Title: Analyze hotel bookings and generate insights and trends.

Data source: Kaggle

Q.1 Data type of columns in the dataset.

```
SELECT column_name, data_type
FROM INFORMATION_SCHEMA.COLUMNS
WHERE table_schema ='hotel_b'
AND table_name = 'hotel_bookings';
```

COLUMN_NAME	DATA_TYPE
hotel	text
is_canceled	int
lead_time	int
arrival_date_year	int
arrival_date_month	text
arrival_date_week_number	int
arrival_date_day_of_month	int
stays_in_weekend_nights	int
stays_in_week_nights	int
adults	int

Q.2 Time period for which the data is given.

```
SELECT MIN(arrival_date_year) AS start_year, MAX(arrival_date_year) AS last_year
FROM hotel_b.hotel_bookings;
```

start_year	last_year
2015	2017

Q.3 How many distinct countries, hotels, and years are given in the data.

```
SELECT COUNT(DISTINCT country) AS total_country, COUNT(DISTINCT hotel) AS total_hotel,
COUNT(DISTINCT arrival_date_year) AS total_year
FROM hotel_b.hotel_bookings;
```

total_country	total_hotel	total_year
177	2	3

Q.4 Total hotel bookings occurred in each year.

```
SELECT arrival_date_year, COUNT(arrival_date_year) AS total_arrival_each_year
FROM hotel_b.hotel_bookings
GROUP BY arrival_date_year;
```

arrival_date_year	total_arrival_each_year
2015	21992
2016	56707
2017	40687

Q.5 Which 2 types of hotels are given in the dataset and how many bookings each have.

```
SELECT hotel AS hotel_types, COUNT(hotel) AS hotel_total_bookings
FROM hotel_b.hotel_bookings
GROUP BY hotel;
```

hotel_types	hotel_total_bookings
Resort Hotel	40060
City Hotel	79326

Q.6 In each hotel_type each year how many hotel bookings occurred.

```
SELECT hotel, arrival_date_year, COUNT(arrival_date_year) AS total_arrival_c
FROM hotel_b.hotel_bookings
GROUP BY hotel, arrival_date_year;
```

hotel	arrival_date_year	total_arrival_c
Resort Hotel	2015	8314
Resort Hotel	2016	18567
Resort Hotel	2017	13179
City Hotel	2015	13678
City Hotel	2016	38140
City Hotel	2017	27508

Q.7 Write a query to create a view of month-on-month analysis of hotel bookings and show results.

```
CREATE VIEW months_total_bookings AS
SELECT arrival_date_year, arrival_date_month, COUNT(arrival_date_month) AS bookings_in_month
FROM hotel_b.hotel_bookings
GROUP BY arrival_date_year, arrival_date_month;
```

Action

CREATE VIEW months_total_bookings AS SELECT arrival_date_year, arrival_date_month, COUNT(arrival_date_month) A...

```
SELECT arrival_date_year, arrival_date_month, bookings_in_month
FROM months_total_bookings;
```

arrival_date_year	arrival_date_month	bookings_in_month
2015	July	2776
2015	August	3885
2015	September	5114
2015	October	4957
2015	November	2340
2015	December	2920
2016	January	2248
2016	February	3891
2016	March	4824
2016	April	5428

Q.8 Perform the month-on-month analysis of hotel bookings for each hotel type.

```
WITH mnth AS (

SELECT hotel, arrival_date_year, arrival_date_month, COUNT(arrival_date_month) AS
month_bookings_count

FROM hotel_b.hotel_bookings

GROUP BY hotel, arrival_date_year, arrival_date_month

)

SELECT hotel, arrival_date_year AS arrival_year, arrival_date_month AS arrival_month,
month_bookings_count

FROM mnth;
```

hotel	arrival_year	arrival_month	month_bookings_count
Resort Hotel	2015	July	1378
Resort Hotel	2015	August	1409
Resort Hotel	2015	September	1585
Resort Hotel	2015	October	1571
Resort Hotel	2015	November	1105
Resort Hotel	2015	December	1266
Resort Hotel	2016	January	884
Resort Hotel	2016	February	1520
Resort Hotel	2016	March	1778
Resort Hotel	2016	April	1867

Q.9 Find the hotel bookings for each week.

```
SELECT arrival_date_week_number, COUNT(arrival_date_week_number) AS count_week_bookings

FROM hotel_b.hotel_bookings

GROUP BY arrival_date_week_number

ORDER BY arrival_date_week_number;
```

arrival_date_week_number	count_week_bookings
1	1047
2	1218
3	1319
4	1487
5	1387
6	1508
7	2109
8	2216
9	2117
10	2149

Q.10 Find the hotel bookings count for each week for each hotel type.

```
SELECT hotel, arrival_date_week_number, COUNT(arrival_date_week_number) AS week_bookings
FROM hotel_b.hotel_bookings
GROUP BY hotel, arrival_date_week_number
ORDER BY arrival_date_week_number;
```

hotel	arrival_date_week_number	week_bookings
Resort Hotel	1	343
City Hotel	1	704
Resort Hotel	2	457
City Hotel	2	761
Resort Hotel	3	553
City Hotel	3	766
Resort Hotel	4	519
City Hotel	4	968
Resort Hotel	5	501
City Hotel	5	886

Q.11 While booking the hotel which type of hotel room were reserved.

```
SELECT reserved_room_type, COUNT(reserved_room_type) AS reserved_roomtype_count
FROM hotel_b.hotel_bookings
GROUP BY reserved_room_type
ORDER BY reserved_roomtype_count DESC;
```

reserved_room_type	reserved_roomtype_count
A	85994
D	19201
E	6535
F	2897
G	2094
B	1114
C	932
H	601
P	12
L	6

Q.12 For each hotel type during hotel bookings which room type were reserved.

```
SELECT hotel, reserved_room_type, COUNT(reserved_room_type) AS total_reserved_roomtype
FROM hotel_b.hotel_bookings
GROUP BY hotel, reserved_room_type
ORDER BY hotel, total_reserved_roomtype DESC;
```

hotel	reserved_room_type	total_reserved_roomtype
City Hotel	A	62595
City Hotel	D	11768
City Hotel	F	1791
City Hotel	E	1553
City Hotel	B	1111
City Hotel	G	484
City Hotel	C	14
City Hotel	P	10
Resort Hotel	A	23399
Resort Hotel	D	7433

Q.13 Write a query to find the total count of reservations where assigned room type is different than reserved room type.

```
SELECT COUNT(assigned_room_type) AS diff_roomtype_assigned
FROM (
SELECT reserved_room_type, assigned_room_type
FROM hotel_b.hotel_bookings
WHERE reserved_room_type <> assigned_room_type) AS diff_room_type_assigned;
```

diff_roomtype_assigned
14917

Q.14 Write a query to find the total count of reservations where assigned room type is different than reserved room type for each hotel type.

```
SELECT hotel, COUNT(assigned_room_type) AS different_roomtype_assigned
FROM (
SELECT hotel, reserved_room_type, assigned_room_type
FROM hotel_b.hotel_bookings
WHERE reserved_room_type <> assigned_room_type) AS dif_room_ty_assig
GROUP BY hotel;
```

hotel	different_roomtype_assigned
Resort Hotel	7725
City Hotel	7192

Q.15 Write a query to find the average days in waiting list.

```
SELECT ROUND(AVG(days_in_waiting_list)) AS average_waiting_days
FROM hotel_b.hotel_bookings;
```

average_waiting_days
2

Q.16 For each hotel type find the average days in waiting list.

```
SELECT hotel, ROUND(AVG(days_in_waiting_list)) AS average_waiting_days
FROM hotel_b.hotel_bookings
GROUP BY hotel;
```

hotel	average_waiting_days
Resort Hotel	1
City Hotel	3

Q.17 Write a query to find how many hotel bookings are for each country.

```
SELECT country, COUNT(country) AS count_country_bookings
FROM hotel_b.hotel_bookings
GROUP by country
ORDER BY count_country_bookings DESC;
```

country	count_country_bookings
PRT	48586
GBR	12129
FRA	10415
ESP	8568
DEU	7287
ITA	3766
IRL	3375
BEL	2342
BRA	2224
NLD	2104

Q.18 Find out how many hotel bookings there are for each hotel type in each country.

```
SELECT hotel, country, COUNT(country) AS country_bookings
FROM hotel_b.hotel_bookings
GROUP by hotel, country
ORDER BY hotel, country_bookings DESC;
```

hotel	country	country_bookings
City Hotel	PRT	30956
City Hotel	FRA	8804
City Hotel	DEU	6084
City Hotel	GBR	5315
City Hotel	ESP	4611
City Hotel	ITA	3307
City Hotel	BEL	1894
City Hotel	BRA	1794
City Hotel	USA	1618
City Hotel	NLD	1590

Q.19 Write a query to find hotel deposit type count.

```
SELECT deposit_type, COUNT(deposit_type) count_deposit_type
FROM hotel_b.hotel_bookings
GROUP BY deposit_type;
```

deposit_type	count_deposit_type
No Deposit	104637
Refundable	162
Non Refund	14587

Q.20 Write a query to find out the count of different deposit types for each hotel type.

```
SELECT hotel, deposit_type, COUNT(deposit_type) booking_deposit_type
FROM hotel_b.hotel_bookings
GROUP BY hotel, deposit_type;
```

hotel	deposit_type	booking_deposit_type
Resort Hotel	No Deposit	38199
Resort Hotel	Refundable	142
Resort Hotel	Non Refund	1719
City Hotel	No Deposit	66438
City Hotel	Non Refund	12868
City Hotel	Refundable	20

Q.21 What is the total number of each Customer type.

```
SELECT customer_type, COUNT(customer_type) count_coustomer_type
FROM hotel_b.hotel_bookings
GROUP BY customer_type;
```

customer_type	count_coustomer_type
Transient	89613
Contract	4076
Transient-Party	25120
Group	577

Q.22 Write a query to find out the number of different customer types for each hotel type.

```
SELECT hotel, customer_type, COUNT(customer_type) coustomer_type
FROM hotel_b.hotel_bookings
GROUP BY hotel, customer_type;
```

hotel	customer_type	coustomer_type
Resort Hotel	Transient	30209
Resort Hotel	Contract	1776
Resort Hotel	Transient-Party	7791
Resort Hotel	Group	284
City Hotel	Transient	59404
City Hotel	Transient-Party	17329
City Hotel	Contract	2300
City Hotel	Group	293

Q.23 Write a query to find the percentage of cancellation in hotel bookings for each month in year 2016 and rank higher percentage lower rank and vice versa.

```
SELECT arrival_date_month, percentage_cancelation_in_month,
DENSE_RANK() OVER(ORDER BY percentage_cancelation_in_month DESC)
AS dense_rank_cancel_percent
FROM (SELECT b_cancel.arrival_date_month,
ROUND(no_cancellation/bookings_in_month*100) AS percentage_cancelation_in_month
FROM (SELECT arrival_date_year, arrival_date_month, COUNT(arrival_date_month) AS no_cancellation
FROM hotel_b.hotel_bookings
WHERE is_canceled = 1 AND arrival_date_year = 2016
GROUP BY arrival_date_month) AS b_cancel
LEFT JOIN months_total_bookings AS m_booking
ON b_cancel.arrival_date_year = m_booking.arrival_date_year
AND b_cancel.arrival_date_month = m_booking.arrival_date_month) AS CancelPercent;
```

arrival_date_month	percentage_cancelation_in_month	dense_rank_cancel_percent
October	41	1
June	40	2
April	38	3
September	37	4
November	37	4
August	36	5
December	36	5
May	35	6
February	34	7
July	33	8

Q.24 Write a query to find the percent of no show for each month in the year 2016.

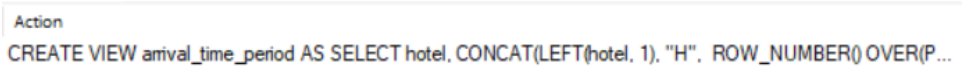
```
SELECT Noshow.arrival_date_month, ROUND(noshow_count/bookings_in_month * 100, 2)
AS percent_noshow
FROM months_total_bookings AS m_Hbooking
RIGHT JOIN (SELECT arrival_date_year, arrival_date_month, COUNT(arrival_date_month)
AS noshow_count
FROM hotel_b.hotel_bookings
WHERE reservation_status = 'No-Show' AND arrival_date_year = 2016
GROUP BY arrival_date_month) AS Noshow
ON m_Hbooking.arrival_date_year = Noshow.arrival_date_year
AND m_Hbooking.arrival_date_month = Noshow.arrival_date_month
ORDER BY percent_noshow DESC;
```

arrival_date_month	percent_noshow
February	3.73
January	2.22
March	1.47
May	1.19
April	0.99
June	0.94
October	0.89
November	0.88
December	0.85
July	0.83

Q.25 Create a view using SQL functions and show the result of the view.

- **Create a single date column using arrival year, month, and date related column**
- **Find the total customer/guest for each booking made for only those customers/guests who did not cancel and showed.**
- **Give unique id for each row.**
- **Return columns hotel, customer_id, arrival_date, total_guest, total_of_special_requests, reservation_status_date, and maximum guest count using window function.**

```
CREATE VIEW arrival_time_period AS
SELECT hotel, CONCAT(LEFT(hotel, 1), "H",
ROW_NUMBER() OVER(PARTITION BY hotel ORDER BY YEAR(arrival_date), MONTH(arrival_date),
DATE(arrival_date))) AS customer_id, arrival_date, total_guest,
total_of_special_requests, reservation_status_date
FROM (SELECT hotel,
STR_TO_DATE(CONCAT(arrival_date_day_of_month, arrival_date_month, arrival_date_year),
"%d %M %Y") AS arrival_date, reservation_status_date, total_of_special_requests,
adults+children+babies AS total_guest, reservation_status
FROM hotel_b.hotel_bookings
WHERE reservation_status = 'Check-Out') AS arrival_d;
```




```
SELECT hotel, customer_id, arrival_date, total_guest, total_of_special_requests,
reservation_status_date, MAX(total_guest) OVER() AS max_guest
FROM arrival_time_period;
```

hotel	customer_id	arrival_date	total_guest	total_of_special_requests	reservation_status_date	max_guest
City Hotel	CH1	2015-07-01	2	0	2015-07-03	12
City Hotel	CH2	2015-07-01	1	0	2015-07-03	12
City Hotel	CH3	2015-07-01	1	0	2015-07-03	12
City Hotel	CH4	2015-07-01	1	0	2015-07-03	12
City Hotel	CH5	2015-07-01	1	0	2015-07-03	12
City Hotel	CH6	2015-07-01	1	0	2015-07-03	12
City Hotel	CH7	2015-07-01	2	0	2015-07-03	12
City Hotel	CH8	2015-07-01	1	0	2015-07-03	12
City Hotel	CH9	2015-07-01	1	0	2015-07-03	12
City Hotel	CH10	2015-07-01	1	0	2015-07-03	12

Q.26 Find the number of days guests/customer had stayed at hotels and return the rows where customer/guest had stayed more than 20 days and rank them by giving lower rank to the higher number of days stayed and vice versa.

```
SELECT customer_id, arrival_date, reservation_status_date, days_stayed,
RANK() OVER(ORDER BY days_stayed DESC) AS rank_stayed
FROM (SELECT customer_id, arrival_date, reservation_status_date,
DATEDIFF(reservation_status_date, arrival_date) AS days_stayed
FROM arrival_time_period
HAVING DATEDIFF(reservation_status_date, arrival_date)>20) AS stay_time;
```

customer_id	arrival_date	reservation_status_date	days_stayed	rank_stayed
RH13252	2016-07-05	2016-09-12	69	1
RH1080	2015-08-01	2015-09-30	60	2
CH24359	2016-09-22	2016-11-18	57	3
RH20098	2017-01-10	2017-03-07	56	4
CH10744	2016-03-11	2016-04-29	49	5
CH20003	2016-07-23	2016-09-09	48	6
RH2334	2015-09-07	2015-10-23	46	7
RH20987	2017-02-03	2017-03-20	45	8
CH31062	2017-01-15	2017-02-27	43	9
RH6819	2016-01-26	2016-03-08	42	10

Q.27 Write a query using total_of_special_requests and create bins for no requests, less than 3 requests, and for greater than 3 requests. Then using window function count the overall requests for each bin in one column and hotel wise special requests for each bin in another column.

```
SELECT hotel, customer_id, special_requests,
COUNT(customer_id) OVER(PARTITION BY special_requests) AS overall_request_count,
COUNT(customer_id) OVER(PARTITION BY hotel, special_requests) AS hotel_wise_special_request
FROM (SELECT customer_id, hotel,
CASE WHEN total_of_special_requests = 0 THEN 'No_Request'
WHEN total_of_special_requests >= 1 AND total_of_special_requests <= 3 THEN '<=3_Requests'
WHEN total_of_special_requests > 3 THEN '>3 requests'
END AS special_requests
FROM arrival_time_period) AS special_req
ORDER BY hotel, CAST(RIGHT(customer_id, LENGTH(customer_id)-2) AS UNSIGNED);
```

hotel	customer_id	special_requests	overall_request_count	hotel_wise_special_request
City Hotel	CH1	No_Request	36762	21617
City Hotel	CH2	No_Request	36762	21617
City Hotel	CH3	No_Request	36762	21617
City Hotel	CH4	No_Request	36762	21617
City Hotel	CH5	No_Request	36762	21617
City Hotel	CH6	No_Request	36762	21617
City Hotel	CH7	No_Request	36762	21617
City Hotel	CH8	No_Request	36762	21617
City Hotel	CH9	No_Request	36762	21617
City Hotel	CH10	No_Request	36762	21617

Insights based on analysis:

For the year 2015 the data starts from the month of July. Entire 12 months of data for 2016 was present. And for 2017 the data was up to the month of August. We should consider them before interpreting the results of the analysis.

Insights on overall data:

- 14 text and 18 int columns are present in the table hotel bookings.
- The given data starts from the year 2015 and it is up to 2017.
- Total distinct countries were 177, distinct hotel type was 2, and distinct years in the dataset were 3.
- Total number of hotel bookings for the year 2015 was 21992, and for year 2016 it was 56707, and for the year 2017 it was 40686.
- Resort Hotel and City Hotel were 2 types of hotels given in the table.
- In 2015, September and then October had the highest hotel bookings.
- In 2016 October month and in 2017 May had the highest number of hotel bookings.
- In 2015 and 2016 October had more hotel bookings than November and December.
- In 2016 and 2017 January had less hotel bookings than February.
- Except for week 51 each week had hotel bookings greater than 1000.
- Room type A, D and E had the highest reservations and room type L and P had the lowest number of reservations.
- 14917 was the count of reservations where assigned room type was different than reserved room type.
- The average days on the waiting list was 2 days.
- PRT, GBR, and FRA countries had the highest number of hotel bookings.
- Countries SLE, ATF, etc. had the lowest number of hotel bookings.
- 14587 hotel bookings had Non Refund deposit type and 162 hotel bookings had Refundable deposit type.
- Transient had the highest count and Group had the lowest count in the customer type.
- In 2016 October had the highest percent of cancellation and January had the lowest percent of cancellation.
- In 2016 February had the highest percent of no show which was 3.73 and September had the lowest no show percent which was 0.54.
- The maximum number of total guests for one hotel booking was 12.
- The maximum number a customer had stayed was 69 days.

Resort Hotel and City Hotel:

- Resort Hotel had 40060 and City Hotel had 79326 number of bookings.
- In all three years City Hotels had higher hotel bookings than Resort Hotel.
- In 2016 Resort Hotel had considerably lower bookings than City Hotel bookings.
- For 2015 data was given from July to December, Resort Hotel: September month had highest and November month had lowest number of hotel bookings.
- Each week City Hotel seems to have had more reservations than Resort Hotel.
- In City Hotel and Resort Hotel type A and D room type had the highest number of reservations and room type P had the lowest number of reservations.
- In Resort Hotel 7725 and City Hotel 7192 were the counts of reservations where assigned room type is different than reserved room type.
- The average days on the waiting list for the Resort Hotel was 1 day and for the City Hotel it was 3 days.
- In City hotel, countries PRT and FRA had the highest number of hotel bookings.
- In Resort hotels, the countries PRT and GBR had the highest number of hotel bookings.
- In City Hotel customer's deposit type Non Refund was 12868 and Refundable was 20.
- In Resort Hotel customer's deposit type Refundable was 142 and Non Refund was 1719.
- In both Resort and City Hotel customer type Group had the lowest count and Transient customer type had the highest count.

Trends

- In both Resort Hotel and City Hotel customers/guests seem to prefer Non Refund over Refundable deposit type.
- City Hotels had more bookings than Resort Hotels.
- The average days on the waiting list was 2 days.
- Room type A and D had the highest number of hotels booking reservation.
- Room type P had the lowest number of hotels bookings reservation.

Recommendations:

- Resort hotel need to advertise more on multiple platforms.
- Considering A and D room type have more bookings a larger number of rooms should be made available.
- And considering P had the lowest number of reservations there should be lesser number of rooms in the hotels.
- Should consider reducing the average waiting days by implementing things like above 2 points.