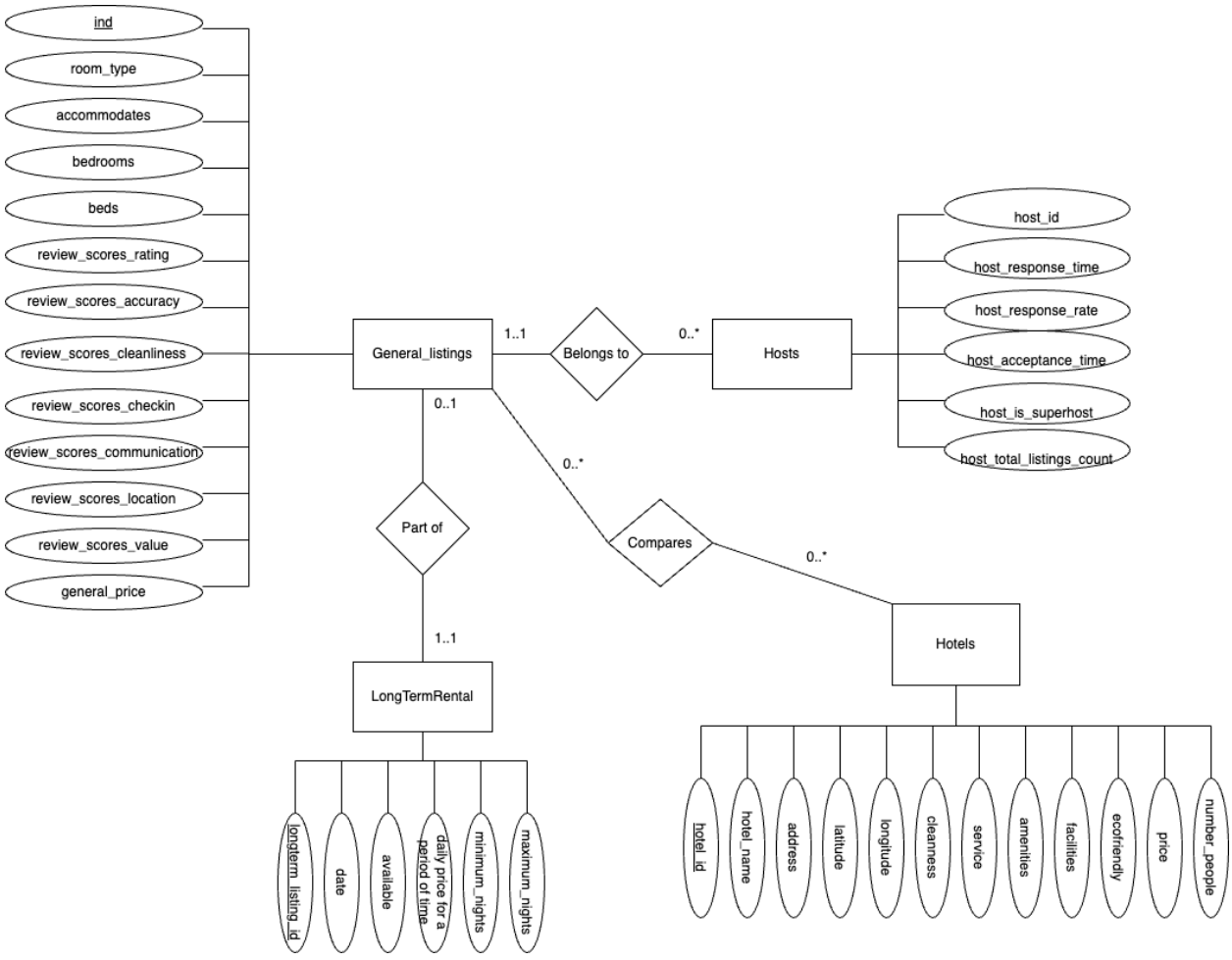
List of Pages:

1. *Front page*: On this page, users could see a short paragraph about the purpose of the website.
2. *Property type webpage*: On this webpage, users could search property features by room type (entire home/apt, private room, hotel room, shared room), No. of guest accommodated (1-2, 3-4, 5-8, 9-16), No. of bedrooms (1-2, 3-4, 5-8, 9-24), and No. of beds (1-2, 3-4, 5-8, 9-50). A textual summary of the search result and a graph of pricing ranges based on the query are demonstrated.
3. *Host information webpage*: On this webpage, users could search host information by host response time (within one hour, within a few hours, within a day, with a few days or

more), response rate (0-0.50, 0.51-0.79, 0.80-1.00), acceptance rate (0-0.79, 0.80-0.89, 0.90-1.00), superhost or not (Yes, No), total listing count (only 1, 2-5, 6-10, 11 or more).

4. *Airbnb vs. hotels webpage*: On this webpage, there is a drop-down menu which users could select Los Angeles (as a general big area). And three districts of Los Angeles also could be clicked (i.e., City of Los Angeles, other cities, unincorporated areas). It is expected that users could see the enlarged map of the specific area selected. And a table with comparison between hotels and Airbnb listings are included.
5. *Top listings by reviews*: On this webpage users could search top listings by reviews by review type (overall review, review on accuracy, review on cleanness, review on check-in, review on communication, review on location, and review on value) and No. of listings (50, 100).

ER Diagram



Relational Schema

Host_information (<u>host_id</u>, host_response_time, host_response_rate, host_acceptance_rate, host_is_superhost, host_total_listings_count)

PRIMARY KEY host_id

General_listings (ind, host_id, room_type, accommodates, bedrooms, beds, review_scores_rating, review_scores_accuracy, review_scores_cleanliness, review_scores_checkin, review_scores_communication, review_scores_location, review_scores_value, general_price)

PRIMARY KEY ind

FOREIGN KEY host_id REFERENCES Host_information(host_id)

LongTermRental (longterm_listing_id, ind, date, available, price, minimum_nights, maximum_nights)

PRIMARY KEY longterm_listing_id

FOREIGN KEY ind REFERENCES General_listings(id)

Hotels (hotel_id, hotel_name, address, latitude, longitude, cleanness, service, amenities, facilities, ecofriendly, price, number_people)

PRIMARY KEY hotel_id

SQL DDL for creating the database

```
CREATE DATABASE AirbnbVSHotels;

USE AirbnbVSHotels;
CREATE TABLE Host_information (
   host_id BigInt,
   host_response_time int,
   host_response_rate Float,
   host_acceptance_rate Float,
   host_is_superhost int,
   host_total_listings_count  int,
   PRIMARY KEY (host_id));

CREATE TABLE General_listings (
   ind BigInt,
   host_id BigInt,
   room_type int,
   accommodates int,
   bedrooms int,
   beds int,
   review_scores_rating Float,
   review_scores_accuracy Float,
   review_scores_cleanliness Float,
   review_scores_checkin Float,
   review_scores_communication Float,
```

```
    review_scores_location Float,
    review_scores_value Float,
    general_price int,
    PRIMARY KEY (ind)
    FOREIGN KEY host_ind REFERENCES Host_information (host_ind));


CREATE TABLE LongTermRental (
    longterm_listing_id BigInt,
    ind BigInt,
    date Date,
    available int,
    price int,
    minimum_nights int,
    maximum_nights int,
    PRIMARY KEY (longterm_listing_id),
    FOREIGN KEY ind REFERENCES General_listings (ind));

CREATE TABLE hotels(
     hotel_id int,
     hotel_name varchar(255),
     address varchar(255),
     latitude float,
     longitude float,
     cleanness float,
     service float,
     amenities float,
     facilities float,
     ecofriendly float,
     price int,
     number_people int,
     PRIMARY KEY (hotel_id));
```

Data Cleaning and Pre-process

We used Pandas to clean and pre-processing data. We listed the primary work as follows:

1) Replace missing data. We use mean to replace missing data for the categories of reviews and property features, and price in the dataset of listings. For reviews (i.e., review_scores_rating, review_scores_accuracy, review_scores_cleanliness, review_scores_checkin, review_scores_communication, review_scores_location, review_scores_value), we round up the results to 2 decimal places. For room-types, accommodates, bedrooms, beds, and price, we round to the nearest integer. However, for host information (i.e., host_response_rate, host_acceptance_rate), we intentionally kept the null values according to TA's guidance on ed, in order to practice the complex query.

2) Replace categorical variables with indicators. We did so for host_response_time and room_type. Host_response_time is classified into 1) within an hour(0), 2) within a few hours (1), 3) within a day (2), 4) a few days or more (3). Room_type: 1) entire home/apt(0), 2) private room(1), 3) hotel room(2), 4) shared room(3).
3) Prepare three tables named with general_listings, Host_informaton, and LongtermRental.

We also collect geographic information from hotels in Los Angeles. The hotels without latitude and longitude are removed.

List of Technologies You will Use

Front-end: React

Back-end: NodeJs Express

DataBase: MySQL hosted on AWS

Data Processing: Pandas

Description of Team Member Responsibility

Rebecca: project overseeing, database, data cleaning, doc writing, website content design

Zhongyin: Front-end, connecting test between front-end and server

Xiaofei: server, website aesthetic design, github setup and maintainence, connecting test between front-end and server

Dian: data scraping of hotels, map design, assistant of Zhongyin on front-end, ppt and design

Appendix Information

In Appendix, we include the codebook for the attributes as well as website demo. The website demo demonstrates a general idea of the information and aesthetic design of our website. Modifications of the website design may occur with the progress of the project.

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.

Appendix A. Codebook

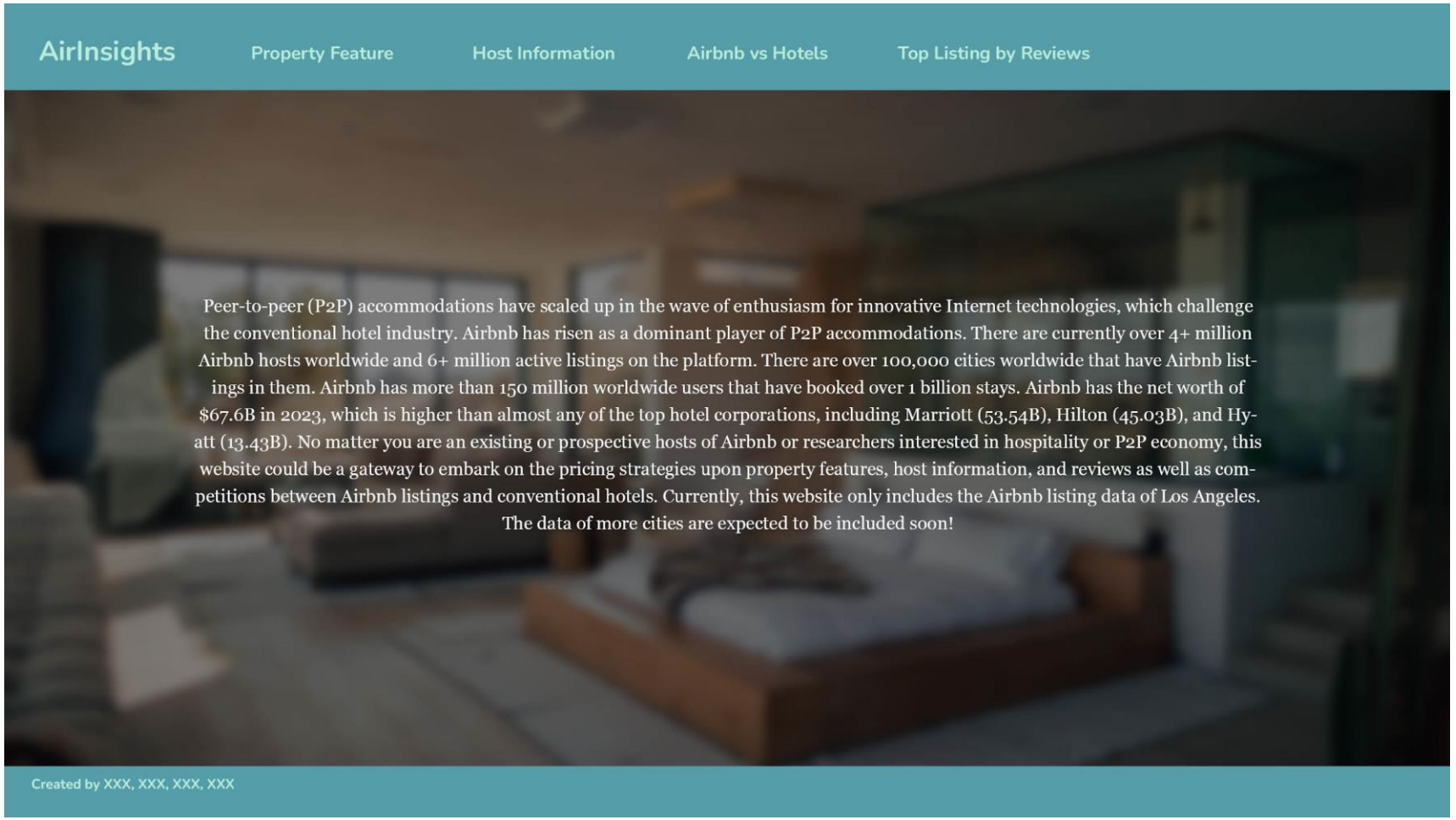
Attribute	Mean	Deviation	Max	Min	Handle Null	Note	
ind	N/A	N/A	N/A	N/A	N/A	key	
host_id	N/A	N/A	N/A	N/A	N/A	A host_id may correspond multiple inds	
host_response_time	0.31	0.6697	3	0	mean	within an hour	0
						within a few hours	1
						within a day	2
						a few days or more	3
						round(0)	
host_response_rate	0.96	0.1210	1	0	mean	Round(0.00)	
host_acceptance_rate	0.96	0.2034	1	0	mean	round(0)	
host_is_superhost	0.34	0.4734	1	0	mean	f	0
						t	1
						Round(0)	
host_total_listings_count	169.61	760.3025	8316	1	mean	round(0)	
neighbourhood_cleansed							
neighbourhood_group_cleansed							
latitude							
longitude							
room_type	0.34	0.5717	3	0	mean	Entire home/apt	0
						Private room	1
						Hotel room	2
						Shared room	3
						Round(0)	
accommodates	3.91	2.7312	16	0	mean	Round(0)	
bedrooms	1.79	1.1331	24	1	mean	Round(0)	
beds	2.16	1.5839	50	1	mean	Round(0)	
price	283.33	1081.1467	99999	0	mean	Round(0)	

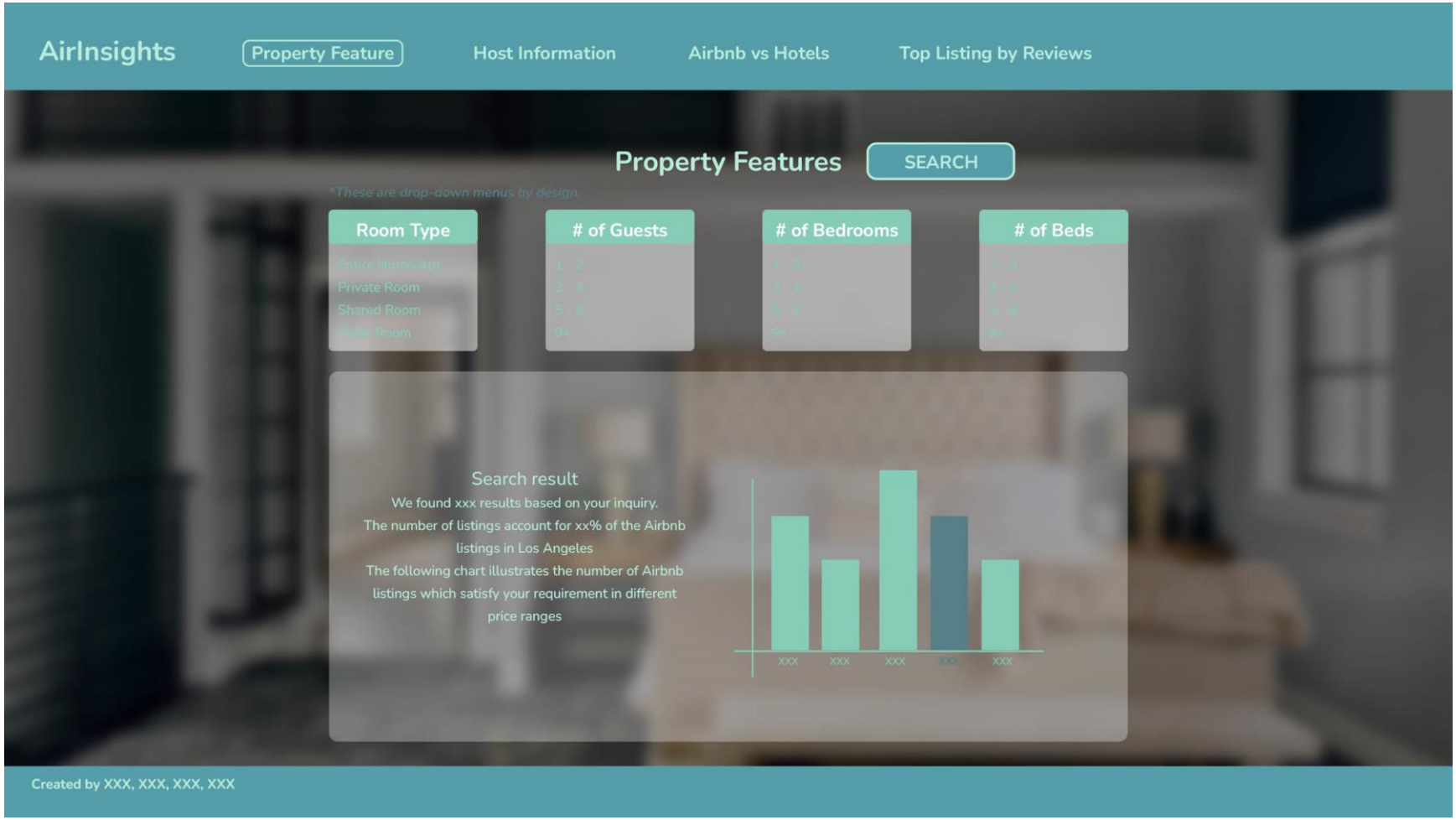
review_scores_rating	4.70	0.5325	5	0	mean	Round(0.00)
review_scores_accuracy	4.77	0.3945	5	0	mean	Round(0.00)
review_scores_cleanliness	4.71	0.4203	5	0	mean	Round(0.00)
review_scores_checkin	4.84	0.3518	5	0	mean	Round(0.00)
review_scores_communication	4.83	0.3690	5	0	mean	Round(0.00)
review_scores_location	4.79	0.3442	5	0	mean	Round(0.00)
review_scores_value	4.67	0.4227	5	0	mean	Round(0.00)

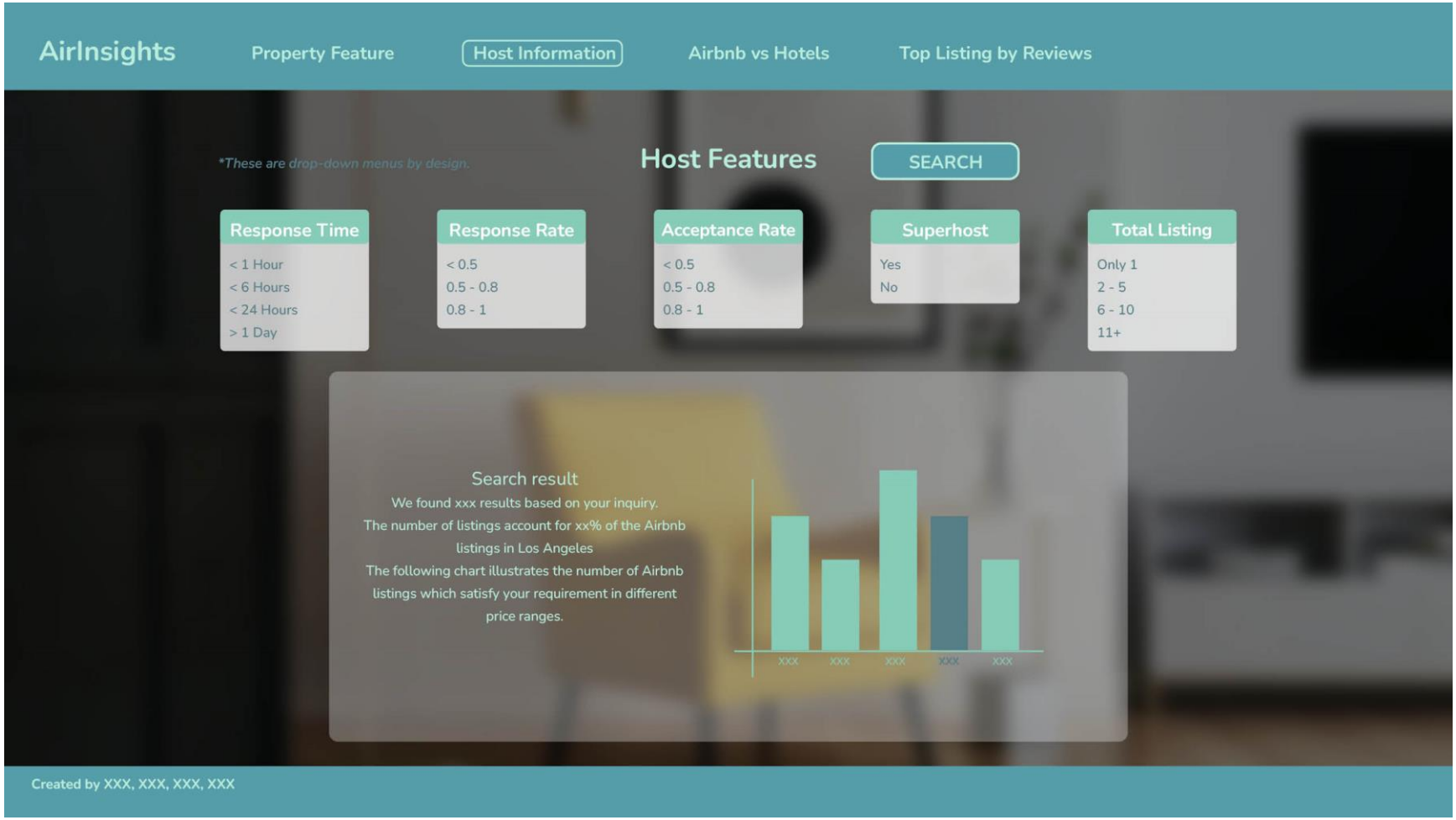
Codebook for Hotels.com Data

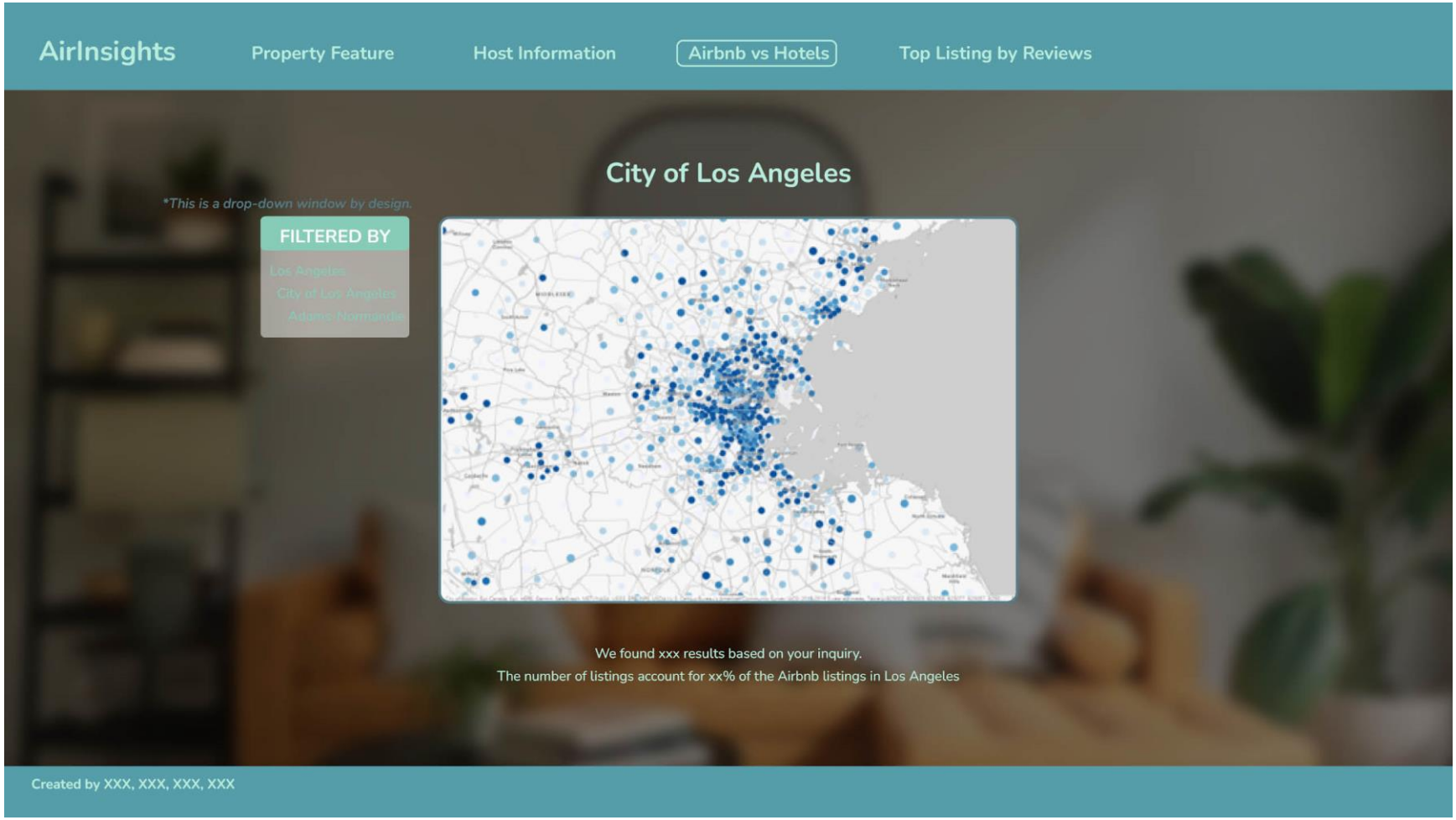
Attribute	Mean	Min	Max	Standard Deviation	Count Valid Values	Note
hotel_id	N/A	N/A	N/A	N/A	N/A	
hotel_name	N/A	N/A	N/A	N/A	N/A	
address	N/A	N/A	N/A	N/A	N/A	
latitude	N/A	N/A	N/A	N/A	N/A	
longitude	N/A	N/A	N/A	N/A	N/A	
cleanness	8.13	3.70	9.90	1.05	2833	out of 10
service	8.19	4.30	9.90	0.85	2833	out of 10
amenities	7.49	2.00	9.40	1.11	2833	out of 10
facilities	7.80	3.50	9.90	1.11	2815	out of 10
ecofriendly	7.78	3.20	10.00	1.06	2790	out of 10
price	272.26	26.00	6000.00	247.86	2850	usd
number_people	3.16	2.00	12.00	1.14	2850	

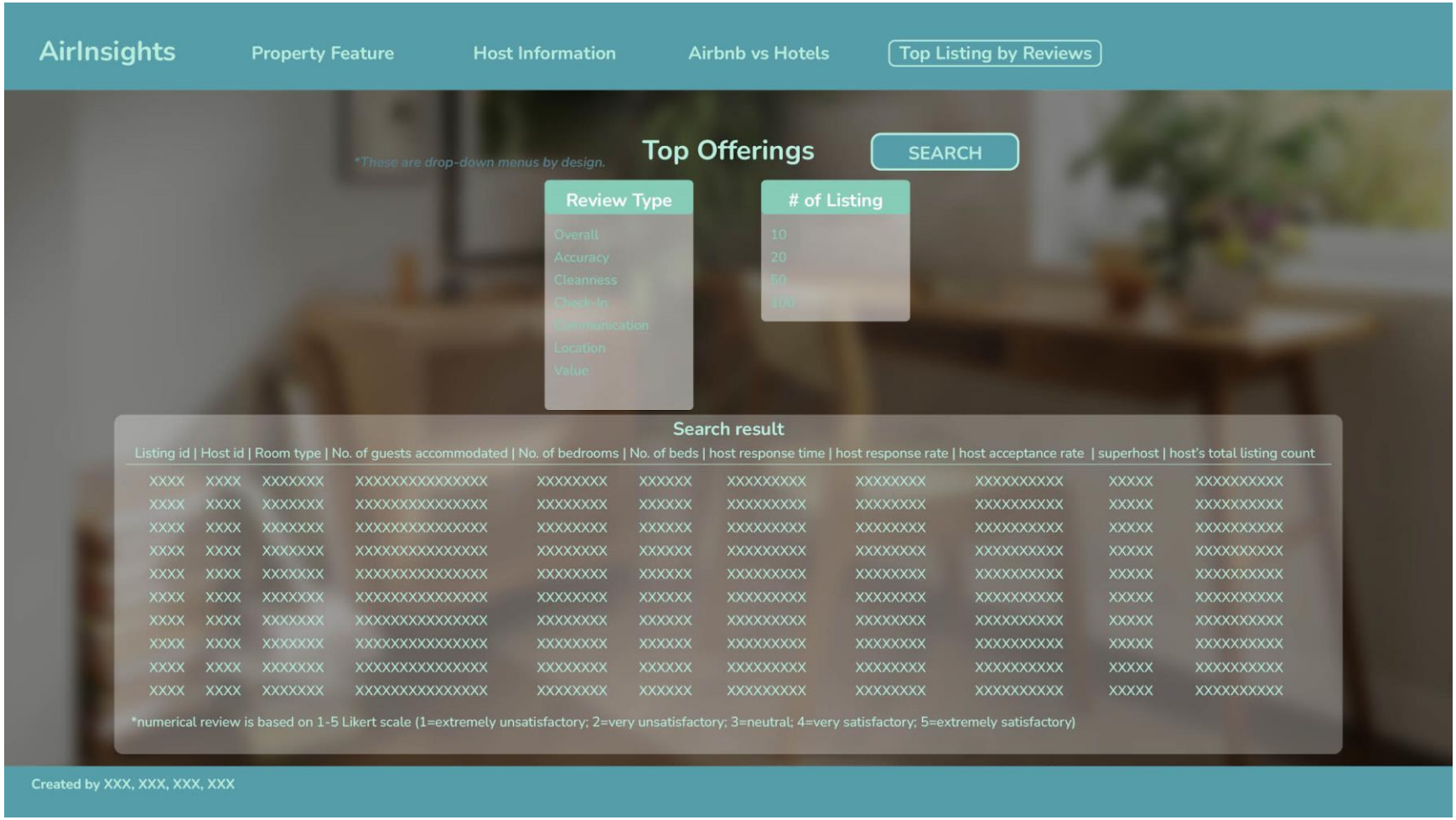
Appendix B. Website Demo

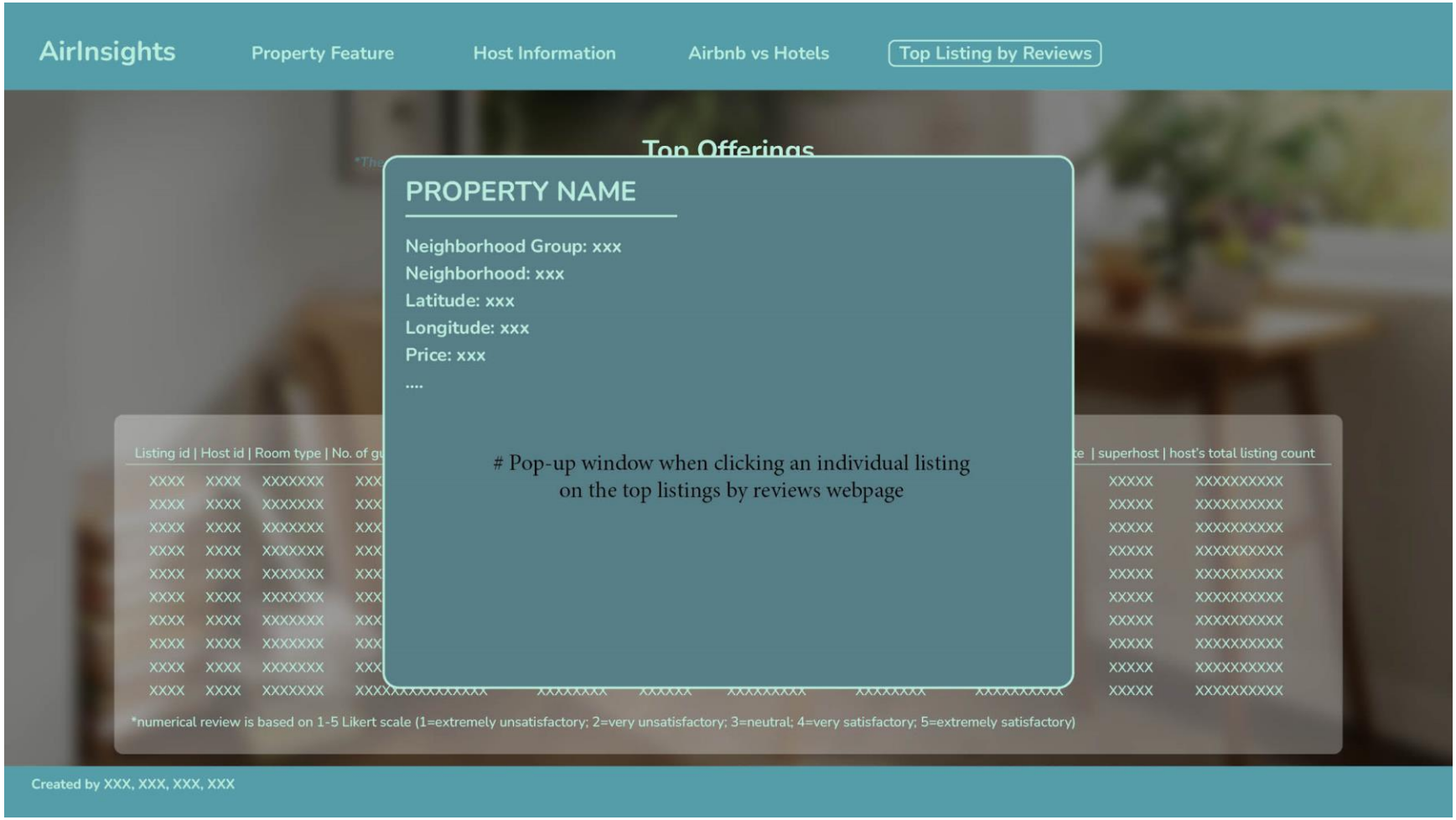



AirInsights
Property Feature
Host Information
Airbnb vs Hotels
Top Listing by Reviews
Property Features
SEARCH
*These are drop-down menus by design
Room Type
Entire Home/Apt
Private Room
Shared Room
Hotel Room
# of Guests
1 - 2
3 - 4
5 - 8
9+
# of Bedrooms
1 - 2
3 - 4
5 - 8
9+
# of Beds
1 - 2
3 - 4
5 - 8
9+
Search result
We found xxx results based on your inquiry.
The number of listings account for xx% of the Airbnb listings in Los Angeles
The following chart illustrates the number of Airbnb listings which satisfy your requirement in different price ranges
XXX
XXX
XXX
XXX
XXX
Created by XXX, XXX, XXX, XXX


AirInsights
Property Feature
Host Information
Airbnb vs Hotels
Top Listing by Reviews
*These are drop-down menus by design.
Host Features
SEARCH
Response Time
< 1 Hour
< 6 Hours
< 24 Hours
> 1 Day
Response Rate
< 0.5
0.5 - 0.8
0.8 - 1
Acceptance Rate
< 0.5
0.5 - 0.8
0.8 - 1
Superhost
Yes
No
Total Listing
Only 1
2 - 5
6 - 10
11+
Search result
We found xxx results based on your inquiry.
The number of listings account for xx% of the Airbnb listings in Los Angeles
The following chart illustrates the number of Airbnb listings which satisfy your requirement in different price ranges.
XXX
XXX
XXX
XXX
XXX
Created by XXX, XXX, XXX, XXX


AirInsights
Property Feature
Host Information
Airbnb vs Hotels
Top Listing by Reviews
City of Los Angeles
*This is a drop-down window by design.
FILTERED BY
Los Angeles
City of Los Angeles
Adams Normandie
We found xxx results based on your inquiry.
The number of listings account for xx% of the Airbnb listings in Los Angeles
Created by XXX, XXX, XXX, XXX


AirInsights
Property Feature
Host Information
Airbnb vs Hotels
Top Listing by Reviews
*These are drop-down menus by design.
Top Offerings
SEARCH
Review Type
Overall
Accuracy
Cleanness
Check-In
Communication
Location
Value
# of Listing
10
20
50
100
Search result
Listing id | Host id | Room type | No. of guests accommodated | No. of bedrooms | No. of beds | host response time | host response rate | host acceptance rate | superhost | host's total listing count
*numerical review is based on 1-5 Likert scale (1=extremely unsatisfactory; 2=very unsatisfactory; 3=neutral; 4=very satisfactory; 5=extremely satisfactory)
Created by XXX, XXX, XXX, XXX


AirInsights
Property Feature
Host Information
Airbnb vs Hotels
Top Listing by Reviews
Top Offerings
PROPERTY NAME
Neighborhood Group: xxx
Neighborhood: xxx
Latitude: xxx
Longitude: xxx
Price: xxx
....
Pop-up window when clicking an individual listing on the top listings by reviews webpage
Listing id | Host id | Room type | No. of gu
te | superhost | host's total listing count
*numerical review is based on 1-5 Likert scale (1=extremely unsatisfactory; 2=very unsatisfactory; 3=neutral; 4=very satisfactory; 5=extremely satisfactory)
Created by XXX, XXX, XXX, XXX