SQL Project

airbnb

Belong Anywhere

Group 007

Aashna, Charlene, Sandhya,
Udisha, Ziqi Liu

Agenda

- Business Scenario
- Swimlane Diagram
- ER Diagram & Dataset Creation
- SQL Queries (Basic & Advanced):
  - Popular Room Type
  - Highest Demand
  - Top 5 properties & Accommodations/Property
  - Payment method
  - Ratings & Amenities
- Tableau Dashboard
- Learnings & Conclusion

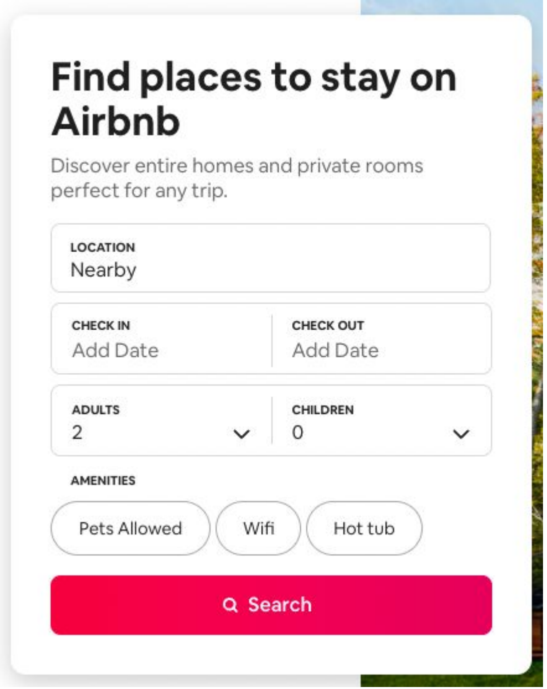
Business Scenario

Kenneth and his friends are going to have a trip to Hawaii during the winter holidays. They plan to go from December 21 to December 27, 2021. They're looking for a house that could fit in 8 people on the Airbnb website. Other than that, they also want to find a house with a balcony and additional amenities such as Wifi and pool. After adding these requirements, a list of choices are shown. They check the images, amenities, room types, reviews, and price range from a few options. They have a few questions regarding the heating situation in the room, so they message the host through the website at 10:30 pm on November 30. Once the host responds to their satisfaction, Kenneth requests for a reservation on the entire condo hosted by Ken on December 2 around 6pm, which has a 4.78 review score with 23 reviews. Each night costs $750 and the total cost for six nights is $5971 including cleaning and service fee before tax. Kenneth inputs his credit card details when requesting a reservation.

Ken receives a notice via Airbnb that there is a reservation request from December 21 to 28 on December 2. He accepts the reservation on the same day right after the notification, and the website deducts money from Kenneth's card. He makes a note on these dates, sends a greeting message to Kenneth and confirms the reservation within 24 hours.

After spending the holidays in Ken's house, Kenneth shared the great thanks and wrote a detailed review with 5 stars for this experience within 14 days. Ken also leaves Kenneth a 5-star review.

Airbnb Booking Process



Find places to stay on Airbnb

Discover entire homes and private rooms perfect for any trip.

LOCATION
Nearby

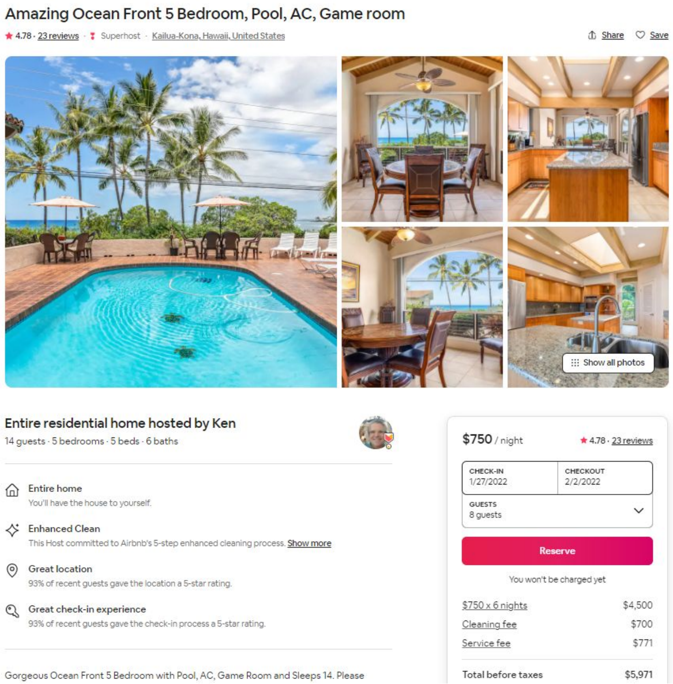
CHECK IN	CHECK OUT
Add Date	Add Date

ADULTS	CHILDREN
2	0

AMENITIES

Pets Allowed | Wifi | Hot tub

Search



Amazing Ocean Front 5 Bedroom, Pool, AC, Game room

★ 4.78 · 23 reviews · Superhost · Kailua-Kona, Hawaii, United States

Share | Save

Show all photos

Entire residential home hosted by Ken

14 guests · 5 bedrooms · 5 beds · 6 baths

Entire home
You'll have the house to yourself.

Enhanced Clean
This Host committed to Airbnb's 5-step enhanced cleaning process. Show more

Great location
93% of recent guests gave the location a 5-star rating.

Great check-in experience
93% of recent guests gave the check-in process a 5-star rating.

Gorgeous Ocean Front 5 Bedroom with Pool, AC, Game Room and Sleeps 14. Please

$750 / night ★ 4.78 · 23 reviews

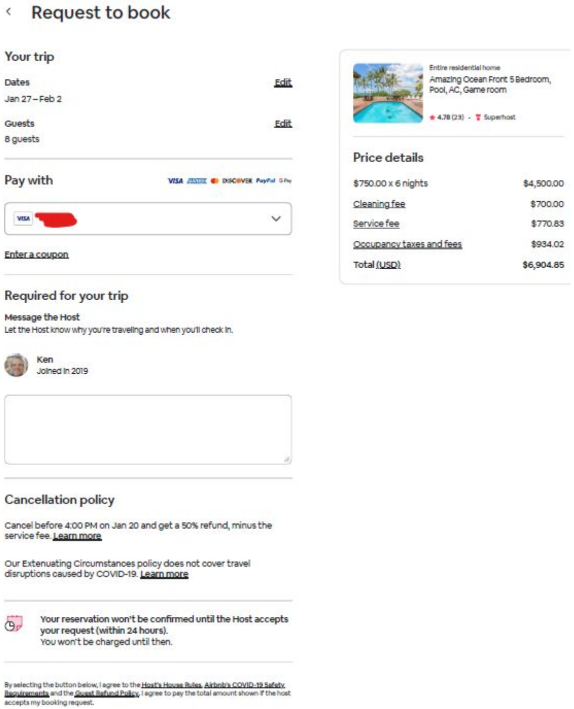
CHECK-IN	CHECKOUT
1/27/2022	2/2/2022

GUESTS
8 guests

Reserve

You won't be charged yet

$750 x 6 nights	$4,500
Cleaning fee	$700
Service fee	$771
Total before taxes	$5,971



Request to book

Your trip

Dates — Edit
Jan 27 – Feb 2

Guests — Edit
8 guests

Pay with

VISA [redacted]

Enter a coupon

Required for your trip

Message the Host
Let the Host know why you're traveling and when you'll check in.

Ken
Joined in 2019

Cancellation policy

Cancel before 4:00 PM on Jan 20 and get a 50% refund, minus the service fee. Learn more

Our Extenuating Circumstances policy does not cover travel disruptions caused by COVID-19. Learn more

Your reservation won't be confirmed until the Host accepts your request (within 24 hours).
You won't be charged until then.

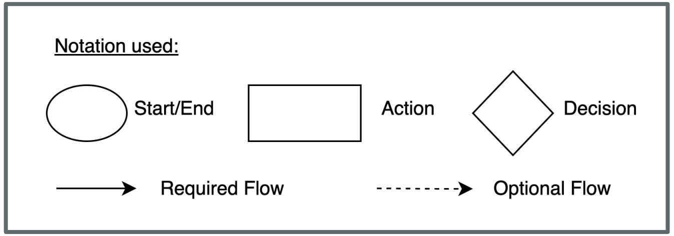
By selecting the button below, I agree to the Host's House Rules, Airbnb's COVID-19 Safety Requirements and the Guest Refund Policy. I agree to pay the total amount shown if the host accepts my booking request.

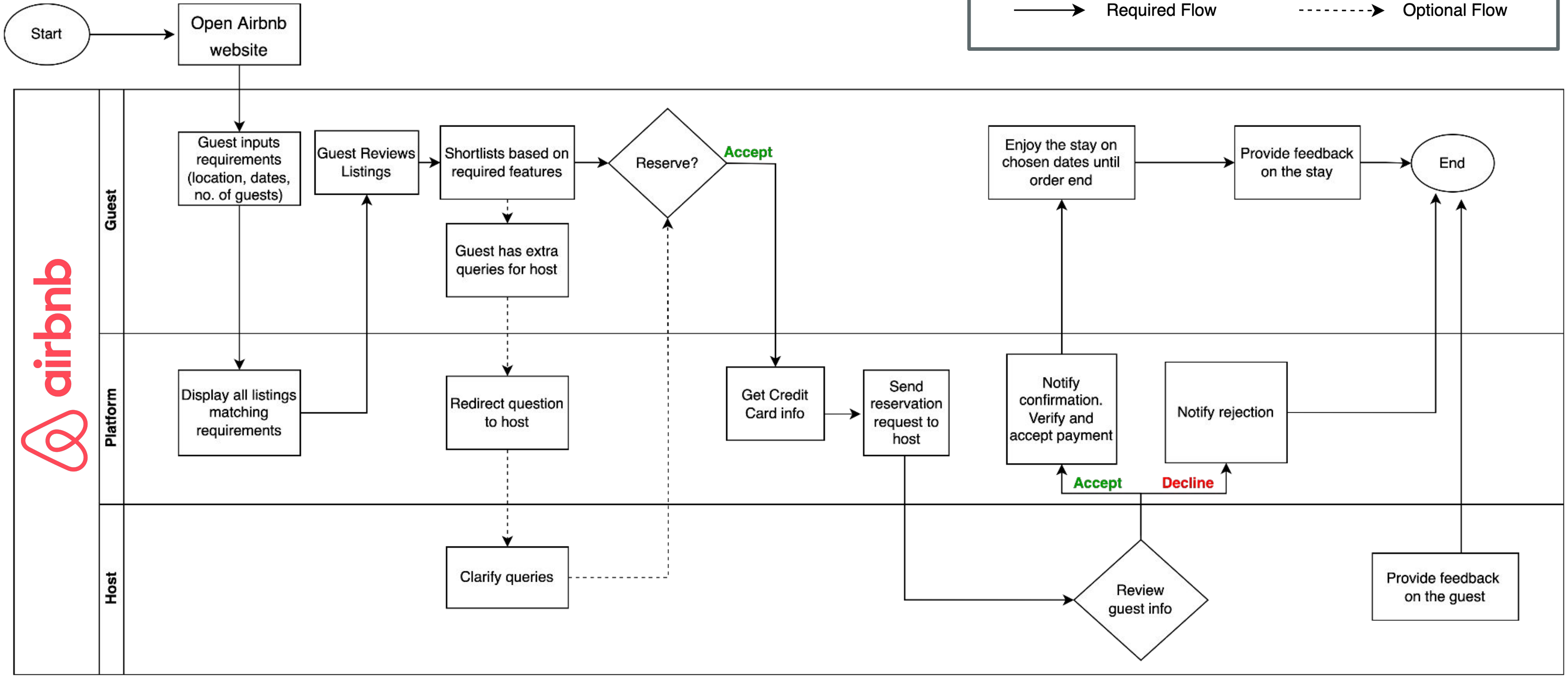
Entire residential home
Amazing Ocean Front 5 Bedroom, Pool, AC, Game room
★ 4.78 (23) · Superhost

Price details

$750.00 x 6 nights	$4,500.00
Cleaning fee	$700.00
Service fee	$770.83
Occupancy taxes and fees	$934.02
Total (USD)	$6,904.85

Swimlane Diagram




Start
Open Airbnb website
airbnb
Guest
Platform
Host
Guest inputs requirements (location, dates, no. of guests)
Guest Reviews Listings
Shortlists based on required features
Reserve?
Accept
Guest has extra queries for host
Enjoy the stay on chosen dates until order end
Provide feedback on the stay
End
Display all listings matching requirements
Redirect question to host
Get Credit Card info
Send reservation request to host
Notify confirmation. Verify and accept payment
Notify rejection
Accept
Decline
Clarify queries
Review guest info
Provide feedback on the guest

Tables Required

Customer

Order

Payment

Payment Method

Property

Host

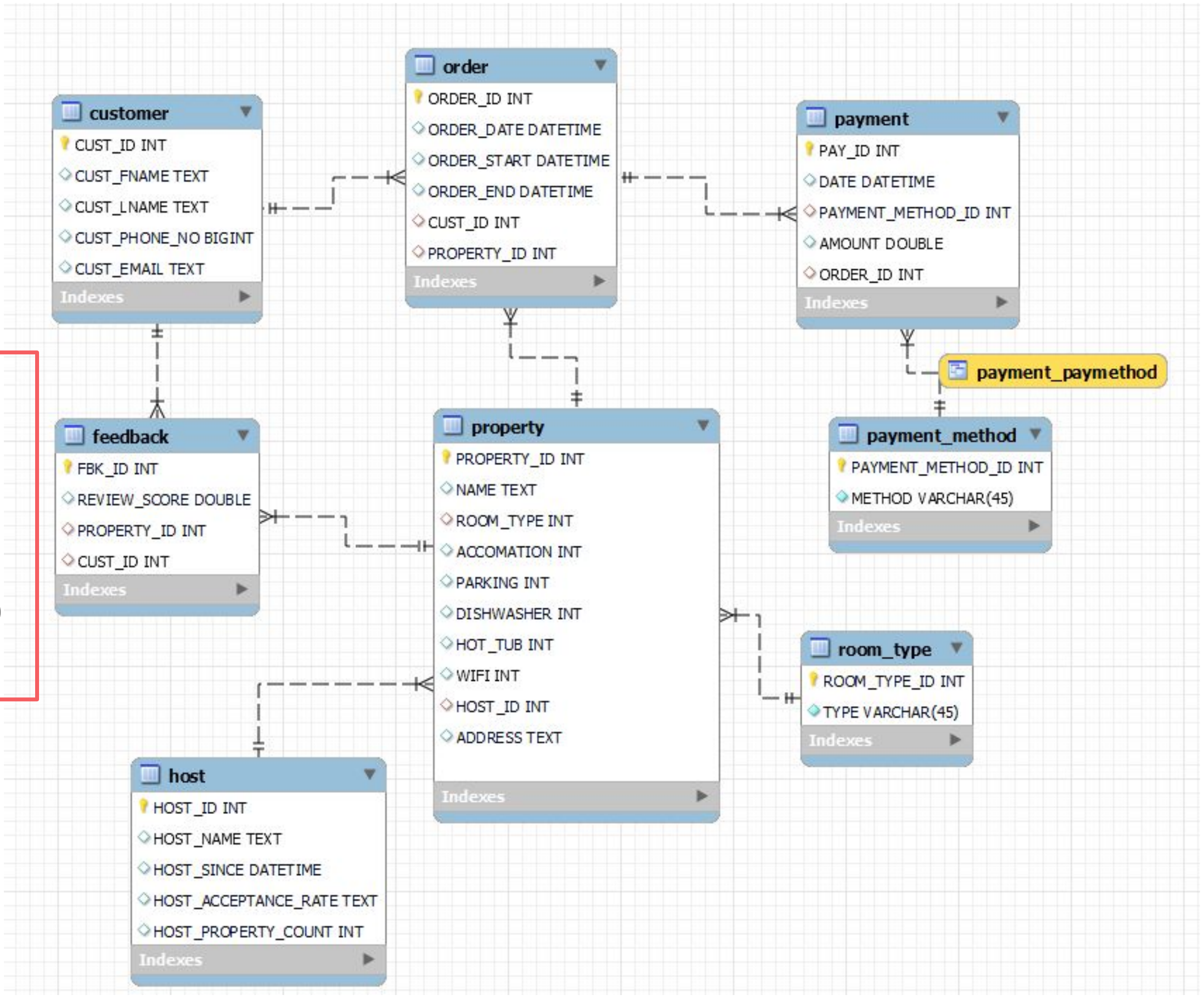
Feedback

Room type

Entity Relationship Diagram

Steps to create ER diagram when we already have a schema:

1. Click on <u>Database</u>
2. Choose <u>Reverse Engineer</u>
3. Select <u>Schemas</u> (airbnb)
4. Select <u>Objects</u> (tables and view)
5. Execute



customer
CUST_ID INT
CUST_FNAME TEXT
CUST_LNAME TEXT
CUST_PHONE_NO BIGINT
CUST_EMAIL TEXT
Indexes

order
ORDER_ID INT
ORDER_DATE DATETIME
ORDER_START DATETIME
ORDER_END DATETIME
CUST_ID INT
PROPERTY_ID INT
Indexes

payment
PAY_ID INT
DATE DATETIME
PAYMENT_METHOD_ID INT
AMOUNT DOUBLE
ORDER_ID INT
Indexes

payment_paymethod

feedback
FBK_ID INT
REVIEW_SCORE DOUBLE
PROPERTY_ID INT
CUST_ID INT
Indexes

property
PROPERTY_ID INT
NAME TEXT
ROOM_TYPE INT
ACCOMATION INT
PARKING INT
DISHWASHER INT
HOT_TUB INT
WIFI INT
HOST_ID INT
ADDRESS TEXT
Indexes

payment_method
PAYMENT_METHOD_ID INT
METHOD VARCHAR(45)
Indexes

room_type
ROOM_TYPE_ID INT
TYPE VARCHAR(45)
Indexes

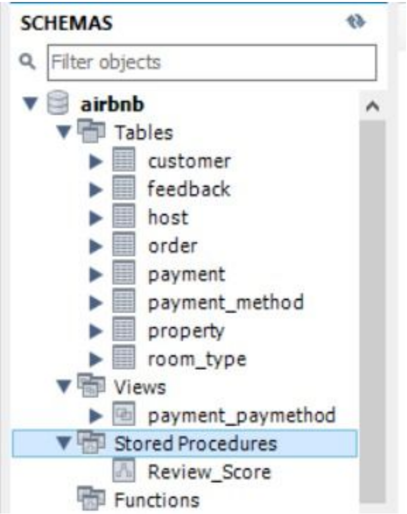
host
HOST_ID INT
HOST_NAME TEXT
HOST_SINCE DATETIME
HOST_ACCEPTANCE_RATE TEXT
HOST_PROPERTY_COUNT INT
Indexes

Database Creation

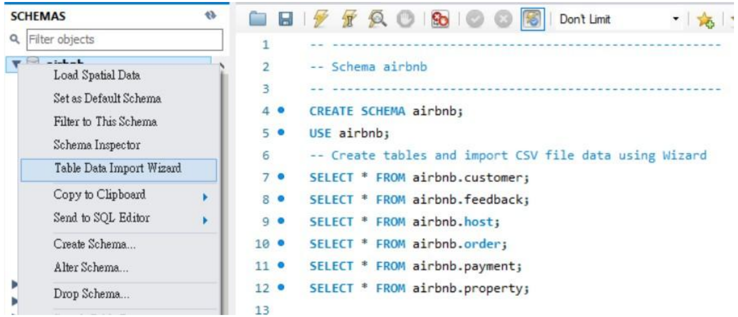
Steps to create Database:

1. Created data and csv tables - spreadsheet with tables
2. Create MySQL tables for csv import
3. Import csv in MySQL using “Table Data Import Wizard”
4. Select files to import, set the destination table and additional options
5. Configure import settings (assign appropriate data type)
6. Add Primary Key and Foreign Key for imported tables
7. For tables only have a few columns and row, creating the tables directly

Resulting schema and tables:

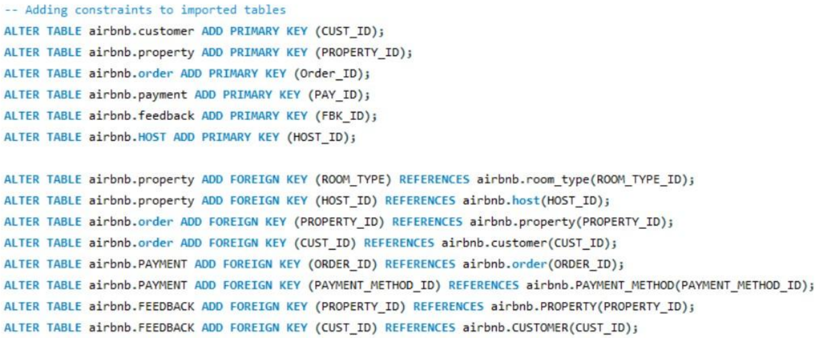


Step 3:



```
-- -----------------------------------------------------
-- Schema airbnb
-- -----------------------------------------------------
CREATE SCHEMA airbnb;
USE airbnb;
-- Create tables and import CSV file data using Wizard
SELECT * FROM airbnb.customer;
SELECT * FROM airbnb.feedback;
SELECT * FROM airbnb.host;
SELECT * FROM airbnb.order;
SELECT * FROM airbnb.payment;
SELECT * FROM airbnb.property;
```

Step 6:

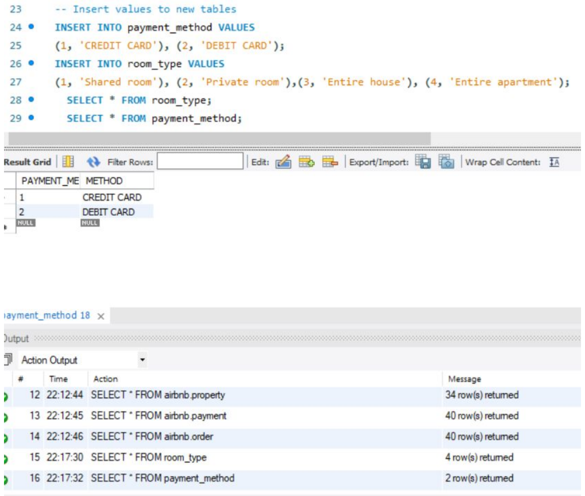


```
-- Adding constraints to imported tables
ALTER TABLE airbnb.customer ADD PRIMARY KEY (CUST_ID);
ALTER TABLE airbnb.property ADD PRIMARY KEY (PROPERTY_ID);
ALTER TABLE airbnb.order ADD PRIMARY KEY (Order_ID);
ALTER TABLE airbnb.payment ADD PRIMARY KEY (PAY_ID);
ALTER TABLE airbnb.feedback ADD PRIMARY KEY (FBK_ID);
ALTER TABLE airbnb.HOST ADD PRIMARY KEY (HOST_ID);

ALTER TABLE airbnb.property ADD FOREIGN KEY (ROOM_TYPE) REFERENCES airbnb.room_type(ROOM_TYPE_ID);
ALTER TABLE airbnb.property ADD FOREIGN KEY (HOST_ID) REFERENCES airbnb.host(HOST_ID);
ALTER TABLE airbnb.order ADD FOREIGN KEY (PROPERTY_ID) REFERENCES airbnb.property(PROPERTY_ID);
ALTER TABLE airbnb.order ADD FOREIGN KEY (CUST_ID) REFERENCES airbnb.customer(CUST_ID);
ALTER TABLE airbnb.PAYMENT ADD FOREIGN KEY (ORDER_ID) REFERENCES airbnb.order(ORDER_ID);
ALTER TABLE airbnb.PAYMENT ADD FOREIGN KEY (PAYMENT_METHOD_ID) REFERENCES airbnb.PAYMENT_METHOD(PAYMENT_METHOD_ID);
ALTER TABLE airbnb.FEEDBACK ADD FOREIGN KEY (PROPERTY_ID) REFERENCES airbnb.PROPERTY(PROPERTY_ID);
ALTER TABLE airbnb.FEEDBACK ADD FOREIGN KEY (CUST_ID) REFERENCES airbnb.CUSTOMER(CUST_ID);
```

Database Creation

Simple data table (<5 rows): Insert data directly for “room_type” & “payment_method”



```sql
-- Insert values to new tables
INSERT INTO payment_method VALUES
(1, 'CREDIT CARD'), (2, 'DEBIT CARD');
INSERT INTO room_type VALUES
(1, 'Shared room'), (2, 'Private room'),(3, 'Entire house'), (4, 'Entire apartment');
  SELECT * FROM room_type;
  SELECT * FROM payment_method;
```

PAYMENT_ME	METHOD
1	CREDIT CARD
2	DEBIT CARD
NULL	NULL

#	Time	Action	Message
12	22:12:44	SELECT * FROM airbnb.property	34 row(s) returned
13	22:12:45	SELECT * FROM airbnb.payment	40 row(s) returned
14	22:12:46	SELECT * FROM airbnb.order	40 row(s) returned
15	22:17:30	SELECT * FROM room_type	4 row(s) returned
16	22:17:32	SELECT * FROM payment_method	2 row(s) returned

Complicated data table(>5 rows): Import data from .csv files (6 main tables)

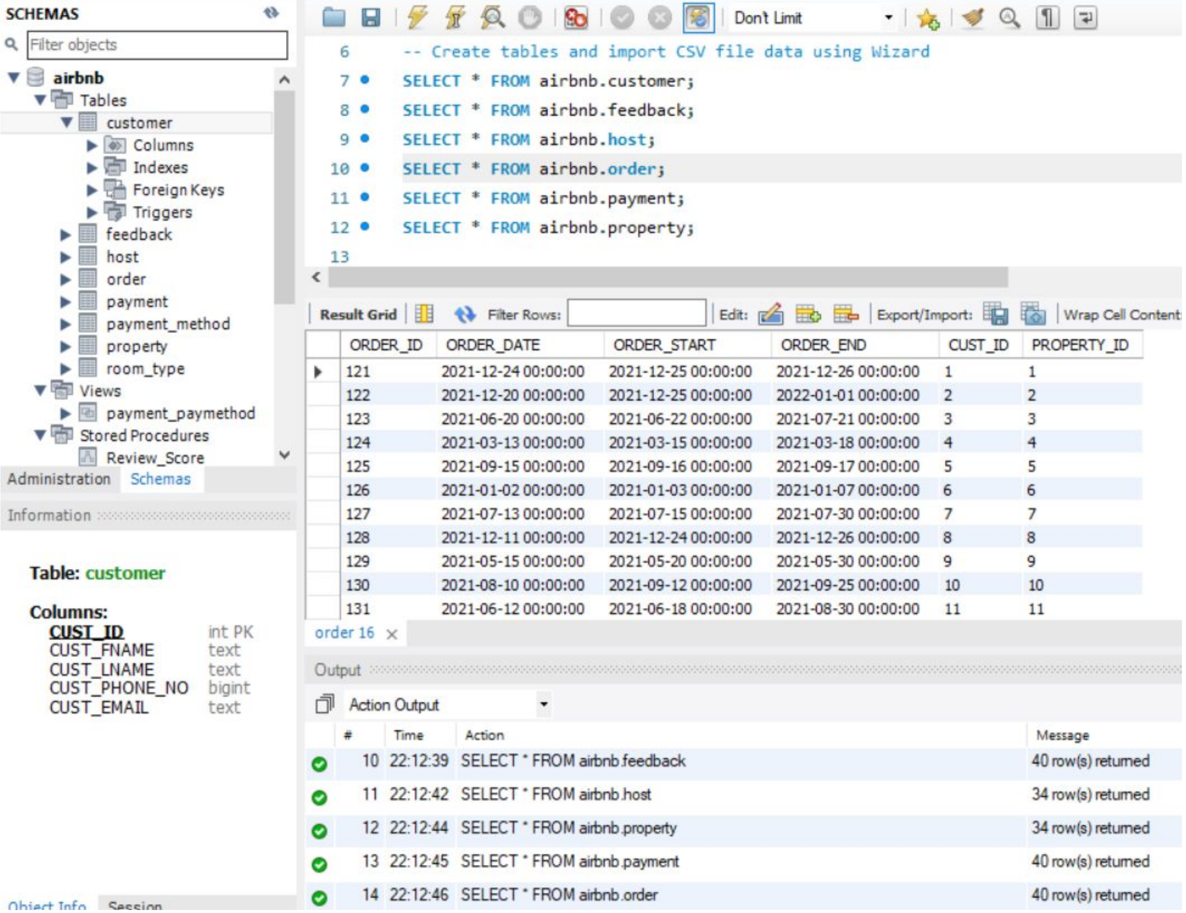


Table: customer

Columns:

CUST_ID	int PK
CUST_FNAME	text
CUST_LNAME	text
CUST_PHONE_NO	bigint
CUST_EMAIL	text

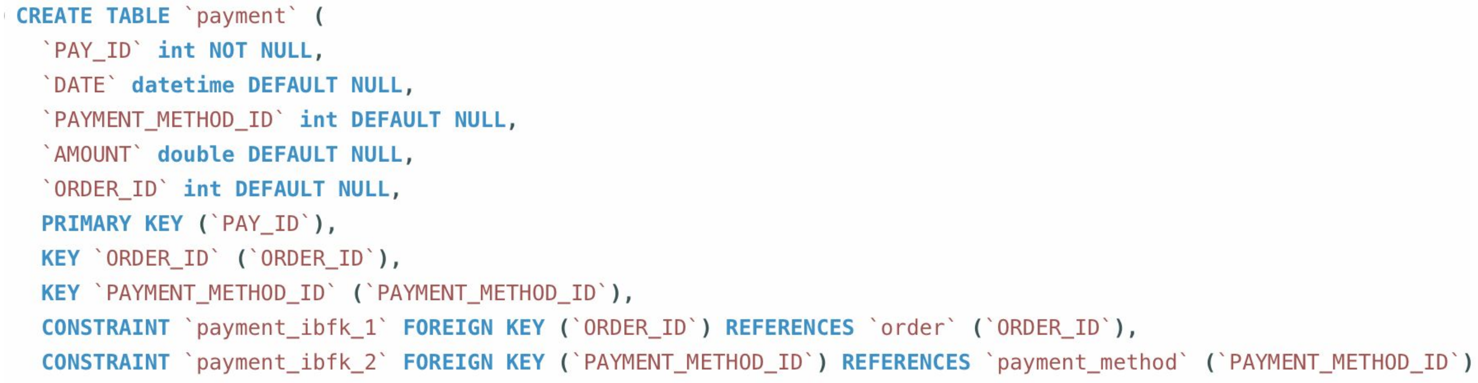
```sql
-- Create tables and import CSV file data using Wizard
SELECT * FROM airbnb.customer;
SELECT * FROM airbnb.feedback;
SELECT * FROM airbnb.host;
SELECT * FROM airbnb.order;
SELECT * FROM airbnb.payment;
SELECT * FROM airbnb.property;
```

ORDER_ID	ORDER_DATE	ORDER_START	ORDER_END	CUST_ID	PROPERTY_ID
121	2021-12-24 00:00:00	2021-12-25 00:00:00	2021-12-26 00:00:00	1	1
122	2021-12-20 00:00:00	2021-12-25 00:00:00	2022-01-01 00:00:00	2	2
123	2021-06-20 00:00:00	2021-06-22 00:00:00	2021-07-21 00:00:00	3	3
124	2021-03-13 00:00:00	2021-03-15 00:00:00	2021-03-18 00:00:00	4	4
125	2021-09-15 00:00:00	2021-09-16 00:00:00	2021-09-17 00:00:00	5	5
126	2021-01-02 00:00:00	2021-01-03 00:00:00	2021-01-07 00:00:00	6	6
127	2021-07-13 00:00:00	2021-07-15 00:00:00	2021-07-30 00:00:00	7	7
128	2021-12-11 00:00:00	2021-12-24 00:00:00	2021-12-26 00:00:00	8	8
129	2021-05-15 00:00:00	2021-05-20 00:00:00	2021-05-30 00:00:00	9	9
130	2021-08-10 00:00:00	2021-09-12 00:00:00	2021-09-25 00:00:00	10	10
131	2021-06-12 00:00:00	2021-06-18 00:00:00	2021-08-30 00:00:00	11	11

#	Time	Action	Message
10	22:12:39	SELECT * FROM airbnb.feedback	40 row(s) returned
11	22:12:42	SELECT * FROM airbnb.host	34 row(s) returned
12	22:12:44	SELECT * FROM airbnb.property	34 row(s) returned
13	22:12:45	SELECT * FROM airbnb.payment	40 row(s) returned
14	22:12:46	SELECT * FROM airbnb.order	40 row(s) returned

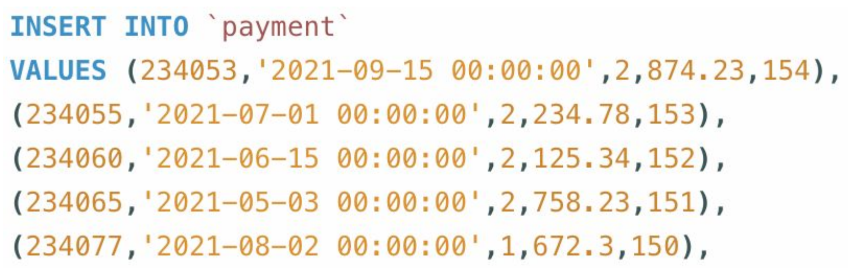
Database Creation Example

Step 1 (**DDL**):



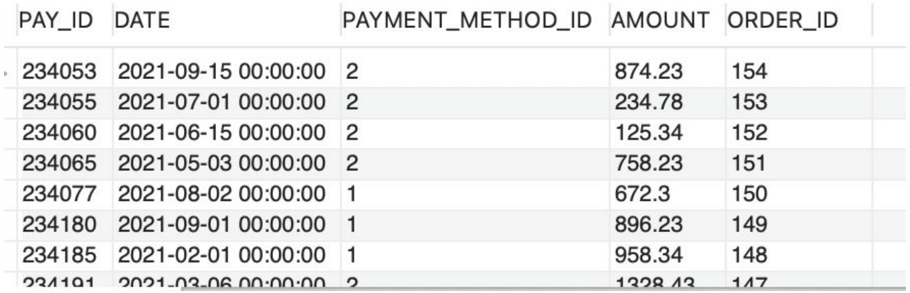
```
CREATE TABLE `payment` (
  `PAY_ID` int NOT NULL,
  `DATE` datetime DEFAULT NULL,
  `PAYMENT_METHOD_ID` int DEFAULT NULL,
  `AMOUNT` double DEFAULT NULL,
  `ORDER_ID` int DEFAULT NULL,
  PRIMARY KEY (`PAY_ID`),
  KEY `ORDER_ID` (`ORDER_ID`),
  KEY `PAYMENT_METHOD_ID` (`PAYMENT_METHOD_ID`),
  CONSTRAINT `payment_ibfk_1` FOREIGN KEY (`ORDER_ID`) REFERENCES `order` (`ORDER_ID`),
  CONSTRAINT `payment_ibfk_2` FOREIGN KEY (`PAYMENT_METHOD_ID`) REFERENCES `payment_method` (`PAYMENT_METHOD_ID`)
```

Step 2 (**DML**):



```
INSERT INTO `payment`
VALUES (234053,'2021-09-15 00:00:00',2,874.23,154),
(234055,'2021-07-01 00:00:00',2,234.78,153),
(234060,'2021-06-15 00:00:00',2,125.34,152),
(234065,'2021-05-03 00:00:00',2,758.23,151),
(234077,'2021-08-02 00:00:00',1,672.3,150),
```

Result with PK and Foreign Key from Payment method table:

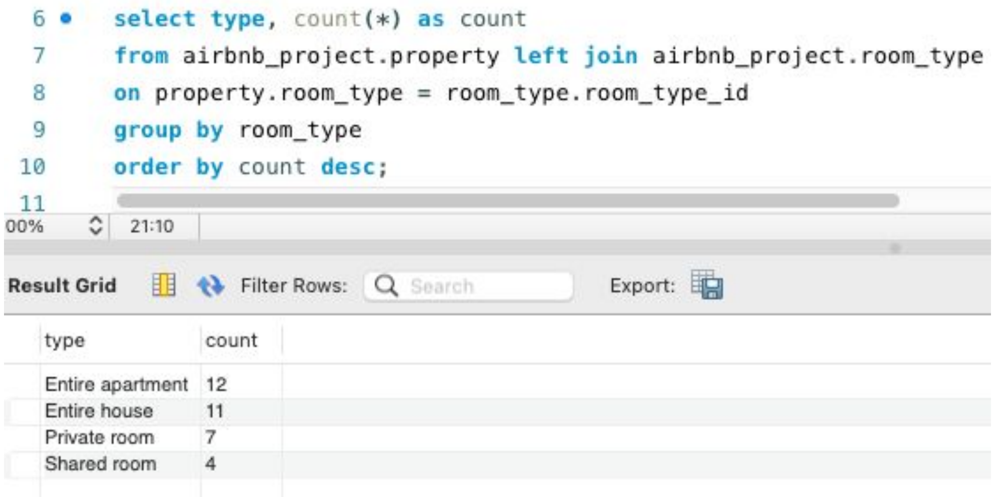


PAY_ID	DATE	PAYMENT_METHOD_ID	AMOUNT	ORDER_ID
234053	2021-09-15 00:00:00	2	874.23	154
234055	2021-07-01 00:00:00	2	234.78	153
234060	2021-06-15 00:00:00	2	125.34	152
234065	2021-05-03 00:00:00	2	758.23	151
234077	2021-08-02 00:00:00	1	672.3	150
234180	2021-09-01 00:00:00	1	896.23	149
234185	2021-02-01 00:00:00	1	958.34	148
234191	2021-03-06 00:00:00	2	1328.43	147

10 Business Queries using the Data Query Language

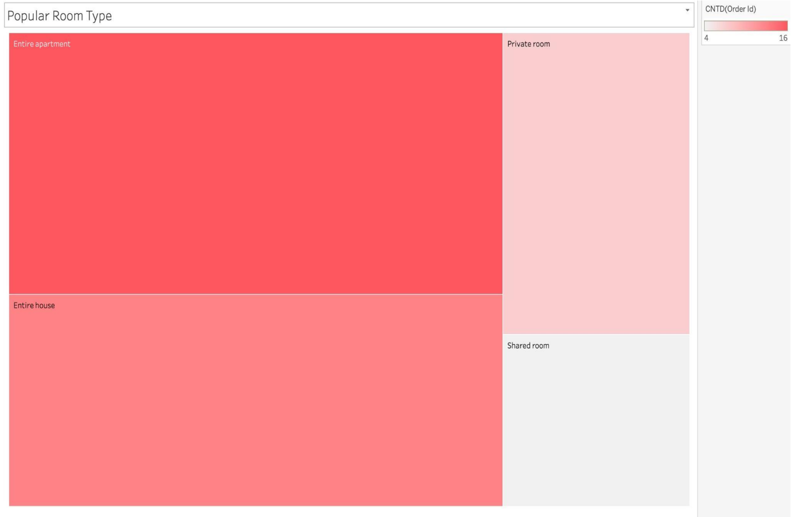
1. What is the most Popular Room Type?

Using Left Join



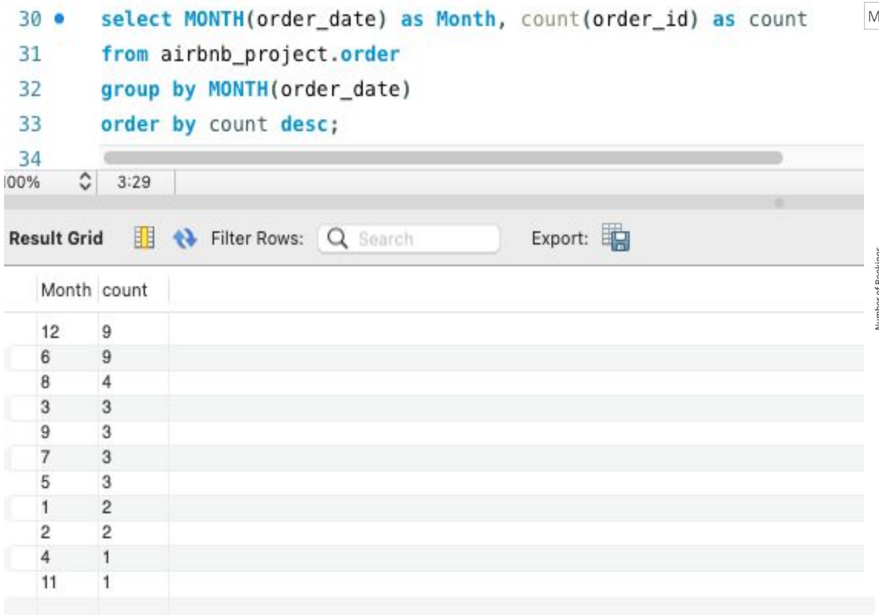
```
select type, count(*) as count
from airbnb_project.property left join airbnb_project.room_type
on property.room_type = room_type.room_type_id
group by room_type
order by count desc;
```

type	count
Entire apartment	12
Entire house	11
Private room	7
Shared room	4



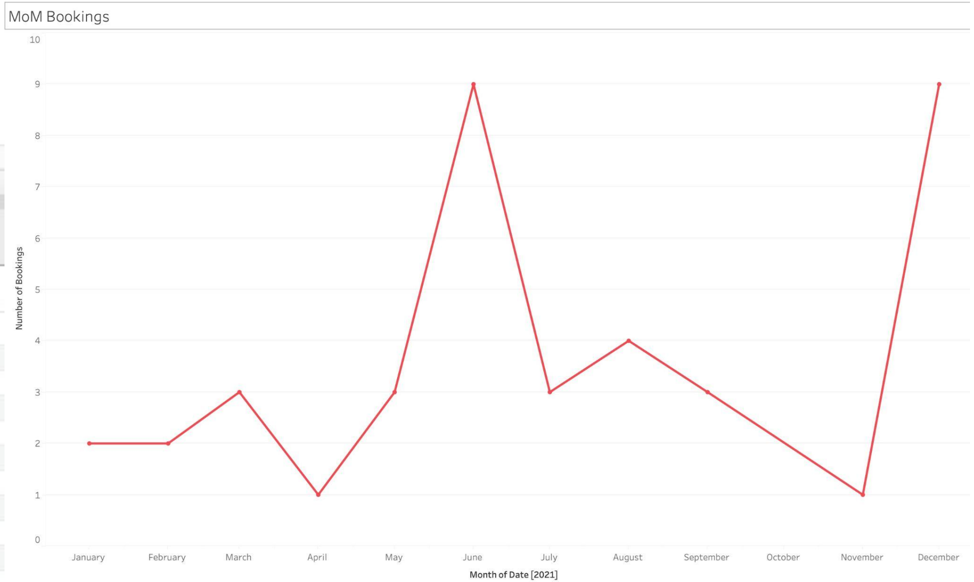
The most popular house types are “Entire Apartment” and “Entire House”, because this is unique to airbnb and more convenient for families, as compared to hotels

2. What time of the year is there highest demand for airbnbs?



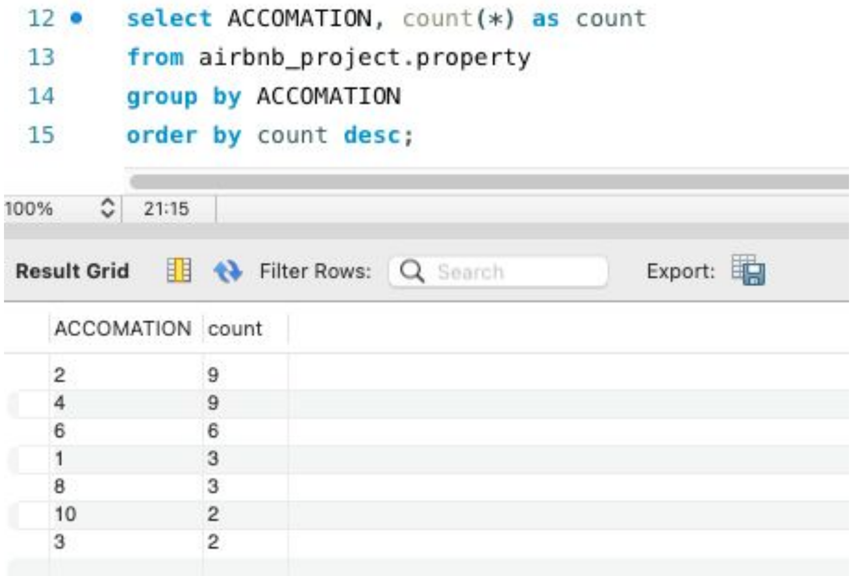
```
select MONTH(order_date) as Month, count(order_id) as count
from airbnb_project.order
group by MONTH(order_date)
order by count desc;
```

```
Month count
12    9
6     9
8     4
3     3
9     3
7     3
5     3
1     2
2     2
4     1
11    1
```



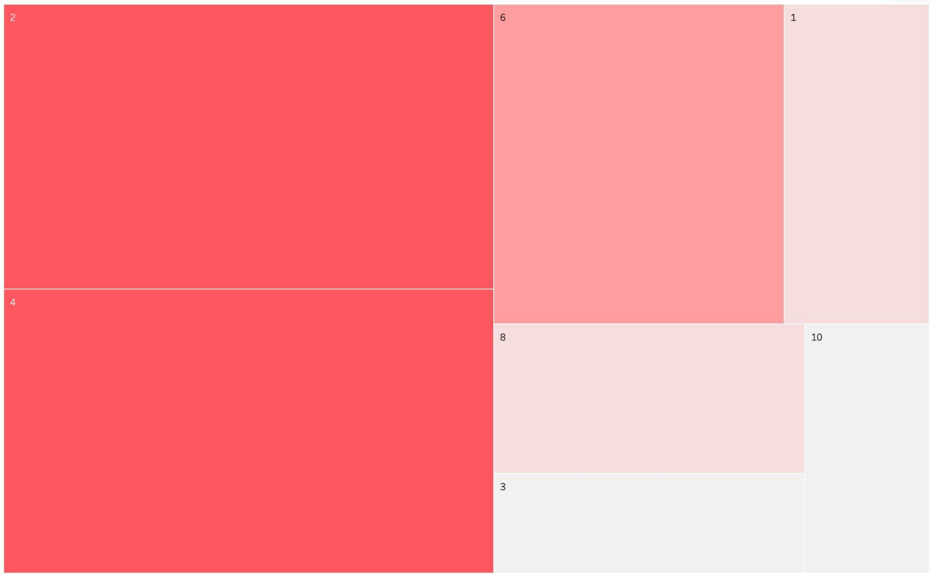
June and December months see the highest demand due to holiday season.

3. What number of accommodations per property is most common?



```sql
select ACCOMATION, count(*) as count
from airbnb_project.property
group by ACCOMATION
order by count desc;
```

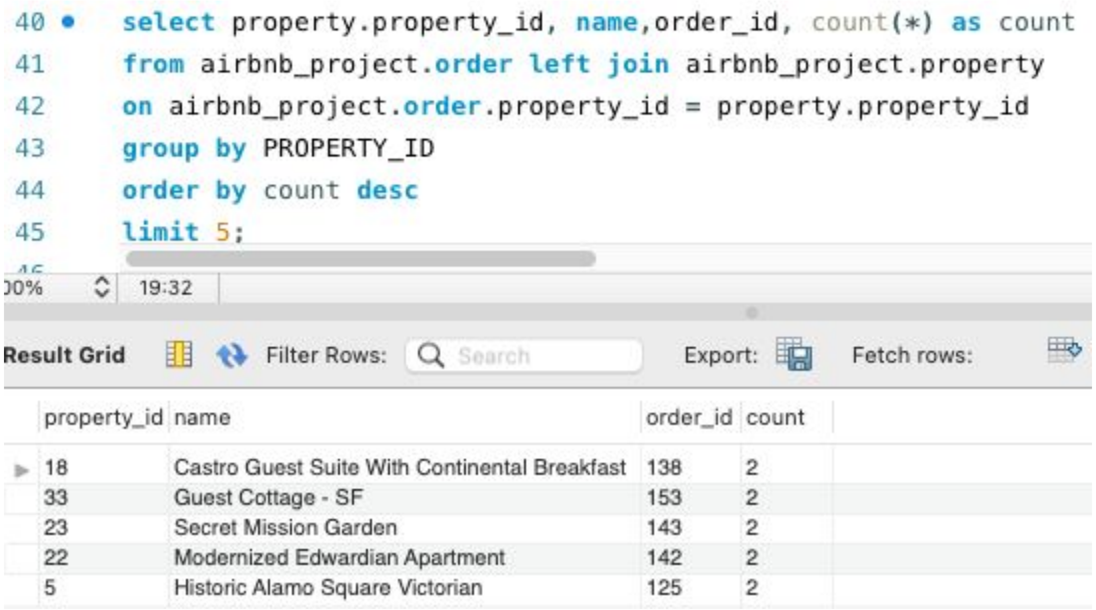
ACCOMATION	count
2	9
4	9
6	6
1	3
8	3
10	2
3	2



Most properties accommodate 2 or 4 people.

4. What are the top 5 properties in the Bay Area?

(Based on number of bookings)



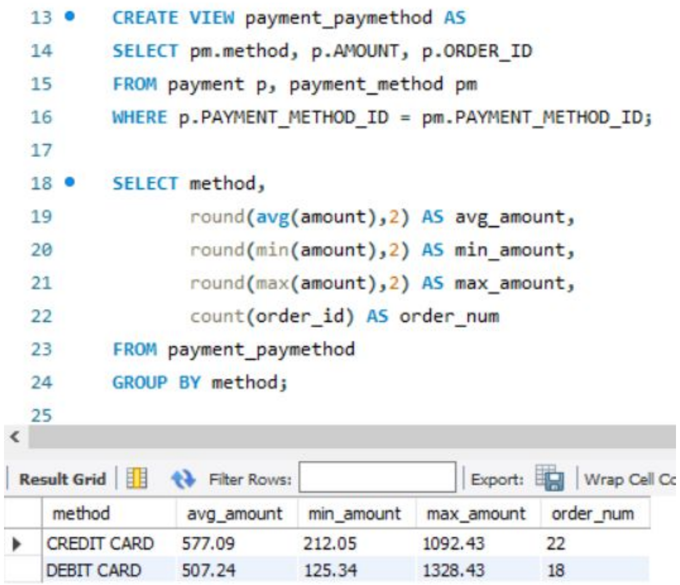
```sql
select property.property_id, name,order_id, count(*) as count
from airbnb_project.order left join airbnb_project.property
on airbnb_project.order.property_id = property.property_id
group by PROPERTY_ID
order by count desc
limit 5;
```

property_id	name	order_id	count
18	Castro Guest Suite With Continental Breakfast	138	2
33	Guest Cottage - SF	153	2
23	Secret Mission Garden	143	2
22	Modernized Edwardian Apartment	142	2
5	Historic Alamo Square Victorian	125	2

With Joins

5. Is payment method correlated with spend? (average, min, max amount per order from different types of payment methods)

Using Views

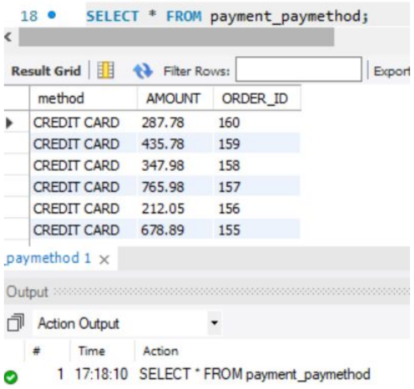


```sql
CREATE VIEW payment_paymethod AS
SELECT pm.method, p.AMOUNT, p.ORDER_ID
FROM payment p, payment_method pm
WHERE p.PAYMENT_METHOD_ID = pm.PAYMENT_METHOD_ID;

SELECT method,
        round(avg(amount),2) AS avg_amount,
        round(min(amount),2) AS min_amount,
        round(max(amount),2) AS max_amount,
        count(order_id) AS order_num
FROM payment_paymethod
GROUP BY method;
```

method	avg_amount	min_amount	max_amount	order_num
CREDIT CARD	577.09	212.05	1092.43	22
DEBIT CARD	507.24	125.34	1328.43	18

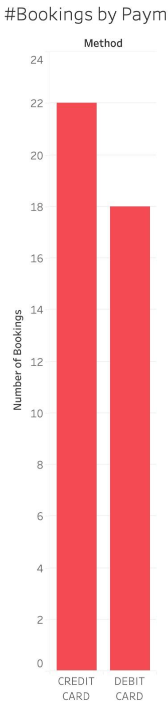
VIEW of payment_paymethod

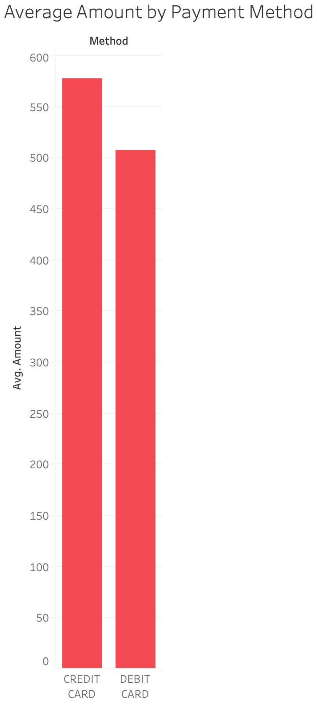


```sql
SELECT * FROM payment_paymethod;
```

method	AMOUNT	ORDER_ID
CREDIT CARD	287.78	160
CREDIT CARD	435.78	159
CREDIT CARD	347.98	158
CREDIT CARD	765.98	157
CREDIT CARD	212.05	156
CREDIT CARD	678.89	155

#	Time	Action
1	17:18:10	SELECT * FROM payment_paymethod

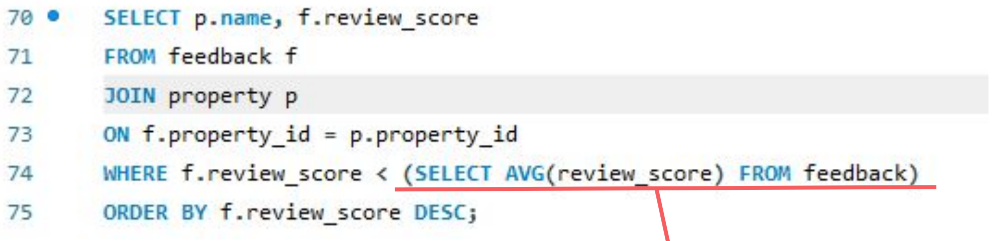

#Bookings by Payment Method


Average Amount by Payment Method

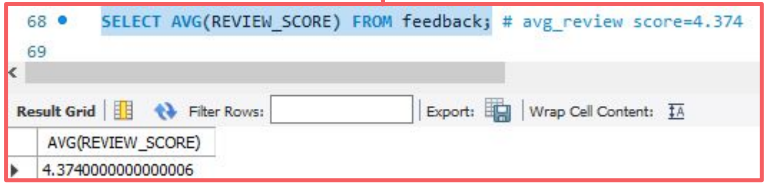
People tend to use Credit card more often while booking. Also, the average spend is higher on Credit Card.

6. Which properties have a review score lower than average review score?

Using Joins, Subqueries

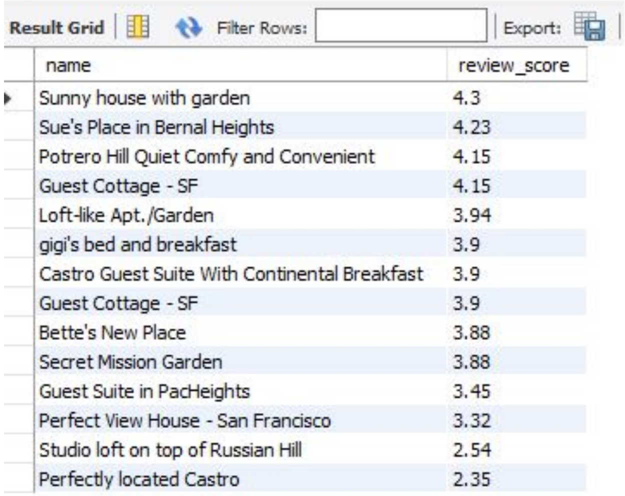


```sql
SELECT p.name, f.review_score
FROM feedback f
JOIN property p
ON f.property_id = p.property_id
WHERE f.review_score < (SELECT AVG(review_score) FROM feedback)
ORDER BY f.review_score DESC;
```



```sql
SELECT AVG(REVIEW_SCORE) FROM feedback; # avg_review score=4.374
```

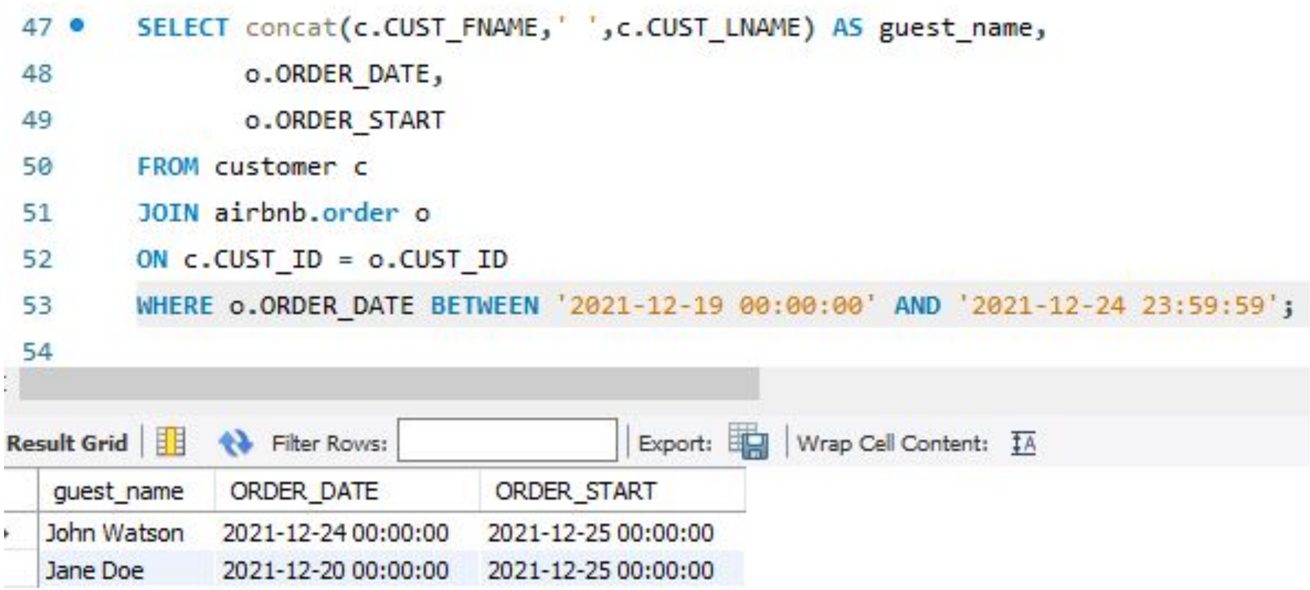
AVG(REVIEW_SCORE)
4.3740000000000006



name	review_score
Sunny house with garden	4.3
Sue's Place in Bernal Heights	4.23
Potrero Hill Quiet Comfy and Convenient	4.15
Guest Cottage - SF	4.15
Loft-like Apt./Garden	3.94
gigi's bed and breakfast	3.9
Castro Guest Suite With Continental Breakfast	3.9
Guest Cottage - SF	3.9
Bette's New Place	3.88
Secret Mission Garden	3.88
Guest Suite in PacHeights	3.45
Perfect View House - San Francisco	3.32
Studio loft on top of Russian Hill	2.54
Perfectly located Castro	2.35

7. Who made the orders just before holidays (Dec 19 to Dec 24, 2021)?

Using Joins

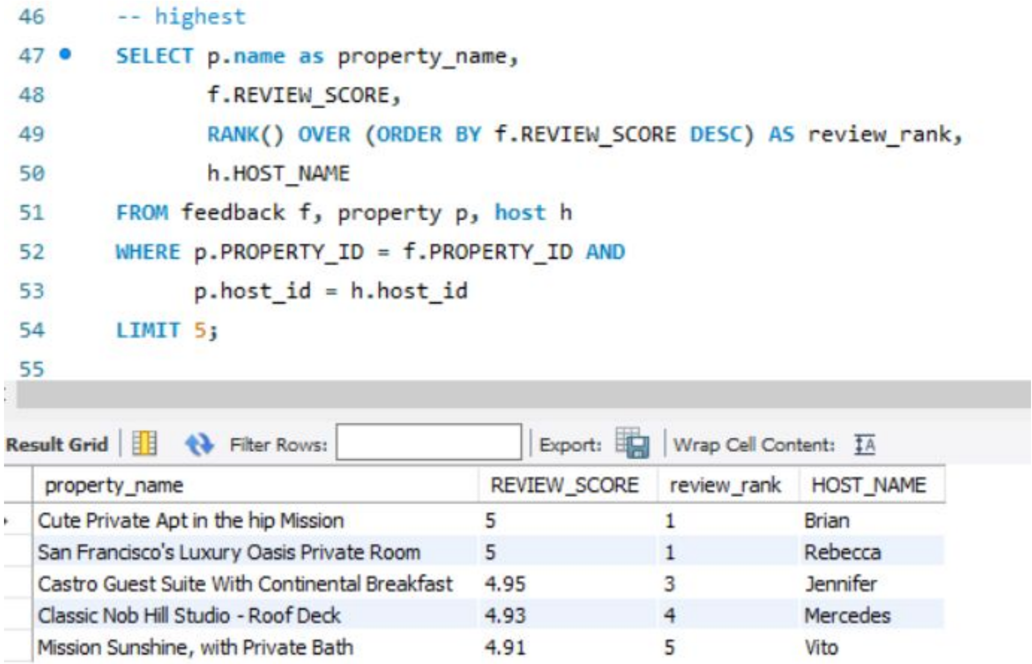


```
SELECT concat(c.CUST_FNAME,' ',c.CUST_LNAME) AS guest_name,
       o.ORDER_DATE,
       o.ORDER_START
FROM customer c
JOIN airbnb.order o
ON c.CUST_ID = o.CUST_ID
WHERE o.ORDER_DATE BETWEEN '2021-12-19 00:00:00' AND '2021-12-24 23:59:59';
```

guest_name	ORDER_DATE	ORDER_START
John Watson	2021-12-24 00:00:00	2021-12-25 00:00:00
Jane Doe	2021-12-20 00:00:00	2021-12-25 00:00:00

8. What are the properties having highest review score, and who's the host for that? (Rank the top 5 properties with review score.)

Using Equi Joins & Rank

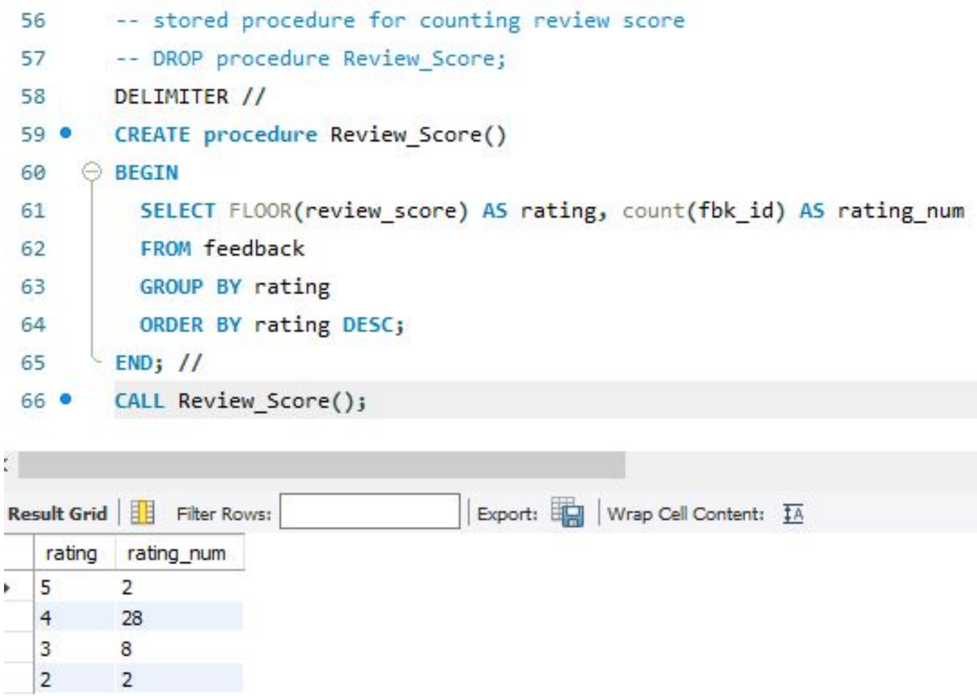


```
-- highest
SELECT p.name as property_name,
       f.REVIEW_SCORE,
       RANK() OVER (ORDER BY f.REVIEW_SCORE DESC) AS review_rank,
       h.HOST_NAME
FROM feedback f, property p, host h
WHERE p.PROPERTY_ID = f.PROPERTY_ID AND
      p.host_id = h.host_id
LIMIT 5;
```

property_name	REVIEW_SCORE	review_rank	HOST_NAME
Cute Private Apt in the hip Mission	5	1	Brian
San Francisco's Luxury Oasis Private Room	5	1	Rebecca
Castro Guest Suite With Continental Breakfast	4.95	3	Jennifer
Classic Nob Hill Studio - Roof Deck	4.93	4	Mercedes
Mission Sunshine, with Private Bath	4.91	5	Vito

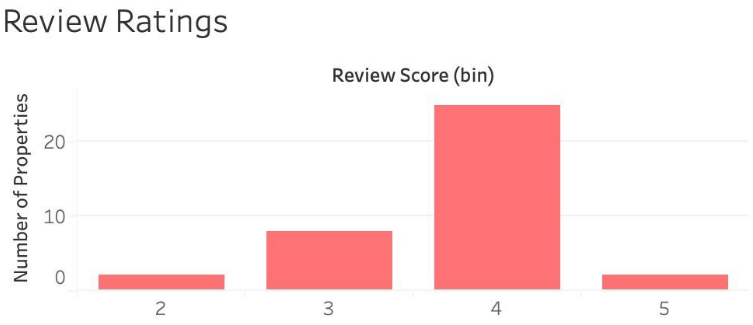
9. What is the frequency of different review ratings?

Using Stored Procedure



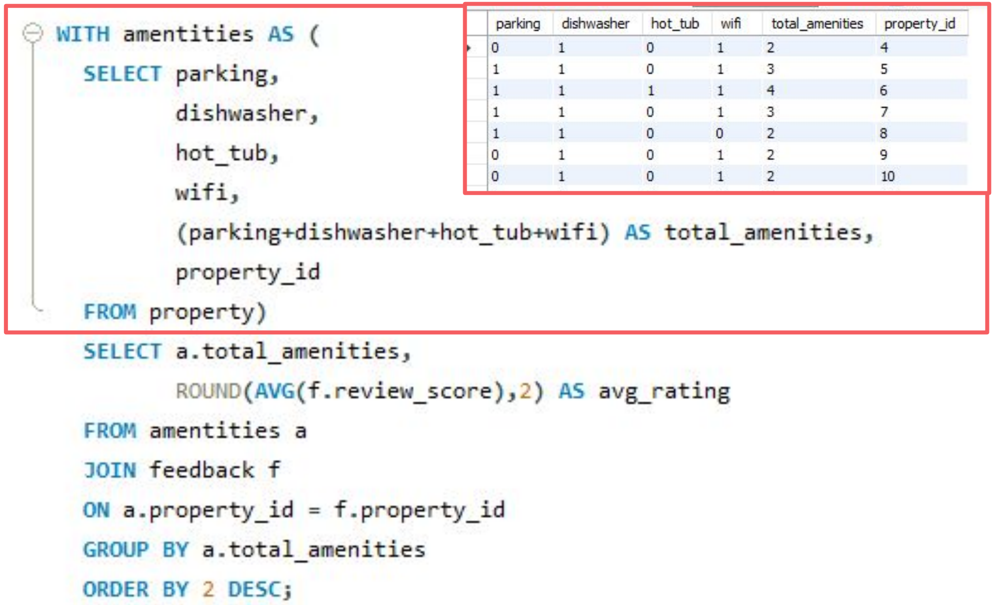
```
-- stored procedure for counting review score
-- DROP procedure Review_Score;
DELIMITER //
CREATE procedure Review_Score()
BEGIN
  SELECT FLOOR(review_score) AS rating, count(fbk_id) AS rating_num
  FROM feedback
  GROUP BY rating
  ORDER BY rating DESC;
END; //
CALL Review_Score();
```

rating	rating_num
5	2
4	28
3	8
2	2



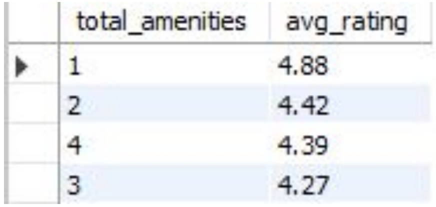
Review Ratings

10. Do properties with more amenities get higher ratings?



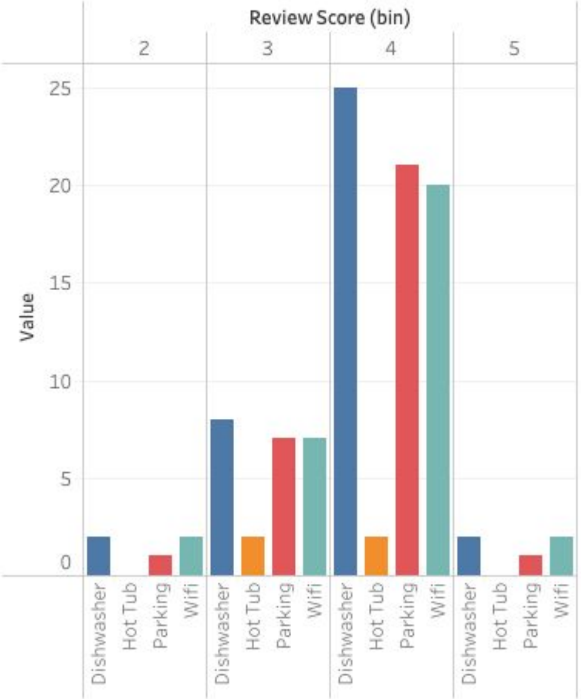
```
WITH amentities AS (
  SELECT parking,
         dishwasher,
         hot_tub,
         wifi,
         (parking+dishwasher+hot_tub+wifi) AS total_amenities,
         property_id
  FROM property)
  SELECT a.total_amenities,
         ROUND(AVG(f.review_score),2) AS avg_rating
  FROM amentities a
  JOIN feedback f
  ON a.property_id = f.property_id
  GROUP BY a.total_amenities
  ORDER BY 2 DESC;
```

parking	dishwasher	hot_tub	wifi	total_amenities	property_id
0	1	0	1	2	4
1	1	0	1	3	5
1	1	1	1	4	6
1	1	0	1	3	7
1	1	0	0	2	8
0	1	0	1	2	9
0	1	0	1	2	10



total_amenities	avg_rating
1	4.88
2	4.42
4	4.39
3	4.27

Using Joins, Subqueries



Data Visualization

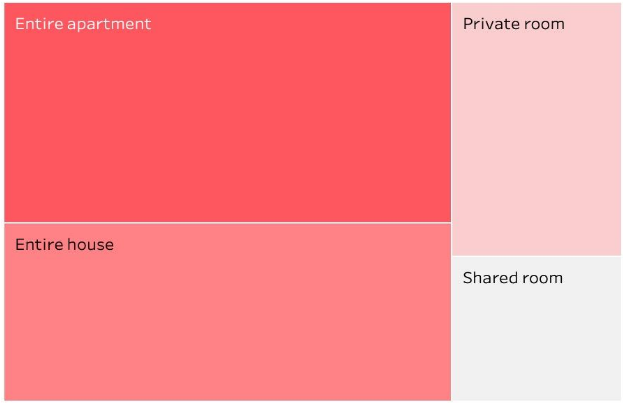
Created Dashboard using Tableau version 2021.4

-Step 1: Upload the **CSVs** on **Tableau** and create ER diagram

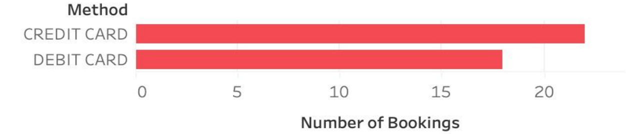
-Step2: Build multiple “Sheets” with graphs visualizing Business **Queries** that we are exploring

-Step 3: Combine different Sheets to form a Business “Dashboard” for Airbnb

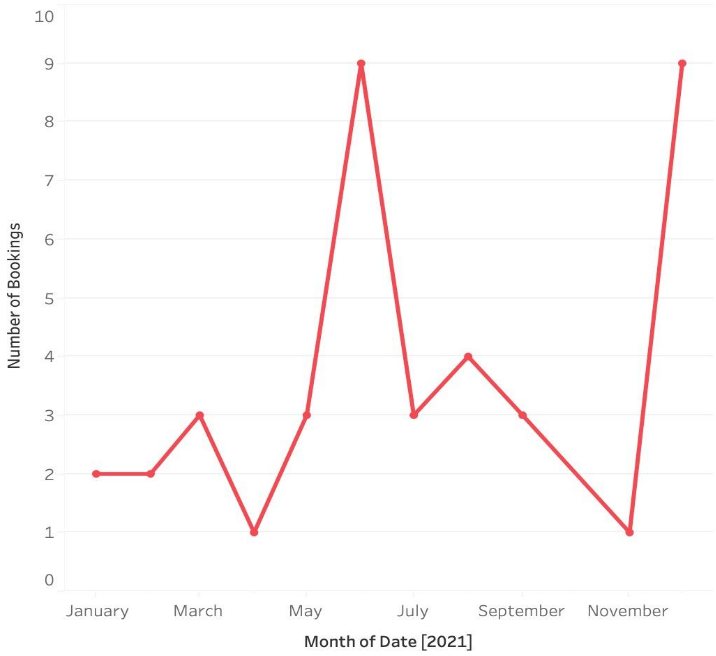
Popular Room Type



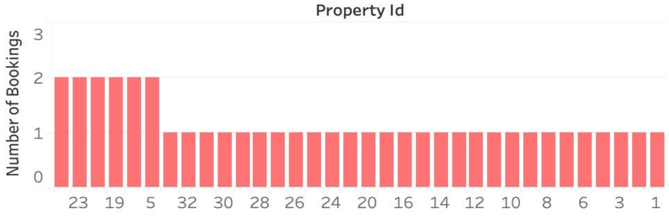
#Bookings by Payment Method



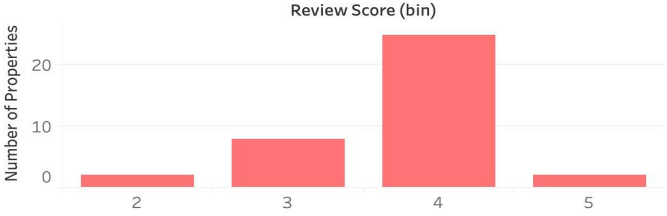
MoM Bookings



Bookings/Property



Review Ratings



Learnings

	Learning	Details
1	Swim Lane Diagram with draw.io	Understanding of the guest and host journey maps
2	ER Diagram with MySQL	Understanding of the relationship between tables in a database
3	Database Creation using different entities involved in an Airbnb	Using CSV upload for large tables (Guest, Host, Property, Feedback, Booking, Payment) & INSERT for small tables (Room type, Payment Method)
4	SQL concepts - DDL, DML & DQL	Used SQL Joins, Views, Subqueries and Stored Procedures
5	Visualization with Tableau	To create dashboard and graphs relevant to business queries

Conclusion

This project helped us analyse

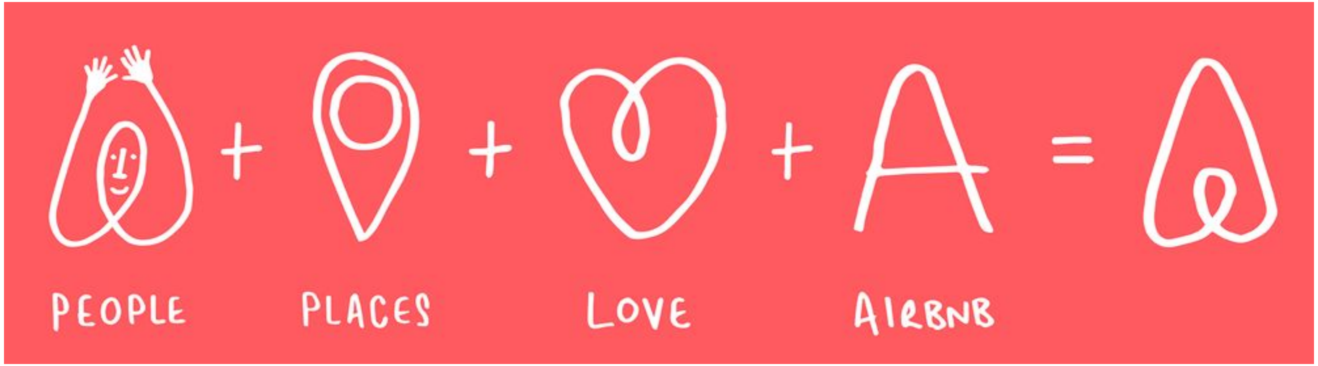
- How Airbnb facilitates the guests and hosts: Various working spects from discovery to sale of an online marketplace
- **Factors** that contribute to a property getting booked (amenities, room type, review ratings, seasonality etc.)
- **Business trends** and **queries** that Airbnb might be interested in tracking

Order Bookings:
When **customers** orders on the Airbnb website for a stay, nhost can simply receive a message and review the customer information to decide whether he want to accept the order.

Airbnb Platform:
Customers select the favorite property to stay and review its rating and amenities. **Hosts** can post their place where they want to rent on the platform. **Payment** will also be issued through Airbnb security website.

Feedback:
After a sweet stay, both **customers** and **host** can leave a feedback/review to each other, which affects the ranking order of the property and the credit on guests.

Thank you



Happy Valentine's! 007