DATABASE FOUNDATIONS FOR BUSINESS ANALYTICS PROJECT

BUAN 6320-002

AIRBNB LISTINGS

UNDER THE ESTEEMED GUIDANCE OF

PROF. DAWN OWENS

TABLE OF CONTENTS

Sr. No.	Title	Pg. No.
1	Project Description	
2	Logical Data Model	
3	Physical Data Model	
4	Data Loading Concept	
5	Insights	
6	Visualization Charts	
7	References	

PROJECT DESCRIPTION

Airbnb is a vacation rental company based out in North America. It operates as an online marketplace, comprising of two parties- guests and hosts, and focuses on short stays and experiences. The business model followed by them is peer to peer platform, they gain service fee from the customers who book the homestays and gain commission from the hosts. The service fee ranges from 5-15% for guests and 3% for hosts. Their platform makes booking easier across the world and expands the industry making it viable to book accommodation all over the globe. The revenue Airbnb generated in 2021 is USD 5.99 Billion a 77% increase from 2021.

This dataset from Airbnb sourced from Kaggle.com has 250000 listings across 10 major cities and includes information on hosts, pricing of the stay, property details, and reviews on multiple criteria.

LOGICAL DATA MODEL

Dataset link
https://www.kaggle.com/datasets/mysarahmadbhat/airbnb-listings-reviews

This dataset has 29 columns and 271792 rows and exists in 2NF form.

Objective
Goal is to derive insights from the data model that has been loaded on MySQL Workbench usinf queries

Explanation of the Attributes
The dataset includes information on –

1. Primary Key – Listing ID
2. Foreign Key –
3. Unique ID's – Listing ID and Host ID
4. Host – Host Since, Host Location, Host/ Superhost, Total Listing by Host, Host Profile Picture, Verified Host Identity
5. Property Location – Name, Neighbourhood, City, Latitude, Longitude
6. Property – Type, Room Type, Accommodates, Bedrooms, Amenities, Price, Maximum and Minimum Nights, Instant Booking Option
7. Reviews – Rating, Accuracy, Cleanliness, Check-In, Communication, Location, Value

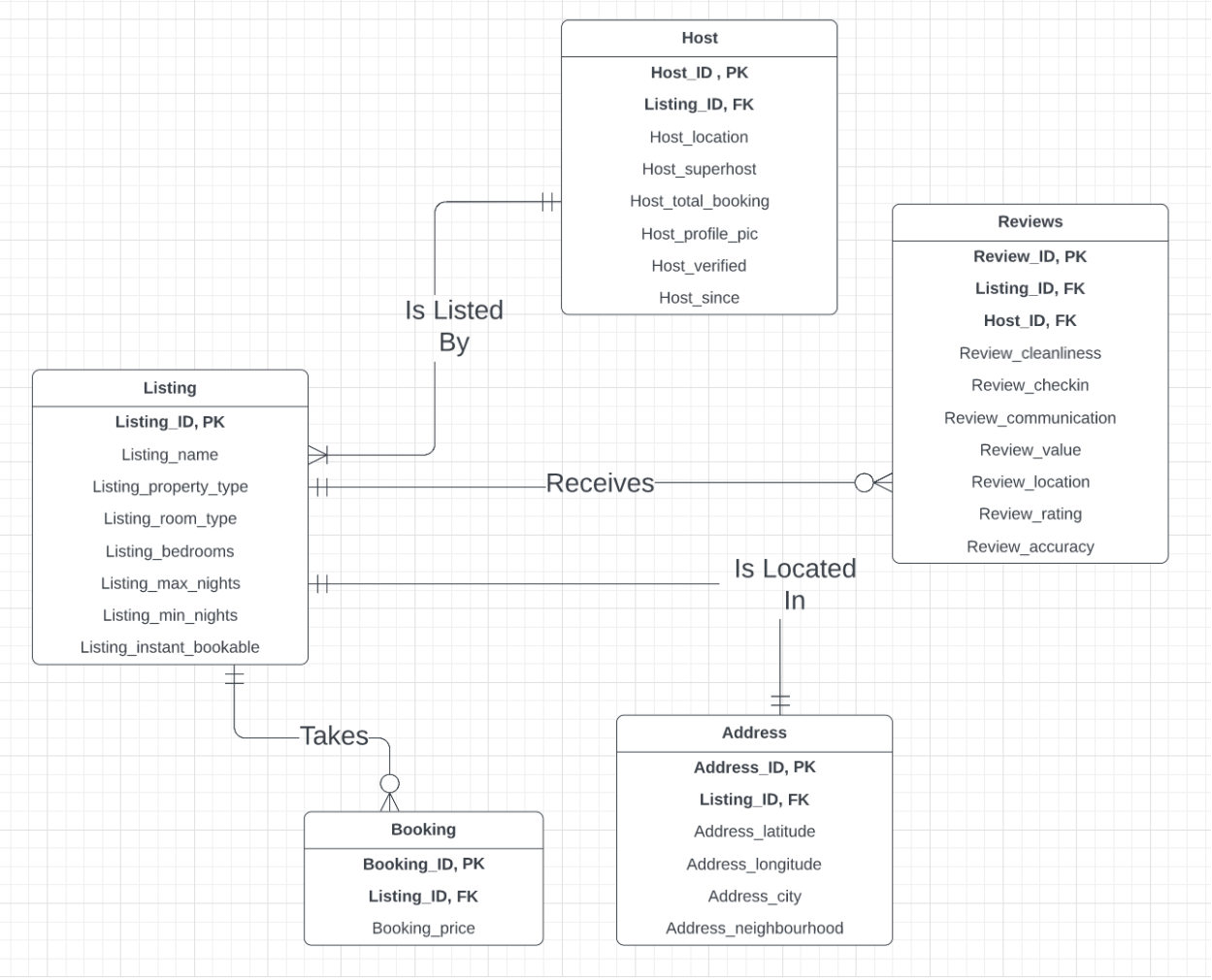
Data Sample Characterization:

1. Listing ID
2. Name
3. Host ID
4. Host Since
5. Host Location
6. Host/ Superhost
7. Total Listing by Host
8. Host Profile Picture

9. Verified Host Identity
10. Neighbourhood
11. City
12. Latitude
13. Longitude
14. Property Type
15. Room Type
16. Accommodates
17. Bedroom
18. Price
19. Minimum Nights
20. Maximum Nights
21. Review Scores Rating
22. Review Score Accuracy
23. Review Score Cleanliness
24. Review Score Check-In
25. Review Score Communication
26. Review Score Location
27. Review Score Value
28. Instant Booking Option

PHYSICAL DATA MODEL

We have identified 5 entities:



Our ERD model has 5 entities- the host, listing, reviews, address, and booking tables. The booking entity includes the booking ID, listing ID, and booking price. It is separate from the listing table as there may be multiple bookings of a single listing throughout the given time period, which can be identified through the booking ID.

All the tables in this ERD have only one primary key, which eliminates the possibility of any partial dependencies. The address table has been separated from the listing table as well, due to the table not being in 2nd normal form. Once done, all our tables are in second normal form.

1. How many listings are there in each city?
   As seen in the output, the city with the highest number of listings is Paris, with more than 19k listings. This is followed by New York, Rio de Janeiro, Sydney, and more. With this insight, Airbnb can collaborate with brands or tourism destinations or airlines that frequently land in Paris, and make sure that the listings are booked throughout the year by giving collaborative deals. It can also do a lot in other cities to promote more hosts to offer their properties as listings.

2. What is the frequency table of the number of bedrooms in the properties?
   As seen in the output, it is concluded that more than 21k of the properties have only 1 bedroom, followed by 10k properties having 2 bedrooms. This signifies that Airbnb has a lot of options for singles or couples travelling together, and can advertise to these parties, while also working towards expanding their options for accommodations with more rooms for larger parties.
3. How many unverified hosts are there who have overall ratings more than 90%?
   There are almost 5k unverified hosts that have an overall rating of more than 90%. Airbnb should reach out to them and urge them to get verified, so their profile can be more attractive and their bookings can increase, generating more commission for the company.
4. How many verified hosts are there?
   There are almost 30k verified hosts on Airbnb. To improve this number, Airbnb should give hosts an incentive to get verified, add a profile picture, and do other things that make their listings and profiles more appealing.
5. What is the maximum booking price in the data?
   The maximum booking price in the data is $9999. This information can be used to study the location and guests of this booking price's listing, and use it to increase the number of such listings and guests, respectively.
6. How many listings allow guests to stay for a long duration (more than 180 nights?)

Almost 43k listings allow guests to stay for longer than 6 months. Airbnb should see that hosts are open to longer stays, and hence incentive guests into booking for longer durations.

CODE:

Address table:

```
CREATE TABLE `address` (
 `Address_id` varchar(50) NOT NULL,
 `Address_Latitude` varchar(50) DEFAULT NULL,
 `Address_Longitude` varchar(50) DEFAULT NULL,
 `Address_City` varchar(45) DEFAULT NULL,
 `Address_Neighbourhood` varchar(45) DEFAULT NULL,
 PRIMARY KEY (`Address_id`),
 CONSTRAINT `Address_Listing_id` FOREIGN KEY (`Address_id`) REFERENCES `listing` (`Listing_id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

Booking Table:

```
CREATE TABLE `booking` (
 `Booking_id` varchar(50) NOT NULL,
 `Booking_Price` varchar(45) DEFAULT NULL,
 PRIMARY KEY (`Booking_id`),
 CONSTRAINT `Booking_Listing_id` FOREIGN KEY (`Booking_id`) REFERENCES `listing` (`Listing_id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

Listing Table:

```
CREATE TABLE `listing` (
 `Listing_id` varchar(100) NOT NULL,
 `Listing_Name` varchar(2000) DEFAULT NULL,
 `Listing_Property_Type` varchar(45) DEFAULT NULL,
 `Listing_Room_Type` varchar(45) DEFAULT NULL,
 `Listing_Bedrooms` varchar(45) DEFAULT NULL,
 `Listing_Max_Nights` varchar(45) DEFAULT NULL,
 `Listing_Instant_Bookable` varchar(45) DEFAULT NULL,
 `Listing_Min_Nights` varchar(45) DEFAULT NULL,  PRIMARY KEY (`Listing_id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

Review Table:

```
CREATE TABLE `review` (
  `Review_id` varchar(100) NOT NULL,
  `Review_Cleanliness` varchar(45) DEFAULT NULL,
  `Review_Communication` varchar(45) DEFAULT NULL,
  `Review_Checkin` varchar(45) DEFAULT NULL,
  `Review_Value` varchar(45) DEFAULT NULL,
  `Review_Location` varchar(45) DEFAULT NULL,
  `Review_Rating` varchar(45) DEFAULT NULL,
  `Review_Accuracy` varchar(45) DEFAULT NULL,
  PRIMARY KEY (`Review_id`),
  CONSTRAINT `Review_Listing_id` FOREIGN KEY (`Review_id`) REFERENCES `listing` (`Listing_id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

Host Table:

DATA LOADING CONCEPT

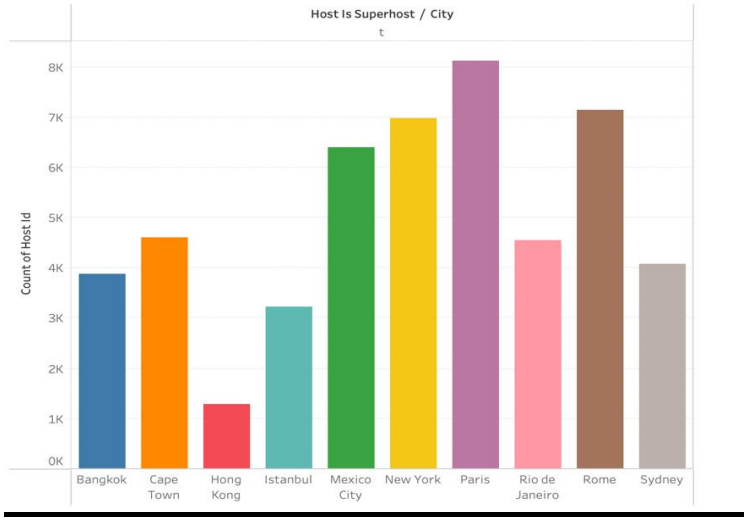
***Swapnil has to add content here**

INSIGHTS

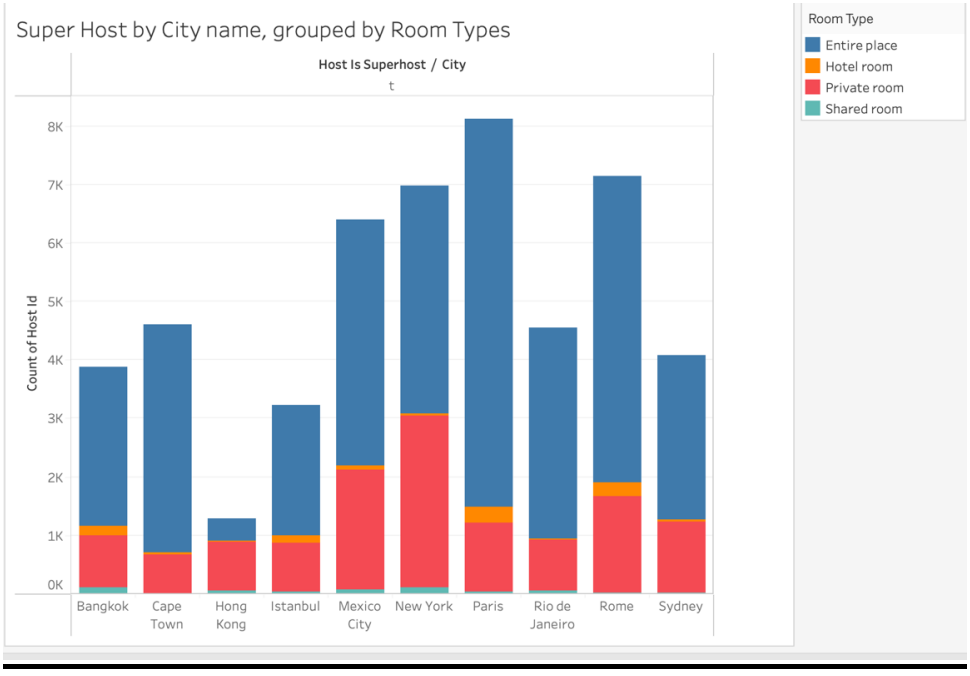
1. Which city has the most affordable stay offered across the listings?
2. Which are the top 3 properties in each of the regions/countries?
3. Which are the top 5 best-reviewed properties across the listings?
4. How many bedrooms do most of the listings have? What is the percentage of these listings?
5. What is the average rating for a super host and the average rating for a non-super host?
6. On average, how much does a non-super host have to improve his rating to become a super host?
7. Which attribute of a listing has the highest chance of getting the lowest rating?
8. What are customers usually least satisfied with during their stay?
9. How many listings has the host stayed in the property itself?
10. What is the variation of prices, on average, per listing?
11. What is the most common property type?
12. How many of the hosts are super hosts?
13. Do the highest booked properties have more variation in their prices?
14. What are the maximum and minimum total listings of a host?
15. How many unverified hosts have listings with ratings above 90?

VISUALIZATION CHARTS

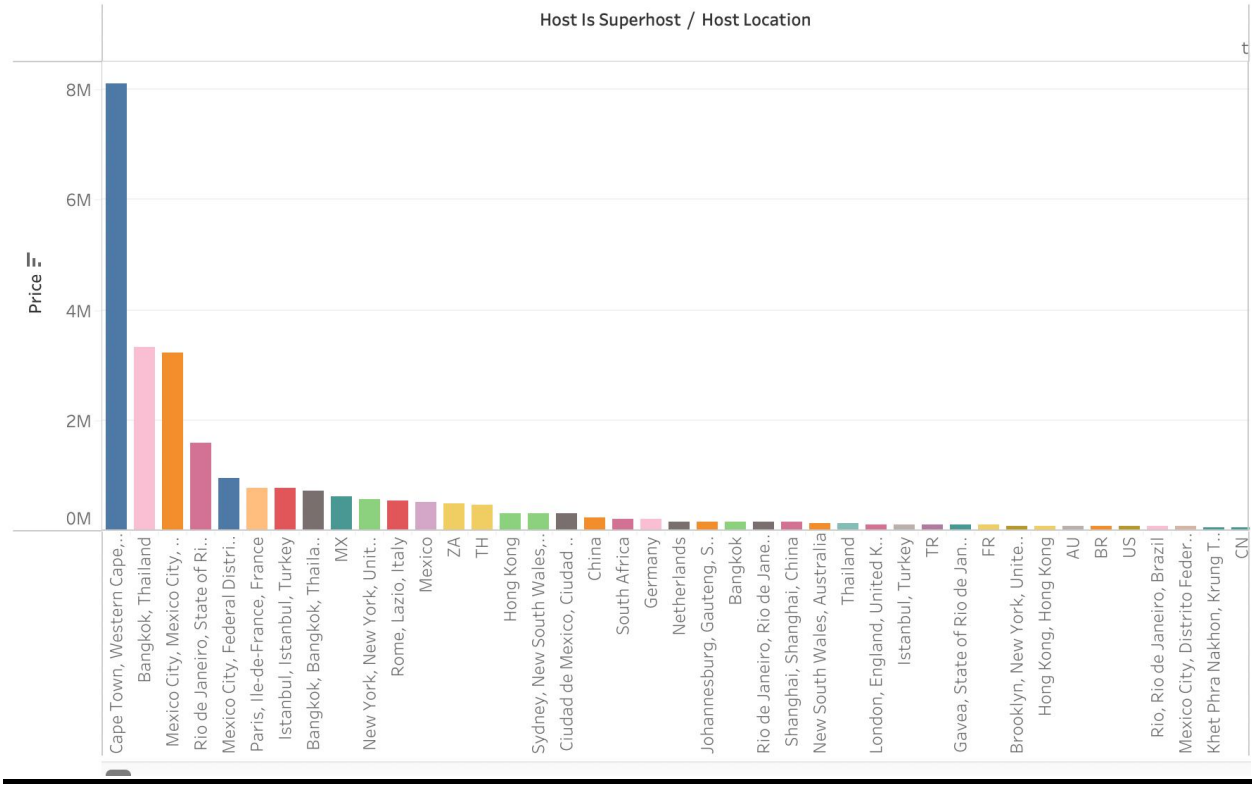
Number of Superhost in each city (Filtered by: Host has Profile Pic = T)



Superhost by City name, grouped by Room Types



Prices of listings in each location



REFERENCES

Dataset link:

https://www.kaggle.com/datasets/mysarahmadbhat/airbnb-listings-reviews

Presentation Link: