

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

Sr. No.	Title	Pg. No.
1	Project Description	3
2	Data Description	4
3	Logical and Physical Data Model	6
4	Data Loading Concept	9
5	Insights	12
6	Visualization Charts	22
7	References	24

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 2020.

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:

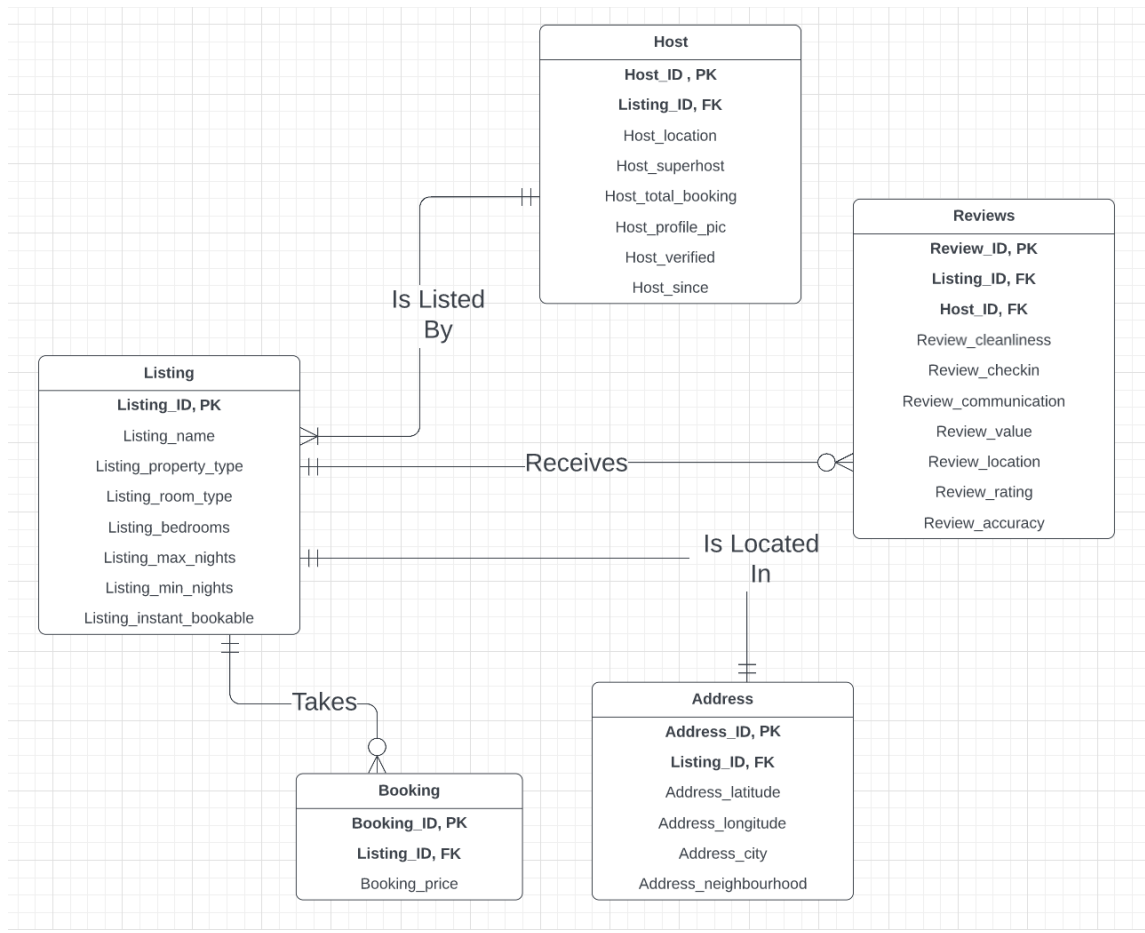
1. Listing ID
2. Name
3. Host ID
4. Host Location
5. Host/ Superhost
6. Total Listing by Host
7. Host has Profile Picture
8. Verified Host Identity
9. Neighbourhood
10. City
11. Latitude
12. Longitude
13. Property Type
14. Room Type
15. Bedroom
16. Price
17. Minimum Nights

- 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:

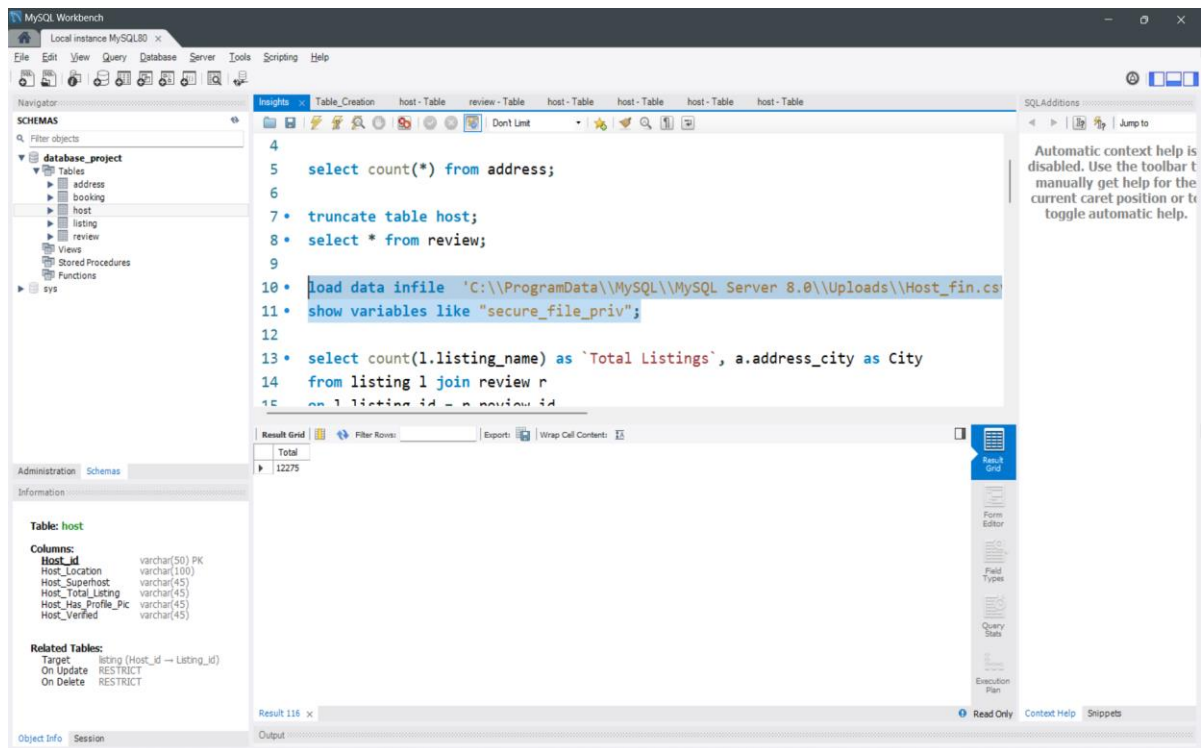
```
CREATE TABLE `review` (  
  `Review_id` varchar(100) NOT NULL,  
  `Review_Cleanliness` varchar(45) DEFAULT NULL,  
  `Review_Communication` varchar(45) DEFAULT NULL,  
  `Review_Checkin` varchar(45) DEFAULT NULL,  
  `Review_Value` varchar(45) DEFAULT NULL,  
  `Review_Location` varchar(45) DEFAULT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

```
8 • select * from host;
9
10 • load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Host_fin.csv'
11 • show variables like "secure_file_priv";
12
```

Result Grid

Host_Id	Host_Location	Host_Superhost	Host_Total_Listing	Host_Has_Profile_Pic	Host_Verified
1000370	United States	f	1	t	f
10001022		f	1	t	f
10002858	Turkey	t	1	t	t
10003514	Italy	f	1	t	t
10003762	France	f	1	t	f
10004368		f	1	t	f
10004492	France	f	1	t	t
10005065	Italy	f	1	t	t
10005780	France	f	1	t	t
10006270	South Africa	f	1	t	t
10006375	Turkey	f	1	t	f
10007224	France	f	1	t	t
10007317	France	f	1	t	t
10007653	Brazil	f	20	t	t
10007788	United States	f	2	t	t
10008627	Brazil	f	1	t	t
10009121	France	f	1	t	t
10009999	Brazil	f	1	t	f

host 102 x

Output

Review Table:

```
8 • select * from review;
9
10 • load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Host_fin.csv'
11 • show variables like "secure_file_priv";
12
```

Result Grid

Review_Id	Review_Cleanliness	Review_Communication	Review_Checkin	Review_Value	Review_Location	Review_Rating	Review_Accuracy
22705737	10	10	10	10	10	98	10
22706825							
22709839	10	10	10	9	9	100	10
22710069	10	10	10	10	8	100	10
22710028	9	10	9	9	9	93	10
22711350	10	10	10	10	10	99	10
22713093	9	10	10	9	10	98	10
22713544	10	10	10	10	10	93	10
22714041	10	10	10	10	9	97	10
2271504							
22715906	10	10	10	10	10	98	10
22715539	9	10	10	10	10	100	10
22717392	8	10	9	9	10	87	9
22718435	10	10	10	10	10	100	10
22719008	10	10	10	10	10	99	10
22720047							
22721120	9	9	9	9	10	93	9
22721136	9	9	9	9	10	88	10

review 106 x

Output

Listing Table:

```
8 • select * from listing;
9
10 • load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Host_fin.csv'
11 • show variables like "secure_file_priv";
12
```

Result Grid

me	Listing_Property_Type	Listing_Room_Type	Listing_Bedrooms	Listing_Max_Nights	Listing_Instant_Bookable	Listing_Min_Nights
burbs/central/bondi	Entire apartment	Entire place	2	1125	f	3
om Apartment In Astoria	Entire apartment	Entire place	1	1125	f	30
Close to Taxim Square	Entire apartment	Entire place	1	1125	t	6
ment in the green	Entire apartment	Entire place	2	1125	f	1
Apartment 15e - trÃ¡Ã¡'s proche T...	Entire apartment	Entire place	2	1125	f	4
Ã¡'s de Montmartre	Entire apartment	Entire place	1	1125	f	1
et Paris Place	Entire apartment	Entire place	1	1125	f	1
me - Uptown Trastevere	Entire apartment	Entire place	1	1125	t	3
at closed to Montmartre	Entire apartment	Entire place	1	1125	f	4
exquisite sea views, discount long s...	Entire apartment	Entire place	1	1125	t	3
ice house by the channel and sea	Entire apartment	Entire place	2	1125	t	1
urious Parisian apartment	Entire apartment	Entire place	4	1125	f	1
LUXURY APP - PARC MONCEAU	Entire apartment	Entire place	3	1125	f	3
duplex Penthouse 4 BDRS	Entire apartment	Entire place	4	1125	f	3
x 1 br near 2/3/4/5 Trans	Entire apartment	Entire place	1	1125	f	30
to moderno e funcional, Silencioso	Entire apartment	Entire place	3	1125	f	3
huit proche de Nation	Entire apartment	Entire place	1	1125	f	1

listing 105 x

Output

Booking Table:

8 • `select * from booking;`
 9
 10 • `load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Host_fin.csv'`
 11 • `show variables like "secure_file_priv";`
 12

Table: host

Columns:

Column Name	Data Type
Host_Id	varchar(50) PK
Host_Location	varchar(100)
Host_Superhost	varchar(45)
Host_Total_Listing	varchar(45)
Host_Has_Profile_Pic	varchar(45)
Host_Verified	varchar(45)

Related Tables:

Target	Relationship
listing	listing (Host_id → Listing_id)
On Update	RESTRICT
On Delete	RESTRICT

Result Grid

Booking_Id	Booking_Price
10000070	45
1000002	49
10000370	180
1000039	200
10000557	219
10000720	500
10000742	930
10001022	105
10001231	145
10001394	207
10001507	181
1000224	200
10002333	500
10002502	132
10002575	190
10002609	90
10002834	60
10002858	715

booking 104 x

Address Table:

8 • `select * from address;`
 9
 10 • `load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\Host_fin.csv'`
 11 • `show variables like "secure_file_priv";`
 12

Table: host

Columns:

Column Name	Data Type
Host_Id	varchar(50) PK
Host_Location	varchar(100)
Host_Superhost	varchar(45)
Host_Total_Listing	varchar(45)
Host_Has_Profile_Pic	varchar(45)
Host_Verified	varchar(45)

Related Tables:

Target	Relationship
listing	listing (Host_id → Listing_id)
On Update	RESTRICT
On Delete	RESTRICT

Result Grid

Address_Id	Address_Latitude	Address_Longitude	Address_City	Address_Neighbourhood
10000070	40.65048	-73.96752	New York	Flatbush
1000002	40.85295	-73.93361	New York	Washington Heights
10000370	-33.8862	151.24245	Sydney	Woollehra
1000039	41.90461	12.48283	Rome	I Centro Storico
10000557	19.4063	-99.16222	Mexico City	Cuauhtemoc
10000720	13.74798	100.49778	Bangkok	Pkra Nakhon
10000742	13.75534	100.52742	Bangkok	Ratchathewi
10001022	40.76047	-73.92288	New York	Astoria
10001231	-33.80275	151.2887	Sydney	Manly
10001394	19.39954	-99.17655	Mexico City	Miguel Hidalgo
10001507	22.28752	114.14208	Hong Kong	Central & Western
1000224	48.8571	2.33694	Paris	Luxembourg
10002333	-33.74124	151.27116	Sydney	Warringah
10002502	22.30087	114.17571	Hong Kong	Yau Tsim Mong
10002575	-33.88842	151.26512	Sydney	Waverley
10002609	-33.87905	151.21301	Sydney	Sydney
10002834	41.90681	12.46152	Rome	I Centro Storico
10002858	41.04654	78.18487	Tshwivil	Sol

address 103 x

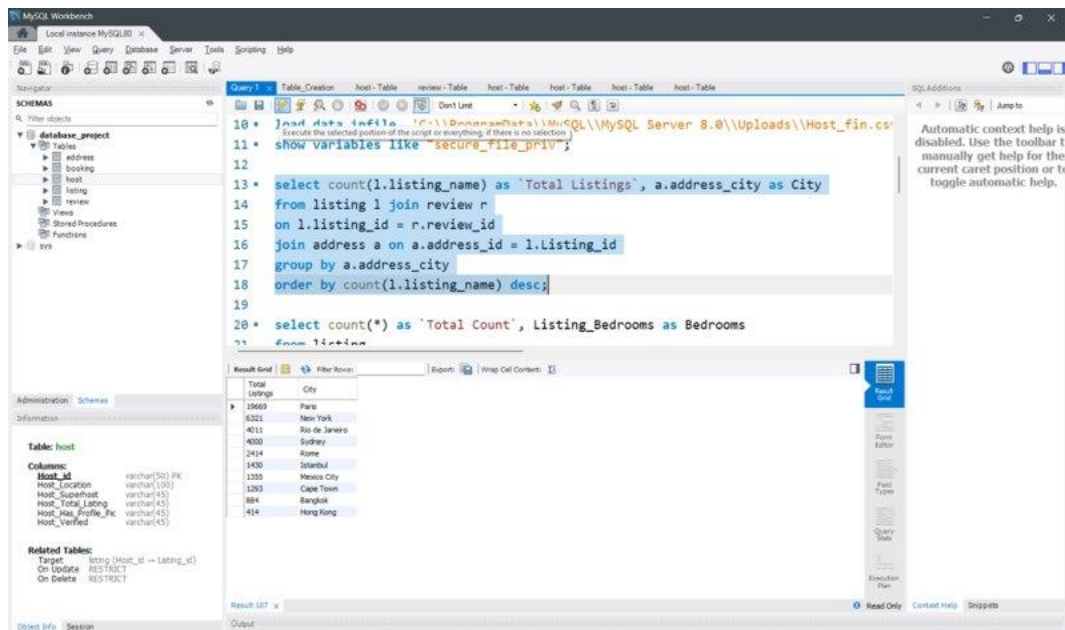
INSIGHTS

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:

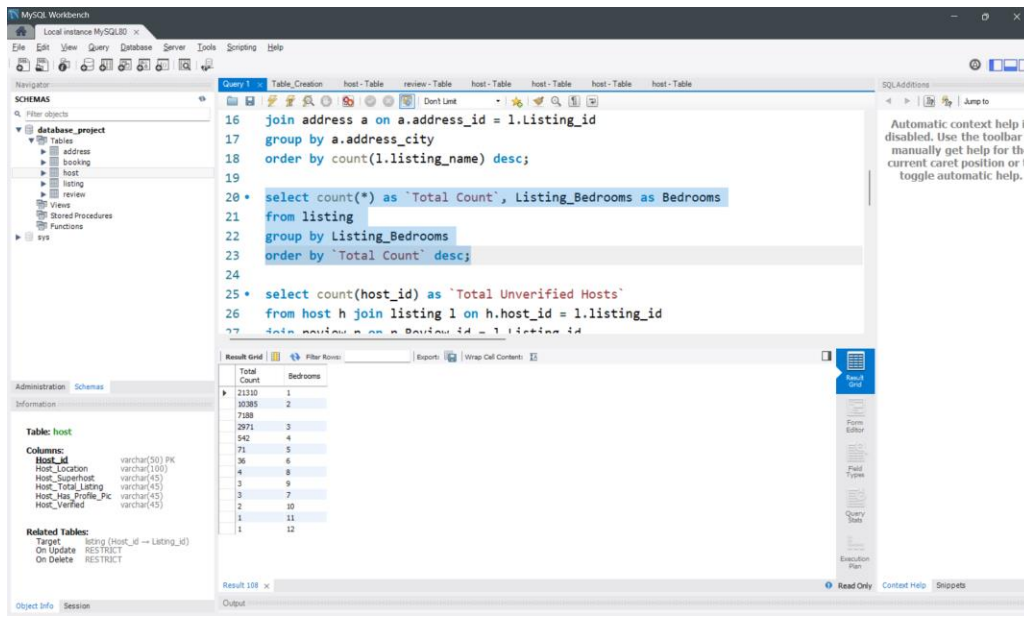
As seen in the output, the city with the highest number of listings is Paris, with more than 19k listings. This is followed by New York, Rio de Janeiro, Sydney, and more. With this insight, Airbnb can collaborate with brands or tourism destinations or airlines that frequently land in Paris, and make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



The screenshot shows the MySQL Workbench interface. The SQL editor contains the following query:

```
16 join address a on a.address_id = l.listing_id
17 group by a.address_city
18 order by count(l.listing_name) desc;
19
20 * select count(*) as `Total Count`, Listing_Bedrooms as Bedrooms
21 from listing
22 group by Listing_Bedrooms
23 order by `Total Count` desc;
24
25 * select count(host_id) as `Total Unverified Hosts`
26 from host h join listing l on h.host_id = l.listing_id
27 join reviews r on r.reviewer_id = l.listing_id
```

The Results grid shows the output of the query:

Total Count	Bedrooms
21110	1
10385	2
7188	3
2971	4
542	5
71	6
36	7
4	8
3	9
3	10
2	11
1	12

The left sidebar shows the Schemas pane with the database project structure. The bottom pane shows the Table structure for the 'host' table, including columns like Host_id, Host_Location, Host_Superhost, Host_Total_Listing, Host_Has_Profile_Pic, and Host_Verified.

Conclusion:

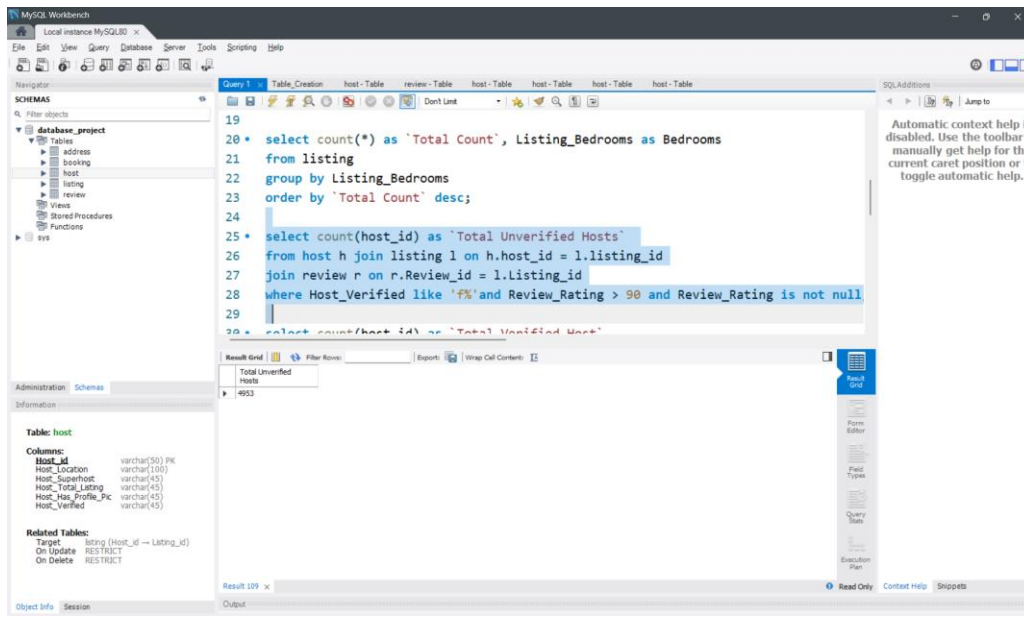
As seen in the output, it is concluded that more than 21k of the properties have only 1 bedroom, followed by 10k properties having 2 bedrooms. This signifies that Airbnb has a lot of options for singles or couples travelling together, and can advertise to these parties, while also working towards expanding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

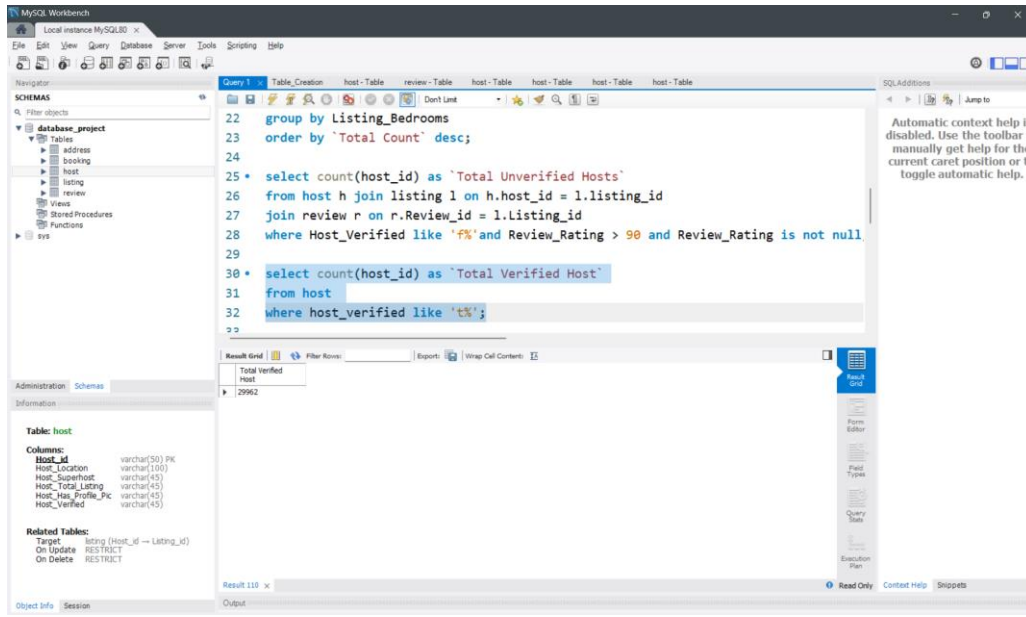
There are almost 5k unverified hosts that have an overall rating of more than 90%. Airbnb should reach out to them and urge them to get verified, so their profile can be more attractive and their bookings can increase, generating more commission for the company.

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:

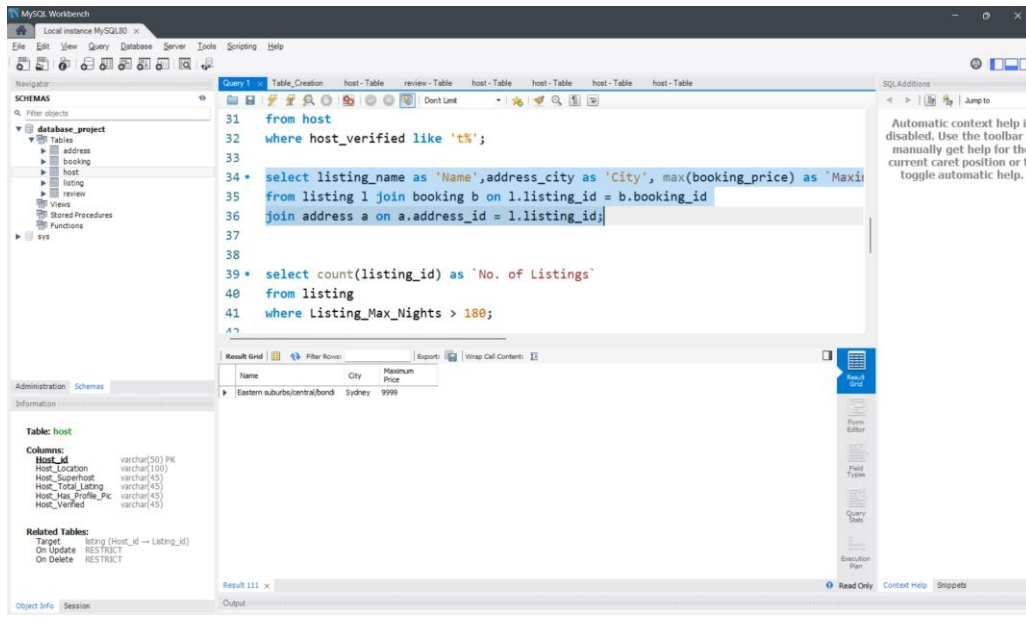
There are almost 30k verified hosts on Airbnb. To improve this number, Airbnb should give hosts an incentive to get verified, add a profile picture, and do other things that make their listings and profiles more appealing.

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:

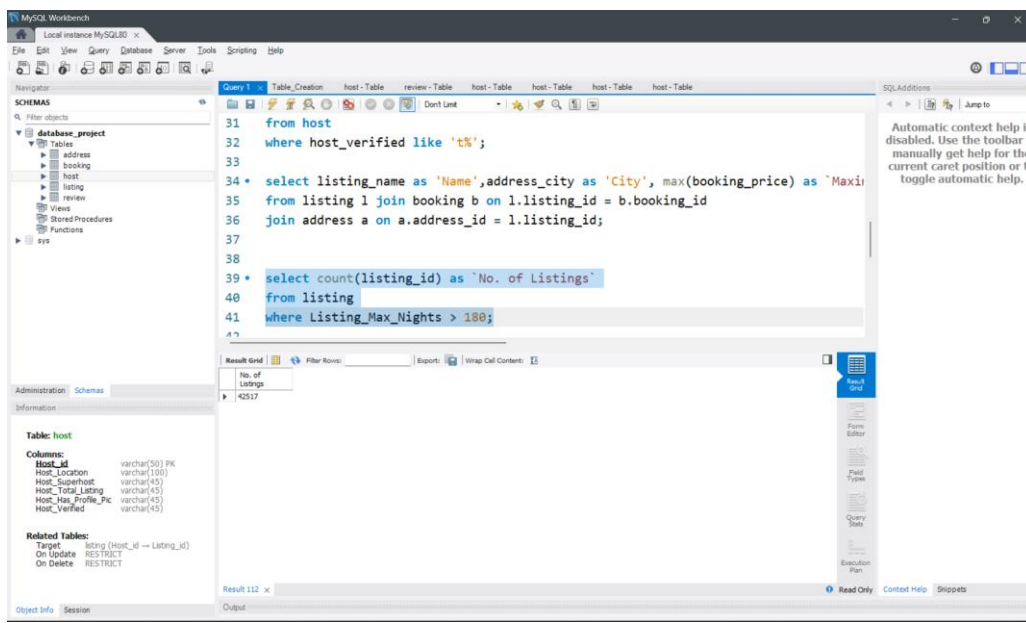
The maximum booking price in the data is \$9999. This information can be used to study the location and guests of this booking price's listing, and use it to increase the number of such listings and guests, respectively.

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:

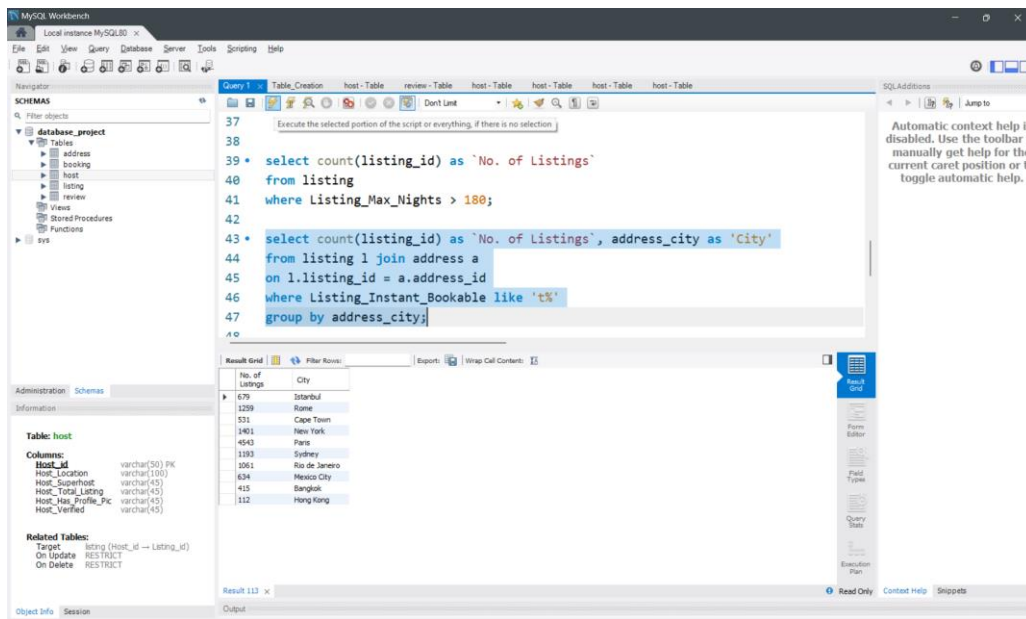
Almost 43k listings allow guests to stay for longer than 6 months. Airbnb should see that hosts are open to longer stays, and hence incentive guests into booking for longer durations.

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:

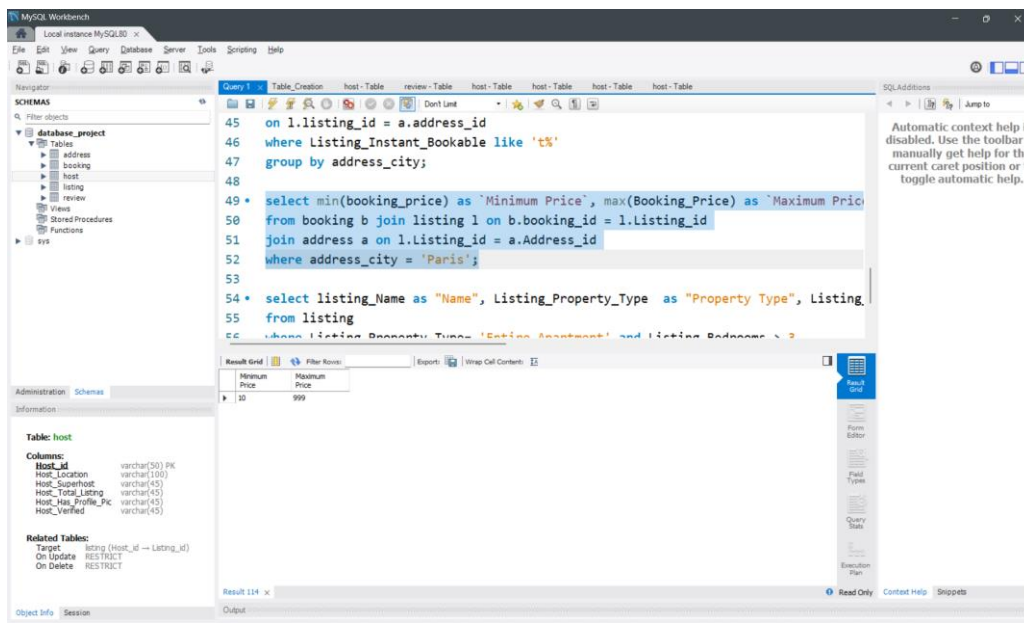
The output shows the number of listings that are instantly bookable in each city. We can see that this number is not very high for cities like Paris, New York etc. Airbnb should incentivize the hosts to provide this feature so that the booking process is quicker and more convenient.

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:

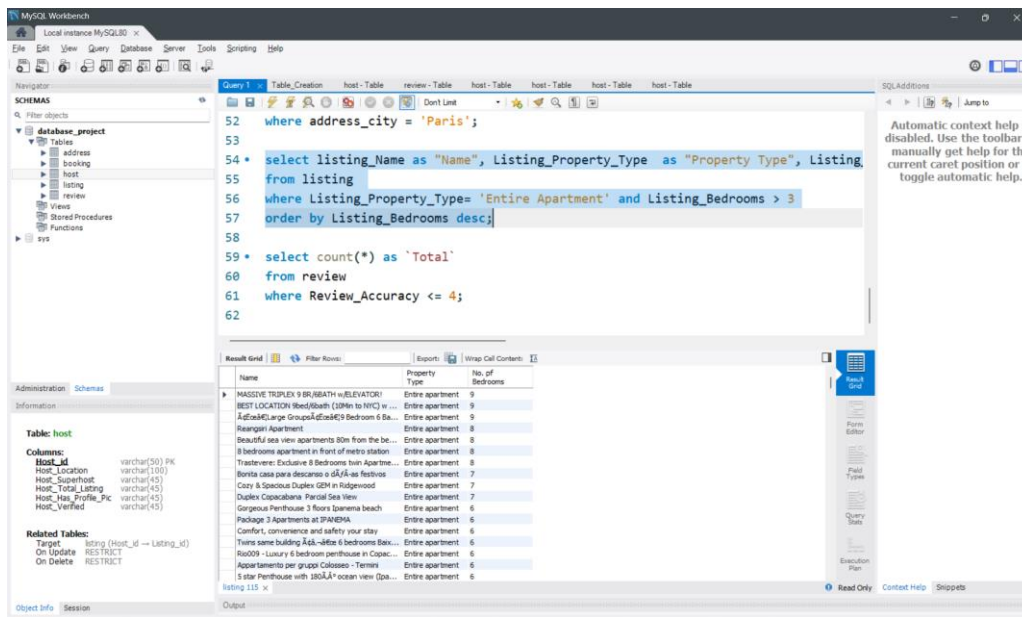
In Paris, a city with a very high number of listings, and a popular tourist destination- the booking price ranges from \$10-\$999. This shows that the city is open to all people within that range and offers a wide variety of experiences to the guests. This should be encouraged at all popular travel destinations.

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. of Bedrooms`  
from listing  
where Listing_Property_Type= 'Entire Apartment' and Listing_Bedrooms > 3  
order by Listing_Bedrooms desc;
```

MySQL Workbench:



The screenshot shows the MySQL Workbench interface. The central pane displays a SQL query for 'Query 1'. The query filters for listings in Paris that are 'Entire Apartment' type with more than 3 bedrooms, ordered by the number of bedrooms in descending order. It also includes a count of these listings. The bottom pane shows the 'Result Grid' with 115 rows of data. The columns are 'Name', 'Property Type', and 'No. of Bedrooms'. The first few rows show properties like 'MASSIVE TRIPLEX 9 BR/4BATH w/ ELEVATOR!' and 'BEST LOCATION 1bed/1bath (30min to NYC) w...'. The right sidebar shows the 'Table: host' structure with columns like 'Host_id', 'Host_Location', 'Host_Superhost', 'Host_Total_Listing', 'Host_Has_Profile_Pic', and 'Host_Verified'.

```
52 where address_city = 'Paris';  
53  
54 * select listing_Name as "Name", Listing_Property_Type as "Property Type", Listing  
55 from listing  
56 where Listing_Property_Type= 'Entire Apartment' and Listing_Bedrooms > 3  
57 order by Listing_Bedrooms desc;  
58  
59 * select count(*) as 'Total'  
60 from review  
61 where Review_Accuracy <= 4;  
62
```

Name	Property Type	No. of Bedrooms
MASSIVE TRIPLEX 9 BR/4BATH w/ ELEVATOR!	Entire apartment	9
BEST LOCATION 1bed/1bath (30min to NYC) w...	Entire apartment	9
Agloa&Large Group&Large 9 Bedrooms 6 Ba...	Entire apartment	9
Reagan's Apartment	Entire apartment	8
Beautiful sea view apartment 80m from the be...	Entire apartment	8
8 bedrooms apartment in front of metro station	Entire apartment	8
Travelers' Exclusive 8 Bedrooms Inn Apartme...	Entire apartment	8
Bonita casa para descansar o disfrutar festiv...	Entire apartment	7
Cosy & Spacious Duplex GEM in Ridgewood	Entire apartment	7
Duplex Capotaormina Partial Sea View	Entire apartment	7
Gorgeous Penthouse 3 Rooms Taormina beach	Entire apartment	6
Package 3 Apartments at IPANEMA	Entire apartment	6
Comfort, convenience and safety your stay	Entire apartment	6
Three same building 4bd-5bdr 6 bedrooms Bar...	Entire apartment	6
Rio009 - Luxury 6 bedroom penthouse in Copac...	Entire apartment	6
Apartmento per grupo Colosseo - Termini	Entire apartment	6
5 star Penthouse with 2000sqm ocean view (Pa...	Entire apartment	6
listing 115		

Conclusion:

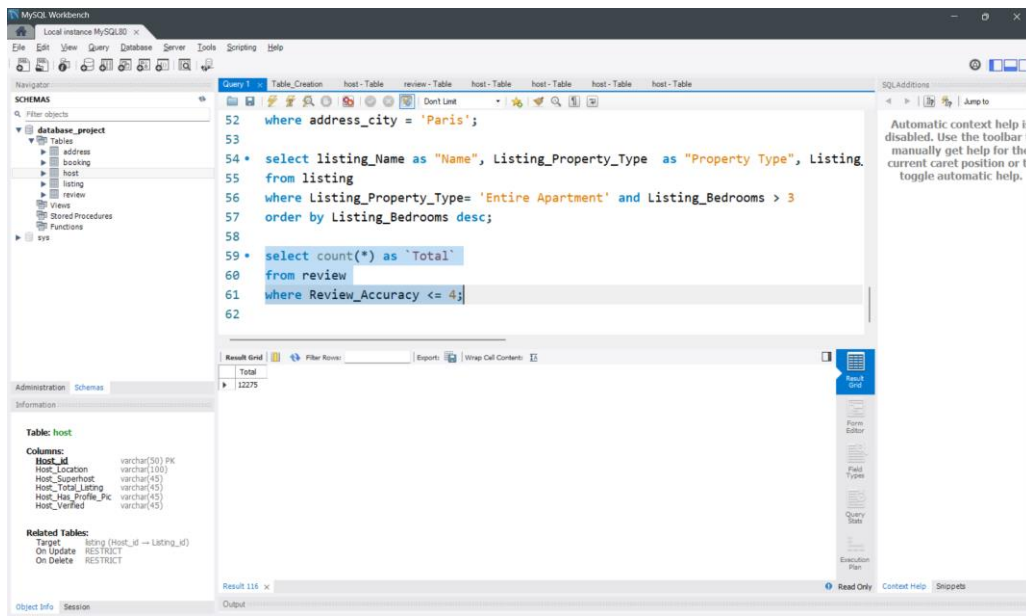
115 listings are there that are available with these features. In case there is a demand for larger properties, Airbnb should incentivize people with larger properties to become hosts, so it can attract bigger parties of people.

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:

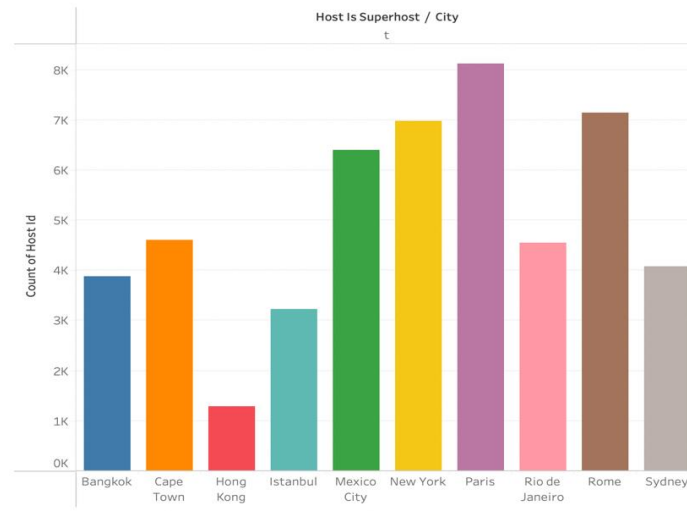


Conclusion:

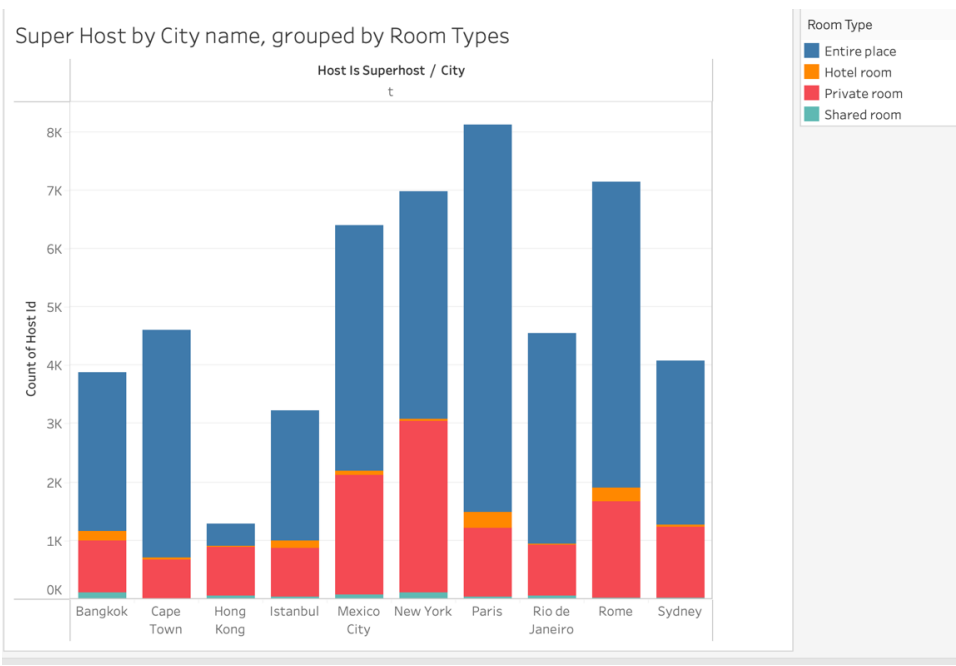
There are around 12,200 listings where the review for accuracy is low. This is not a good look for Airbnb as customers will feel the descriptions are misleading. The company should look into these listings and ask the hosts to be transparent to avoid dissatisfaction, refunds, and cancellations.

VISUALIZATION CHARTS

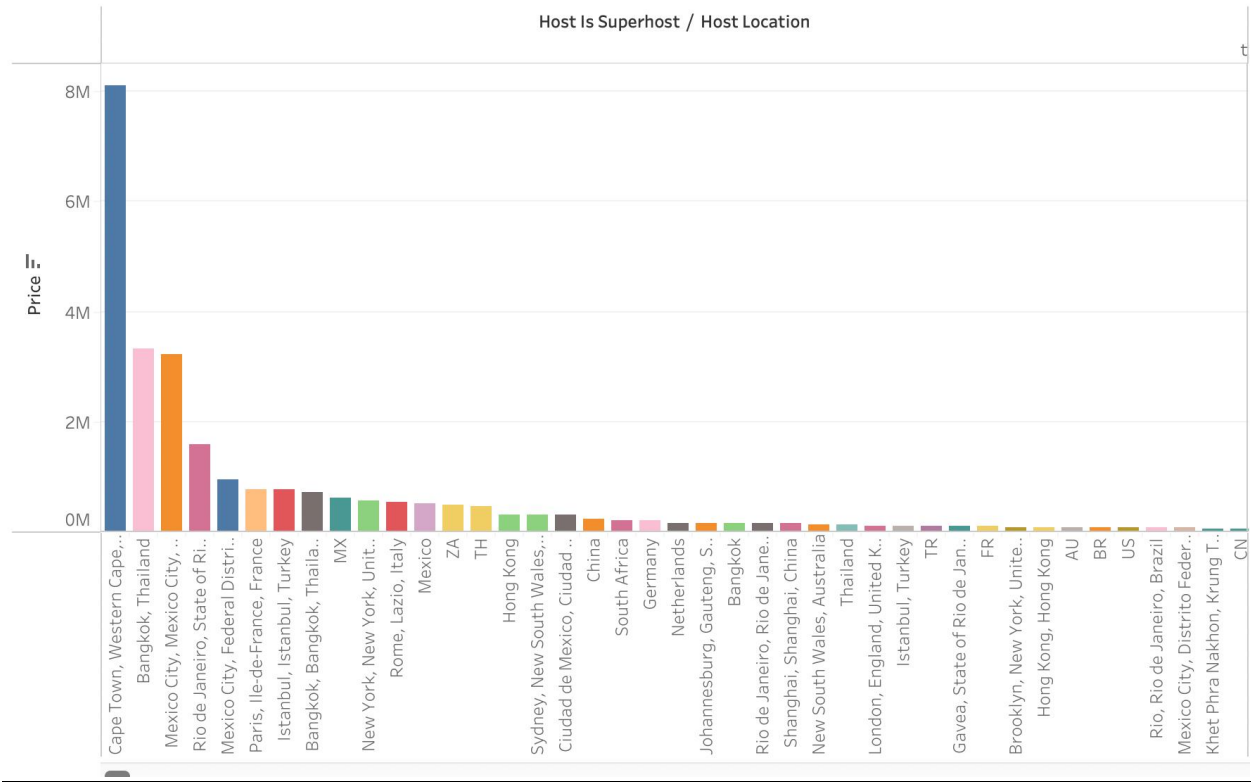
Number of Superhost in each city (Filtered by: Host has Profile Pic = T)



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_uhdallas_edu/_layouts/15/stream.aspx?id=%2Fpersonal%2Faxs220151%5Futdallas%5Fedu%2FDocuments%2FDBProjectPresentation%2Emp4&referrer=Teams%2ETEAMS%2DELECTRON&referrerScenario=p2p%5Fns%2Dbim&ga=1