# How to Build a Data Analyst Portfolio?

# Hotel Revenue Exploratory Analysis

**Note**: This training material is designed for those who already have beginner knowledge of SQL, Excel, and Power BI. It will help you learn more about visualization, reporting, DAX functions, and SQL queries to enhance your analytical skills to add to your portfolio.

In this tutorial, we will follow some steps to build the best portfolio as a data analyst:

1. Create a Database
2. Query and analyse data with SQL
3. Integrate Power BI with a Database
4. Create Data Visualizations Using Power BI

## 1. Create a Database

Let’s start by creating a database in SSMS (SQL Server Management Studio) to analyse Hotel Booking Data. Download the data [here](https://absentdata.com/data-analysis/where-to-find-data/).

* Open SSMS and connect to the Server. Copy the server name for later use.
* Once connected, right-click on the databases on the left side (Object Explorer) and create a new database. Give a name “hotel\_revenue”.
* Now right-click on the newly created database and click on Import data.
* An import and export wizard will be opened.
  + Select the data source – Excel and import the Excel file ‘hotel\_revenue\_historical\_data which was ’downloaded from [here](https://absentdata.com/data-analysis/where-to-find-data/) earlier.
  + Select destination – Microsoft OLE DB Provider for SQL Server
    - Database name – Above created database name, in my case “hotel\_revenue” and click Next.
  + Select Source tables and views and Run Package by clicking Run Immediately.
  + Now you can see the tables in the database “hotel\_revenue”

## Querying data

Let’s now apply some commands to explore the data.

### Fetching Data from Tables

We will use the following commands to fetch data to view from the tables.

The wildcard ‘\*’ operator for retrieving all the records from the table.

**select \* from dbo.['2018$']**

**select \* from dbo.['2019$']**

**select \* from dbo.['2020$']**

Let’s combine the data using the union function.

**select \* from dbo.['2018$']**

**union**

**select \* from dbo.['2019$']**

**union**

**select \* from dbo.['2020$']**

## Exploratory Data Analysis (EDA)

Now, we are going to apply EDA to the data and try to answer the following questions.

1. Is our hotel revenue growing yearly?
2. Should we increase our parking lot size?
3. What trends can we see in the data?

Before answering the questions, we will first create a single temporary table hotel that combines all the data using the following code for easier access and analysis.

Either you create a temporary table or create a new table and insert the data into the table.

1. Open the SQL file name ‘SQLQuery1’ and copy the query under the comment ‘-- Creating the Revenues table to combine all the data’.
2. To insert data into the table, use the SQL query under the comment ‘--Combining the Data and Inserting into revenues table by using union.’

In our dataset, we don’t have revenue, but we do have adr (Average Daily Rate), stays\_in\_week\_nights, and stays\_in \_weekend\_nights.

So, we will create a new column revenue by using the data of these three columns as follows. Use the following Queries.

**-- Total Revenue from the Hotel Bookings only by year**

select r.arrival\_date\_year, sum((r.stays\_in\_week\_nights+r.stays\_in\_weekend\_nights)\*r.adr) as revenue

from dbo.['revenues$'] as r

group by r.arrival\_date\_year;

**-- Total Revenue from the Hotel Bookings only by year and hotel**

select r.arrival\_date\_year, r.hotel, cast(sum((r.stays\_in\_week\_nights+r.stays\_in\_weekend\_nights)\*r.adr) as decimal(10,2)) as revenue

from dbo.['revenues$']as r

group by r.hotel, r.arrival\_date\_year

order by r.hotel;

**-- Total Revenue from the Hotel Bookings only by year and hotel; Percentagae of Parking occupied/left over these years**

select r.arrival\_date\_year, r.hotel, cast(sum((r.stays\_in\_week\_nights+r.stays\_in\_weekend\_nights)\*r.adr) as decimal(10,2)) as revenue,

concat( round( (sum(r.required\_car\_parking\_spaces)/sum(r.stays\_in\_week\_nights + r.stays\_in\_weekend\_nights)) \*100 , 2), '%') as parking\_percentage

from dbo.['revenues$'] as r

group by r.arrival\_date\_year, r.hotel

order by r.hotel;

Before moving forward, we need to preprocess some columns. As we see the revenue is based on historical bookings. We haven’t included the market segment and meal cost here. Market segment involves the discount percentage and Meal cost includes the cost of the meal per person. The total revenue depends on both the data tables.

So, let's join the tables and find the actual revenue.

Using SQL, we will perform two left-join queries on the data.

First Left Join: Combines the hotels table with the market\_segment table by matching the market\_segment column in the “hotels” table with the market\_segment.market\_segment column.

Second Left Join: Combines the hotels table with the meal\_cost table by matching the meal column in the hotels table and the meal\_cost.meal column.

**-- Actual total revenue from the hotel booking, market segmentation and meals**

select \* from dbo.['revenues$']

left join dbo.market\_segment$

on ['revenues$'].market\_segment = [market\_segment$].market\_segment

left join dbo.meal\_cost$

on [meal\_cost$].meal = ['revenues$'].meal;

Now calculate the actual revenue using the following query

select r.arrival\_date\_year, r.hotel,

round (sum(((r.stays\_in\_week\_nights+r.stays\_in\_weekend\_nights) \* r.adr \* (1 - ms.Discount)) +

((r.adults + r.children + r.babies) \* mc.Cost)), 2) as revenue

from dbo.['revenues$'] as r

left join dbo.market\_segment$ as ms

on r.market\_segment = ms.market\_segment

left join dbo.meal\_cost$ as mc

on mc.meal = r.meal

group by r.arrival\_date\_year, r.hotel

order by r. hotel, r.arrival\_date\_year ;

A screenshot of a computer

Description automatically generated

## Create Data Visualizations Using Power BI

Now open Power BI on your PC to connect data base.

* Click on Get Data then select SQL Server.
* A new window titled **SQL Server Database** will appear.
* Enter your **server’s name** and database name in the respective fields.
* Expand the **Advanced** options and paste the following code in the “SQL text” field.
* And click **Ok**.

with ['temp\_revenues$'] as (

select \* from dbo.['2018$']

union

select \* from dbo.['2019$']

union

select \* from dbo.['2020$'])

select \* from dbo.['temp\_revenues$']

left join dbo.market\_segment$

on ['temp\_revenues$'].market\_segment = [market\_segment$].market\_segment

left join dbo.meal\_cost$

on meal\_cost$.meal = ['temp\_revenues$'].meal

* Now you can see that we have brought in our SQL information that we need.
* Hit Load button to load the data.

Before creating measures and Calculated columns, make sure the data types are aligned with the data.

### Calculated Columns

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

### Calculated Measure

* Total Nights = SUM(Query1[stays\_in\_weekend\_nights]) + SUM(Query1[stays\_in\_week\_nights])
* Total Percentage = SUM(Query1[required\_car\_parking\_spaces])/[Total Nights]

A screenshot of a hotel revenue report

Description automatically generated

**What trends can we see in the data?**

We have created some visuals using Power BI that show some possible trends. Here are a few of them:

1. Revenue increased from 2018 to 2019, but it began to decrease from 2019 to 2020.
2. The average daily rate (ADR) has increased from 2019 to 2020, from **$99.53** to **$104.47.**
3. Total number of nights booked by customers decreased from 2019 to 2020.
4. The discount percentage offered by the hotel has increased from 2019 to 2020 to attract more customers.