

Business Analytics – BUSI 5100

Project Proposal **[Airbnb Data Analytics]**

Sweta Patel

Swetha Sridhar

Qiaotong Li

Peter El Khoury

15/10/2023

1 Introduction

The subject of this study is one of the most popular online service businesses: Airbnb. The name is an abbreviation of “Air Bed and Breakfast” and was founded by two friends in 2008. The company provides a digital platform where a homeowner and a person looking for a short-term rental or stay can safely exchange goods and services. The company experienced a slow down during the pandemic due to travel restrictions and stay at home mandates all over the world, but it recovered steadily after 2022 and revenues started increasing quarterly.

As people face increased challenges during the current economic downturn, many Airbnb operators are seeing decline in their revenues, especially with high interest rates that make the upkeep of a property quite expensive, that coupled with the decrease in the purchasing power of the market. Therefore, it is crucial for Airbnb to take into account the elements affecting the customer's choice of properties (price being the significant one) and their satisfaction levels in order to keep more of its current customers and attract new ones thereby maximizing revenue for the hosts and the platform itself.

This study aims to perform data analysis on raw data obtained from the website [insideairbnb.com](https://www.airbnb.com) which uses only the public information available from the official website of [Airbnb.com](https://www.airbnb.com) excluding any private information. The dataset includes a lot of data about Airbnb listings, where each listing includes specific information about hosts, including the type of accommodations and services they provide, reviews, as well as the listing's prices in the Toronto region for the past one year to better understand what makes a successful listing, what people are valuing now versus 2 years ago, and how to help homeowners attract more customers. This will help Airbnb retain both of their customers, where homeowners are seeing returns on their investments and the customers are getting exactly what they want (at a reasonable price).

The results of this study will give insight into the ideal listing in terms of host performance, amenities, the asking price, the type of accommodation, availability, location, customer reviews and their review scores for each listing in order to gain insights about customers' preferences. This will help Airbnb make

some of the informed business decisions such price optimization, introducing flexibility in cancellation policies to adhere to uncertainties, showcasing user reviews and ratings to help customers make confident decisions, improving search rankings and host performance, expanding in new markets and promoting.

2 Available Data

The data available is raw and unnecessary data needs to be filtered out based on the needs of this analysis. There are three separate data sets available:

- Listings, which is information on the host, the accommodation, including customer reviews. The number of observations in this dataset is 18000.
- Reviews, which is a compilation of all reviews the customers have written. This will be challenging to analyze since it's exactly what the customers have written. The number of observations in this dataset is 4,82,000.
- Calendar, which shows the date each listings was available and at what price, as well as the limitations in terms of duration of stay. The number of observations in this dataset is 10,48,000

Table 1 below shows the data available for each listing:

Host Information

Field	Data type	Description
host_name	Text	Name of the host. Usually just the first name(s).
host_since	Date	The date the host/user was created. For hosts that are Airbnb guests this could be the date they registered as a guest.
host_location	Text	The host's self-reported location.
host_about	Text	Description about the host.
host_response_time	Character	How long it takes the host to respond.
host_response_rate	Character	How many times did the host respond in percentage.
host_acceptance_rate		That rate at which a host accepts booking requests.
host_is_superhost	Boolean [t=true; f=false]	Whether the host is a superhost or not.
host_neighbourhood	Text	The neighbourhood name.
host_listings_count	Text	The number of listings the host has.
host_total_listings_count	Text	The number of listings the host has.
host_has_profile_pic	boolean [t=true; f=false]	Whether the host has a profile picture or not.
host_identity_verified	boolean [t=true; f=false]	Whether the host's identity has been verified through Airbnb's platform.

Listing Information

Field	Data Type	Description
listing_id	Integer	Airbnb's unique identifier for the listing
latitude	Numeric	Uses the World Geodetic System (WGS84) projection for latitude and longitude.
longitude	Numeric	Uses the World Geodetic System (WGS84) projection for latitude and longitude.
property_type	Text	Self-selected property type. Hotels and Bed and Breakfasts are described as such by their hosts in this field
room_type	Text	<p>[Entire home/apt Private room Shared room Hotel] All homes are grouped into the following three-room types: Entire place Private room Shared room Entire place Entire places are best if you're seeking a home away from home. With an entire place, you'll have the whole space to yourself. This usually includes a bedroom, a bathroom, a kitchen, and a separate, dedicated entrance. Hosts should note in the description if they'll be on the property or not (ex: "Host occupies first floor of the home"), and provide further details on the listing.</p> <p>Private rooms Private rooms are great for when you prefer a little privacy, and still value a local connection. When you book a private room, you'll have your own private room for sleeping and may share some spaces with others. You might need to walk through indoor spaces that another host or guest may occupy to get to your room.</p> <p>Shared rooms Shared rooms are for when you don't mind sharing a space with others. When you book a shared room, you'll be sleeping in a space that is shared with others and share the entire space with other people. Shared rooms are popular among flexible travelers looking for new friends and budget-friendly stays.</p>
accommodates	Integer	The maximum capacity of the listing
bathrooms	Numeric	The number of bathrooms in the listing
bathrooms_text	String	The number of bathrooms in the listing. On the Airbnb web-site, the bathrooms field has evolved from a number to a textual description. For older scrapes, bathrooms is used.
bedrooms	Integer	The number of bedrooms
beds	Integer	The number of bed(s)
amenities	Json	
price	Currency	daily price in local currency
has_availability	Boolean	[t=true; f=false]
availability_30	Integer	availability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing

		may not be available because it has been booked by a guest or blocked by the host.
availability_60	Integer	availability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
availability_90	Integer	availability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
availability_365	Integer	availability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.

Reviews Information

Field	Data Type	Description
number_of_reviews	Integer	The number of reviews the listing has
number_of_reviews_ltm	Integer	The number of reviews the listing has (in the last 12 months)
number_of_reviews_l30d	Integer	The number of reviews the listing has (in the last 30 days)
first_review	Date	The date of the first/oldest review
last_review	Date	The date of the last/newest review
review_scores_rating		
review_scores_accuracy		
review_scores_cleanliness		
review_scores_checkin		
review_scores_communication		
review_scores_location		
review_scores_value		

Table 2 below shows the data available for reviews:

Field	Data Type	Description
listing_id	integer	Airbnb's unique identifier for the listing
id		
date	datetime	The date on which review is given
reviewer_id		Id of reviewer
reviewer_name	text	Name of reviewer
comments	text	Comments of the reviewer

Table 3 below shows the data available for calendar:

Field	Data Type	Description
listing_id		
date	datetime	The date in the listing's calendar
available	boolean	Whether the date is available for a booking

price	currency	The price listed for the day
adjusted_price		
minimum_nights	integer	Minimum nights for a booking made on this day
maximum_nights	integer	Minimum nights for a booking made on this day