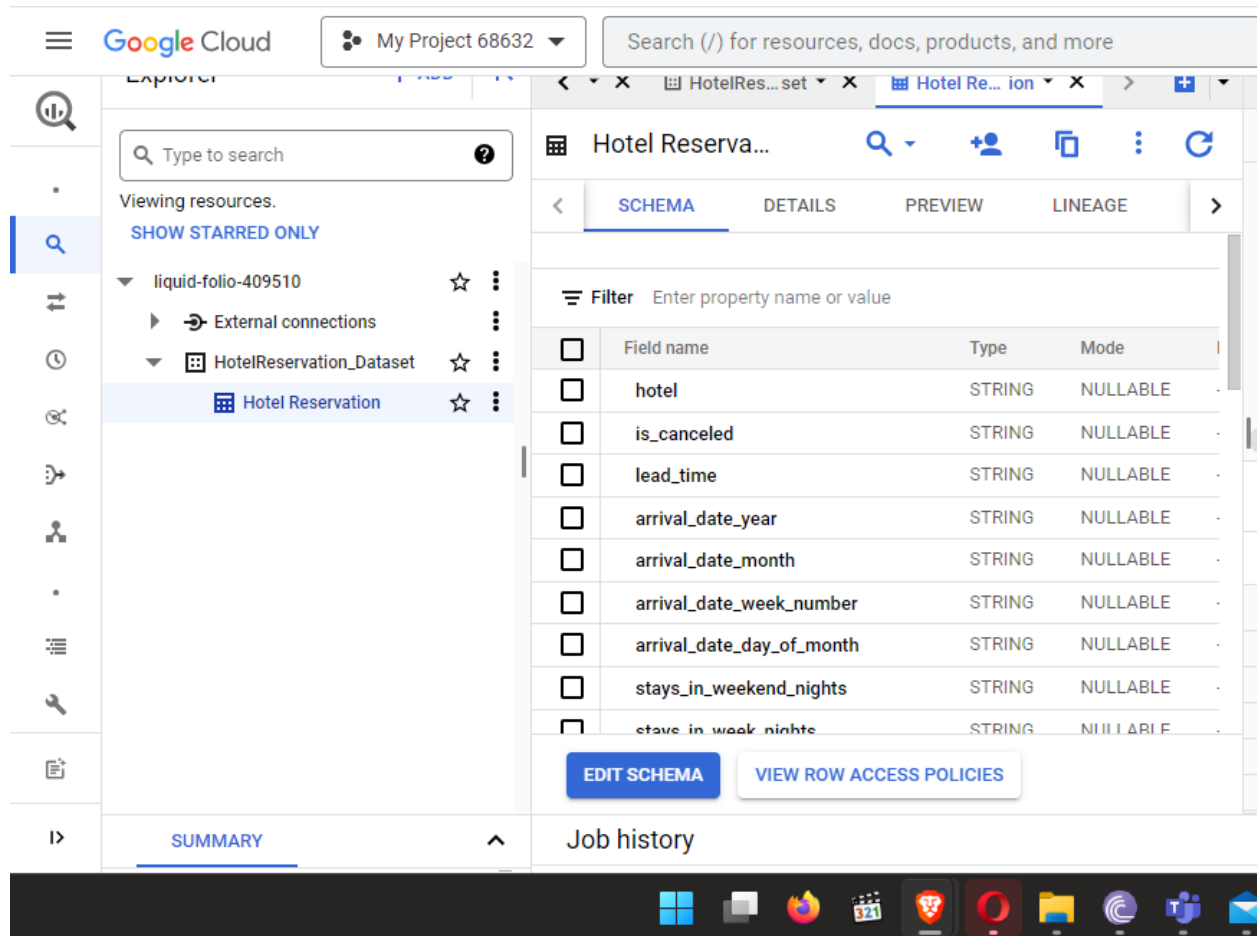# Hotel Reservation Analysis using BigQuery

As a data analyst working on a hotel reservation analysis project, my objective was to extract meaningful insights from a dataset provided by the hotel. Using structured query language (SQL), I aimed to retrieve valuable information that would contribute to improving sales performance.



This table contains data with over 30 columns and 100,000+ rows.

Here, I discovered that the dataset had mixed data types containing both numeric and alphabetical information. Since I did not have access to the primary data source and separating the data into specific numeric and alphabetical types would have been challenging, I made a practical decision to handle all data as strings.
By treating all data as strings, I avoided the hassle of dealing with complex data transformations and conversions. It allowed me to work with the dataset as it was without risking data loss or inaccuracies in the analysis.

This approach provided a suitable workaround, given the constraints of the project. Furthermore, I noticed that some cells in the dataset contained null values, indicating missing or unknown information. To accommodate these null values without causing any issues during analysis, I set the modes of all fields to nullable. This ensured that the dataset could handle and process these missing data points smoothly.
In summary, by using the string data type for all data and setting the modes to nullable, I effectively managed the mixed data types and null values in the dataset. This approach allowed me to work with the available data in a practical and efficient manner, considering the limitations of the project.

```
1  SELECT COUNT(*) AS total_reservations
2  FROM `liquid-folio-409510.HotelReservation_Dataset.Hotel
   Reservation` LIMIT 1000
```

Processing location: US ⊗                        Press Alt+F1 for Accessibility Options.

Query results        ⬇ SAVE RESULTS ▼

JOB INFORMATION        RESULTS        CHART  PREVIEW

| Row | total_reservations |
|-----|--------------------|
| 1   | 119390             |

The purpose of this query is to obtain a quick count of the total number of reservations in the dataset. This information can be useful for basic data analysis, understanding the scale of the dataset, or providing a summary statistic for reporting purposes.

## total_reservations
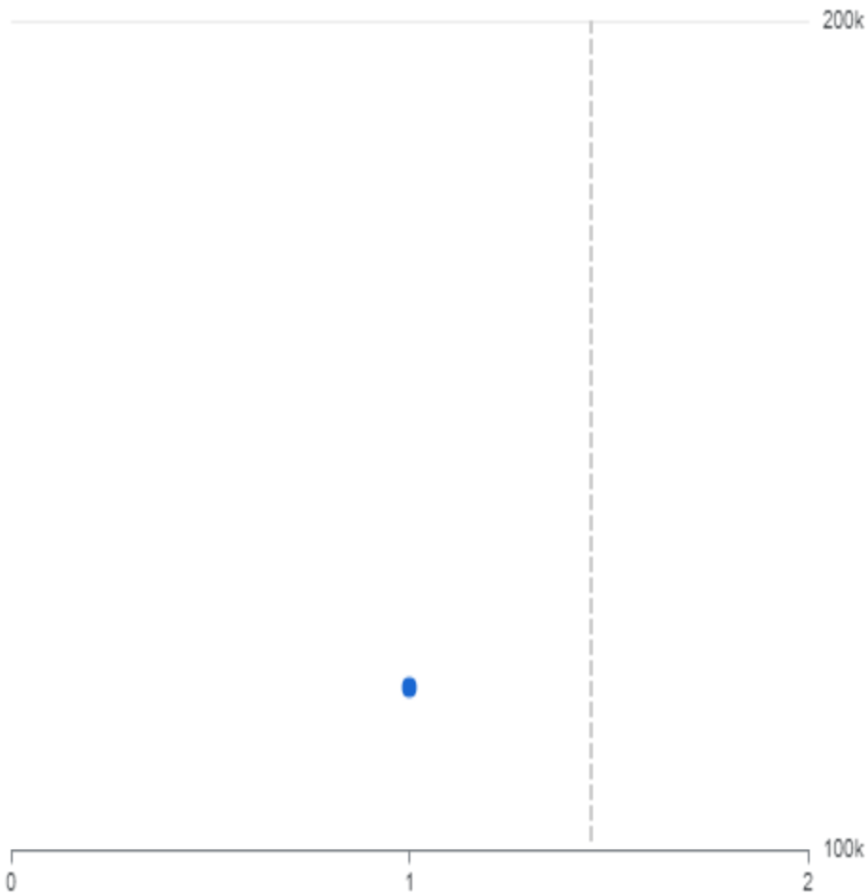


This visualization provides a visual representation of the total number of reservations, making it easier to comprehend the scale and magnitude of the dataset. It gives the hotel a quick overview of the volume of reservations they have received.

Untitled 8    Untitled 9

**Untitled 9**    ▶ RUN

```
1   SELECT market_segment, COUNT(*) AS reservations_count
2   FROM `liquid-folio-409510.HotelReservation_Dataset`.
    `Hotel Reservation`
3   GROUP BY market_segment
```

Processing location: US ⊗                   Press Alt+F1 for Accessibility Options.

**Query results**    ⬇ SAVE RESULTS ▾    📈 ▾    ⬍

| JOB INFORMATION | RESULTS | CHART | PREVIEW |

| Row | market_segment ▾ | reservations_count |
| --- | --- | --- |
| 1 | Offline TA/TO | 24219 |
| 2 | Groups | 19811 |
| 3 | Direct | 12606 |
| 4 | Online TA | 56477 |

The query calculates the count of reservations for each market segment and groups the results by the market_segment column.

## Query results

⬇ SAVE RESULTS ▾    📈 ▾    ⌄⌃

< | JOB INFORMATION | RESULTS | CHART PREVIEW | >

| Row | market_segment ▾ | reservations_count |
|-----|------------------|--------------------|
| 1 | Offline TA/TO | 24219 |
| 2 | Groups | 19811 |
| 3 | Direct | 12606 |
| 4 | Online TA | 56477 |
| 5 | Complementary | 743 |
| 6 | Corporate | 5295 |
| 7 | Aviation | 237 |
| 8 | Undefined | 2 |

🔄 REFRESH    ⌃

## reservations_count by market_segment
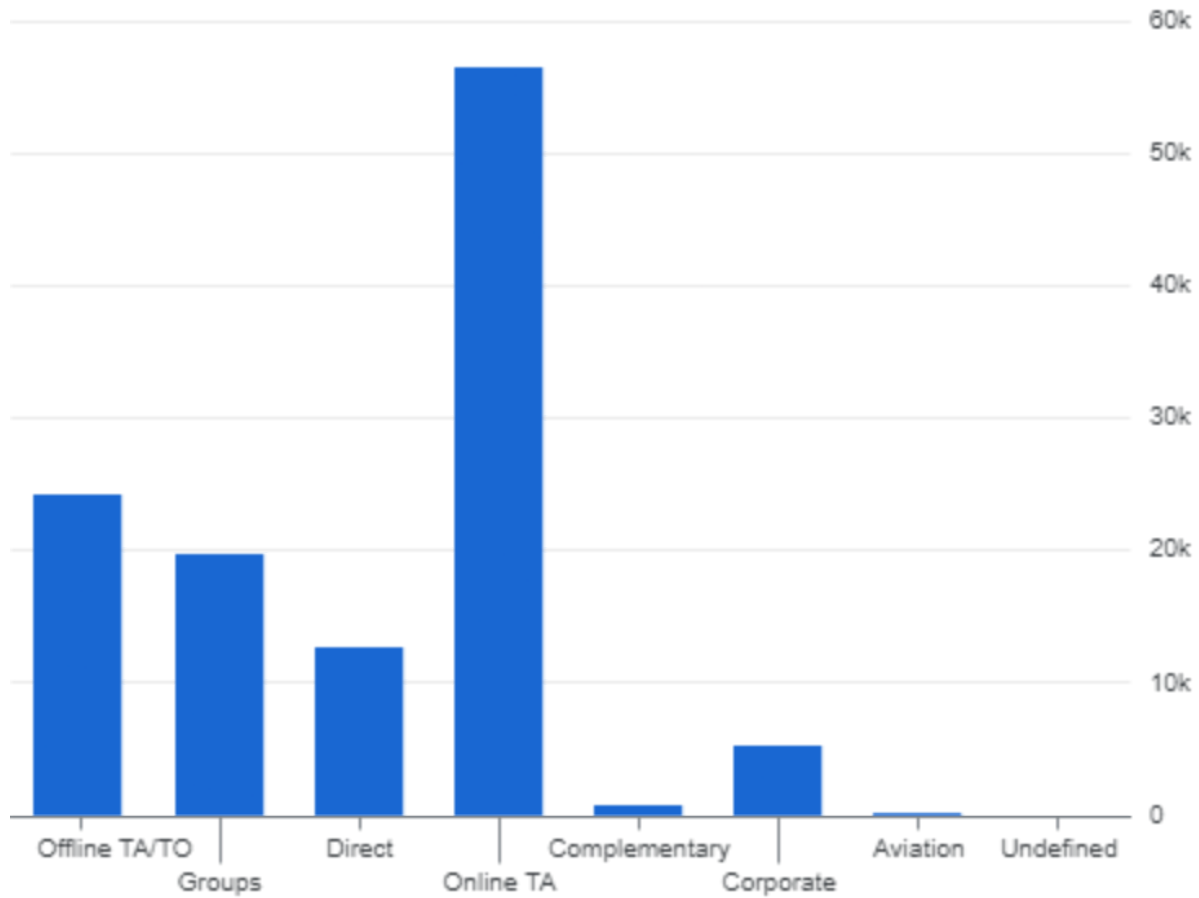


This visualization  provides valuable insights into the distribution of reservations across different market segments. It provides a comprehensive understanding of the hotel's customer base and market dynamics. It guides strategic decision-making, enhances marketing strategies, and supports business growth by enabling the hotel to effectively target and serve different market segments

```
1   SELECT country, COUNT(*) AS reservations_count
2   FROM `liquid-folio-409510.HotelReservation_Dataset.Hotel
    Reservation`
3   GROUP BY country
4   ORDER BY reservations_count DESC
5   LIMIT 10;
```

Processing location: US ⊗

Press Alt+F1 for Accessibility Options.

Query results

⬇ SAVE RESULTS ▾

JOB INFORMATION    RESULTS    CHART  PREVIEW

| Row | country ▾ | reservations_count |
|-----|-----------|--------------------|
| 6   | ITA       | 3766               |
| 7   | IRL       | 3375               |
| 8   | BEL       | 2342               |
| 9   | BRA       | 2224               |

The query calculates the count of reservations for each country and sorts the results based on the number of reservations in descending order. The LIMIT 10 clause is used to restrict the output to the top 10 countries with the highest number of reservations.

## reservations_count by country
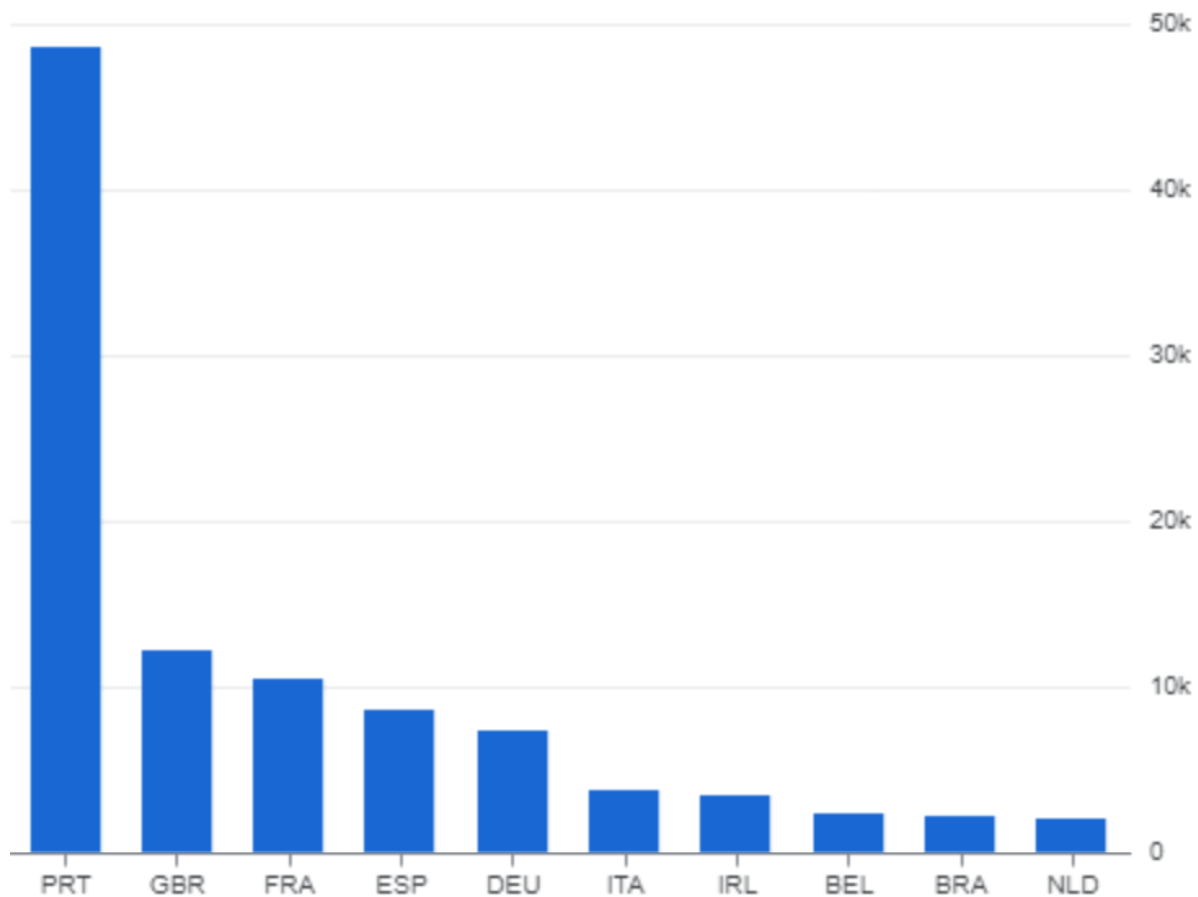


This visualization provides a clear and concise overview of the hotel's international customer base. It provides valuable insights into the distribution of reservations across different countries.

By grouping the data based on the country column and counting the number of reservations for each country, the query identifies the countries with the highest number of reservations. The results are then ordered in descending order, and only the top 10 countries are displayed.

This visualization helps the hotel in several ways:

1. **Understanding Customer Preferences**: The visualization shows which countries have the most reservations, helping the hotel identify their main customer markets. This information is valuable for targeting marketing efforts and understanding where to focus resources.

2. **Better Planning**: Knowing which countries contribute the most reservations allows the hotel to plan their strategies accordingly. They can customize services, promotions, and guest experiences to meet the needs and preferences of these important markets.
3. **Improving Revenue Management**: Analyzing reservations by country helps the hotel optimize their revenue management. They can study demand patterns from different countries and make adjustments to pricing, availability, and distribution strategies to maximize their revenue potential.
4. **Identifying Growth Opportunities**: By visualizing the top countries with the highest reservation counts, the hotel can spot opportunities for expanding into new markets. They can identify untapped markets or emerging trends, which may lead to developing targeted marketing campaigns or forming partnerships.

In summary, this visualization helps the hotel understand their primary markets, plan their business strategies more effectively, optimize revenue management, and identify potential opportunities for growth.

```
@  Untitled 3        ▶ RUN                          ⋮

1   SELECT COUNT(*) AS canceled_reservations
2   FROM `liquid-folio-409510.HotelReservation_Dataset.Hotel
    Reservation`
3   WHERE is_canceled = '1';
```

Processing location: US ⊗                    Press Alt+F1 for Accessibility Options.

Query results          ⬇ SAVE RESULTS ▾      📈 ▾      ⬍

‹    JOB INFORMATION      RESULTS      CHART   PREVIEW      ›

| Row | canceled_reservation |
|-----|----------------------|
| 1   | 44224                |

This query is used to calculate the count of canceled reservations from a specific table named "Hotel Reservation" in the dataset "HotelReservation_Dataset" within the project "liquid-folio-409510"

## canceled_reservations
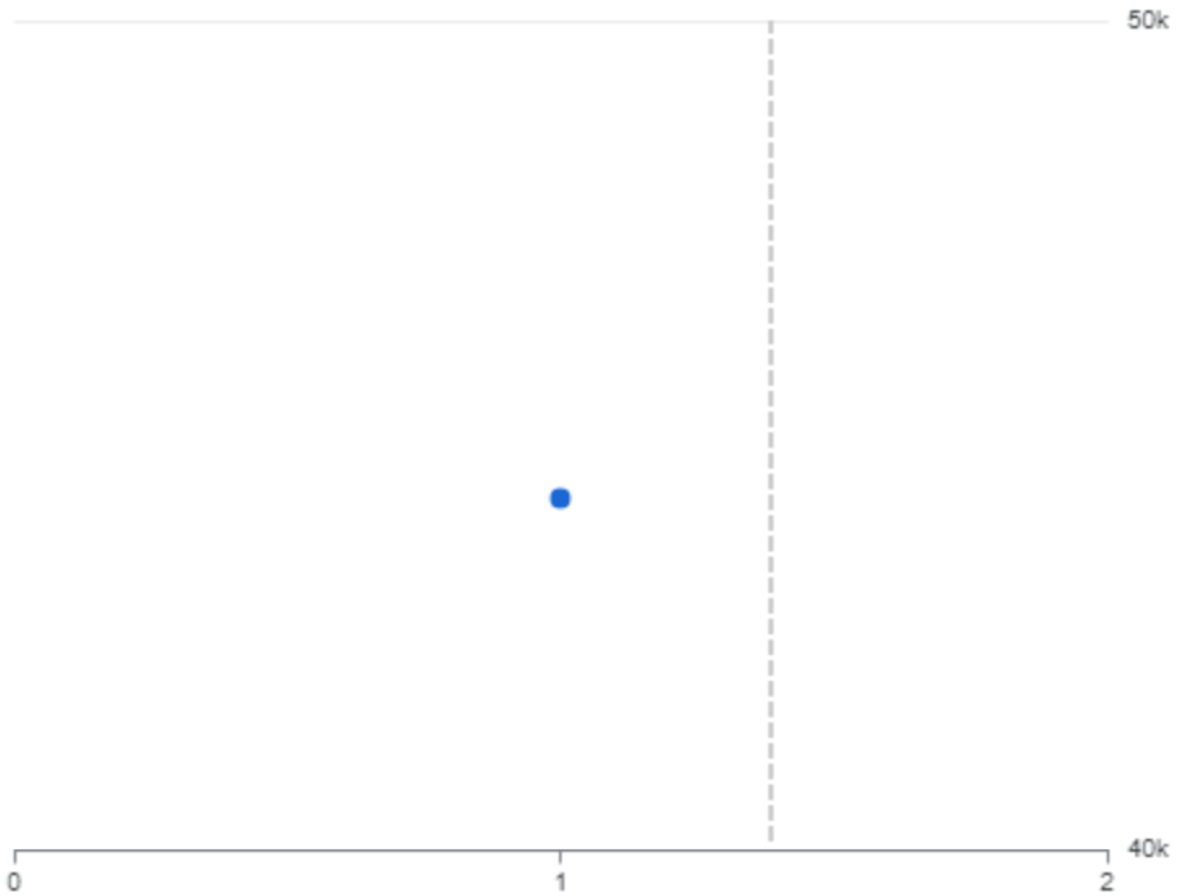


This visualization aids in decision-making processes by providing a comprehensive view of the count of canceled reservations. This visual representation can help hotel managers or analysts identify areas for improvement.

```
1  SELECT reservation_status, SUM(CAST
   (stays_in_weekend_nights AS INT64) + CAST
   (stays_in_week_nights AS INT64)) AS total_nights_stayed
2  FROM `liquid-folio-409510.HotelReservation_Dataset.Hotel
   Reservation`
3  GROUP BY reservation_status;
```

Processing location: US ⊗          Press Alt+F1 for Accessibility Options.

Query results          ⬇ SAVE RESULTS ▾

| Row | reservation_status ▾ | total_nights_stayed |
|-----|----------------------|---------------------|
| 1   | Check-Out            | 255040              |
| 2   | Canceled             | 149740              |
| 3   | No-Show              | 4477                |

The query calculates the total number of nights stayed by adding the stays_in_weekend_nights and stays_in_week_nights columns for each reservation status, grouped by reservation_status.

## total_nights_stayed by reservation_status
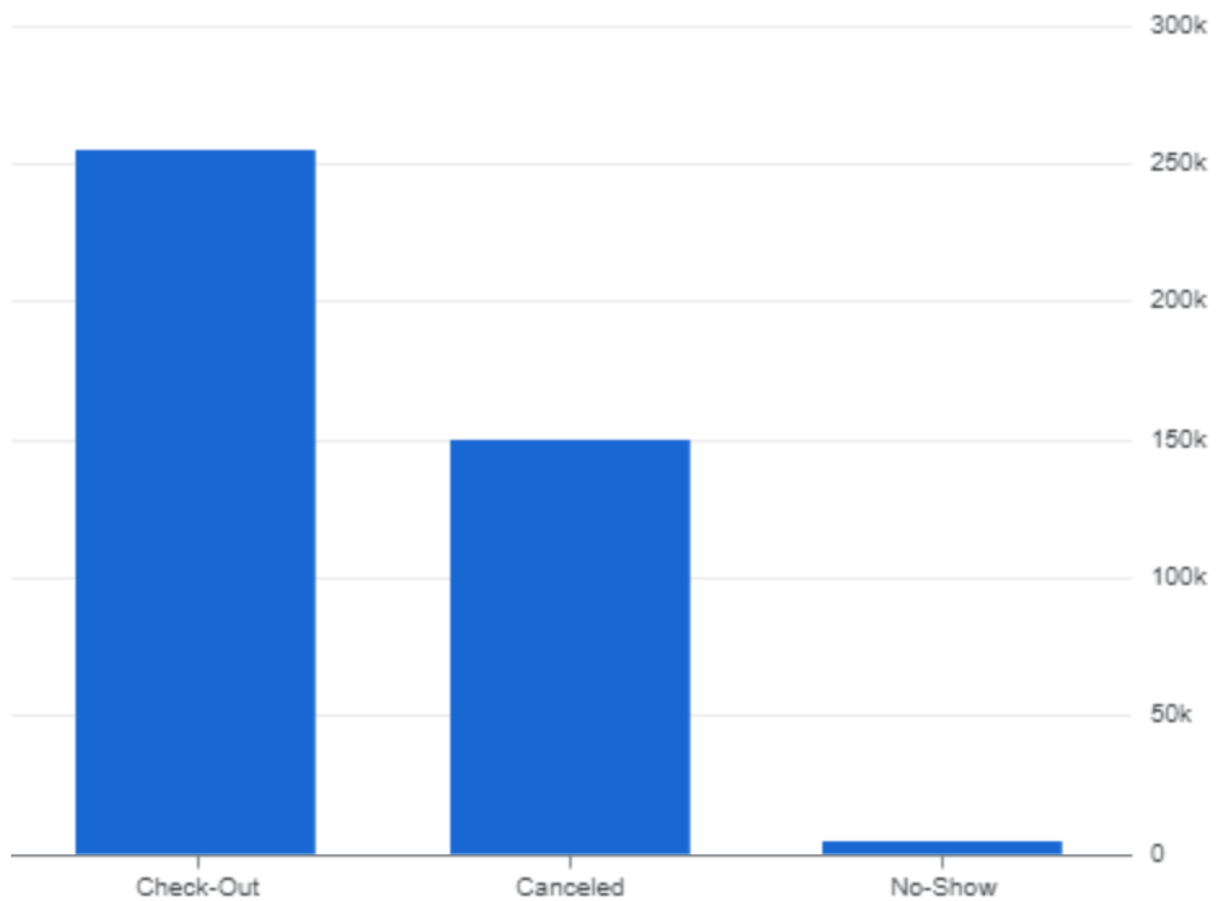


This visualization helps the hotel analyze and compare the total nights stayed across different reservation statuses.
Visualizing this data helps the hotel understand the distribution of total nights stayed for each reservation status.

⊕ Untitled 5    ▶ RUN                                      ⋮

```
1   SELECT arrival_date_year, COUNT(*) AS reservations_count
2   FROM `liquid-folio-409510.HotelReservation_Dataset`.
    `Hotel Reservation`
3   GROUP BY arrival_date_year
4   ORDER BY arrival_date_year;
```

Processing location: US ⊗               Press Alt+F1 for Accessibility Options.

Query results          ⬇ SAVE RESULTS ▾      ⋔ ▾    ⇕

<   JOB INFORMATION      RESULTS      CHART  PREVIEW    >

| Row | arrival_date_year ▾ | reservations_count |
|-----|---------------------|--------------------|
| 1   | 2015                | 21996              |
| 2   | 2016                | 56707              |
| 3   | 2017                | 40687              |

This query is used to retrieve the count of reservations grouped by the year of arrival from a table named "Hotel Reservation" in the dataset "HotelReservation_Dataset" within the project "liquid-folio-409510.

## reservations_count by arrival_date_year
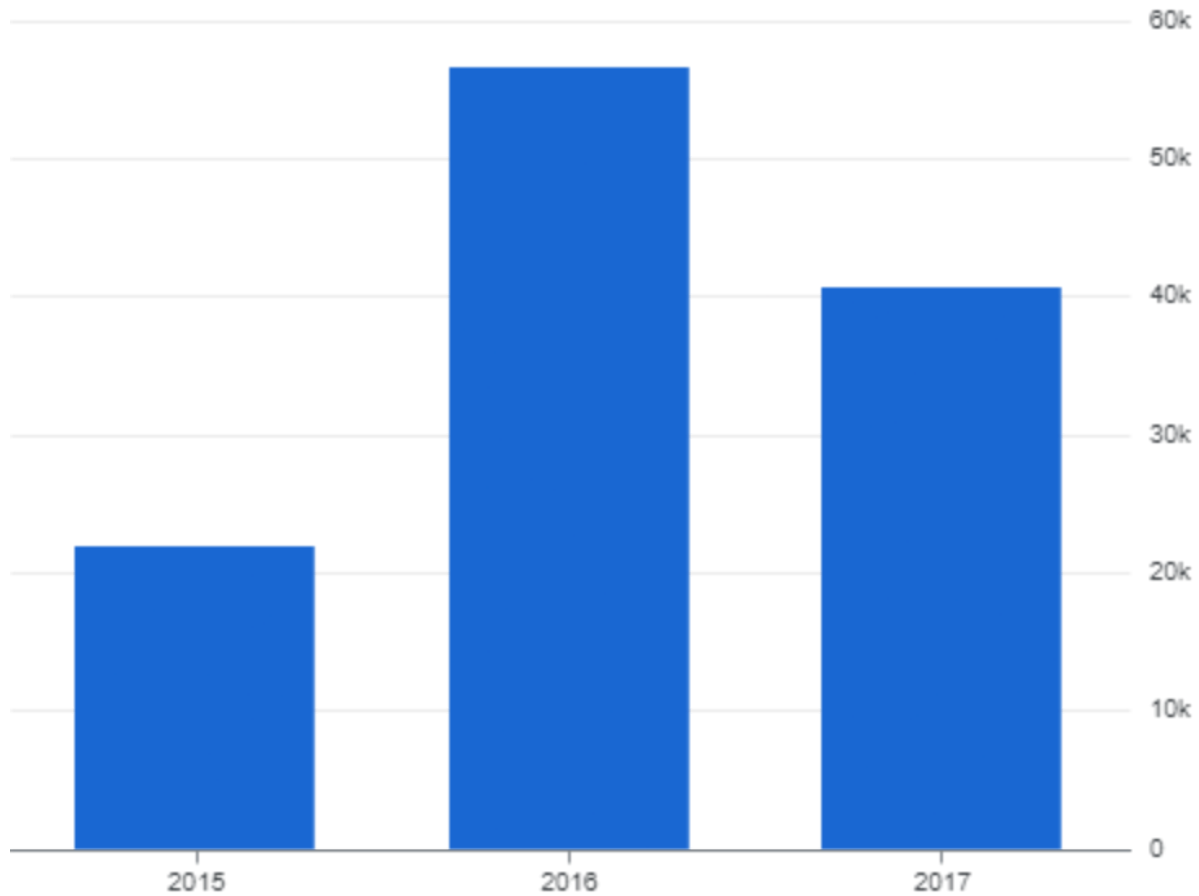


This visualization helps the hotel identify booking trends and patterns. It allows them to understand which years experienced high or low levels of demand, enabling them to make informed decisions about pricing, staffing, and resource allocation. It also

```sql
SELECT arrival_date_month, AVG(CAST(adr AS FLOAT64)) AS average_daily_rate
FROM `liquid-folio-409510.HotelReservation_Dataset`.`Hotel Reservation`
GROUP BY arrival_date_month
ORDER BY CASE arrival_date_month
            WHEN 'January' THEN '01'
            WHEN 'February' THEN '02'
            WHEN 'March' THEN '03'
            WHEN 'April' THEN '04'
            WHEN 'May' THEN '05'
```

Processing location: US  ⊗                        Press Alt+F1 for Accessibility Options.

Query results                    ⬇ SAVE RESULTS ▾      📈 ▾      ↕

JOB INFORMATION        RESULTS        CHART  PREVIEW

| Row | arrival_date_month ▾ | average_daily_rate |
|-----|----------------------|--------------------|
| 1   | January              | 70.36124135604…    |
| 2   | February             | 73.58227565691…    |
| 3   | March                | 80.67964570144…    |
| 4   | April                | 100.3807899720…    |

This query calculates the average daily rate for each month of the year in the hotel reservation dataset.
In this query, the CASE statement is used to map the month names to their respective month numbers in the "YYYY-MM" format. The query then orders the results based on the transformed values.

The query's ordering of the months ensures that the visualization displays the data in chronological order, making it easier to see any seasonal patterns that may exist.

## Query results

⬇ SAVE RESULTS ▾        📈 ▾        ⌄⌄⌃

<    JOB INFORMATION      **RESULTS**      CHART   PREVIEW   Expand

| Row | arrival_date_month ▾ | average_daily_rate |
|-----|----------------------|--------------------|
| 1 | January | 70.36124135604… |
| 2 | February | 73.58227565691… |
| 3 | March | 80.67964570144… |
| 4 | April | 100.3807899720… |
| 5 | May | 108.6955160715… |
| 6 | June | 116.6721921565… |
| 7 | July | 126.7880127951… |
| 8 | August | 140.1115226633… |
| 9 | September | 105.0496574038… |
| 10 | October | 87.90887903225… |
| 11 | November | 73.79496173093… |
| 12 | December | 81.07677581120… |

The query's ordering of the months ensures that the visualization displays the data in chronological order, making it easier to see any seasonal patterns that may exist.

## average_daily_rate by arrival_date_month
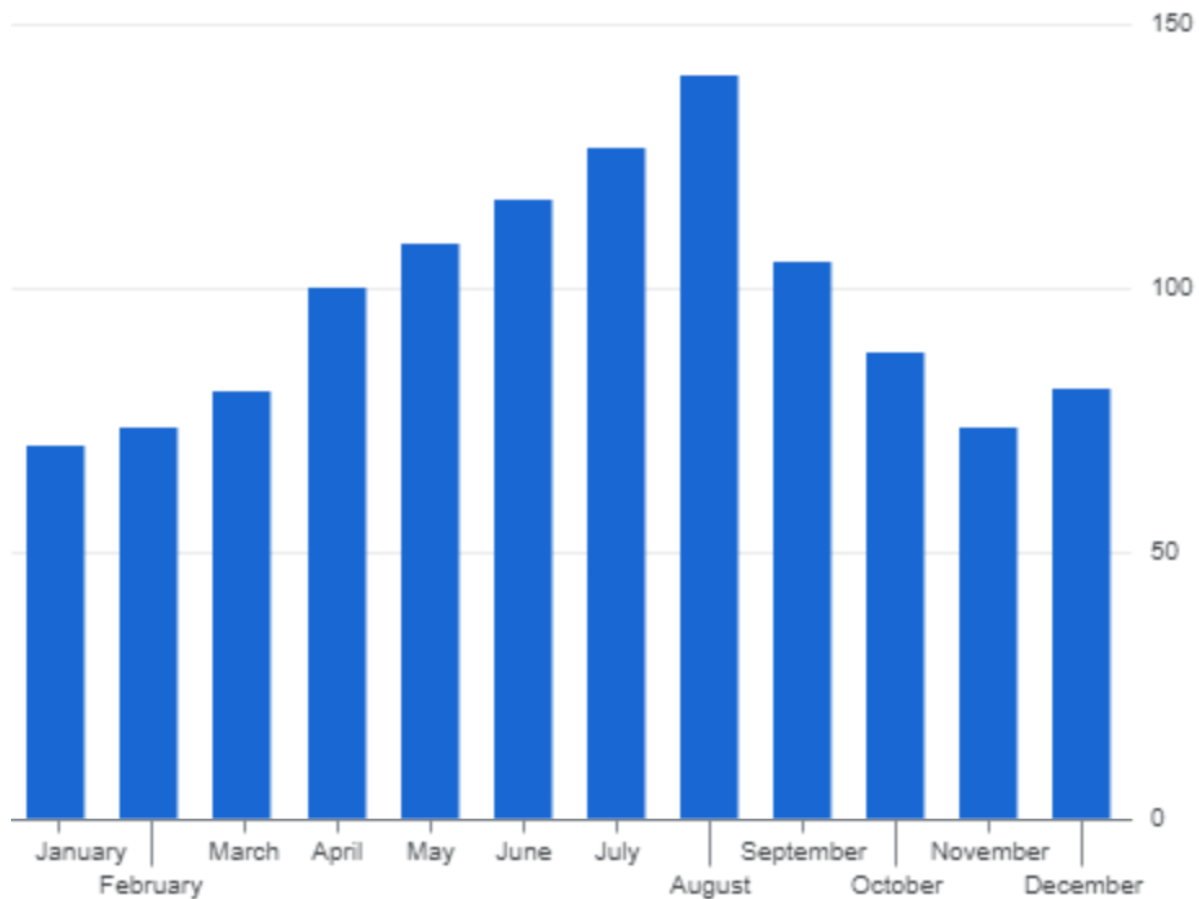


This data helps the hotel understand pricing trends and differences throughout the year.
With the visualization, the hotel can easily compare the average rates across different months and see if there are any noticeable patterns or unusual changes.

With this visualization, the hotel can make informed decisions about pricing strategies, managing revenue, and forecasting demand.

They can adjust prices based on the trends they observe and plan marketing campaigns or special offers to attract guests during slower periods or maximize revenue during busy seasons.

Overall, this visualization helps the hotel optimize their revenue management, identify opportunities to adjust prices, and plan marketing strategies to attract guests at different times of the year.

```
Q  Untitled 10          ▶ RUN                         ⋮

1   SELECT meal, COUNT(*) AS meal_count
2   FROM `liquid-folio-409510.HotelReservation_Dataset`.
    `Hotel Reservation`
3   GROUP BY meal;
```

Processing location: US ⊗          Press Alt+F1 for Accessibility Options.

Query results          ⬇ SAVE RESULTS ▾        📈 ▾    ↕

‹    JOB INFORMATION        RESULTS        CHART  PREVIEW    ›

| Row | meal ▾ | meal_count ▾ | |
|-----|--------|--------------|---|
| 1 | BB | 92310 | |
| 2 | HB | 14463 | |
| 3 | SC | 10650 | |
| 4 | FB | 798 | |
| 5 | Undefined | 1169 | |

↻ REFRESH    ⌃

This query counts how many times each meal option appears in the dataset and shows the count of reservations for each meal type. It helps us see which meal options are popular and how many reservations are associated with each meal type.

## meal_count by meal



The visualization displays the meal options on the x-axis and the corresponding count of reservations on the y-axis.

This allows the hotel management to easily compare the popularity of different meal options and understand which ones are in high demand or potentially underutilized.

By visualizing the distribution of meal options, the hotel can make informed decisions related to menu planning, staffing, and resource allocation. It can help them optimize their food and beverage operations, tailor promotional offers, and enhance the overall dining experience for guests.

```
Q  Untitled 11        ▶ RUN          ⋮  ✓ Query completed.
  1  SELECT COUNT(*) AS total_repeated_guests
  2  FROM `liquid-folio-409510.HotelReservation_Dataset`.
     `Hotel Reservation`
  3  WHERE CAST(is_repeated_guest AS INT64) = 1;
```

Press Alt+F1 for Accessibility Options.

Query results          ⬇ SAVE RESULTS ▾      📈 ▾      ⬍

<  |  JOB INFORMATION  |  RESULTS  |  CHART  PREVIEW  |  >

| Row | total_repeated_guest |
|-----|---------------------|
| 1   | 3810                |

In this query, the CAST function is used to convert the is_repeated_guest column from a string to an integer (INT64). Then, the condition checks if the converted value equals 1, indicating a repeated guest. The query counts the number of rows that satisfy this condition and returns the result as total_repeated_guests.

## total_repeated_guests
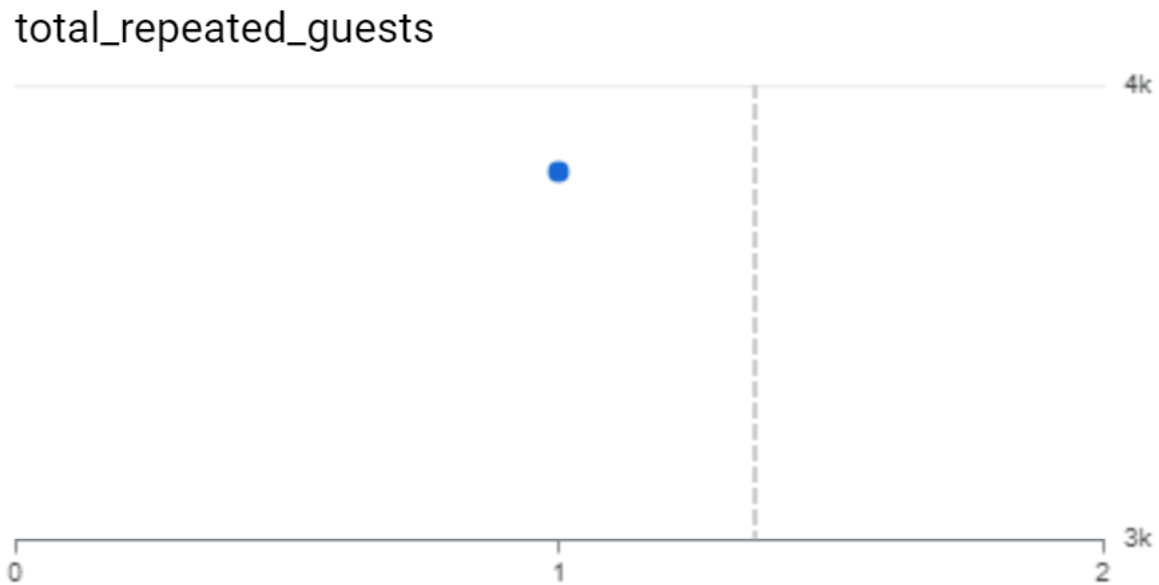


Visualizing the total number of repeated guests helps the hotel understand how many guests return to stay again. This information is valuable for the hotel's growth and customer retention strategies. It shows how many customers are loyal and satisfied with their previous experiences. By visualizing this data, the hotel can track how well their efforts to keep customers coming back are working.