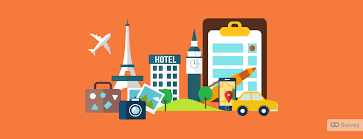
**Introduction:**

In today’s highly competitive environment, analysis of business in a company is an essential element so that it can build constructive plan accordingly. Through which they can find the insights of the business which can also answer the business questions/problems.



Hotel ABC is the chain hotel which is providing the finest hospitality services in different countries. The business management of the hotel wants to understand the business performance over the period through the visual dashboard. The hotel has provided the historical data for the analysis team.

Implementing the different phases of data analysis process:

**1. Ask:** Ask the questions to identify, understand and define the problem.

Key tasks –

**Business task:** Hotel ABC have two segments of hotels, city hotel and resort hotel. They are looking for a visual dashboard which can answer the following questions:

1. Is the hotel revenue growing over period of time?
2. Should the hotel increase parking lot size?
3. What trends can be seen in the data?

**Key stakeholders:**

Hotel business management team.

**2. Prepare:** Collecting the data, understanding the metrics of data and storing data.

* Using theHotel ABC previous 24 months historical data.
* The data is organized in the CSV file in the year wise with 32 attributes and more than 100000 records. Data related to market segment and meal cost also provided.
* This data set contains booking information for a city hotel and a resort hotel and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

Limitations found in the data:

* So many parameters have been included in the dataset which may not be needed for answering particular business questions.
* The data does not have the unique id related to customers, which would be helpful for customer preferences analysis in future.

**3. Process:** We have to clean the data before we transform and analyze it.

The data when merged is around 15KB. SQL and PowerBI tools are used for data anlysis.

*STEP:1- Loading the data*

For data cleaning and transforming MySQL server Workbench is used. As the data provided is in .xlsx, have to convert the file into .CSV file and then these files are imported into MySQL workbench.

*STEP: 2- Wrangling the data*

The data for year 2018 and 2019 are in different tables. Discounts, food pricing details are in market segment and meal cost tables respectively.

I) Checking the inconsistencies between the tables:

with comparing\_tble as

(select table\_name, column\_name, ordinal\_position, data\_type, column\_type, count(\*) as cnt

from information\_schema.columns

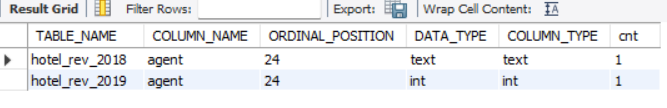
WHERE table\_schema='hotel\_db'

AND table\_name in ('hotel\_rev\_2018', 'hotel\_rev\_2019')

group by column\_name, ordinal\_position, data\_type, column\_type)

select \* from comparing\_tble

where comparing\_tble.cnt = 1



Shows agent column has different data type text and int.

As the data in the agent column is a number, converting the its data type in table hotel\_rev\_2018 into int.

Alter table hotel\_rev\_2018

Modify column agent int

II) Performed the LEFT JOIN operation to combine the entire data (four tables) into one table.

with hotels as

(select \* from hotel\_rev\_2018

union

select \* from hotel\_rev\_2019)

select \*

from hotels h

left join market\_segment ms on h.market\_segment = ms.market\_segment

left join meal\_cost mc on h.meal = mc.meal

Result: 65853 rows returned

*STEP:3- Cleaning the data*

* Finding the duplicate records: As there is no unique ID, tried used SQL CONCAT() function over the attributes arrival\_date\_day\_of\_month, arrival\_date\_month, arrival\_date\_year to create unique identification of the records. But noticed there are multiple records on each date. So, for finding the duplicate records more data is required. With help of PowerQuery feature we can check if there are any duplicate records.
* After the join operation there are 36 columns are formed. But as per the business task, instead only following 20 columns are considered:

hotel, arrival\_date\_year, arrival\_date\_month, arrival\_date\_week\_number, arrival\_date\_day\_of\_month, stays\_in\_weekend\_nights, stays\_in\_week\_nights, meal, country, market\_segment, reserved\_room\_type, assigned\_room\_type, agent, customer\_type, adr, required\_car\_parking\_spaces, reservation\_status,

reservation\_status\_date, Discount, Cost

A new table temp\_tb is created with the result set.

create table temp\_tb as

with hotels as

(select \* from hotel\_rev\_2018

union

select \* from hotel\_rev\_2019)

select

h.hotel, h.arrival\_date\_year, h.arrival\_date\_month, h.arrival\_date\_week\_number, h.arrival\_date\_day\_of\_month,

h.stays\_in\_weekend\_nights, h.stays\_in\_week\_nights, h.meal, h.country, h.market\_segment, h.reserved\_room\_type,

h.assigned\_room\_type, h.agent, h.customer\_type, h.adr, h.required\_car\_parking\_spaces, h.reservation\_status,

h.reservation\_status\_date, ms.Discount, mc.Cost

from hotels h

left join market\_segment ms on h.market\_segment = ms.market\_segment

left join meal\_cost mc on h.meal = mc.meal

*STEP:4- Using the view to find NULL values and percentages in the columns:*

SELECT

100.0 \* SUM(CASE WHEN hotel IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS hotelPercent,

100.0 \* SUM(CASE WHEN arrival\_date\_year IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS arrival\_date\_yearPercent,

100.0 \* SUM(CASE WHEN arrival\_date\_month IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS arrival\_date\_monthPercent,

100.0 \* SUM(CASE WHEN arrival\_date\_day\_of\_month IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS arrival\_date\_day\_of\_monthPercent,

100.0 \* SUM(CASE WHEN arrival\_date\_week\_number IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS arrival\_date\_week\_numberPercent,

100.0 \* SUM(CASE WHEN stays\_in\_week\_nights IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS stays\_in\_week\_nightsPercent,

100.0 \* SUM(CASE WHEN stays\_in\_weekend\_nights IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS stays\_in\_weekend\_nightsPercent,

100.0 \* SUM(CASE WHEN meal IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS mealPercent,

100.0 \* SUM(CASE WHEN country IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS countryPercent,

100.0 \* SUM(CASE WHEN market\_segment IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS market\_segmentPercent,

100.0 \* SUM(CASE WHEN reserved\_room\_type IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS reserved\_room\_typePercent,

100.0 \* SUM(CASE WHEN assigned\_room\_type IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS assigned\_room\_typePercent,

100.0 \* SUM(CASE WHEN agent IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS agentPercent,

100.0 \* SUM(CASE WHEN customer\_type IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS customer\_typePercent,

100.0 \* SUM(CASE WHEN adr IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS adrPercent,

100.0 \* SUM(CASE WHEN required\_car\_parking\_spaces IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS required\_car\_parking\_spacesPercent,

100.0 \* SUM(CASE WHEN reservation\_status IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS reservation\_statusPercent,

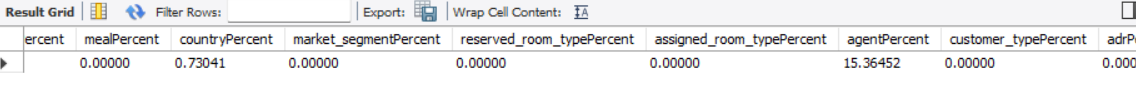
100.0 \* SUM(CASE WHEN reservation\_status\_date IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS reservation\_status\_datePercent,

100.0 \* SUM(CASE WHEN Discount IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS DiscountPercent,

100.0 \* SUM(CASE WHEN Cost IS NULL THEN 1 ELSE 0 END) / COUNT(\*) AS CostPercent

FROM temp\_tb

2 columns agent, country found to have NULL values with 15.36 % and 0.73% respectively.



*STEP:5- Handling the NULL values:*

Updating the column null values with the default values.

UPDATE temp\_tb

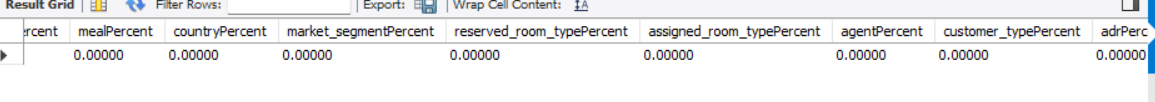
SET agent=0

WHERE agent is null

UPDATE temp\_tb

SET country = 'NA'

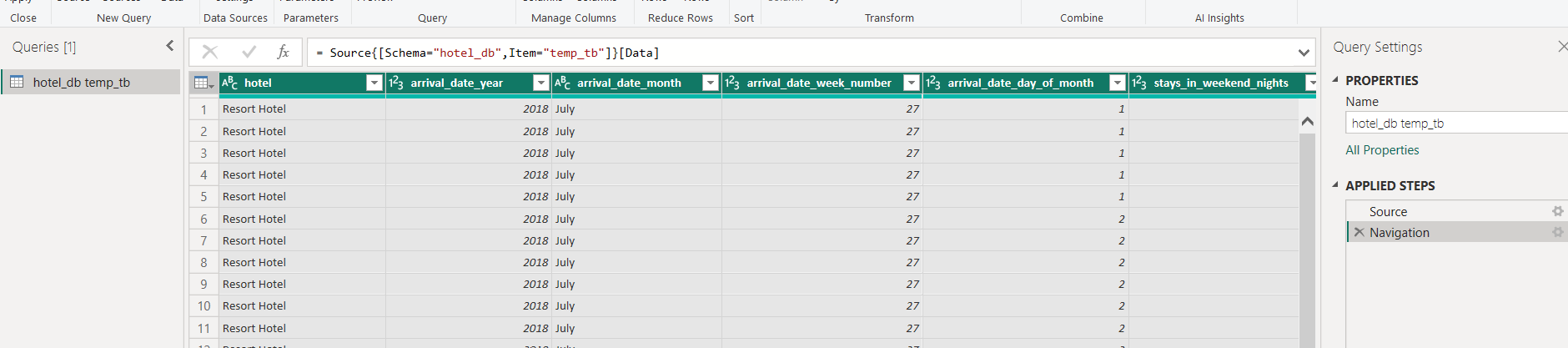
WHERE country is null



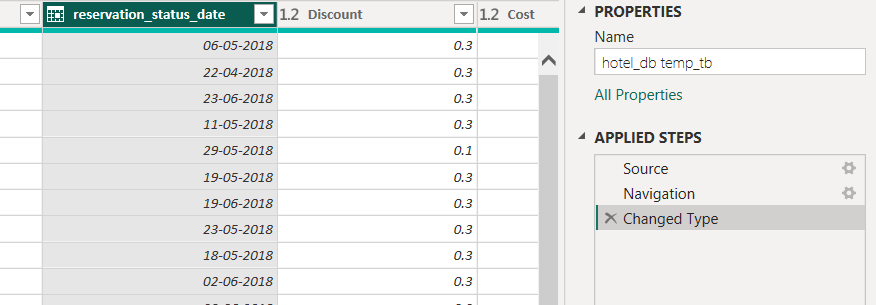
*STEP:6 - Data transformation in PowerBI:*

. Transferring the merged dataset table to PowerBI for further data transformation, analysis and visualization.

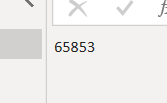
. Transforming the data in the PowerQuery



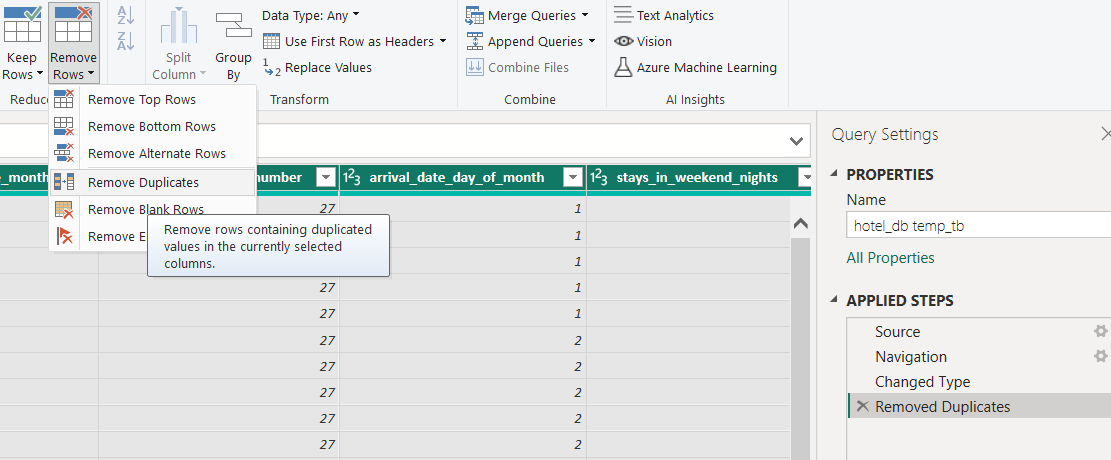
. Changed the datatype of reservation\_status\_date from text to date:



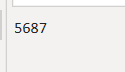
. Checking the total row count:



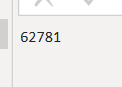
. Checking duplicate values:



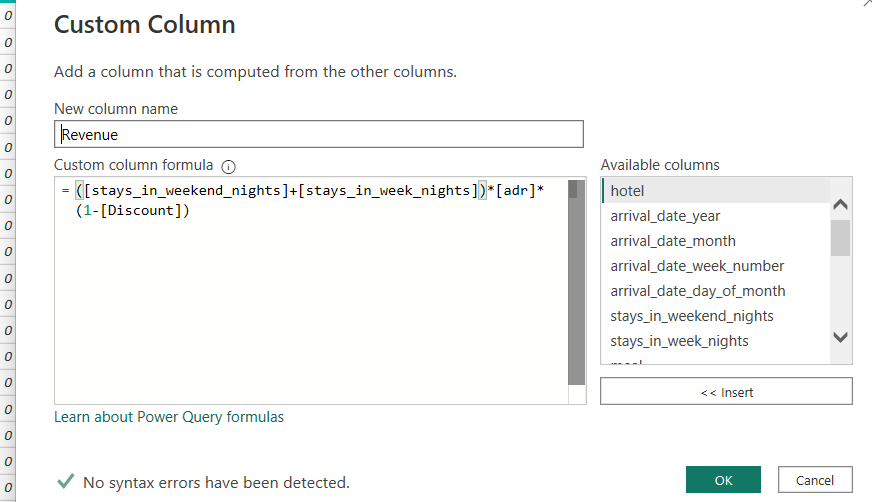
Total duplicate values found are 5687:



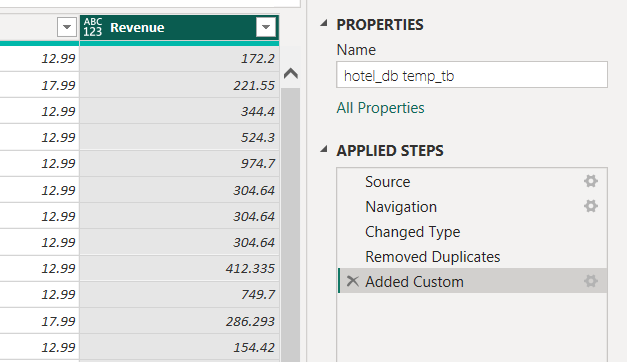
After removing the duplicate records, the dataset remained with 62781 records.



. Created a new Custom column Revenue:



Revenue = ([stays\_in\_weekend\_nights]+[stays\_in\_week\_nights])\*[adr]\*(1-[Discount])



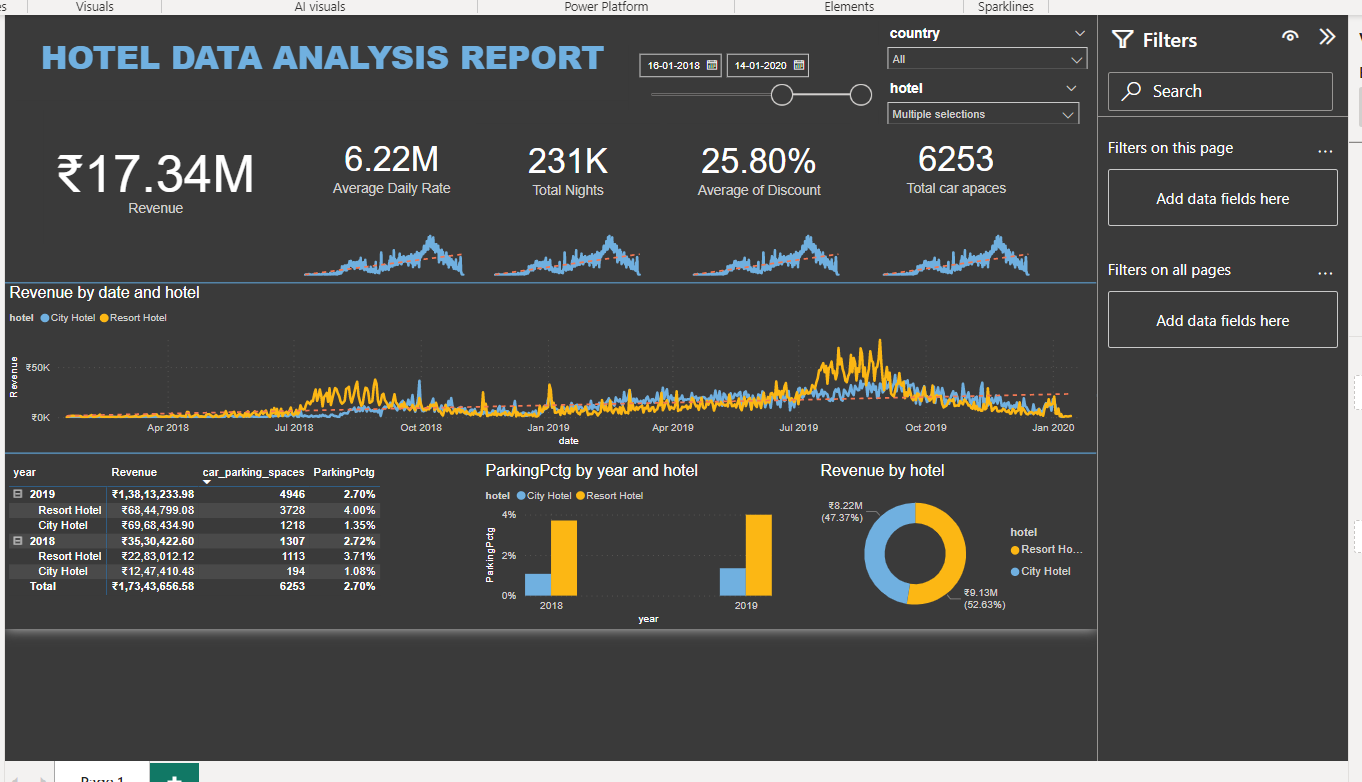
**4. Analyze:**

. In the fileds of PowerBI, added two new measures:

To understand the rate of bookings, total nights is used. For analyzing the parking space utilization parking percentage is being calculated.

Total Nights = sum(Query1[stays\_in\_week\_nights])+sum(Query1[stays\_in\_weekend\_nights])

Parking\_Pctg = sum(Query1[required\_car\_parking\_spaces])/[Total Nights]



**5. Key questions to answer:**

1. *Is the hotel revenue growing over period of time?*

Yes, in the Revenue by date chart there is clear indicator of upward trend. Particularly, Resort hotels are performing better.

1. *Should the hotel increase parking lot size?*

With the limited data provided regarding the parking, parking percentages over the period reveal that there is slight increase as per the bookings rate. Even though the bookings are increased over the time, increase in parking slot is very low.

With the existing scenarios parking space is well utilized with current booking rate.

1. *What trends can be seen in the data?*

The linear trend line shows positive with respective to major attributes. We can see that Revenue, Average discount rate, Bookings are showing up trends.