DATABASE FOUNDATIONS FOR BUSINESS ANALYTICS PROJECT

BUAN 6320-002

AIRBNB LISTINGS

UNDER THE ESTEEMED GUIDANCE OF

PROF. DAWN OWENS

Group Members:

- **Apala Mishra**
- **Arushi Saxena**
- **Sukanya Ghosh**
- **Swapnil Mungi**

TABLE OF CONTENTS

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.

DATA DESCRIPTION

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 using queries

Explanation of the Attributes
The dataset includes information on –

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

Data Sample Characterization:

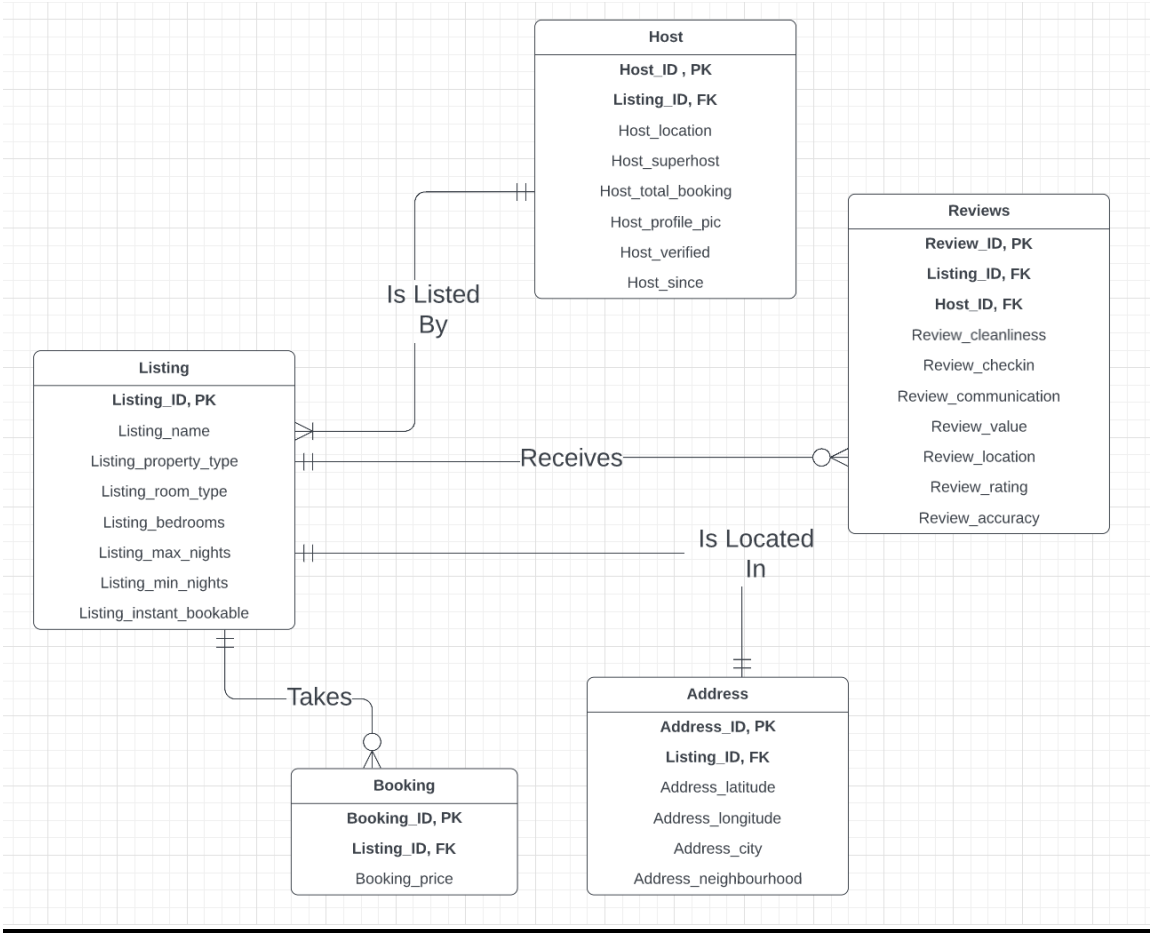
1. Listing ID
2. Name
3. Host ID
4. Host Location
5. Host/ Superhost
6. Total Listing by Host
7. Host has Profile Picture
8. Verified Host Identity
9. Neighbourhood
10. City
11. Latitude
12. Longitude
13. Property Type
14. Room Type
15. Bedroom
16. Price
17. Minimum Nights

18. Maximum Nights
19. Review Scores Rating
20. Review Score Accuracy
21. Review Score Cleanliness
22. Review Score Check-In
23. Review Score Communication
24. Review Score Location
25. Review Score Value
26. Instant Booking Option

LOGICAL AND PHYSICAL DATA MODEL

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 3rd normal form. Once done, all our tables are in second normal form.



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:

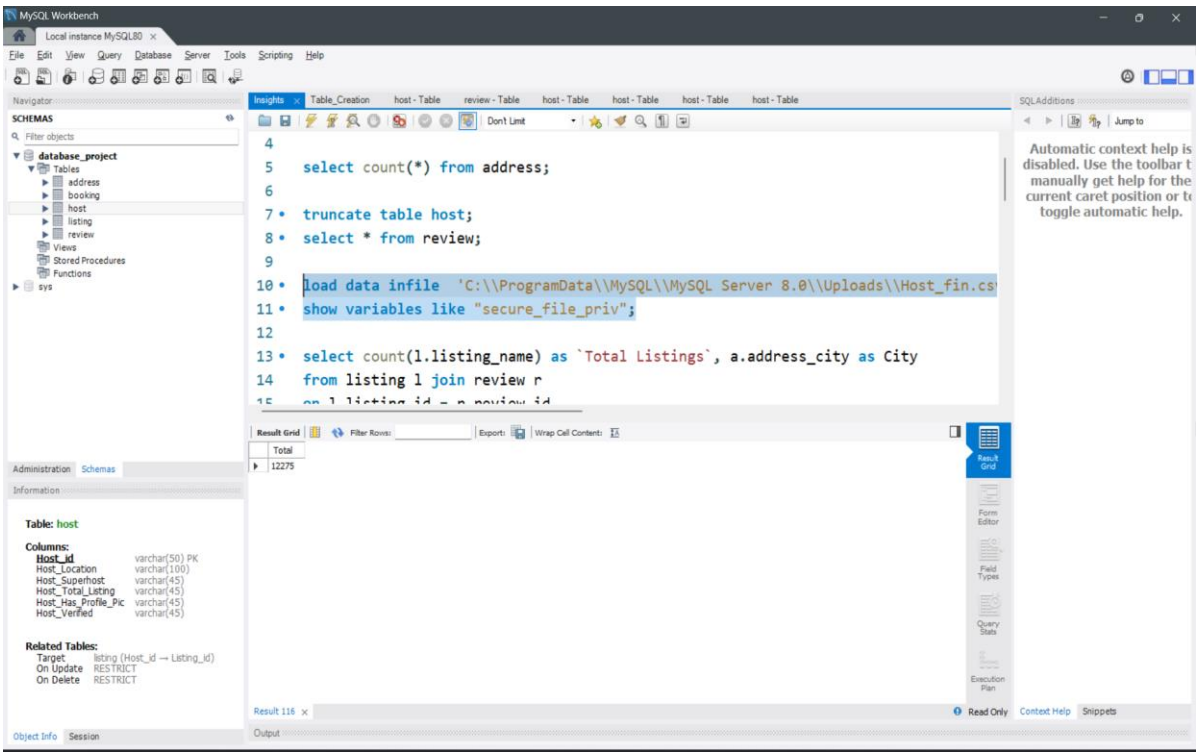
```
CREATE TABLE `review` (
  `Review_id` varchar(100) NOT NULL,
  `Review_Cleanliness` varchar(45) DEFAULT NULL,
  `Review_Communication` varchar(45) DEFAULT NULL,
  `Review_Checkin` varchar(45) DEFAULT NULL,
  `Review_Value` varchar(45) DEFAULT NULL,
  `Review_Location` varchar(45) DEFAULT NULL,
  `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:

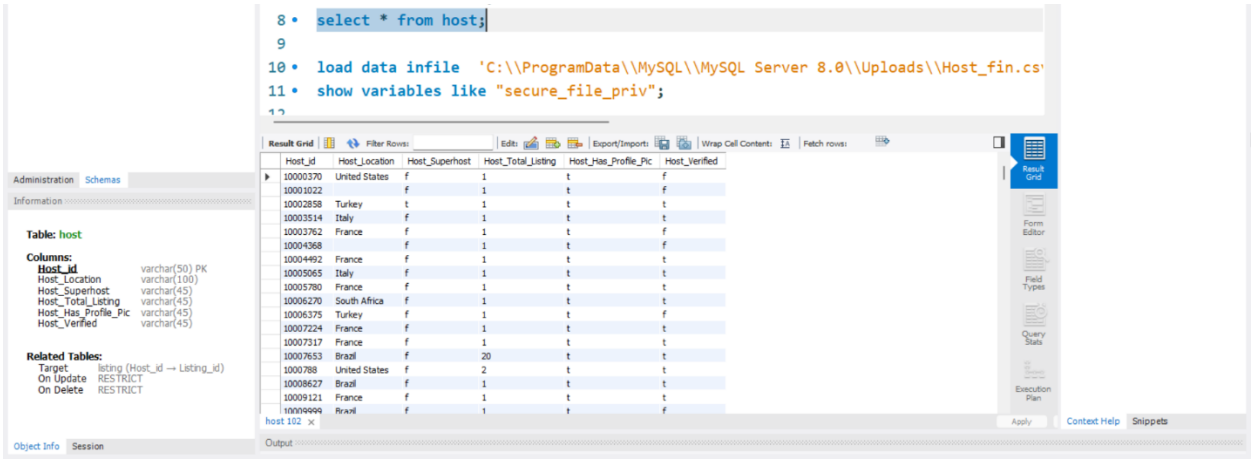
```
CREATE TABLE `host` (
  `Host_id` varchar(50) NOT NULL,
  `Host_Location` varchar(100) DEFAULT NULL,
  `Host_Superhost` varchar(45) DEFAULT NULL,
  `Host_Total_Booking` varchar(45) DEFAULT NULL,
  `Host_Has_Profile_Pic` varchar(45) DEFAULT NULL,
  `Host_Verified` varchar(45) DEFAULT NULL,
  PRIMARY KEY (`Host_id`),
  CONSTRAINT `Host_Listing_id` FOREIGN KEY (`Host_id`) REFERENCES `listing` (`Listing_id`)
ON DELETE   RESTRICT ON UPDATE RESTRICT
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

DATA LOADING CONCEPT

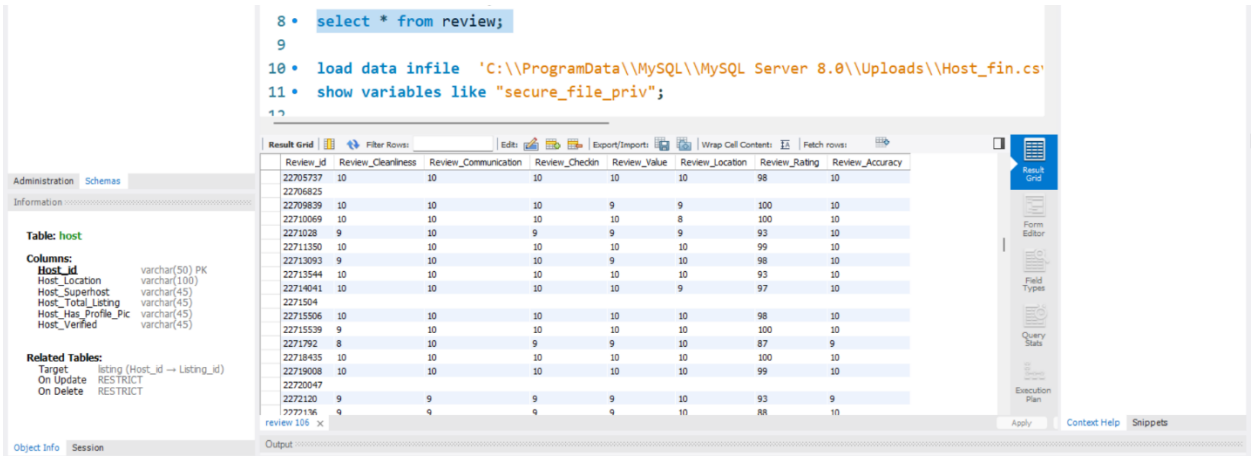
- Tool used is **MySQL Workbench**
- **Data Loading Command:**
  load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Host_fin.csv' into table host fields terminated by ',' enclosed by "" lines terminated by '\n' ignore 1 rows;
  (*This particular command represents the data loading for Host table and the similar command was used to load data for the rest of the four tables)



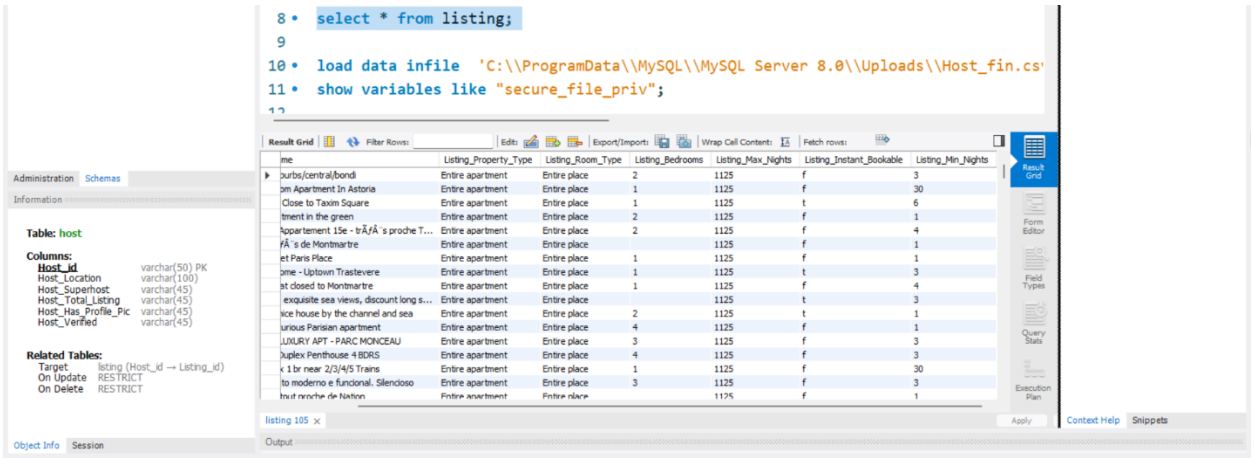
Host Table:



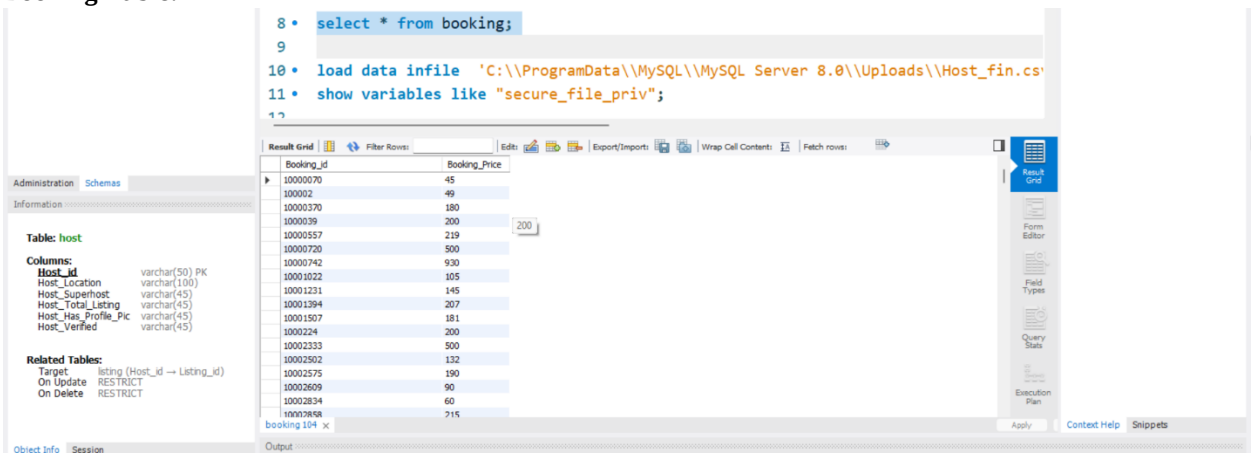
Review Table:



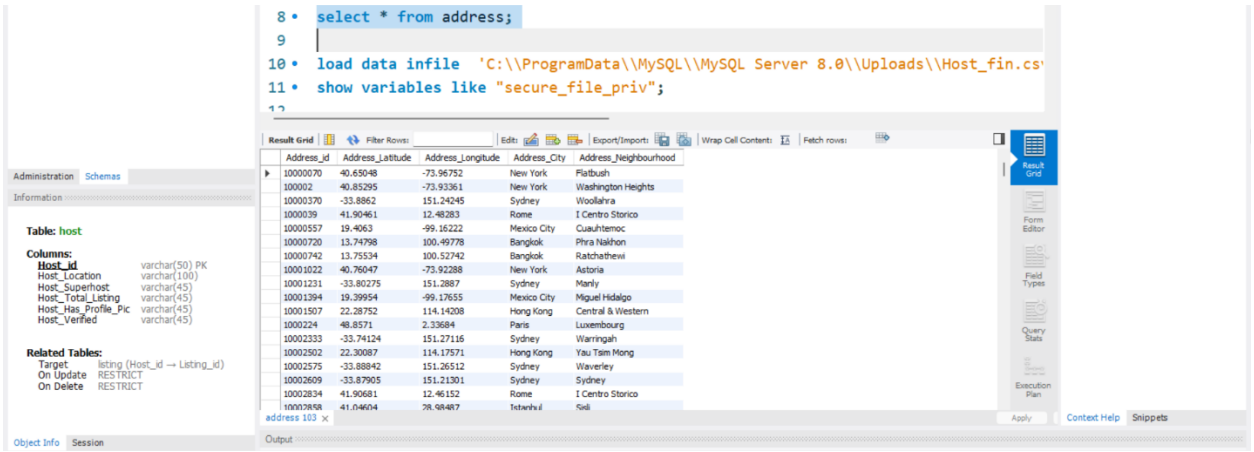
Listing Table:



Booking Table:



Address Table:



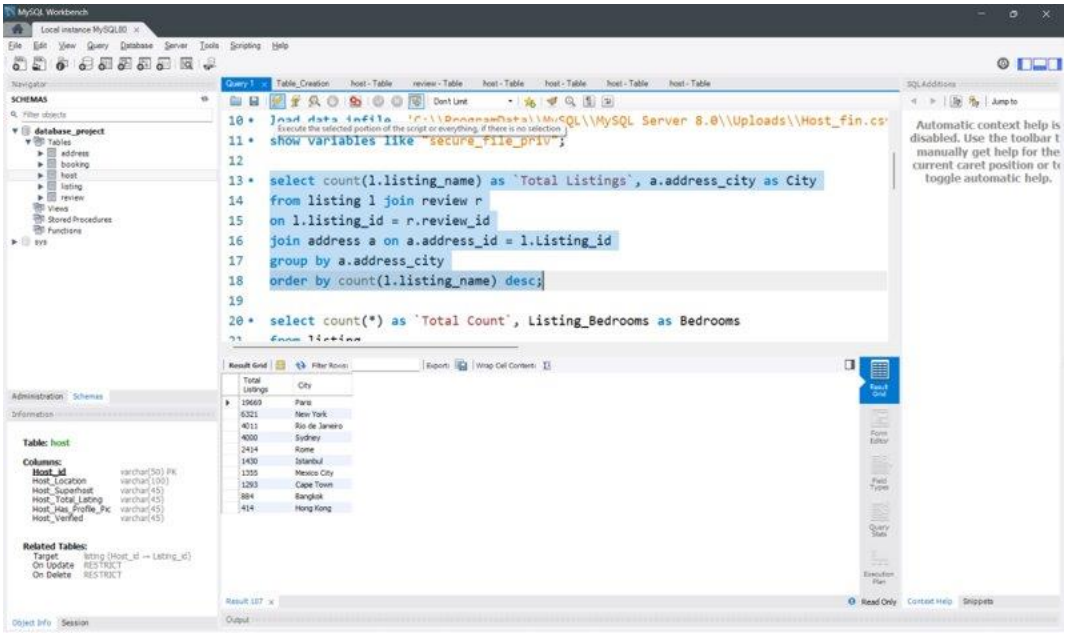
INSIGHTS

Insight #1: How many listings are there in each city?

Code:

```
select count(l.listing_name) as `Total Listings`, a.address_city as City
from listing l join review r
on l.listing_id = r.review_id
join address a on a.address_id = l.Listing_id
group by a.address_city
order by count(l.listing_name) desc;
```

MySQL Workbench:



Conclusion:

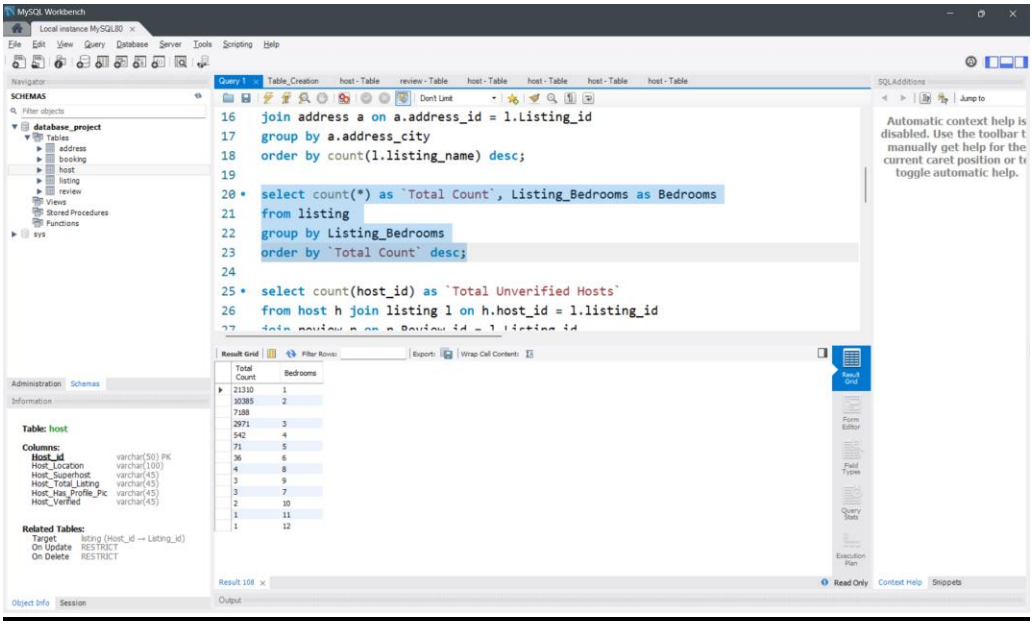
As seen in the output, the city with the highest number of listings is Paris, with more than 19k listings. This is followed by New York, Rio de Janeiro, Sydney, and more. With this insight, Airbnb can collaborate with brands or tourism destinations or airlines that frequently land in Paris, and make 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.

Insight #2: What is the frequency table of the number of bedrooms in the properties?

Code:

```
select count(*) as `Total Count`, Listing_Bedrooms as Bedrooms
from listing
group by Listing_Bedrooms
order by `Total Count` desc;
```

MySQL Workbench:



Conclusion:

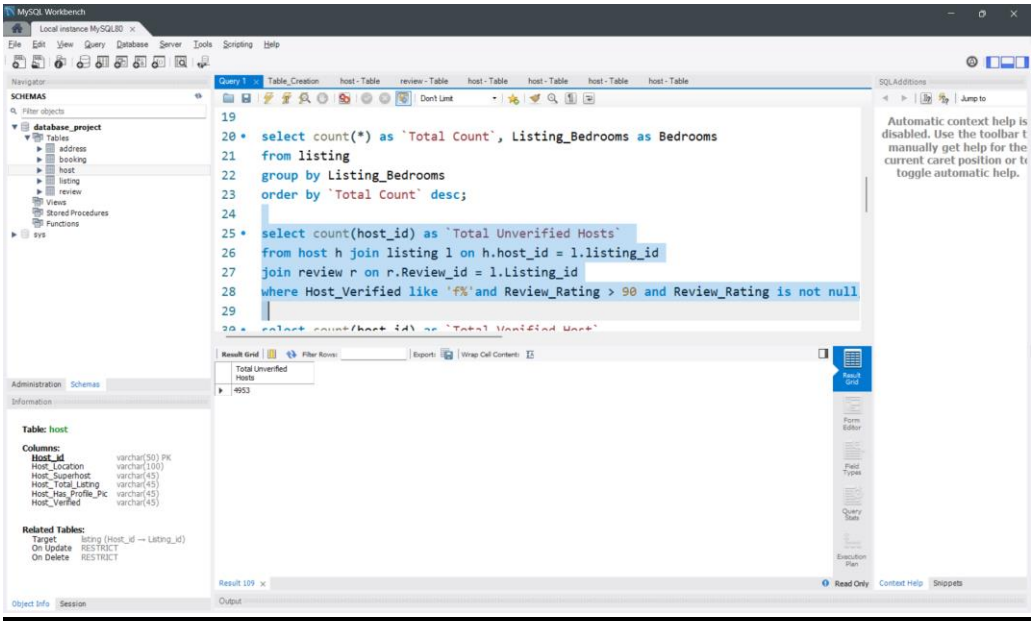
As seen in the output, it is concluded that more than 21k of the properties have only 1 bedroom, followed by 10k properties having 2 bedrooms. This signifies that Airbnb has a lot of options for singles or couples travelling together, and can advertise to these parties, while also working towards expanding their options for accommodations with more rooms for larger parties.

Insight #3: How many unverified hosts are there who have overall ratings more than 90%?

Code:

```
select count(host_id) as `Total Unverified Hosts`
from host h join listing l on h.host_id = l.listing_id
join review r on r.Review_id = l.Listing_id
where Host_Verified like 'f%'and Review_Rating > 90 and Review_Rating is not null;
```

MySQL Workbench:



Conclusion:

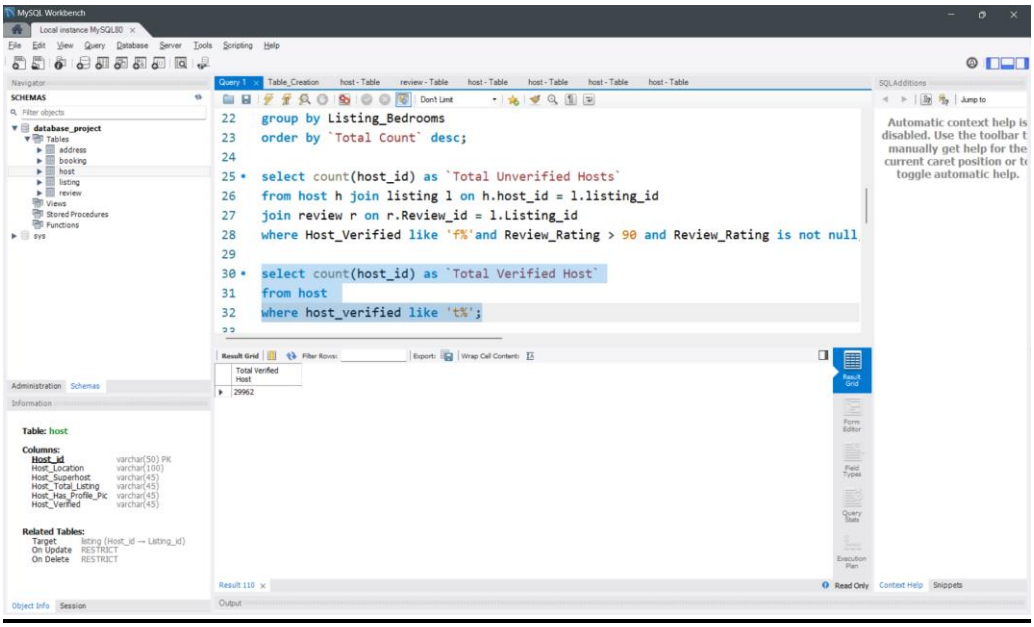
There are almost 5k unverified hosts that have an overall rating of more than 90%. Airbnb should reach out to them and urge them to get verified, so their profile can be more attractive and their bookings can increase, generating more commission for the company.

Insight #4: How many verified hosts are there?

Code:

```
select count(host_id) as `Total Verified Host`
from host
where host_verified like 't%';
```

MySQL Workbench:



Conclusion:

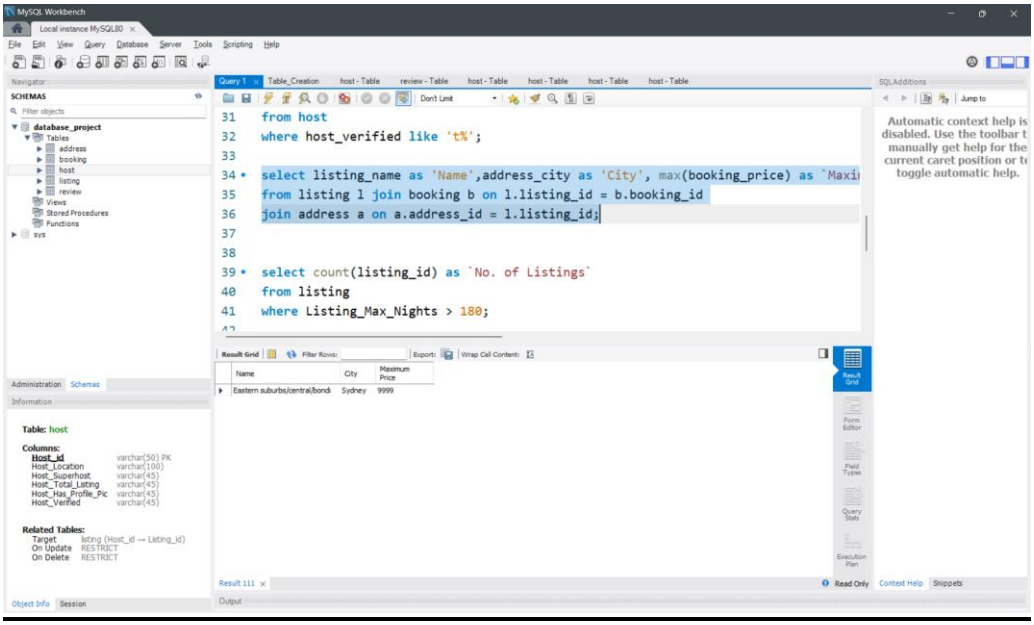
There are almost 30k verified hosts on Airbnb. To improve this number, Airbnb should give hosts an incentive to get verified, add a profile picture, and do other things that make their listings and profiles more appealing.

Insight #5: What is the maximum booking price in the data?

Code:

```
select listing_name as 'Name',address_city as 'City', max(booking_price) as `Maximum Price`
from listing l join booking b on l.listing_id = b.booking_id
join address a on a.address_id = l.listing_id;
```

MySQL Workbench:



Conclusion:

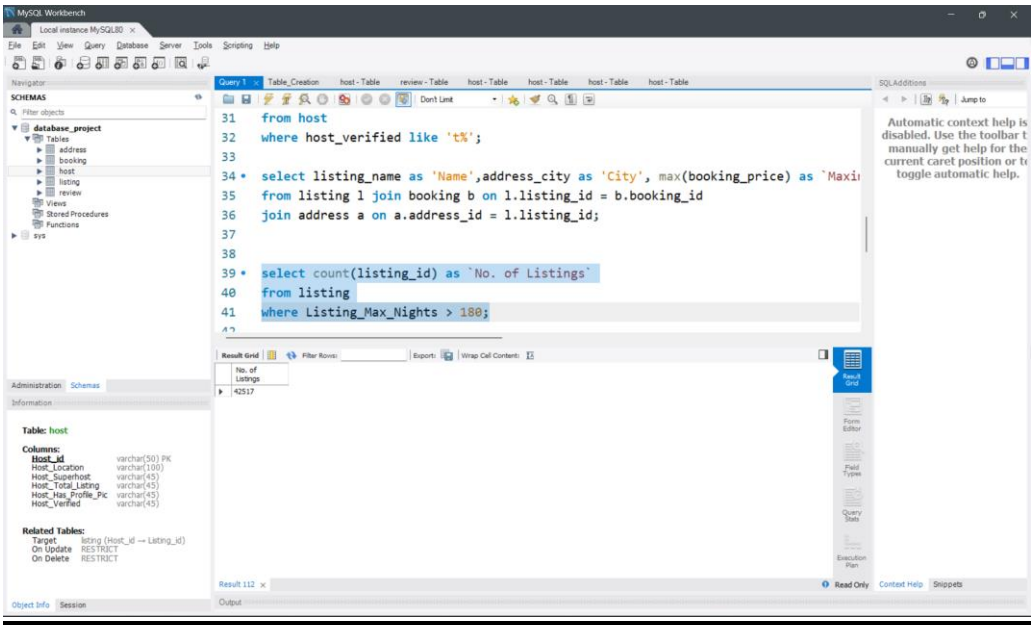
The maximum booking price in the data is $9999. This information can be used to study the location and guests of this booking price's listing, and use it to increase the number of such listings and guests, respectively.

Insight #6: How many listings allow guests to stay for a long duration (more than 180 nights?

Code:

```
select count(listing_id) as `No. of Listings`
from listing
where Listing_Max_Nights > 180;
```

MySQL Workbench:



Conclusion:

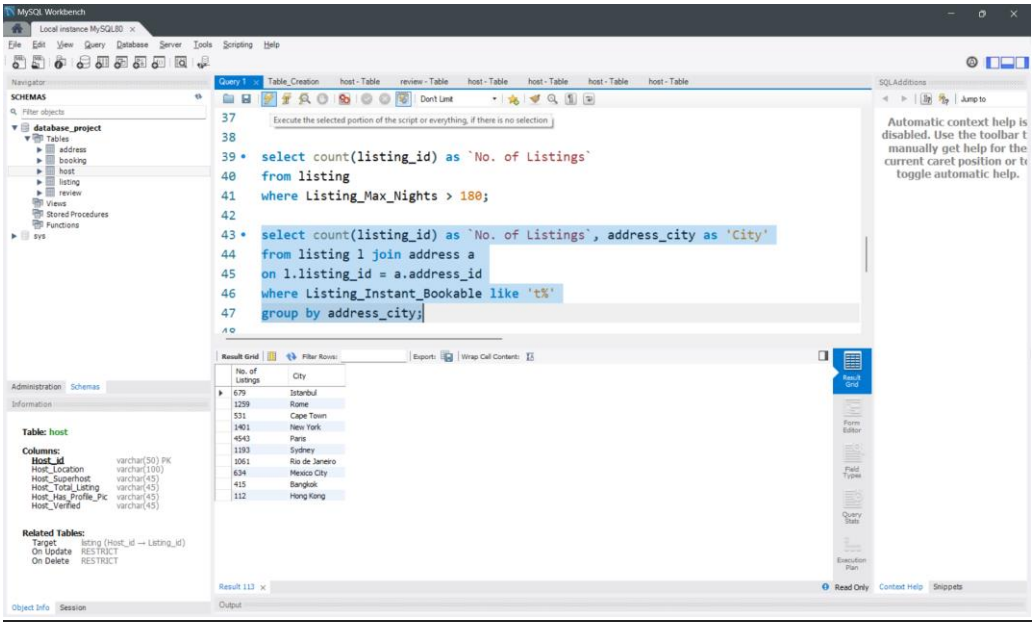
Almost 43k listings allow guests to stay for longer than 6 months. Airbnb should see that hosts are open to longer stays, and hence incentive guests into booking for longer durations.

Insight #7: How many listings in each city are instantly bookable?

Code:

```
select count(listing_id) as `No. of Listings`, address_city as 'City'
from listing l join address a
on l.listing_id = a.address_id
where Listing_Instant_Bookable like 't%'
group by address_city;
```

MySQL Workbench:



Conclusion:

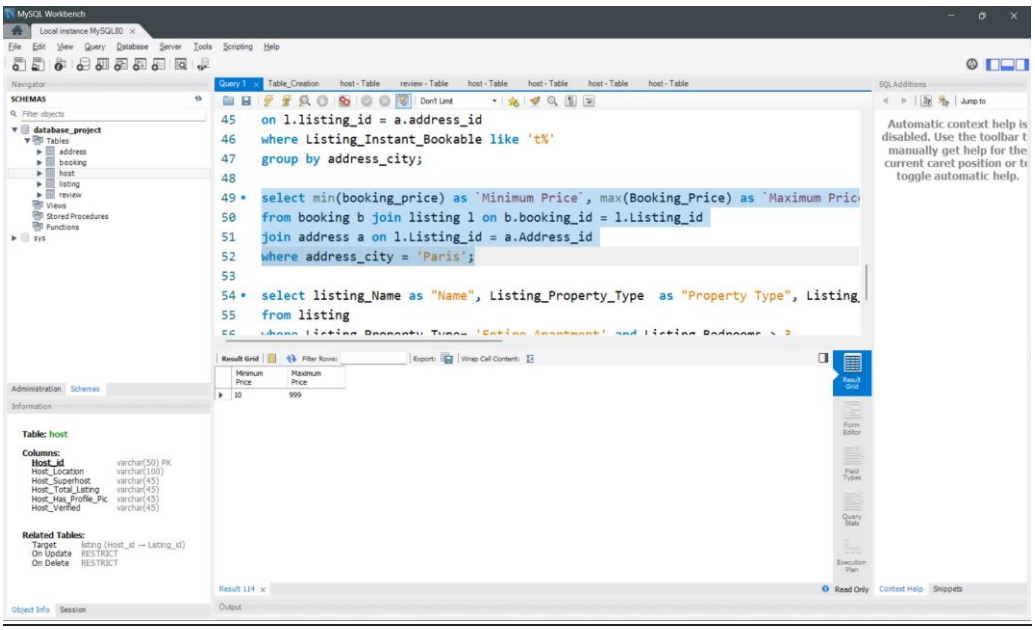
The output shows the number of listings that are instantly bookable in each city. We can see that this number is not very high for cities like Paris, New York etc. Airbnb should incentivize the hosts to provide this feature so that the booking process is quicker and more convenient.

Insight #8: What is the range of the booking prices of listings in Paris?

Code:

```
select min(booking_price) as `Minimum Price`, max(Booking_Price) as `Maximum Price`
from booking b join listing l on b.booking_id = l.Listing_id
join address a on l.Listing_id = a.Address_id
where address_city = 'Paris';
```

MySQL Workbench:



Conclusion:

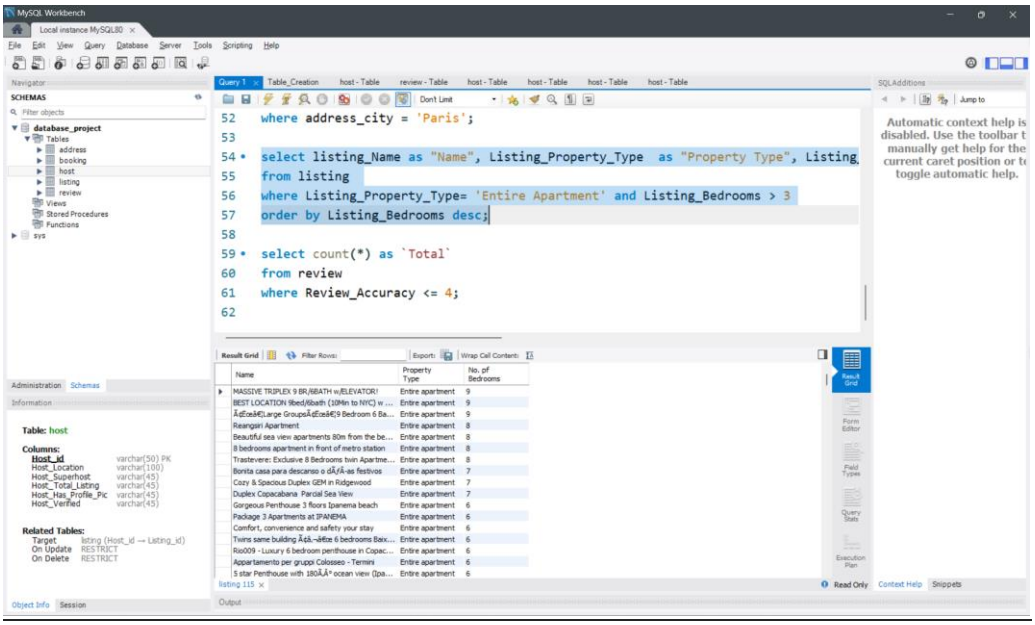
In Paris, a city with a very high number of listings, and a popular tourist destination- the booking price ranges from $10-$999. This shows that the city is open to all people within that range and offers a wide variety of experiences to the guests. This should be encouraged at all popular travel destinations.

Insight #9: How many listings are there where the entire place is available and has more than 3 bedrooms? How appealing is Airbnb to larger groups?

Code:

```
select listing_Name as "Name", Listing_Property_Type  as "Property Type", Listing_Bedrooms as `No. pf Bedrooms`
from listing
where Listing_Property_Type= 'Entire Apartment' and Listing_Bedrooms > 3
order by Listing_Bedrooms desc;
```

MySQL Workbench:



Conclusion:

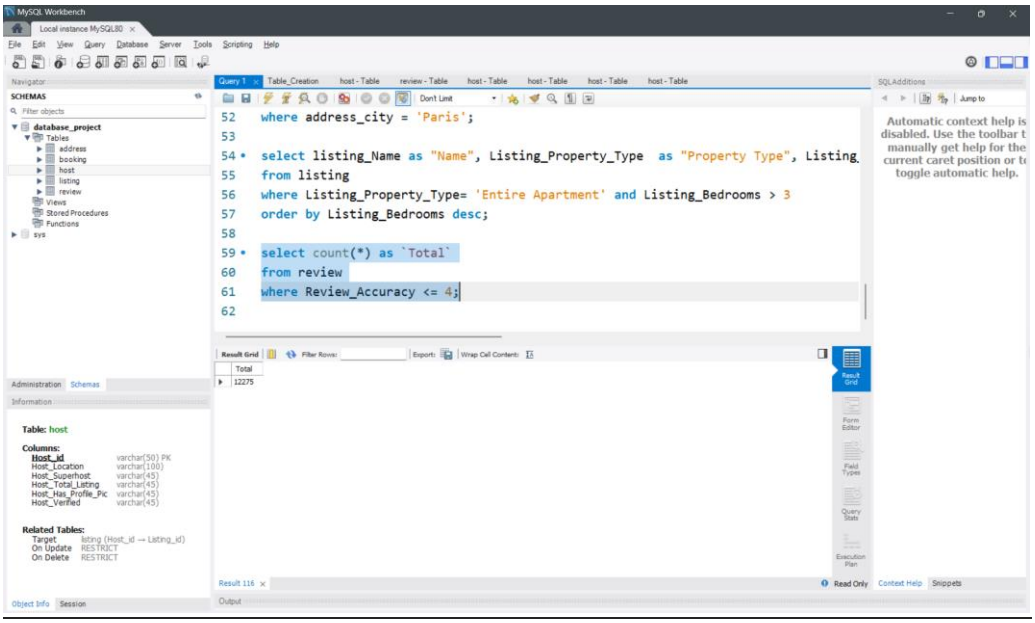
115 listings are there that are available with these features. In case there is a demand for larger properties, Airbnb should incentivize people with larger properties to become hosts, so it can attract bigger parties of people.

Insight #10: How many listings have reviews with review accuracy being rated less than or equal to only 40%?

Code:

```
select count(*) as `Total`
from review
where Review_Accuracy <= 4;
```

MySQL Workbench:

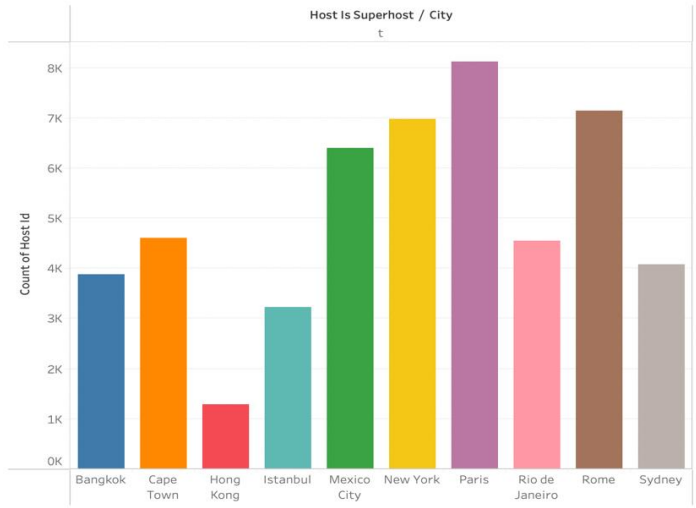


Conclusion:

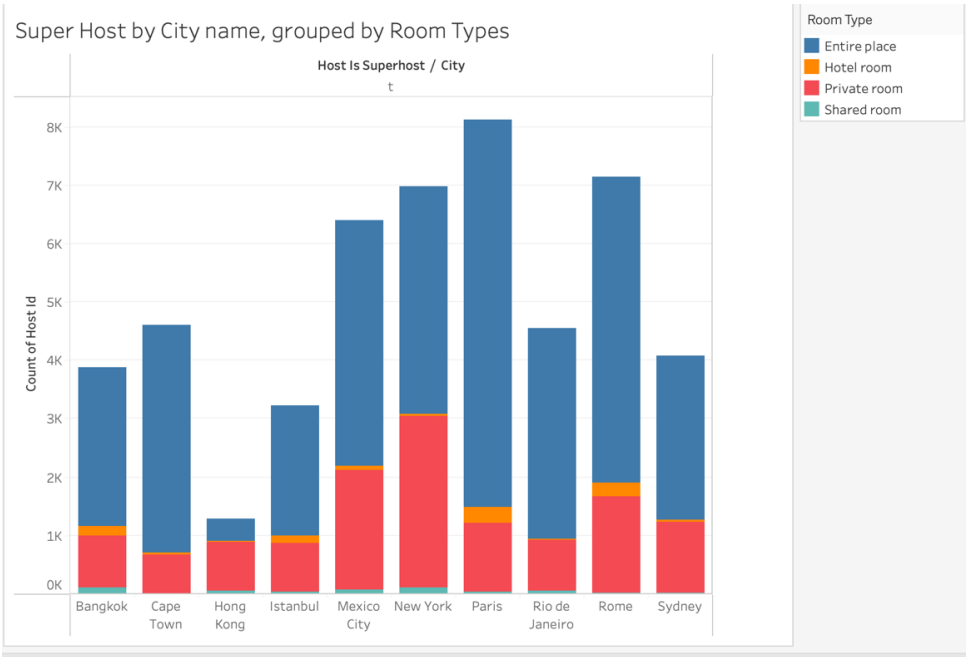
There are around 12,200 listings where the review for accuracy is low. This is not a good look for Airbnb as customers will feel the descriptions are misleading. The company should look into these listings and ask the hosts to be transparent to avoid dissatisfaction, refunds, and cancellations.

VISUALIZATION CHARTS

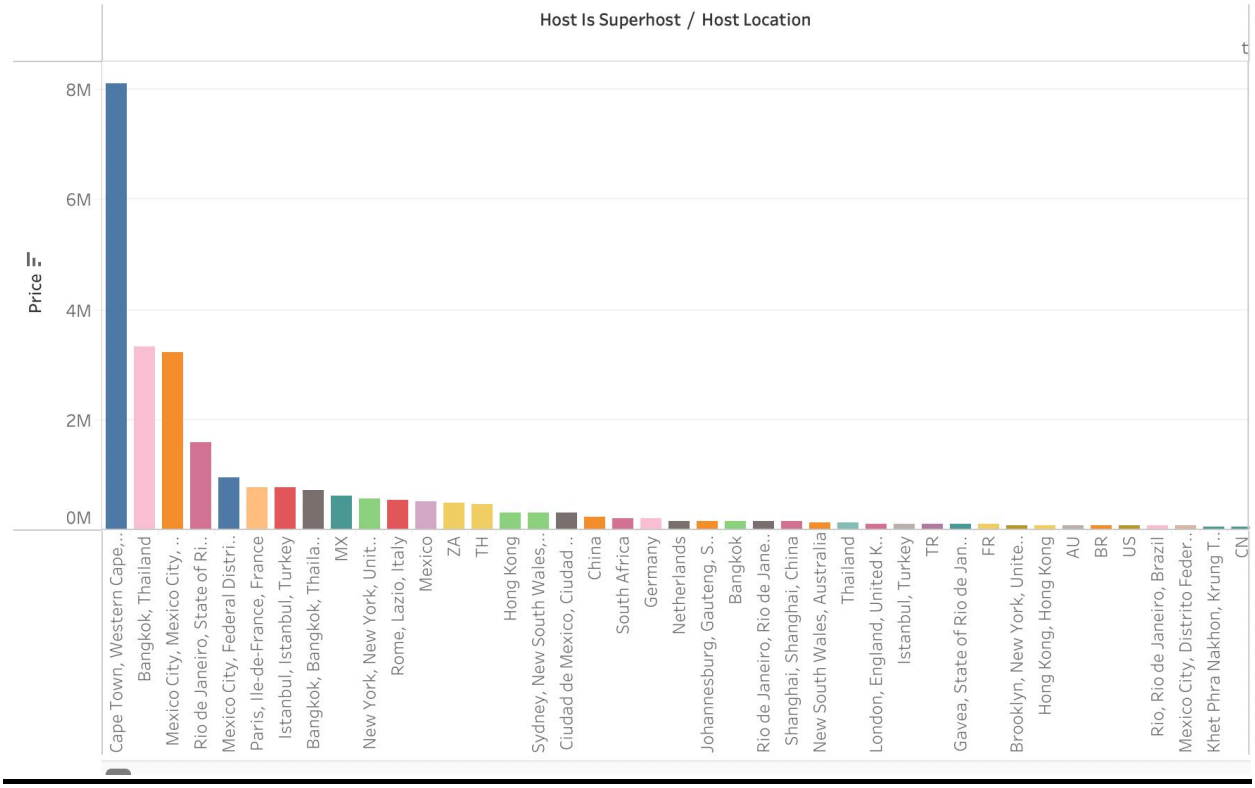
Number of Superhost in each city (Filtered by: Host has Profile Pic = T)



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:

https://cometmail-my.sharepoint.com/personal/axs220151_utdallas_edu/_layouts/15/stream.aspx?id=%2Fpersonal%2Faxs220151%5Futdallas%5Fedu%2FDocuments%2FDBProjectPresentation%2Emp4&referrer=Teams%2ETEAMS%2DELECTRON&referrerScenario=p2p%5Fns%2Dbim&ga=1