Hotel Data Analysis

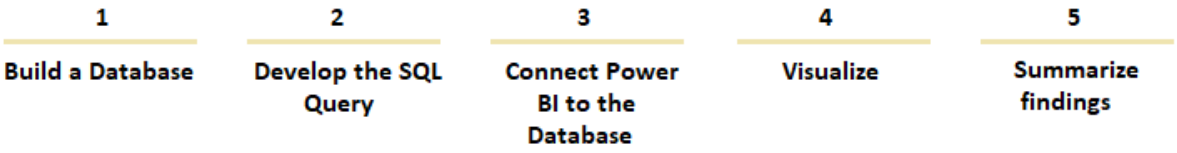
Goal: Develop a Database to analyze & visualize Hotel Booking Data

Requirements: Build a data story or dashboard using Power BI to present to your stakeholders

Questions from stakeholders:

1. Is our hotel revenue growing by year?
   (We have two hotel types so it would be good to segment revenue by hotel type.)
2. Should we increase our parking lot size?
   (We want to understand if there is a trend for guests with personal cars.)
3. What trends can we see in the data?
   (Focus on average daily rate and guests to explore seasonality.)

Data Analysis Project Pipeline to analyze and visualize Hotel Booking Data can be seen below.

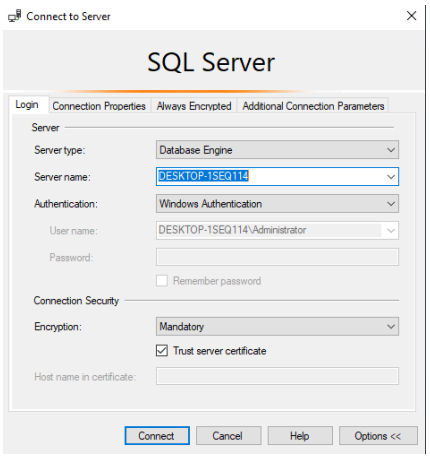


Step 1: Build a SQL database

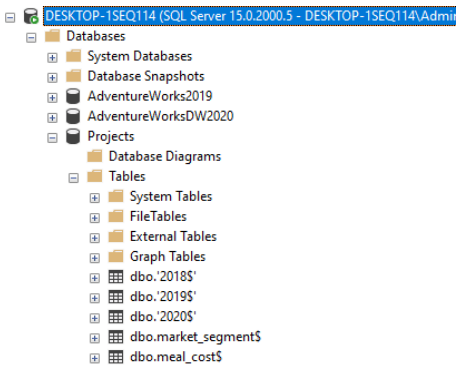
'Hotel Revenue Dataset' for the project was retrieved from Absent Data (Dataset URL). There are five tables in the MS Excel workbook.

1. 2018 (21996 rows)
2. 2019 (79264 rows)
3. 2020 (40687 rows)
4. market_segment (8 rows)
5. meal_cost (5 rows)

A connection to SQL Server was made in Microsoft SQL Server Management Studio.

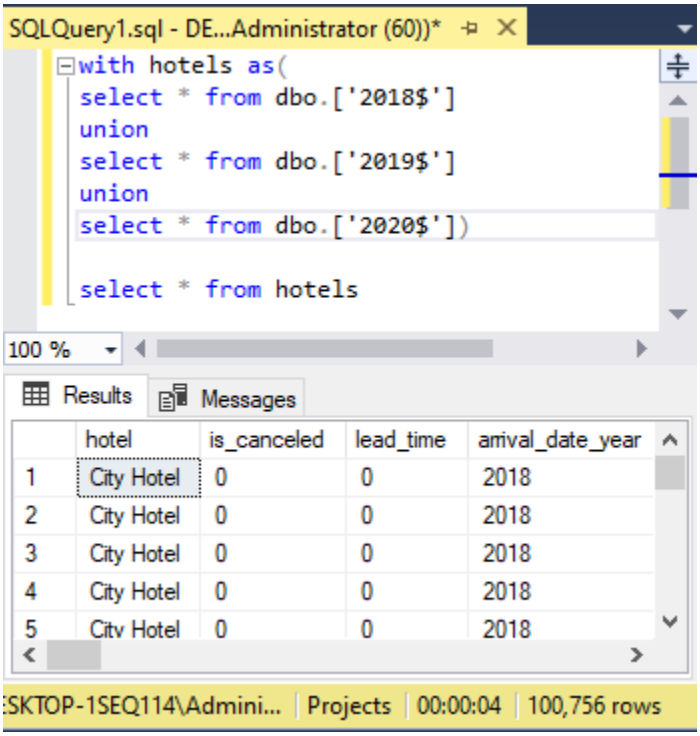


A new database named 'Projects' was created and data was imported into the 'Projects' database.



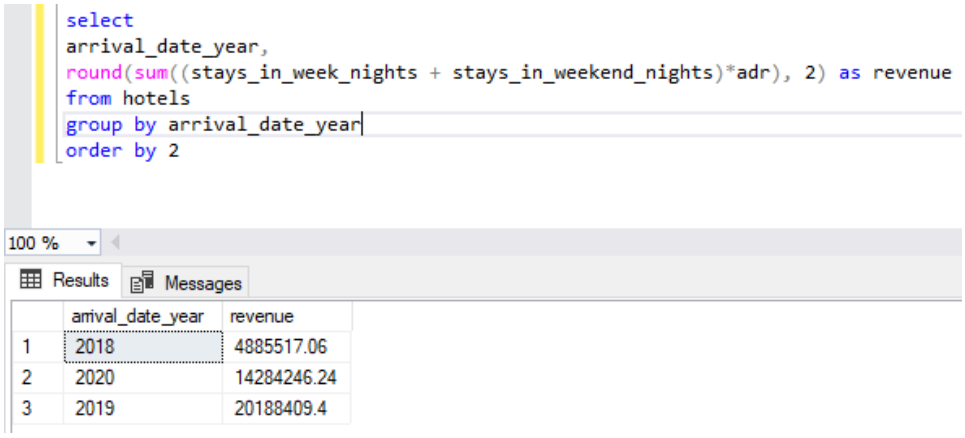
Step 2: Develop the SQL Query

First, hotel data from three tables (2018, 2019, 2020) was unified into one table using UNION.



In order to answer the first question ('Is our hotel revenue growing by year?'), 'revenue' was calculated using three columns (stays_in_week_nights, stays_in_weekend_nights and adr). 'revenue' by year was additionally calculated using GROUP BY and SUM. It can be stated that hotel revenue is growing by year according to our query result.

Note: *Data was only available for 9 months (Jan – Sept) for year 2020, therefore, revenue in 2019 is higher than revenue in 2020.*



There are two different hotel types in the dataset, therefore, 'hotel' column for each hotel type was added in the result set.

```sql
select
arrival_date_year, hotel,
round(sum((stays_in_week_nights + stays_in_weekend_nights)*adr), 2) as revenue
from hotels
group by arrival_date_year, hotel
order by 3
```

100 %

Results | Messages

	arrival_date_year	hotel	revenue
1	2018	City Hotel	1764667.57
2	2018	Resort Hotel	3120849.49
3	2020	Resort Hotel	6266123.81
4	2020	City Hotel	8018122.43
5	2019	Resort Hotel	9432430.29
6	2019	City Hotel	10755979.11

In order to create a dashboard in Power BI, two additional tables ('market_segment' and 'meal_cost') were combined with 'hotels' table using LEFT JOIN.

```sql
select *
from hotels h
left join dbo.market_segment$ s
on h.market_segment = s.market_segment
left join dbo.meal_cost$ c
on h.meal = c.meal
```

100 %

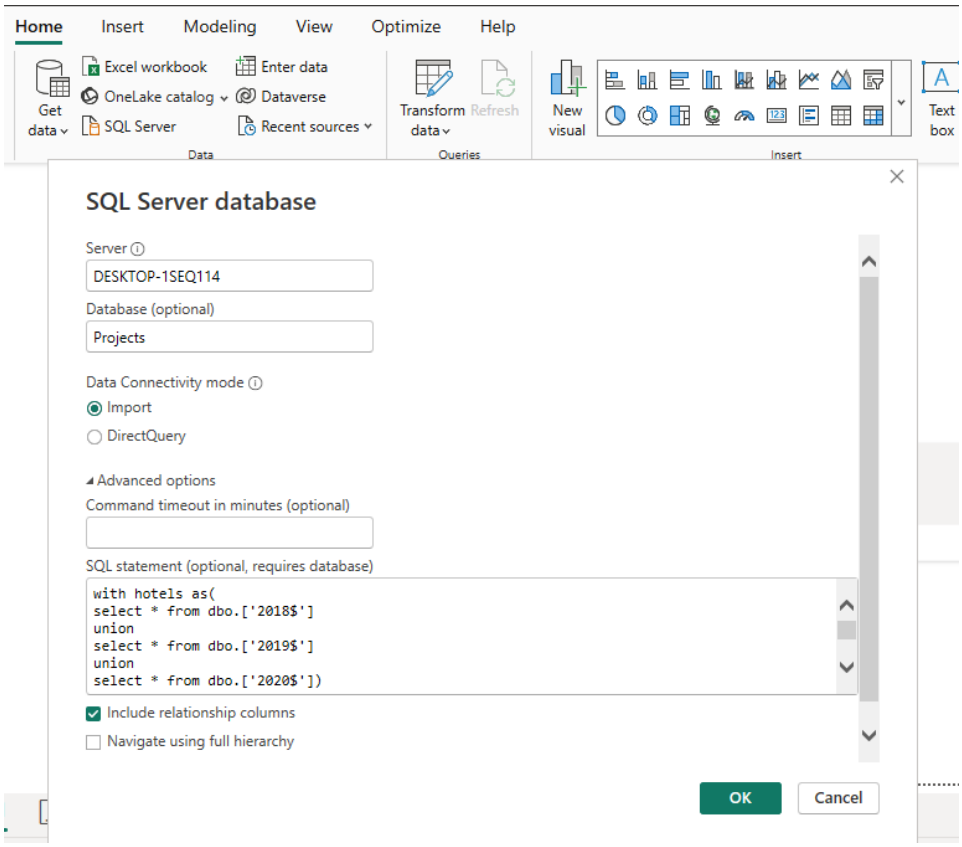
Results | Messages

	hotel	is_canceled	lead_time	arrival_date_year	arrival
1	City Hotel	0	0	2019	Febru
2	City Hotel	0	0	2019	July
3	City Hotel	0	0	2019	July
4	City Hotel	0	0	2019	July
5	City Hotel	0	0	2019	June

Up to this point, a required SQL query was developed to visualize data in Power BI.

Step 3: Connect Power BI to the Database

A connection to SQL server was configured in Power BI desktop, and data was imported from SQL database into Power BI using SQL query from the previous step.



Step 4: Visualize

Additional data transformation was performed before visualization process in Power BI. A custom column, 'Revenue', was added to Power BI model with the following calculation.

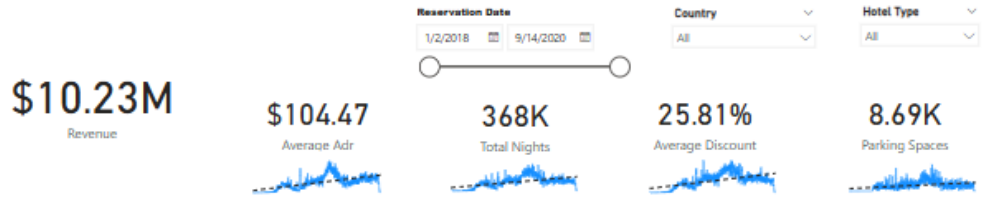
Revenue = ([stays_in_week_nights] + [stays_in_weekend_nights]) * ([adr]*[Discount])

A measure, 'total nights', was added into the model with the following calculation.

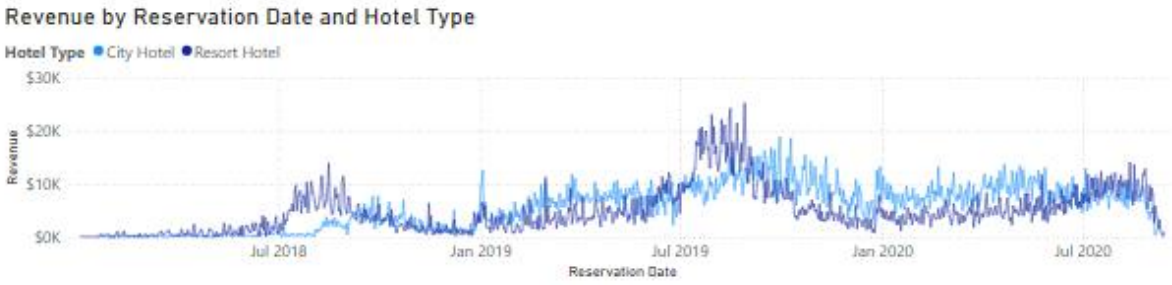
total nights = SUM(Query1[stays_in_week_nights])+SUM(Query1[stays_in_weekend_nights])

Creating visually pleasing layout was the first step in building the dashboard. Report page was divided into an invisible grid formed by two equally spaced horizontal lines. Therefore, report objects can be placed in three equally sized horizontal regions.

The upper region was used to show the KPis which includes 'Revenue', 'Average Adr', 'Total Nights', 'Average Discount' and 'Parking Spaces'. Five card visuals were added to display these data. Three slicer visuals ('Reservation Date', 'Country', 'Hotel Type') were added in this region. These slicers allow users to dynamically filter data displayed in visuals, enabling exploration and analysis by focusing on specific subsets of data for 'Reservation Date', 'Country' and 'Hotel Type'.



The middle region was used to show a line chart with 'Reservation Date' in X-axis and 'Revenue' in Y-axis. A page level filter was added to display data only after Jan 01, 2018 which removed unnecessary data on the line chart visual. 'Hotel Type' was additionally added in Legend of the line chart which allows report users perceive each hotel type's performance. This visual will help stakeholders answer their question ('What trends can we see in the data?').



In the lower region of the page, a matrix visual which displays four columns ('Year', 'Revenue', 'Parking Spaces Required', 'Parking Percentage') was added. This visual will help stakeholders answer their question whether they should increase parking lot spaces.

Year		Parking Spaces Required	Parking Percentage
⊟ 2018	3.511500014	1,283.00	2.49%
City Hotel	3.0550000015	185.00	0.95%
Resort Hotel	70.456500002	1,098.00	3.43%
⊟ 2019	51.455500021	5,161.00	2.50%
City Hotel	49.013499984	1,384.00	1.29%
Resort Hotel	12.441999978	3,777.00	3.80%
⊟ 2020	2.125499912	2,248.00	2.05%
City Hotel	58.283499975	532.00	0.88%
Resort Hotel	63.842000013	1,716.00	3.51%
Total	37.09249995	8,692.00	2.36%

The lower region contains a donut visual which displays performance percentage of each hotel type in revenue (i.e. a requirement from the stakeholder).