Group Project Report

Understanding the tourism and hospitality market through AirBnB services in the US from 2013 to 2018

17/06/2022
CISM2010
Professor Le Duy Dung

Team 1

Lam Nguyen Thanh Thao
Nguyen Mai Phuong
Ha Phuong Thao

A. Background & Context

Since its birth, Airbnb has transformed the hospitality sector with a low-cost solution for travel accommodation seekers. In 2022, Airbnb has 2.9 million hosts with over 7 million listings in 100,000 cities (Steve, 2022). To learn about this innovative business model and put forward suggestions for Airbnb service, our team analyzes the price and review ratings of different types of Airbnb property offerings. Besides, we obtain insights into customers' feedback based on Airbnb hosts, listings, and review details.

Our data was found on Kaggle, a free and open-source website for datasets and information. More than 30000 data points about Boston (USA) listings were recorded from 2013 to 2018. The link to the data was embedded here.

B. Objectives

Our objective is to utilize this AirBnB information dataset to look at meaningful statistics and results to gain better insights into the tourism and hospitality industry.

- Firstly, the information on Airbnb hosts and listings will demonstrate a big picture of the current offerings within the hospitality market, thereby showing whether the market has diverse offerings and potential growth or not.
- Secondly, the information on the corresponding property and room of each AirBnB listing above will allow our group to identify the scale and characteristics of each listing, as well as the differences in properties offered between cities and countries.
- Finally, the review details which consist of reviewers' name and comments by date will provide insights into customers' feedback on the services. From that, we can identify which aspects of the services that customers like or dislike so as to make appropriate recommendations for host' improvements.

Overall, the objective of the project is to build a SQL database that provides insightful information about the tourism industry and assists different stakeholders including businesses, customers and investors in making better choices.

This database will also allow our group to see a relationship between listings, hosts, reviewers and how they connect to each other. Understanding this interconnected relationship will help us to provide helpful, appropriate recommendations for our stakeholders.

C. Original database & improvements

The original dataset is about Airbnb - an online platform that connects homestay hosts to people who seek for temporary accomodation. This data records the detailed information of each listing from 2013 to 2018. There are 3 big tables: listings (including homestay information and review scores), calendars (listing ID, price and daily availability), and reviews (detailed comments of each listing).

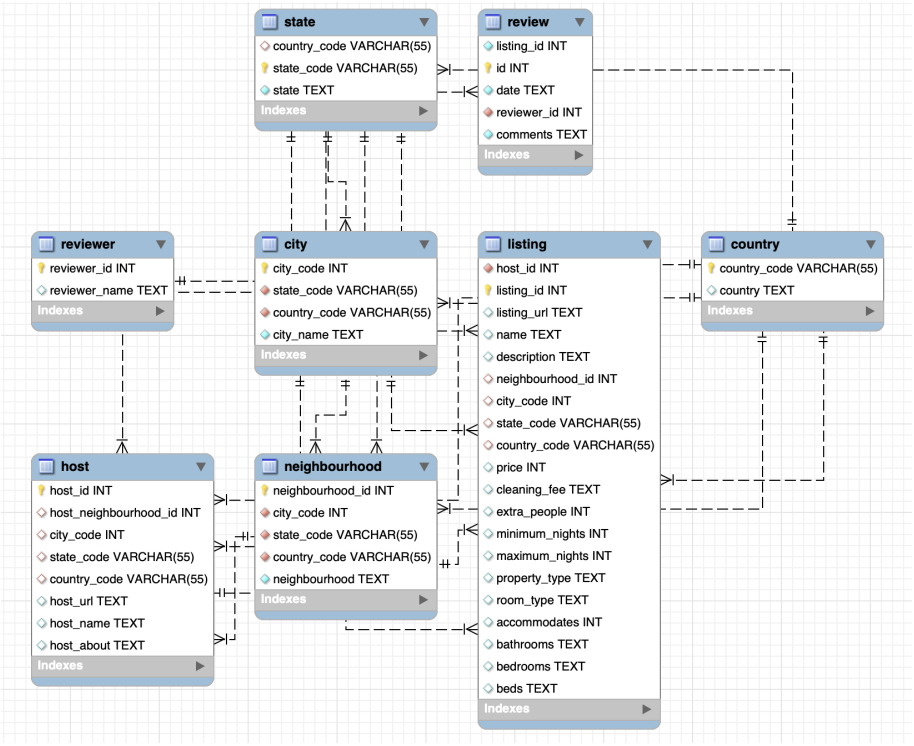
To better define the relationship, we made those following amendments:

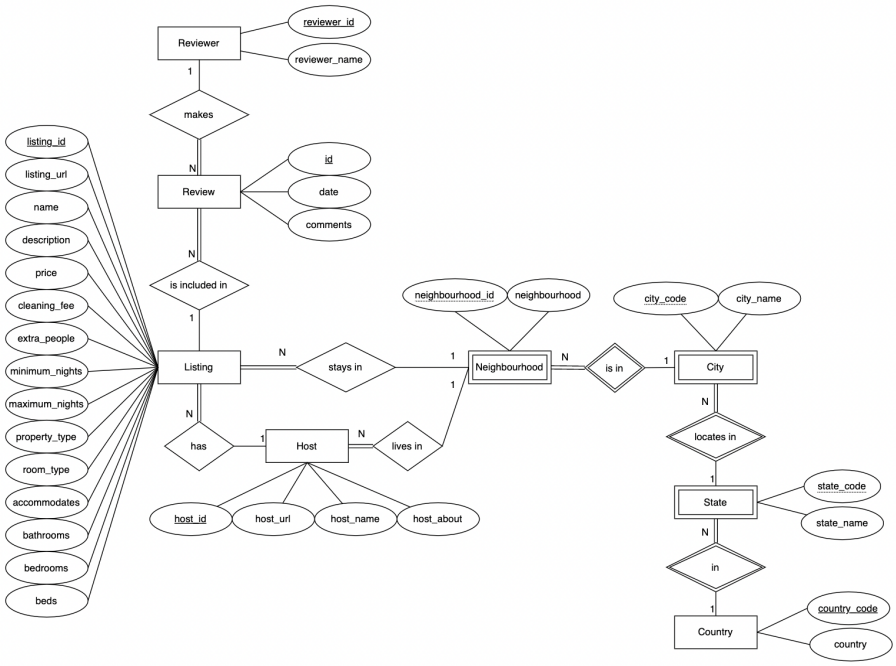
- **We reorganized the listing table and limited the attributes of "Listing".** The original table contains 95 columns, relating to property, host, neighborhood information, etc. However, we only retain essential information related to the listing, such as its id, host_id, listing_url, etc.
- **We set up the "Host" entity from the original "Listing" table.** This is to avoid too much information stacked in the listing attribute. By separating the big table and diversifying into smaller tables of "Property" and "Host", we can simplify the process to get the information for "Property" and "Host" solely, and easily see the relationship of these two entities compared to other entities in the database. "Property" is a weak entity, included in "Listing", while "Host" is a strong entity, and has a one to many relationship with "Listing"
- **"Neighbourhood", "City", "State" and "Country" were separated from Listing table as separated entities and added index numbers.** By doing this way, it would be much more convenient for future data modification and expansion.
- **We divided the "review" table into "Reviewer" and "Review".** We normalized this into two different tables to better show the relationship between reviewer and their review, as "Reviewer" and "Review" have a one to many relationships. Moreover, having 2 separated tables will ease the future data modification about "reviewer" or "review".
- **Original table of "Calendar" is not included in the final database to simplify queries.** Through the "Calendar", we can know the availability and the price of one listing in a specific day. However, in this scope of the project, we only want to look at the big scope of the business, rather than the details into the availability at each listing. By skipping calendar, we can focus on other objectives of the project.

The new database design will benefit the system in the following way:

- Ease the data modification as there are more separated entity
- Allow more flexible and convenient data retrieveing process
- Minimize data inconsistency

D. Entities & Attributes





Entity: Reviewer

- This contains all the necessary information about reviewers.
- Contains attributes:
    - reviewer_id (unique integer identifier for reviewer)
    - reviewer_name (reviewer's name)
- reviewer_id is the primary key for the entity

Entity: Review

- This contains all the necessary information about each comments/ review from customers.
- Contains attributes:
    - id (unique integer identifier for each review)
    - date (review date)
    - comments (the content of each review)
    - listing_id (unique integer identifier for listing)
    - reviewer_id (unique integer identifier for reviewer)
- id is the primary key for the entity
- listing_id and reviewer_id are foreign keys

Entity: Listing

- This contains all the necessary information about an online entry for one host's place
- Contains attributes:
    - listing_id (unique integer identifier for listing)
    - host_id
    - listing_url (the unique address of the listing on the Internet)
    - name (the title of the listing)
    - description (listing description)
    - neighbourhood_id (unique key of neighbourhood entity)
    - city_code (unique key of city entity)
    - state_code (unique key of state entity)
    - country_code (unique key of country entity)
    - price (listing price)
    - cleaning_fee (the cleaning fee of the listing)
    - extra_people (the extra fee charged for one extra people)

    - minimum_nights (the minimum number of staying nights for a listing)
    - maximum_nights (the maximum number of staying nights for a listing)
    - property_type (housing type)
    - room_type (room type)
    - accommodates (the number of people allowed to stay in a property)
    - bathrooms (the number of bathrooms in a property)
    - bedrooms (the number of bedrooms in a property)
    - beds (the number of beds in a property)
- listing_id is the primary key for the entity
- Foreign keys:
    - host_id
    - neighbourhood_id (unique key of neighbourhood entity)
    - city_code (unique key of city entity)
    - state_code (unique key of state entity)
    - country_code (unique key of country entity)

Entity: Host

- This contains all the necessary information for the host.
- Contains attributes:
    - host_id (unique integer identifier for host)
    - host_url (the unique address of the host on the Internet)
    - host_name (the name of the host)
    - host_about (the detailed description of the host)
    - host_neighbourhood_id (the ID of the neighbourhood where the host lives)
- host_id is the primary key for the entity.
- host_neighbourhood_id is the foreign key

Entity: Neighbourhood

- This contains all the necessary information about the surrounding areas of different homestays.
- Contains attributes:
    - neighbourhood_id (unique integer identifier for neighborhood)
    - neighbourhood (the name of the surrounding neighborhood)
    - city_code (unique key of city entity)

    - state_code (unique key of state entity)
    - country_code (unique key of country entity)
- neighbourhood_id is the primary key for the entity
- Foreign keys:
    - city_code
    - state_code
    - country_code

Entity: City

- This contains all the necessary information about each city that the listings belong to.
- Contains attributes:
    - city_code (unique integer identifier for each city)
    - city_name (name of the city of the listing)
    - state_code (unique key of state entity)
    - country_code (unique key of country entity)
- city_code is the primary key for the entity
- Foreign keys:
    - state_code
    - country_code

Entity: State

- This contains all the necessary information about each state that the listings belong to.
- Contains attributes:
    - state_code (unique integer identifier for each state)
    - state_name (name of the state of the listing)
    - country_code (unique key of country entity)
- state_code is the primary key for the entity
- country_code is the foreign key of the entity

Entity: Country

- This contains all the necessary information about each country that the listings belong to.
- Contains attributes:
    - country_code (numeric geographical code for each country)
    - country (name of the state of the country)

- country_code is the primary key for the entity

E. Relationships Description

- One host can have several listings while a listing is made exactly by one host id.
- Hosts live in different neighborhoods, which belong to different cities, states and countries.
- Listings also stay in different neighborhood, which belong to different cities, states and countries.
- One listing could include more than 1 review and comments from reviewers.
- One reviewer could make many reviews and comments.
- Each country has several states and each state have different cities which contain different neighborhood.

F. Query Design & Results

In order to test the usability of this dataset, we design a total of 15 queries which belong to three purposes as follows:

F.1 For investor-use: To understand market insights, trends and key players so as to make profitable investment in the homestay market in Boston.

1. Count how many hosts whose neighborhood in Jamaica Plain

- *Access*

```
SELECT COUNT(*) AS NumOfHostInJamica
FROM HOST
WHERE host_neighbourhood_id = 2;
```

- *MySQL*

```
SELECT COUNT(*) AS NumOfHostInJamica
FROM HOST
WHERE host_neighbourhood_id = 2;
```

→ **Purpose:** Determine whether the market for tourism and hospitality is potential in Jamaica Plain.

2. Find all the listings that could accommodate more than 8 people

   ```
   SELECT listing_id, accommodates
   FROM listing
   WHERE accommodates > 8;
   ```

   → **Purpose:** Assist in estimating the market size for medium-sized homestays.

3. How many listings whose property type is an apartment?

   ```
   SELECT COUNT(property_type) AS NumOfListingIsApartment
   FROM listing
   WHERE property_type = 'Apartment';
   ```

   → **Purpose:** Estimate the market size for the apartment tourism segment.

4. Count the average accommodates of an entire home/ apt room type.

- *Access:*

  ```
  SELECT AVG(accommodates)
  FROM listing
  WHERE room_type = "Entire home/apt";
  ```

- *MySQL:*

  ```
  SELECT AVG(accommodates)
  FROM listing
  WHERE room_type = "Entire home/apt";
  ```

  → **Purpose:** Estimate the average capacity that the segment of entire home/ apt room has.

5. Find the property type and room type of all listings in Boston City.

- *Access:*

  ```
  SELECT listing_id, property_type, room_type
  FROM listing
  WHERE city_code IN (SELECT city_code
                      FROM city
                      WHERE city_name = "Boston")
  ```

- *MySQL:*

```
SELECT listing_id, property_type, room_type
FROM listing
WHERE city_code IN (SELECT city_code
                    FROM city
                    WHERE city_name = "Boston")
;
```

→ **Purpose:** Determine the overall trend of property type and room type within the homestay market in Boston City.

6. What is the state_code and country code for all listings whose price is above $100?

```
SELECT city_code, state_code, country_code, price
FROM listing
WHERE Price > 100;
```

→ **Purpose:** Identify the diversity of locations for the medium-sized homestay market.

E.2 For customer-use: To obtain the necessary information regarding price, reviews and capacity so as to make appropriate choices for themselves.

INNER JOIN

7. Return all the property type, room type and related listing name, city, host name, sorting the results by the listing price from the most expensive to the least ones.

```
SELECT property_type, room_type, name, city_name, host_name, price
FROM listing
INNER JOIN host ON listing.host_id = host.host_id
INNER JOIN city ON listing.city_code = city.city_code
ORDER BY price DESC;
```

→ **Purpose:** As customers, they can look for listings that could accommodate their needs.

8. Return the listing name, listing neighborhood and their corresponding comments and reviewer names.

- *Access*

```
SELECT listing.name, neighbourhood.neighbourhood, review.comments,
reviewer.reviewer_name
```

```
FROM ((review
INNER JOIN listing ON listing.listing_id = review.listing_id)
INNER JOIN neighbourhood ON listing.neighbourhood_id =
neighbourhood.neighbourhood_id)
INNER JOIN reviewer on review.reviewer_id = reviewer.reviewer_id;
```

- *MySQL*

```
SELECT listing.name, neighbourhood.neighbourhood, review.comments,
reviewer.reviewer_name
FROM (review
INNER JOIN listing ON listing.listing_id = review.listing_id
INNER JOIN neighbourhood ON listing.neighbourhood_id =
neighbourhood.neighbourhood_id
INNER JOIN reviewer on review.reviewer_id = reviewer.reviewer_id);
```

→ **Purpose:** As customers, they can look for the listings with good reviews.

OUTER JOIN

9. Return the listing description, review comments and review dates. Sort the results by the latest review dates.

- *Access*

```
SELECT A.description, B.comments, B.date
FROM Listing AS A
LEFT JOIN review AS B
ON A.listing_id = B.listing_id
ORDER BY B.date DESC;
```

- *MySQL*

```
SELECT A.description, B.comments, B.date
FROM Listing AS A
LEFT JOIN review AS B
ON A.listing_id = B.listing_id
ORDER BY B.date DESC;
```

→ **Purpose:** As customers, they can look for listings with updated, good reviews.

10. What are the top 5 listings with the lowest price?

- *Access*
  SELECT TOP 5 listing_id, price
  FROM LISTING
  ORDER BY price ASC;

- *MySQL*

  SELECT listing_id, price
  FROM LISTING
  ORDER BY price ASC
  Limit 5;
  → **Purpose:** Help customers with tight budgets to identify AirBnB options in the low-end segment.

11. Find the actual number of bathrooms, bedrooms, and beds for the listing whose ID is 8548176

- *Access:*
  SELECT bathrooms, bedrooms, beds
  FROM LISTING
  WHERE listing_id = '8548176';

- *MySQL:*
  SELECT bathrooms, bedrooms, beds
  FROM LISTING
  WHERE listing_id = '8548176';

  → **Purpose:** As customers, they can identify the number of different types of rooms in a homestay, thereby making appropriate choices for themselves.

SUB-QUERIES

12. Find the host information of all listings that allow more than 3 minimum nights.
    SELECT host_name, host_about
    FROM HOST
    WHERE host_id in (SELECT host_id
    FROM listing
    WHERE minimum_nights > 3);

    → **Purpose**: As customers, they can look for the listings that could accommodate their needs.

E.3 For business-use: To perform benchmarking with other homestay providers in terms of price, services and offerings as well as to understand customer' preferences.

OUTER JOIN

13. Return the cleaning fee and listing name corresponding with their host name and host neighborhood.

```
SELECT  A.name, A.cleaning_fee, B.host_name, B.host_neighbourhood_id
FROM Listing AS A
RIGHT JOIN host AS B
ON B.host_id = A.host_id;
```

→ **Purpose:** Identify the fluctuations of the cleaning fee as one source of cost for the homestay owners.

14. What is the listing that has the second most expensive cleaning fee?

- *Access:*

```
SELECT TOP 1 listing_id, name, cleaning_fee
FROM LISTING WHERE listing_id IN (SELECT TOP 2 listing_id
FROM LISTING
ORDER BY cleaning_fee DESC)
ORDER BY cleaning_fee;
```

- *Mysql:*

```
SELECT listing_id, name, cleaning_fee
FROM LISTING
ORDER BY cleaning_fee DESC
LIMIT 2,1;
```

→ **Purpose:** As homeowners, they can identify the range of the cleaning fee to calculate appropriate costs.

15. Find the average price of listings in Boston.

- *Access:*

```
SELECT AVG(price)
FROM LISTING
WHERE CITY_CODE = 1;
```

- *Mysql:*

```
SELECT AVG(price)
FROM LISTING
WHERE CITY_CODE = 1;
```

→ **Purpose:** Identify the average market price of homestay listings in Boston to benchmark with.

G. Conclusions

After analyzing the datasets and performing query design, it can be concluded that this database acts as a handy tool for three types of stakeholders, which are businesses, customers and investors. By using this database, these stakeholders could discover insightful information regarding the whole tourism and hospitality industry. The benefits of each of these stakeholder are detailed as below:

- For AirBnB businesses (or hosts):
    - To identify customers' pain points and gain points, thus, further enhance their service offerings by providing tailor-made services for customers.
    - To improve customer satisfaction, generate higher retention and greater profits.
- For potential AirBnB customers (tourists):
    - To identify the capacity, price and reviews of different offerings in the market
    - To make appropriate choices based on their needs.
- For investors:
    - To understand the market growth and identify the potential segment to invest in.
    - To generate higher return on their investment.

REFERENCE

Deane, S. (2022, March 28). *2022 Airbnb Statistics: Usage, Demographics, and Revenue growth*. Stratos Jet Charters, Inc. Retrieved June 1, 2022, from https://www.stratosjets.com/blog/airbnb-statistics/