TEAM 4 DELIVERABLE 1 : PROPOSAL

"Success in Hosting"

I. Problem Statement & Stakeholders

Airbnb, established in 2008, is one of the most popular online housing platforms offering cheaper vacation rental alternatives for travelers as compared to the more expensive hotel accommodations. Despite being hardly hit because of the pandemic travel restrictions, Airbnb continues to be a valued company, and even went public last December 2020. As economies reopen, we believe that the demand for travel will flourish and will bring in a great opportunity to generate revenue in rental hosting. To maximize potential revenue, Airbnb hosts can make use of attributes that drive booking price and improve these features for the benefit of travelers.

Hosts pay commissions to Airbnb when customers book accommodations via Airbnb's online platform. Customers rely mainly on Airbnb's past customer ratings when choosing a place to stay. The ratings therefore reflect overall customer satisfaction. To leverage this, identifying rental property features that significantly affect customers' satisfaction rating during their visits could help the host improve their service and therefore generate more revenue. Hosts will have the opportunity to explore potential areas for improvement, shape their reputation and increase satisfaction of the customers. This also drives higher commissions for Airbnb, opening a potential goal of increasing revenue by 5%.

II. Data-set review and Methodology

The Airbnb data set shown in Table 1 covers Airbnb house/room ratings between August 2009 and September 2020 in the Greater Toronto Area. Data is sourced from Kaggle[1]. Since our objective is to identify the major factors that affect customer rating, we plan to implement multiple regression as our main machine learning algorithm. A clustering analysis will be implemented if the multiple regression provides results that require further explanation.

We will make use of 73 explanatory variables including but not limited to - room type, number of accommodation and host response rate. A data dictionary provided in Table 1 consists of the variables used, with corresponding descriptions and data type. For example, 'host_response_time' refers to the time to respond to customer booking inquiry, ranging from 1 hour to a few days. The dependent variables to be used are (1) price and (2) customer rating ('review_scores_rating').

In terms of data cleaning, we will first select the features that are relevant for our machine learning algorithm through linear regression. We will then proceed to the standard data exploration for selection of relevant features and data cleaning process such as removal of NaN values, feature engineering and data normalization. After data cleaning, we will build two multiple regression models to identify major factors that will predict price and predict customer rating.

Appendix

Table 1. Airbnb Data Set with variable description and data type

Number of Variables	Column Name	Description	Feature type(e.g., Numeric, String)
	id	Unique listing id (Primary Key)	Numeric
1	listing_url	Link to the rental property listing on Airbnb	String
2	scrape_id	Identifier for scraper	Numeric
3	last_scraped	Last date listing was scraped	Numeric
4	name	Title of Posting	String
5	description	Description of Posting	String
6	neighborhood_overview	Overview of neighborhood	String
7	picture_url	Link to the main vacation rental listing image on Airbnb	Image
8	host_id	Unique id for each host	Numeric
9	host_url	Link to the host profile on Airbnb	String
10	host_name	Host Name	String
11	host_since	Date an individual became a host	Numeric
12	host_location	Location of Listing	String
13	host_about	Description of host - relationship status, interests and hobbies	String
14	host_response_time	Time to respond to customer booking inquiry; ranges from 1hour to few days	Numeric
15	host_response_rate	Ranges from 0% to 100% for reply to booking inquiries	Numeric
16	host_acceptance_rate	Ranges from 0% to 100% response for acceptance of booking	Numeric
17	host_is_superhost	Is either t(true) or f(false)	Boolean
18	host_thumbnail_url	Host thumbnail URL	Image
19	host_picture_url	Host Picture URL	Image
20	host_neighbourhood	Description of neighborhood listing	String
21	host_listings_count	Current number of host listings	Numeric
22	host_total_listings_count	Total number of listings made by host	Numeric
23	host_verifications	Identifies how the host has completed the identity verification process	String

24	host_has_profile_pic	Host Profile Pic	Image
25	host_identity_verified	Identifies if the host has completed the verification process by indicating true or false	Boolean
26	neighbourhood	Specific location of Toronto area	String
27	neighbourhood_cleansed	Represents one of boroughs in Toronto in which a listing resides	String
28	neighbourhood_group_cleansed	No values (N/A)	N/A
29	latitude	The angular distance of a location or object north or south of the Earth's celestial equator	String
30	longitude	The angular distance of a location or object east or west of the meridian	String
31	property_type	Type of property (e.g., Entire house)	String
32	room_type	Specific type of room (e.g., Entire home/apt)	String
33	accommodates	How many people can stay (e.g., 6)	Numeric
34	bathrooms	No values, NA	N/A
35	bathrooms_text	Number of bathrooms in property	Numeric
36	bedrooms	Number of bedrooms in property	Numeric
37	beds	Number of beds in property	Numeric
38	amenities	List of amenities such as shampoo available in property	String
39	price	Price of stay per night	Numeric
40	minimum_nights	Minimum nights can be booked by same individual	Numeric
41	maximum_nights	Maximum nights can be booked by same individual	Numeric
42	minimum_minimum_nights	Same values as Minimum nights	Numeric
43	maximum_minimum_nights	Same values as Maximum_nights	Numeric
44	minimum_maximum_nights	Same values as Maximum_nights	Numeric
45	maximum_maximum_nights	Same values as Maximum_nights	Numeric
46	minimum_nights_avg_ntm	Average minimum nights can be booked by same individual	Numeric
47	maximum_nights_avg_ntm	Average maximum nights can be booked by same individual	Numeric
48	calendar_updated	No values (N/A)	N/A
49	has_availability	True or False if listing is available	Boolean
50	availability_30	Availability of Property in 30 days	Numeric
51	availability_60	Availability of Property in 60 days	Numeric
52	availability_90	Availability of Property in 90 days	Numeric

53	availability_365	Availability of Property in 365 days	Numeric
54	calendar_last_scraped	Date last scraped	Numeric
55	number_of_reviews	Total number of reviews that a listing has received from customers	Numeric
56	number_of_reviews_ltm	The number of reviews that a listing has received last twelve month	Numeric
57	number_of_reviews_l30d	The number of reviews that a listing has received per 130 days	Numeric
58	first_review	Date of first review by customer	Numeric
59	last_review	Date of last review by customer	Numeric
60	review_scores_rating	Customer-provided score rating (0% to 100%); A customer-provided review score attributed to a listing based on overall experience and satisfaction	Numeric
61	review_scores_accuracy	Accuracy of review scores (0 to 10)	Numeric
62	review_scores_cleanliness	Cleanliness score (0 to 10)	Numeric
63	review_scores_checkin	Over-all check in score (0 to 10)	Numeric
64	review_scores_communication	Score on communication with host (0 to 10)	Numeric
65	review_scores_location	Score on location based on factors such as nearby transportation, noise level (0 to 10)	Numeric
66	review_scores_value	Over- all value/quality/experience (0 to 10)	Numeric
67	license	No values (N/A)	N/A
68	instant_bookable	True or False if customer can instantly book	BOolean
69	calculated_host_listings_count	Calculated number of listings by host	Numeric
70	calculated_host_listings_count_entire_homes	Number of host listings which are entire homes	Numeric
71	calculated_host_listings_count_private_rooms	Number of host listings which are private rooms	Numeric
72	calculated_host_listings_count_shared_rooms	Number of host listings which are shared homes	Numeric
73	reviews_per_month	Number of Customer reviews of the host accommodation or accommodations per month	Numeric

[1]https://www.kaggle.com/robinkongninglo/toronto-airbnb-dataset

Other Reference :

https://www.phocuswire.com/airbnb-q1-2021-earnings