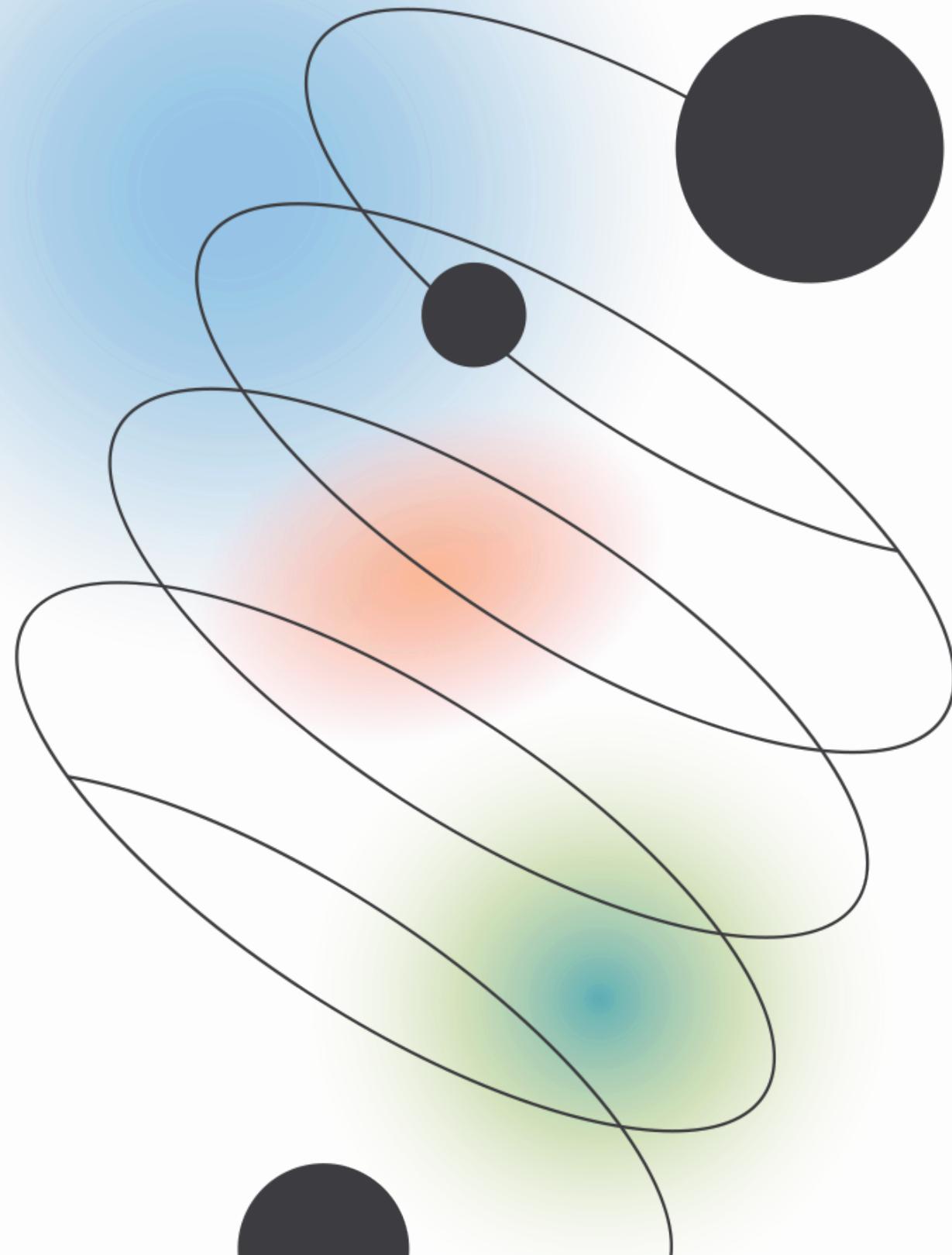


Hotel Reservation SQL Analysis

DATA ANALYST INTERN

Reggi Ahmad Fauzi - MIP-DA-10

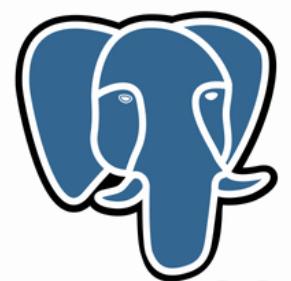


MENTORNESS

Objective Overview

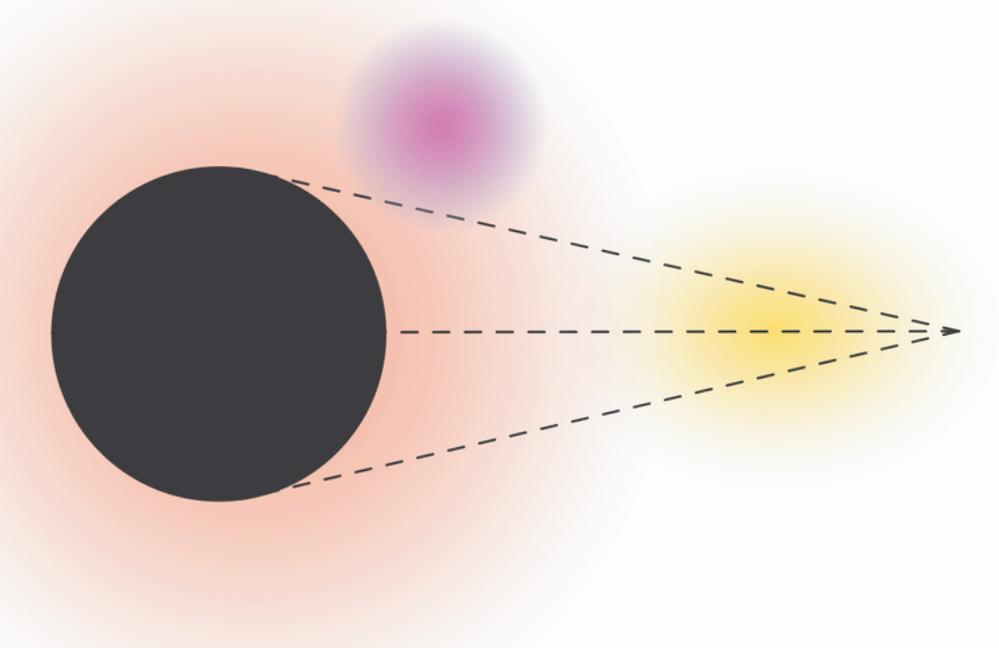
The hotel industry relies on data to make informed decisions and provide a better guest experience. In this internship, you will work with a hotel reservation dataset to **gain insights into guest preferences, booking trends, and other key factors that impact the hotel's operations**. You will use SQL to query and analyze the data, as well as answer specific questions about the dataset.

Tools Used



PostgreSQL

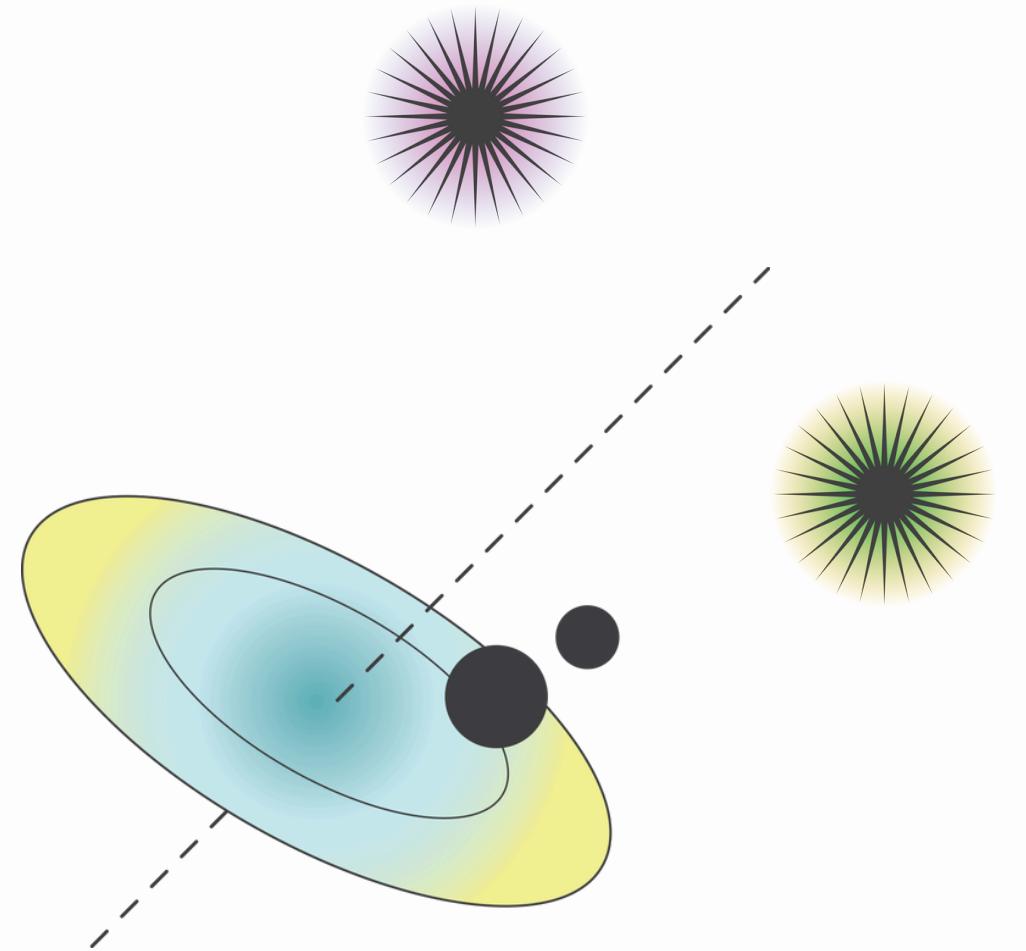
from: [SQLite Online](#)



Dataset Overview

The dataset includes the following columns:

- **booking_id** : A unique identifier for each hotel reservation.
- **no_of_adults** : The number of adults in the reservation.
- **no_of_children** : The number of children in the reservation.
- **no_of_weekend_nights** : The number of nights in the reservation that fall on weekends.
- **no_of_week_nights** : The number of nights in the reservation that fall on weekdays.
- **type_of_meal_plan** : The meal plan chosen by the guests.
- **room_type_reserved** : The type of room reserved by the guests.
- **lead_time** : The number of days between booking and arrival.
- **arrival_date** : The date of arrival.
- **market_segment_type** : The market segment to which the reservation belongs.
- **avg_price_per_room** : The average price per room in the reservation.
- **booking_status** : The status of the booking.



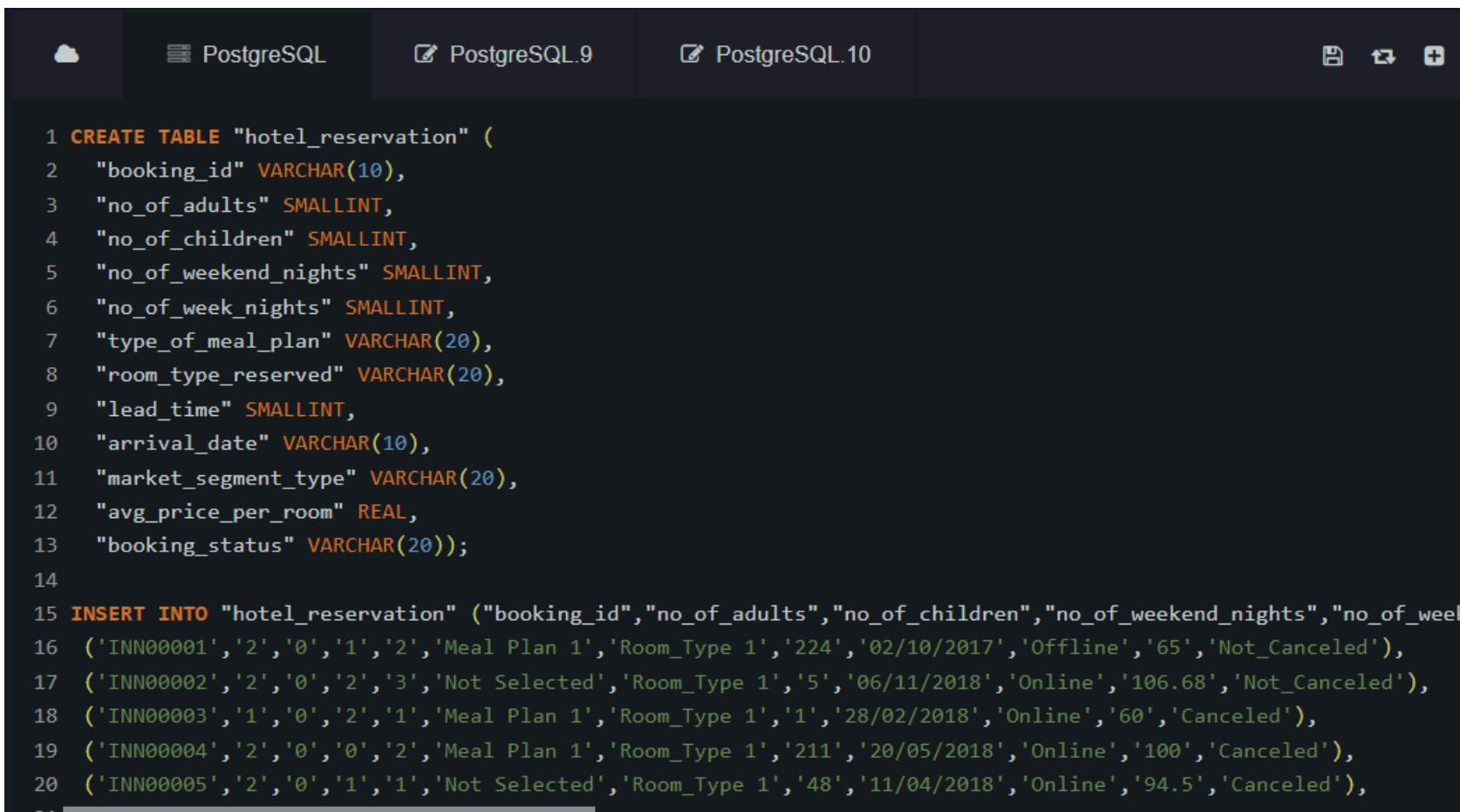


Assessing & Cleaning

This is how we checked and got rid of the 'dirty' from data.

Prepare the Data

- Step 1. Import dataset



```

1 CREATE TABLE "hotel_reservation" (
2   "booking_id" VARCHAR(10),
3   "no_of_adults" SMALLINT,
4   "no_of_children" SMALLINT,
5   "no_of_weekend_nights" SMALLINT,
6   "no_of_week_nights" SMALLINT,
7   "type_of_meal_plan" VARCHAR(20),
8   "room_type_reserved" VARCHAR(20),
9   "lead_time" SMALLINT,
10  "arrival_date" VARCHAR(10),
11  "market_segment_type" VARCHAR(20),
12  "avg_price_per_room" REAL,
13  "booking_status" VARCHAR(20),
14
15 INSERT INTO "hotel_reservation" ("booking_id", "no_of_adults", "no_of_children", "no_of_weekend_nights", "no_of_week_nights", "type_of_meal_plan", "room_type_reserved", "lead_time", "arrival_date", "market_segment_type", "avg_price_per_room", "booking_status")
16  ('INN00001', '2', '0', '1', '2', 'Meal Plan 1', 'Room_Type 1', '224', '02/10/2017', 'Offline', '65', 'Not_Canceled'),
17  ('INN00002', '2', '0', '2', '3', 'Not Selected', 'Room_Type 1', '5', '06/11/2018', 'Online', '106.68', 'Not_Canceled'),
18  ('INN00003', '1', '0', '2', '1', 'Meal Plan 1', 'Room_Type 1', '1', '28/02/2018', 'Online', '60', 'Canceled'),
19  ('INN00004', '2', '0', '0', '2', 'Meal Plan 1', 'Room_Type 1', '211', '20/05/2018', 'Online', '100', 'Canceled'),
20  ('INN00005', '2', '0', '1', '1', 'Not Selected', 'Room_Type 1', '48', '11/04/2018', 'Online', '94.5', 'Canceled'),
21

```

- Step 2. Check columns data type

hotel_reservation	
	Column
1	booking_id character varying(10)
2	no_of_adults smallint
3	no_of_children smallint
4	no_of_weekend_nights smallint
5	no_of_week_nights smallint
6	type_of_meal_plan character varying(20)
7	room_type_reserved character varying(20)
8	lead_time smallint
9	arrival_date character varying(10)
10	market_segment_type character varying(20)
11	avg_price_per_room real
12	booking_status character varying(20)

there's an incorrect data type in '**arrival_date**' column. It should be a **DATE** type instead of **VARCHAR**. So we'll going to fix it!

Prepare the Data

- Step 3. Change the incorrect data type

```

PostgreSQL PostgreSQL.9 PostgreSQL.10

--DATA CLEANING STEP
-- Step 1. Add a temporary column with type DATE.
ALTER TABLE hotel_reservation ADD COLUMN arrival_date_temp DATE;

-- Step 2. Update the temporary column with the converted value of the text column using TO_DATE.
UPDATE hotel_reservation
SET arrival_date_temp = TO_DATE(arrival_date, 'DD-MM-YYYY');

-- Step 3. Verify that the data has been successfully converted and there are no NULL values indicating a failed conversion
SELECT * FROM hotel_reservation WHERE arrival_date_temp IS NULL;

-- Step 4. If all data has been successfully converted, delete the original text column and rename the column
ALTER TABLE hotel_reservation
DROP COLUMN arrival_date;

ALTER TABLE hotel_reservation
RENAME COLUMN arrival_date_temp TO arrival_date

```

Text	Binary																											
hotel_reservation <table border="1"> <thead> <tr> <th>Column</th></tr> </thead> <tbody> <tr> <td>booking_id character varying(10)</td></tr> <tr> <td>no_of_adults smallint</td></tr> <tr> <td>no_of_children smallint</td></tr> <tr> <td>no_of_weekend_nights smallint</td></tr> <tr> <td>no_of_week_nights smallint</td></tr> <tr> <td>type_of_meal_plan character varying(20)</td></tr> <tr> <td>room_type_reserved character varying(20)</td></tr> <tr> <td>lead_time smallint</td></tr> <tr> <td>arrival_date character varying(10)</td></tr> <tr> <td>market_segment_type character varying(20)</td></tr> <tr> <td>avg_price_per_room real</td></tr> <tr> <td>booking_status character varying(20)</td></tr> <tr> <td>arrival_date date</td></tr> </tbody> </table>	Column	booking_id character varying(10)	no_of_adults smallint	no_of_children smallint	no_of_weekend_nights smallint	no_of_week_nights smallint	type_of_meal_plan character varying(20)	room_type_reserved character varying(20)	lead_time smallint	arrival_date character varying(10)	market_segment_type character varying(20)	avg_price_per_room real	booking_status character varying(20)	arrival_date date	hotel_reservation <table border="1"> <thead> <tr> <th>Column</th></tr> </thead> <tbody> <tr> <td>booking_id character varying(10)</td></tr> <tr> <td>no_of_adults smallint</td></tr> <tr> <td>no_of_children smallint</td></tr> <tr> <td>no_of_weekend_nights smallint</td></tr> <tr> <td>no_of_week_nights smallint</td></tr> <tr> <td>type_of_meal_plan character varying(20)</td></tr> <tr> <td>room_type_reserved character varying(20)</td></tr> <tr> <td>lead_time smallint</td></tr> <tr> <td>market_segment_type character varying(20)</td></tr> <tr> <td>avg_price_per_room real</td></tr> <tr> <td>booking_status character varying(20)</td></tr> <tr> <td>arrival_date date</td></tr> </tbody> </table>	Column	booking_id character varying(10)	no_of_adults smallint	no_of_children smallint	no_of_weekend_nights smallint	no_of_week_nights smallint	type_of_meal_plan character varying(20)	room_type_reserved character varying(20)	lead_time smallint	market_segment_type character varying(20)	avg_price_per_room real	booking_status character varying(20)	arrival_date date
Column																												
booking_id character varying(10)																												
no_of_adults smallint																												
no_of_children smallint																												
no_of_weekend_nights smallint																												
no_of_week_nights smallint																												
type_of_meal_plan character varying(20)																												
room_type_reserved character varying(20)																												
lead_time smallint																												
arrival_date character varying(10)																												
market_segment_type character varying(20)																												
avg_price_per_room real																												
booking_status character varying(20)																												
arrival_date date																												
Column																												
booking_id character varying(10)																												
no_of_adults smallint																												
no_of_children smallint																												
no_of_weekend_nights smallint																												
no_of_week_nights smallint																												
type_of_meal_plan character varying(20)																												
room_type_reserved character varying(20)																												
lead_time smallint																												
market_segment_type character varying(20)																												
avg_price_per_room real																												
booking_status character varying(20)																												
arrival_date date																												

Now the dataset we will be using is clean and ready to be analyzed.

Prepare the Data

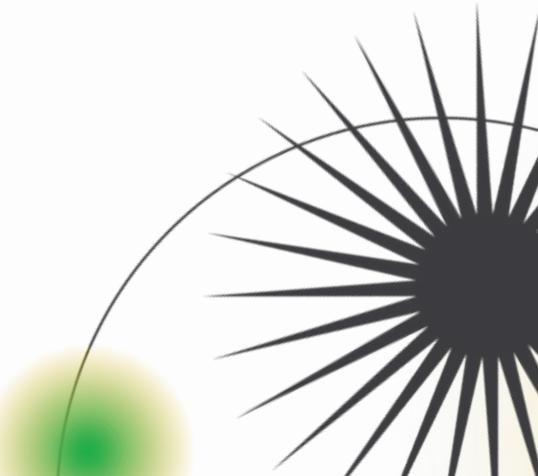
- Step 4. Scope the dataset

```
SELECT  
    MIN(arrival_date) oldest_reservation,  
    MAX(arrival_date) newest_reservation  
FROM hotel_reservation;
```

Output

	oldest_reservation	newest_reservation
	2017-07-01	2018-12-31

Now we know that the dataset was collected between July 1, 2017 and December 31, 2018. **About 18 Months**





The Task!

WE WILL WRITE SQL QUERIES TO GET INSIGHTS RELATED TO THE GIVEN TASK.



TASK 1:

What is the total number of reservations in the dataset?

```
5 SELECT
6     COUNT(booking_id) numb_of_reservation
7 FROM hotel_reservation;
```

numb_of_reservation
700



There is a total of 700 reservations.

(01/07/2017 - 31/12/2018)



TASK 2:

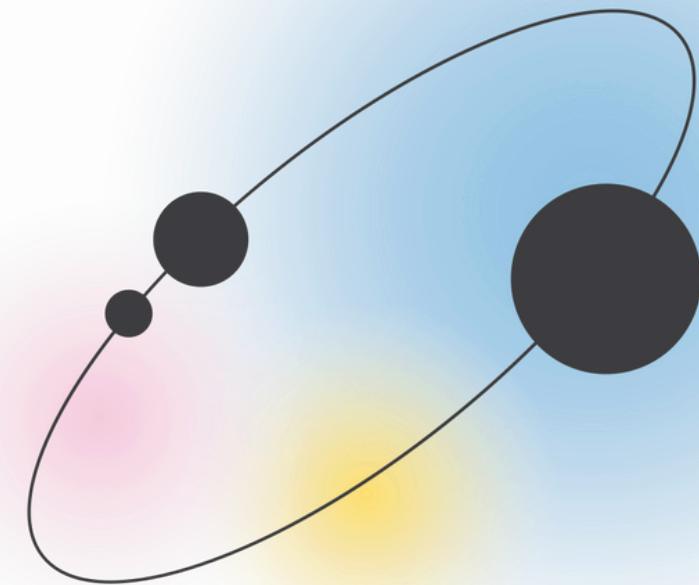
Which meal plan is the most popular among guests?

```
SELECT
    type_of_meal_plan,
    COUNT(type_of_meal_plan)
FROM hotel_reservation
GROUP BY 1
ORDER BY 2 DESC;
```

type_of_meal_plan	count
Meal Plan 1	527
Not Selected	109
Meal Plan 2	64



Meal Plan 1 is the most popular





TASK 3:

What is the average price per room for reservations involving children?

```
SELECT
    ROUND(AVG(avg_price_per_room)::numeric, 2)
    AS avg_price_per_room_with_children
FROM hotel_reservation
WHERE no_of_children > 0;
```

	avg_price_per_room_with_children
	144.57



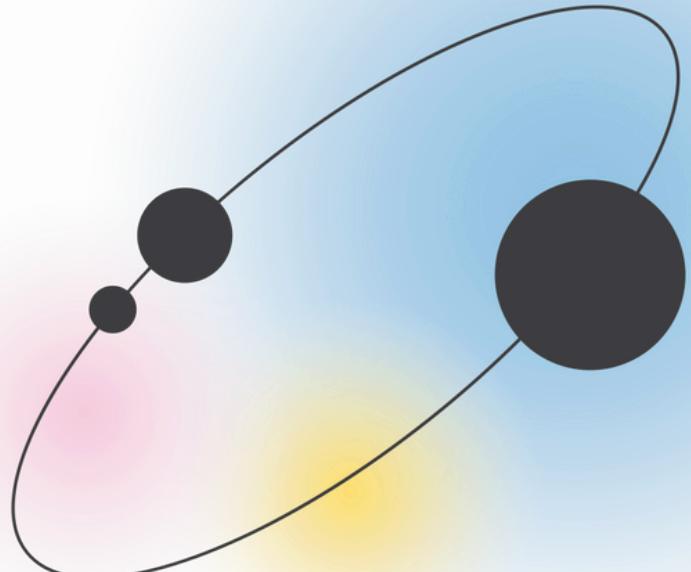
TASK 4:

How many reservations were made for the year ?

```
WITH r_2017 AS (
    SELECT
        booking_id reservation_2017
    FROM hotel_reservation
    WHERE EXTRACT (YEAR FROM arrival_date) = 2017
),
r_2018 AS (
    SELECT
        booking_id reservation_2018
    FROM hotel_reservation
    WHERE EXTRACT (YEAR FROM arrival_date) = 2018
)
SELECT
    COUNT(reservation_2017) num_reservation_2017,
    COUNT(reservation_2018) num_reservation_2018
FROM r_2017
FULL OUTER JOIN r_2018
    ON r_2017.reservation_2017 = r_2018.reservation_2018;
```

	num_reservation_2017	num_reservation_2018
	123	577

reservations were made for the year
2017 and 2018





TASK 5:

What is the most commonly booked room type?

```
SELECT
    room_type_reserved,
    COUNT(room_type_reserved) num_room_type_reserved
FROM hotel_reservation
GROUP BY 1
ORDER BY 2 DESC;
```

room_type_reserved	num_room_type_reserved
Room_Type 1	534
Room_Type 4	130
Room_Type 6	18
Room_Type 2	8
Room_Type 7	6
Room_Type 5	4



Room Type 1 is the most commonly booked



TASK 6:

How many reservations fall on a weekend
(no_of_weekend_nights > 0)?

```
SELECT
    COUNT(booking_id) reservations_on_weekend
FROM hotel_reservation
WHERE no_of_weekend_nights > 0;
```

reservations_on_weekend
383



TASK 7:

What is the highest and lowest lead time
for reservations?

```
SELECT
    MIN(lead_time) lowest_leadtime,
    MAX(lead_time) highest_leadtime
FROM hotel_reservation;
```

lowest_leadtime	highest_leadtime
0	443





TASK 8:

What is the most common market segment type for reservations?

```
SELECT
    market_segment_type,
    COUNT(market_segment_type) num_market_segment_type
FROM hotel_reservation
GROUP BY 1
ORDER BY 2 DESC;
```

market_segment_type	num_market_segment_type
Online	518
Offline	140
Corporate	27
Complementary	14
Aviation	1

Online is most common market segment type



TASK 9:

How many reservations have a booking status of "Confirmed"?

```
SELECT
    COUNT(booking_status) confirmed_booking_status
FROM hotel_reservation
WHERE booking_status = 'Not_Canceled';
```

confirmed_booking_status
493



TASK 10:

What is the total number of adults and children across all reservations?

```
SELECT
    SUM(no_of_adults) total_adults,
    SUM(no_of_children) total_children
FROM hotel_reservation;
```

total_adults	total_children
1316	69





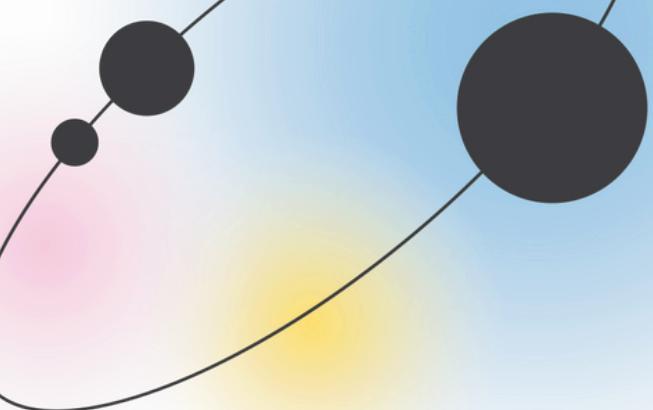
TASK 11:

What is the average number of weekend nights for reservations involving children?

```
SELECT
    ROUND(AVG(no_of_weekend_nights), 2)
        AS avg_weekend_nights_with_children
FROM hotel_reservation
WHERE no_of_children > 0;
```



```
avg_weekend_nights_with_children
1.00
```



TASK 12:

How many reservations were made in each month of the year?

```
WITH reservations AS (
    SELECT
        EXTRACT(MONTH FROM arrival_date) AS month_number,
        TO_CHAR(arrival_date, 'Month') AS MONTH,
        EXTRACT(YEAR FROM arrival_date) AS YEAR,
        booking_id
    FROM hotel_reservation
    WHERE EXTRACT(YEAR FROM arrival_date) IN (2017, 2018)
)
SELECT
    MONTH,
    month_number,
    COUNT(CASE WHEN YEAR = 2017 THEN booking_id END) AS total_reservation_2017,
    COUNT(CASE WHEN YEAR = 2018 THEN booking_id END) AS total_reservation_2018
FROM reservations
GROUP BY 1, 2
ORDER BY month_number ASC;
```



month	total_reservation_2017	total_reservation_2018
January	0	11
February	0	28
March	0	52
April	0	67
May	0	55
June	0	84
July	8	36
August	14	56
September	35	45
October	40	63
November	13	41
December	13	39



TASK 13:

What is the average number of nights (both weekend and weekday) spent by guests for each room type?

```
SELECT
    room_type_reserved,
    ROUND(AVG(no_of_weekend_nights + no_of_week_nights), 2)
        AS avg_nights_spent
FROM hotel_reservation
GROUP BY 1
ORDER BY 1;
```

room_type_reserved	avg_nights_spent
Room_Type 1	2.88
Room_Type 2	3.00
Room_Type 4	3.80
Room_Type 5	2.50
Room_Type 6	3.61
Room_Type 7	2.67



TASK 14:

For reservations involving children, what is the most common room type, and what is the average price for that room type?

```
SELECT
    room_type_reserved,
    COUNT(room_type_reserved) num_of_room_type_reserved,
    ROUND(AVG(avg_price_per_room)::numeric, 2) AS avg_price_per_room
FROM hotel_reservation
WHERE no_of_children > 0
GROUP BY 1
ORDER BY 1 ASC;
```

room_type_reserved	num_of_room_type_reserved	avg_price_per_room
Room_Type 1	24	123.12
Room_Type 2	5	112.08
Room_Type 4	1	86.32
Room_Type 6	17	185.33
Room_Type 7	1	187.04



Room Type 1 is most common room type with average price is 123.12



TASK 15:

Find the market segment type that generates the highest average price per room.

```
SELECT
    market_segment_type,
    ROUND(AVG(avg_price_per_room)::numeric, 2)
        AS avg_price_per_market_segment
FROM hotel_reservation
GROUP BY 1
ORDER BY 2 DESC;
```

market_segment_type	avg_price_per_market_segment
Online	112.46
Aviation	110.00
Offline	89.98
Corporate	82.40
Complementary	2.54



Recommendations

Adjust Pricing Based on Room Type Popularity:

- Dynamic Pricing: Implement a dynamic pricing strategy that adjusts rates based on demand, room type popularity, and booking lead times.
- Seasonal Adjustments: Adjust pricing based on seasonal demand fluctuations to ensure maximum occupancy and revenue

Enhance Guest Experience:

- Personalized Services: Offer personalized services such as welcome drinks, room preferences, and tailored recommendations for activities.
- Guest Communication: Maintain regular communication with guests before, during, and after their stay to ensure satisfaction and gather feedback.

Sustainability Initiatives:

- Eco-Friendly Practices: Adopt eco-friendly practices such as energy-efficient lighting, water conservation measures, and waste reduction programs.
- Promote Sustainability: Highlight your hotel's sustainability efforts in your marketing materials to attract eco-conscious guests.

Implement Strategies to Reduce Cancellations

- Flexible Booking Policies: Offer flexible booking and cancellation policies to give guests confidence when making reservations.
- Non-Refundable Rates: Provide a non-refundable booking option at a discounted rate to secure commitments.



**Thank you
for attention!**

