

HOTEL RESERVATIONS DATA ANALYSIS USING SQL

MENTORNESS INTERNSHIP PROJECT

Overview

This analysis of the hotel reservation dataset uncovers key insights into guest preferences, booking trends, and other key factors that impact the hotel's operations and makes recommendations for improving business performance.



Data Sources and Data

SQL data analysis was conducted on the
[Hotel Reservation Dataset.csv](#) from
Mentorness Internship.

Data cleaning and Database creation

For the data cleaning process:

- I used Microsoft Excel to remove all empty cells and duplicates
- I created new columns to separate the Date column, by creating new columns for Day, Month, Year, and Time.

Database creation on Mysql:

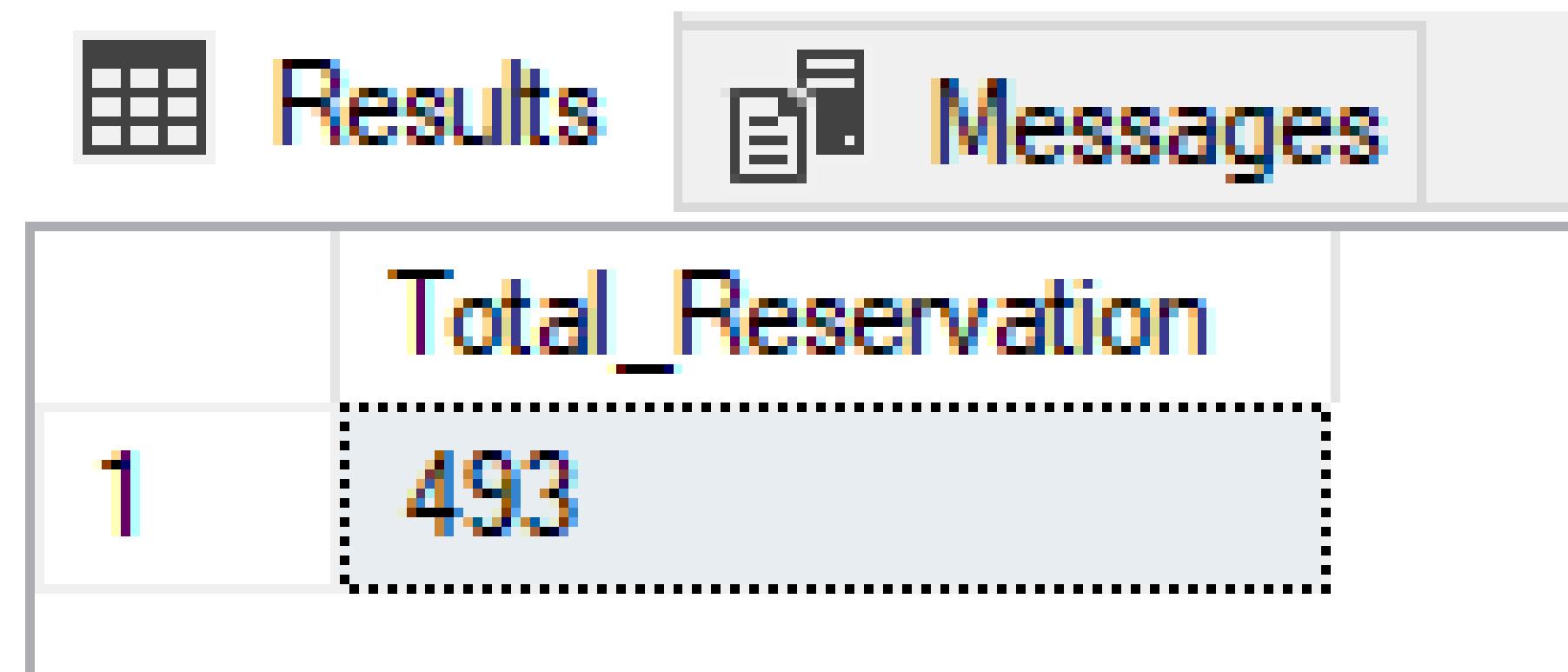
I used the Sql server 2019 import and export Data option to import the Dataset (CSV file) into Mysql.

15 questions for which SQL queries are written to gain insights:

1. What is the total number of reservations in the dataset?
2. Which meal plan is the most popular among guests?
3. What is the average price per room for reservations involving children?
4. How many reservations were made for the year 20XX (replace XX with the desired year)?
5. What is the most commonly booked room type?
6. How many reservations fall on a weekend (`no_of_weekend_nights > 0`)?
7. What is the highest and lowest lead time for reservations?
8. What is the most common market segment type for reservations?
9. How many reservations have a booking status of "Confirmed"?
10. What is the total number of adults and children across all reservations?
11. What is the average number of weekend nights for reservations involving children?
12. How many reservations were made in each month of the year?
13. What is the average number of nights (both weekend and weekday) spent by guests for each room type?
14. For reservations involving children, what is the most common room type, and what is the average price for that room type?
15. Find the market segment type that generates the highest average price per room

Total Hotel Reservations

```
Select count(Booking_ID) Total_Reservation  
from ['Hotel Reservation$']  
where booking_status = 'not_canceled'
```

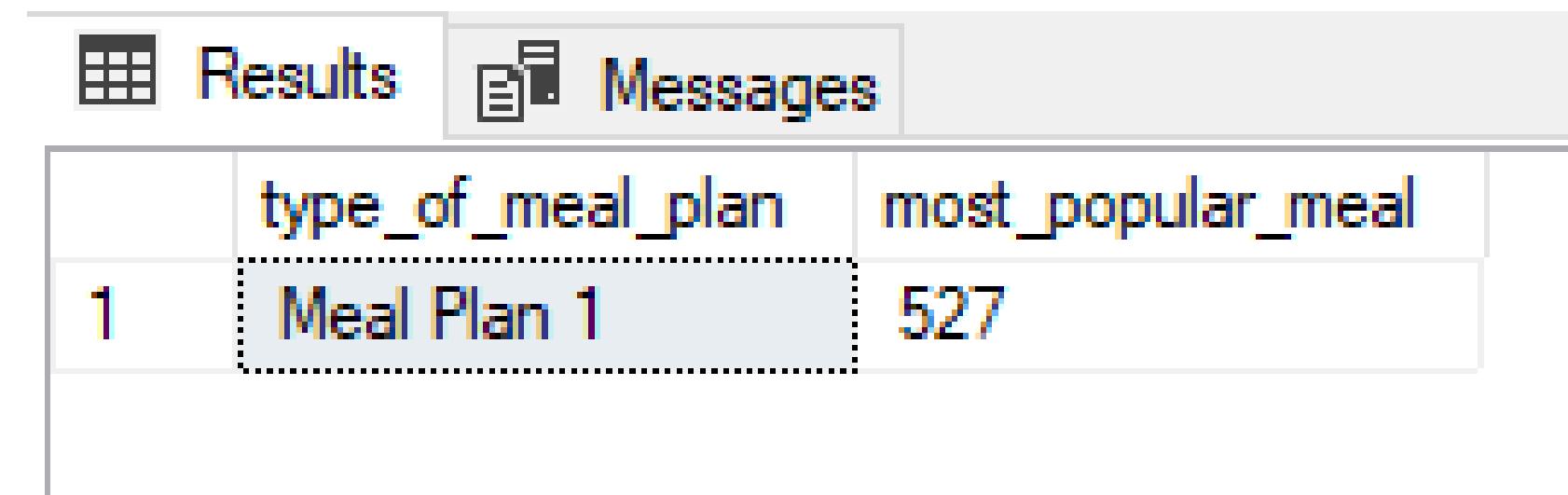


The screenshot shows a user interface for viewing database query results. At the top, there are two tabs: "Results" (selected) and "Messages". Below the tabs, the query results are displayed in a table with one row. The table has two columns: the first column contains the value "1", and the second column contains the value "493", which is highlighted with a dotted border. The column header is "Total_Reservation".

	Total_Reservation
1	493

Most popular meal among guests

```
WITH CTE_Popular AS (
Select
    max(meal_plans) popular_meal,
    type_of_meal_plan
from (
Select type_of_meal_plan,
    count(type_of_meal_plan) as meal_plans
from ['Hotel Reservation$']
group by type_of_meal_plan
) popular_Meal
group by type_of_meal_plan
)
Select Top 1
    type_of_meal_plan,
    max(popular_meal) as most_popular_meal
from CTE_Popular
group by type_of_meal_plan
order by most_popular_meal desc
```

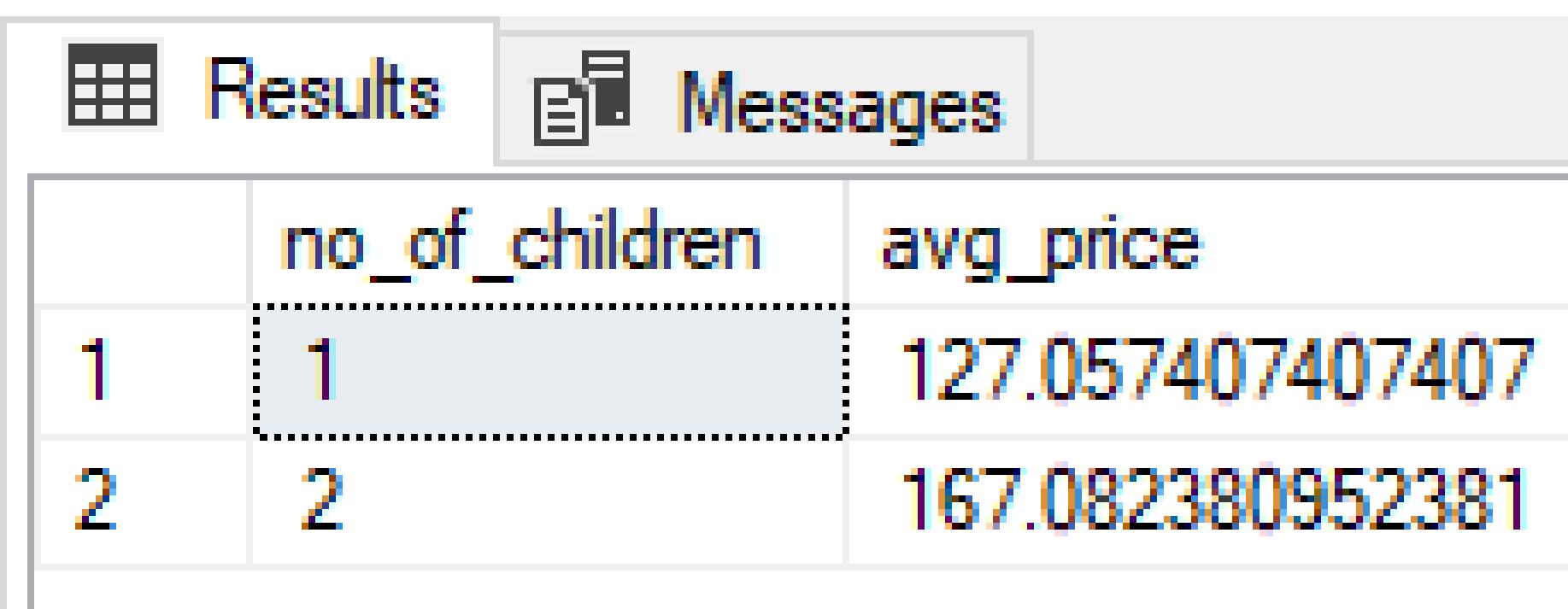


The screenshot shows a SQL query results window with three tabs: 'Results' (selected), 'Messages', and 'Results'. The 'Results' tab displays a table with two columns: 'type_of_meal_plan' and 'most_popular_meal'. There is one row with the value 'Meal Plan 1' in the first column and '527' in the second column.

	type_of_meal_plan	most_popular_meal
1	Meal Plan 1	527

Average price per room for reservation involving children

```
select no_of_children,  
avg(avg_price_per_room) avg_price from ['Hotel Reservations']  
where no_of_children > 0  
group by no_of_children
```

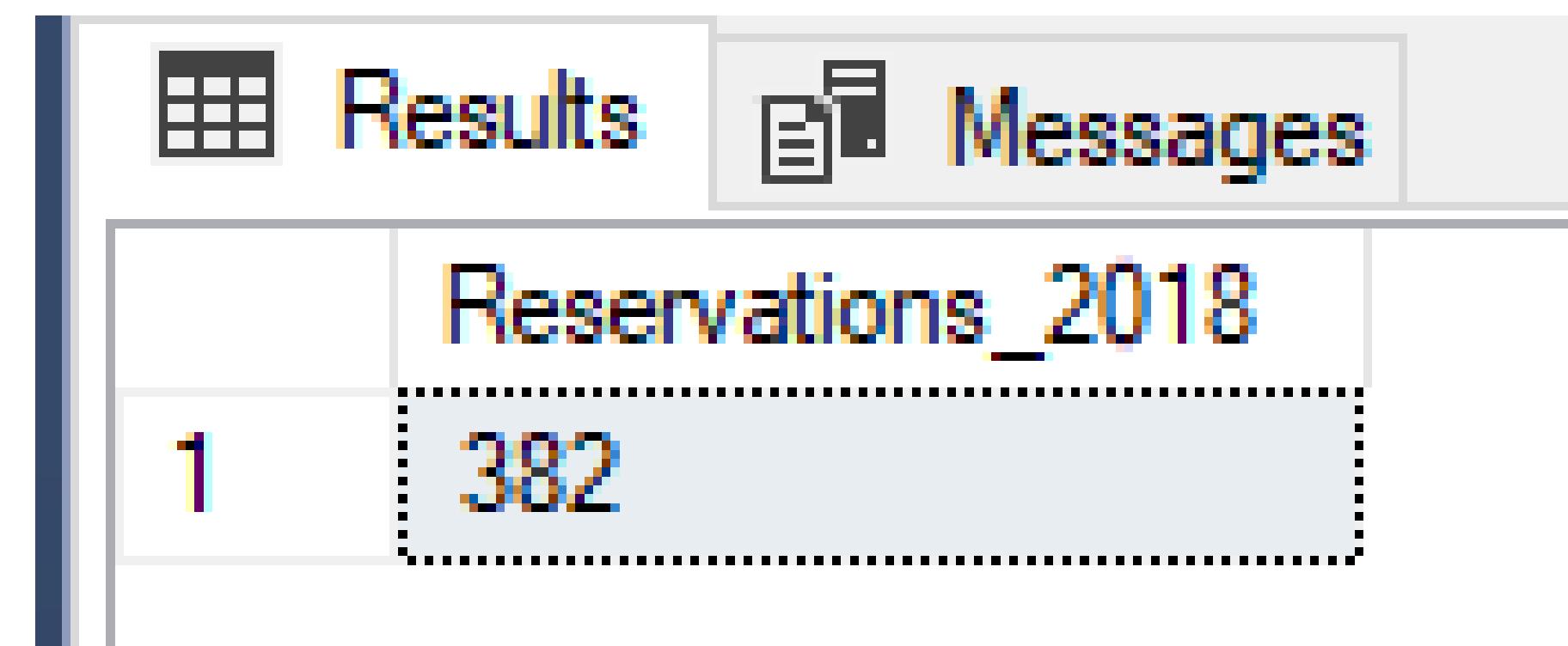


The screenshot shows a database interface with two tabs: 'Results' and 'Messages'. The 'Results' tab is selected, displaying a table with two rows of data. The table has two columns: 'no_of_children' and 'avg_price'. The first row, labeled '1', has a value of '1' in the 'no_of_children' column and '127.057407407407' in the 'avg_price' column. The second row, labeled '2', has a value of '2' in the 'no_of_children' column and '167.082380952381' in the 'avg_price' column.

	no_of_children	avg_price
1	1	127.057407407407
2	2	167.082380952381

Reservations made for year 2018

```
    select count(booking_status) as Reservations_2018
from ['Hotel Reservation$']
where booking_status = 'not_canceled'
    and arrival_date >= '2018-01-01'
    and arrival_date <= '2018-12-31'
```

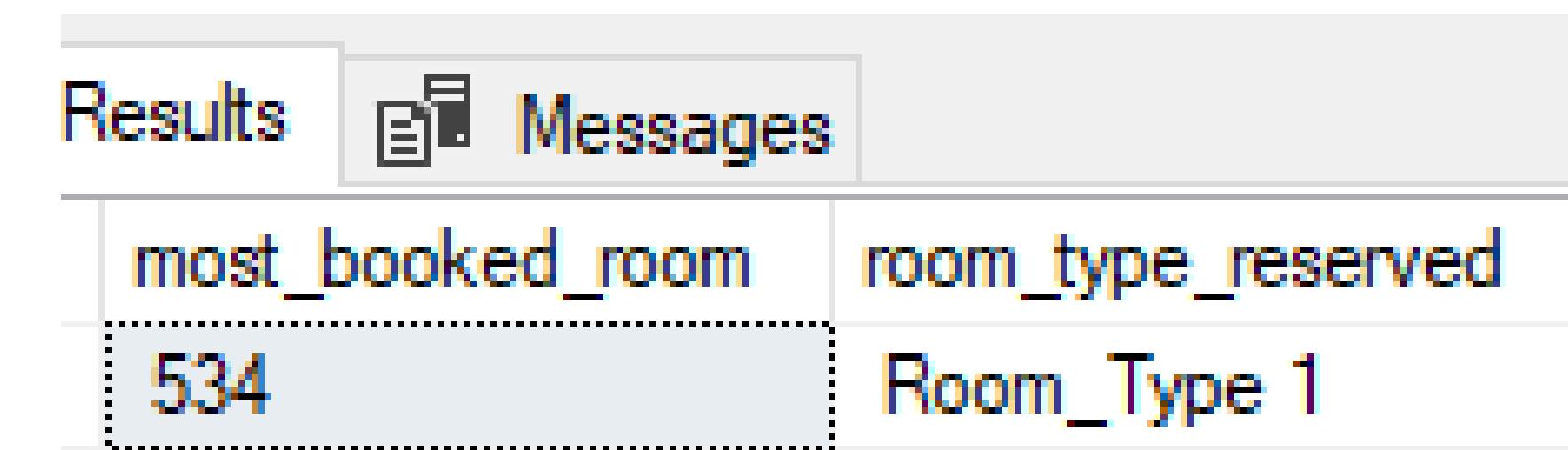


The screenshot shows a database query results interface. At the top, there are two tabs: "Results" (selected) and "Messages". Below the tabs, the query results are displayed in a table with one row. The table has two columns: the first column contains the value "1", and the second column, which is highlighted with a dashed border, contains the value "382". The column header is "Reservations_2018".

	Reservations_2018
1	382

Most commonly booked room type

```
Select room_type_reserved,  
      max(room_count) Commonly_Booked_room  
from (  
Select  
      room_type_reserved,  
      count(room_type_reserved) as room_count  
from ['Hotel Reservation$']  
group by room_type_reserved  
) room  
group by room_type_reserved  
)  
  
Select top 1  
      max(Commonly_Booked_room) as most_booked_room,  
      room_type_reserved  
from Book_Room  
group by room_type_reserved  
order by most booked room desc
```

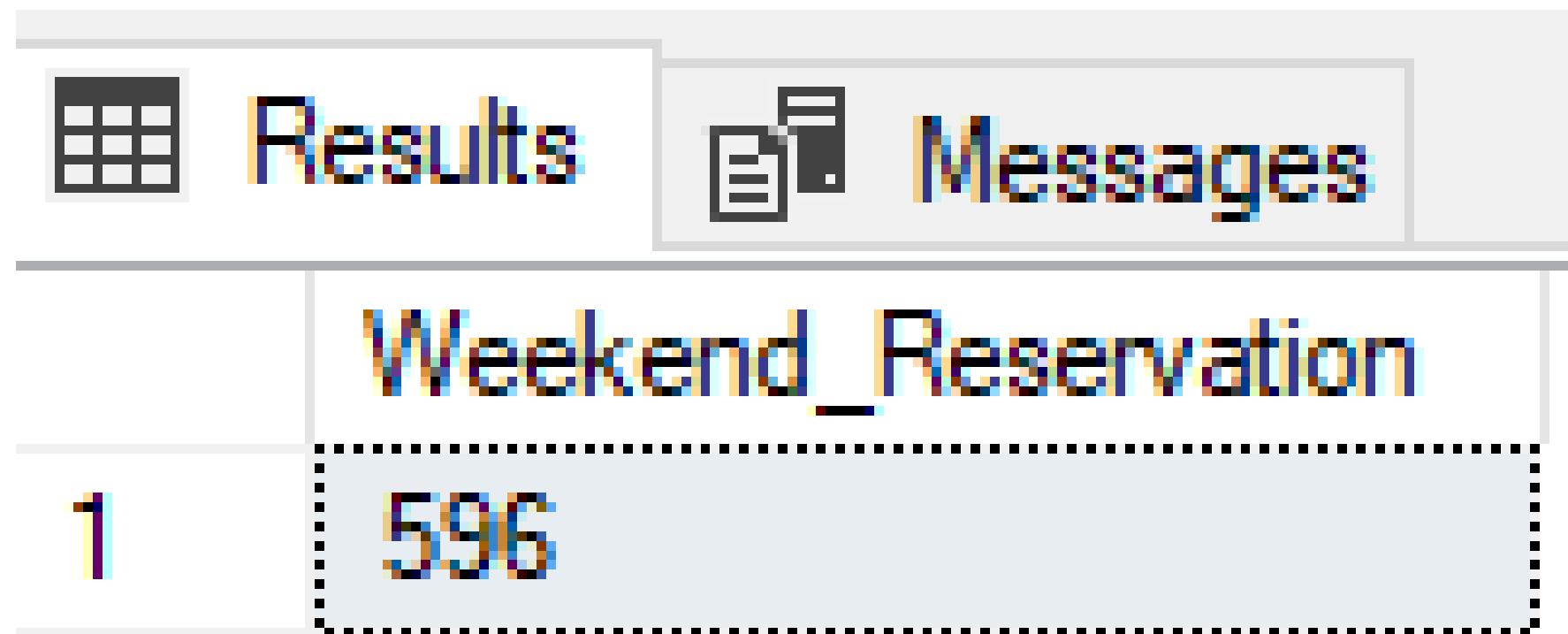


The screenshot shows a SQL query results window with two tabs: 'Results' and 'Messages'. The 'Results' tab is selected, displaying a single row of data. The data consists of two columns: 'most_booked_room' and 'room_type_reserved'. The value for 'most_booked_room' is 534, and the value for 'room_type_reserved' is Room_Type 1.

most_booked_room	room_type_reserved
534	Room_Type 1

Reservation that falls on weekend night

```
Select sum(no_of_weekend_nights) as Weekend_Reservation  
from ['Hotel Reservation$']  
where no_of_weekend_nights > 0
```



The screenshot shows a database interface with two tabs at the top: 'Results' (selected) and 'Messages'. Below the tabs is a table with one row. The table has two columns: the first column contains the value '1', and the second column, which is highlighted with a dotted border, contains the value '596'. The column header is 'Weekend_Reservation'.

	Weekend_Reservation
1	596

Highest and Lowest Lead time for reservation

```
Select max(Lead_time_2) max_lead_time,  
      min(Lead_time_2) min_lead_time  
from ['Hotel Reservation$']
```



The screenshot shows a database interface with a results table. The table has two columns: 'max_lead_time' and 'min_lead_time'. The first row contains the value '1' under the 'max_lead_time' column and '1899-12-30 00:00:00.000' under the 'min_lead_time' column. The second row contains the value '2' under the 'max_lead_time' column and '1899-12-30 04:43:00.000' under the 'min_lead_time' column. The 'Results' tab is selected.

	max_lead_time	min_lead_time
1	1899-12-30 04:43:00.000	1899-12-30 00:00:00.000
2	1899-12-30 04:43:00.000	1899-12-30 04:43:00.000

Most common market segment type for reservations

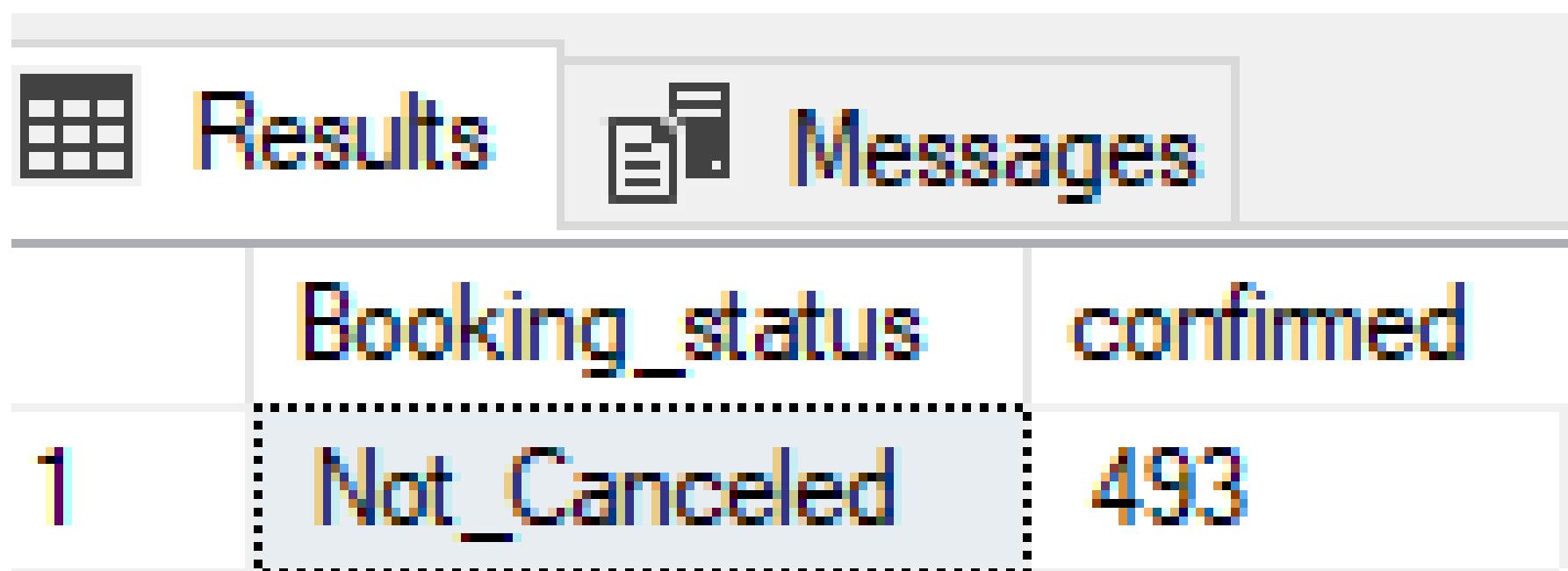
```
Select market_segment_type,  
      count(market_segment_type) as Most_common_market  
from ["Hotel Reservation$"]  
group by market_segment_type  
order by Most_common_market desc
```

Results Messages

	market_segment_type	Most_common_market
1	Online	518
2	Offline	140
3	Corporate	27
4	Complementary	14
5	Aviation	1

Reservations that have a booking status of Confirmed

```
Select Booking_status,  
      count(Booking_status) confirmed  
from ['Hotel Reservations']  
where booking_status = 'Not_canceled'  
group by Booking_status
```

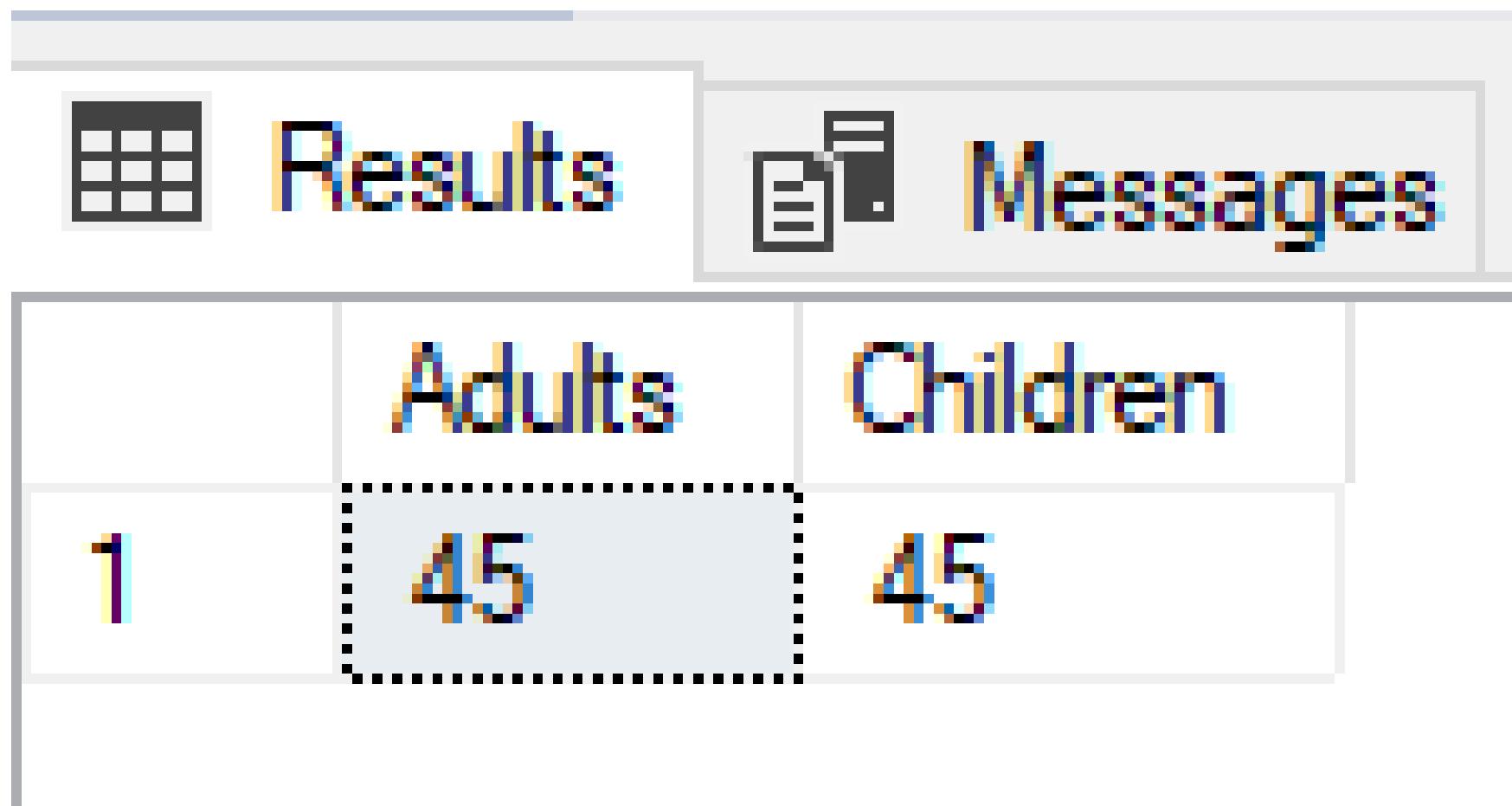


The screenshot shows a database query results interface. At the top, there are two tabs: "Results" (selected) and "Messages". The main area displays a table with the following data:

	Booking_status	confirmed
1	Not_Canceled	493

Total number of adults and children across all reservations

```
Select count(no_of_adults) Adults,  
      count(no_of_children) Children  
from ['Hotel_Reservation$']  
where no_of_children > 0 and no_of_adults > 0
```

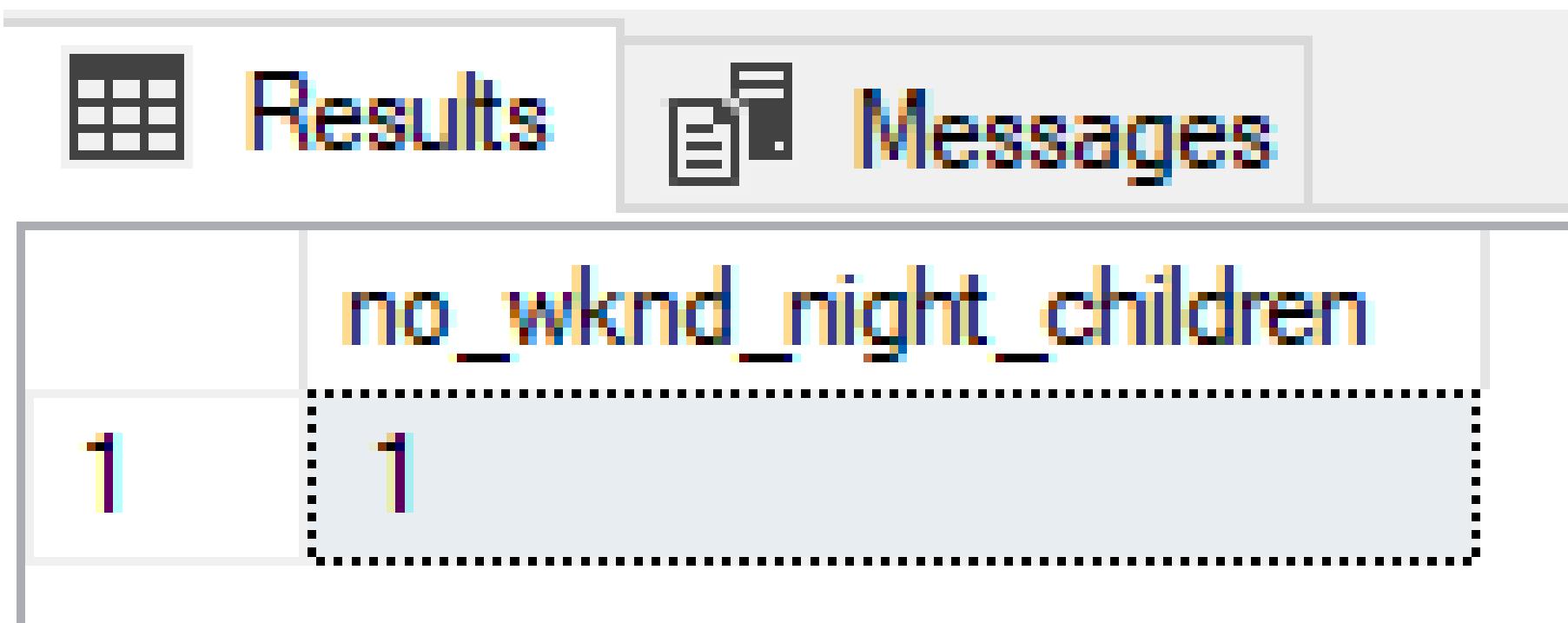


The screenshot shows a data visualization interface with a toolbar at the top. The 'Results' tab is selected, indicated by a blue border around its icon and text. Next to it is the 'Messages' tab. Below the toolbar is a table with two columns: 'Adults' and 'Children'. The 'Adults' column contains the value '45', which is highlighted with a dashed border, suggesting it is the current focus or a selected value.

	Adults	Children
1	45	45

The average number of weekend nights for reservations involving children

```
Select Avg(no_of_weekend_nights) as no_wknd_night_children  
from ["Hotel Reservation$"]  
where no_of_children > 0
```



A screenshot of a database query results interface. At the top, there are two tabs: "Results" (selected) and "Messages". The "Results" tab shows a single row of data in a table format. The table has one column labeled "no_wknd_night_children" and one row with the value "1". The cell containing the value "1" is highlighted with a dashed border.

	no_wknd_night_children
1	1

Reservations made in each month of the year

SELECT

```
    Year AS reservation_year,  
    Months AS reservation_month,  
    COUNT(booking_status) AS reservations_count  
    from ['Hotel Reservation$']  
    where booking_status = 'not_canceled'
```

GROUP BY

```
    Year,  
    Months
```

ORDER BY

```
    Year  
    Months;
```

Results Messages

	reservation_year	reservation_month	reservations_count
1	2017	Apr	2
2	2017	Aug	13
3	2017	Dec	11
4	2017	Feb	7
5	2017	Jan	2
6	2017	Jul	4
7	2017	Jun	2
8	2017	Mar	1
9	2017	May	4
10	2017	Nov	9
11	2017	Oct	29

Results Messages

	reservation_year	reservation_month	reservations_count
12	2017	Sep	27
13	2018	Apr	32
14	2018	Aug	41
15	2018	Dec	21
16	2018	Feb	28
17	2018	Jan	19
18	2018	Jul	20
19	2018	Jun	49
20	2018	Mar	42
21	2018	May	41
22	2018	Nov	28

The average number of nights (both weekend and weekday) spent by guests for each room type

```
Select room_type_reserved,  
      Avg(no_of_weekend_nights) avg_weekend_night,  
      Avg(no_of_week_nights) avg_weekdays_night  
from ['Hotel Reservation$']  
      Where no_of_weekend_nights > 0 and no_of_week_nights > 0  
      group by room_type_reserved
```

Results Messages

	room_type_reserved	avg_weekend_night	avg_weekdays_night
1	Room_Type 1	1.53781512605042	2.43697478991597
2	Room_Type 2	1.6	2.2
3	Room_Type 4	1.6547619047619	2.78571428571429
4	Room_Type 6	1.9	2.9
5	Room_Type 7	1.6666666666666667	1.6666666666666667

The most common room type, and the average price for that room type,for reservations involving children

```
WITH CommonRoom AS (
    SELECT Top 1 room_type_reserved,
           COUNT(*) AS reservation_count
    FROM ['Hotel Reservation$'] r
   WHERE no_of_children > 0
   GROUP BY room_type_reserved
   ORDER BY reservation_count DESC
)
SELECT
    r.room_type_reserved,
    AVG(r.avg_price_per_room) AS average_price
FROM
    ['Hotel Reservation$'] r
JOIN
    CommonRoom cr ON r.room_type_reserved = cr.room_type_reserve
WHERE
    r.no_of_children > 0
GROUP BY
    r.room_type_reserved;
```



A screenshot of a SQL query results window. The window has tabs for 'Results' and 'Messages'. The 'Results' tab is selected and shows a single row of data in a table format. The table has two columns: 'room_type_reserved' and 'average_price'. The value for 'room_type_reserved' is 'Room_Type 1' and the value for 'average_price' is '123.122916666667'.

	room_type_reserved	average_price
1	Room_Type 1	123.122916666667

The market segment type that generates the highest average price per room

```
Select Top 1 market_segment_type,  
      max(avg_price_per_room) Highest_Price  
from ['Hotel_Reservation$']  
Group by market_segment_type  
Order by Highest_Price desc
```

Results Messages

	market_segment_type	Highest_Price
1	Online	258

Recommendations

Based on the data analysis, the following key recommendations are made:

- Enhance Popular Meal Plans: Invest in improving the most popular meal plans to further enhance guest satisfaction.
- Focus on Family Packages: Offer special packages for families with children to leverage the average price per room.
- Optimize Room Types: Promote the most commonly booked room types and analyze why other types are less popular.
- Lead Time Strategy: Develop strategies to manage reservations with high lead times effectively.
- Market Segment Targeting: Focus marketing efforts on the most common and high-value market segments.



Summary: The analysis of hotel reservation data provides valuable insights into guest preferences and behavior.

Thank you

BABATUNDE O. AKANBI

Email: abutunde43@gmail.com

Phone No: 08166778866

Github: <https://github.com/abutunde/>

Linkedin: <http://www.linkedin.com/in/babatunde-akanbi-60682a13b>