

Hotel Booking Analysis – SQL Project

Data Analysis Portfolio

By

Muhammad Sohaib Ul Hassan

Date

Saturday, 09 August 2025

Table Of Contents

1. Introduction	3
2. Dataset Overview	3
3. Tools Used	4
4. Database Creation	4
5. Data Cleaning	5
6. Visualization	5
7. Key Insights	10
8. Conclusion	11

1. Introduction

This project utilizes SQL to clean, transform, and analyze hotel booking data. The primary objective is to identify key booking patterns, understand cancellation drivers, and provide actionable insights that can help hotels reduce cancellations and optimize their revenue management. By leveraging data analytics, this study aims to support decision-making processes that improve operational efficiency and customer satisfaction within the hospitality industry.

2. Dataset Overview

The dataset used in this project was downloaded from Kaggle and contains detailed hotel booking records spanning multiple years. It includes **42,000+ rows and 32 columns**, capturing information such as booking dates, customer demographics, stay duration, and cancellation status.

Key columns in the dataset include:

Column Name	Description
hotel	Type of hotel (e.g., Resort Hotel, City Hotel)
is_canceled	Booking cancellation status (0 = Not canceled, 1 = Canceled)
lead_time	Number of days between booking date and arrival date
arrival_date_year	Year of arrival
arrival_date_month	Month of arrival
arrival_date_day_of_month	Day of the month of arrival
stays_in_weekend_nights	Number of weekend nights booked
stays_in_week_nights	Number of weeknights booked
adults	Number of adults in the booking
children	Number of children in the booking
meal	Type of meal booked (e.g., BB, HB, FB)
market_segment	Market segment designation (e.g., Online TA, Offline TA)

Column Name	Description
distribution_channel	Booking distribution channel
previous_cancellations	Number of previous cancellations by the customer
booking_changes	Number of changes made to the booking
deposit_type	Type of deposit (No Deposit, Refundable, Non Refund)
customer_type	Type of customer (Transient, Contract, Group)
adr	Average Daily Rate (price per night)

This rich dataset allows for an in-depth analysis of booking trends, cancellation patterns, and customer behavior, providing valuable insights for hotel management to optimize revenue and improve guest satisfaction.

3. Tools Used

This project leveraged a suite of tools to manage, analyze, and present hotel booking data effectively. The core of the analysis was conducted using **SQL**, which enabled robust data cleaning, database creation, and deep analytical querying to uncover actionable insights. SQL was used to handle missing values, normalize data formats, and perform complex queries to identify booking trends, cancellation rates, and customer segmentation.

Excel played a key role in visualizing the results, allowing for the creation of clear and informative charts that supported data storytelling. Finally, **Microsoft Word** was used to compile the findings into a structured and professional document, ensuring clarity and coherence in the presentation of results.

Tools Summary:

- **SQL (MySQL Workbench)** – Data cleaning, database creation, and analytical queries
- **Excel** – Data visualization and chart creation
- **Microsoft Word** – Documentation and report writing

4. Database Creation

To facilitate structured analysis and efficient querying, a dedicated database named `hotel_booking` was created using MySQL Workbench. The cleaned dataset was imported into this database, allowing for seamless execution of SQL queries and better data management. This setup provided a scalable and organized environment for conducting deep analysis and extracting insights.

5. Data Cleaning

Before conducting any analysis, it was essential to ensure the dataset was clean, consistent, and reliable. Data cleaning was performed using SQL to address missing values, correct data types, and remove any inconsistencies that could affect the accuracy of insights. These preprocessing steps laid the foundation for meaningful analysis and ensured the integrity of the database.

6. Visualization

To enhance the interpretability of the SQL analysis, query outputs were exported to CSV format and visualized using Excel. Each chart was designed to highlight key patterns and trends in the hotel booking data, making the insights more accessible and actionable. The visualizations use consistent formatting, clear labeling, and high-resolution graphics to ensure clarity and professionalism.

Below are the charts included in this report:

Figure 1. Total Bookings by Hotel

This chart compares the total number of bookings between city and resort hotels, revealing which type attracts more guests.



Figure 2. Cancelation rate by Month

This chart illustrates the monthly trend in booking cancellations throughout the year. By visualizing cancellation rates per month, it highlights seasonal patterns and potential operational challenges. Peaks in cancellation rates may correspond to specific holidays, weather conditions, or promotional periods, offering valuable insights for improving booking policies and customer retention strategies.

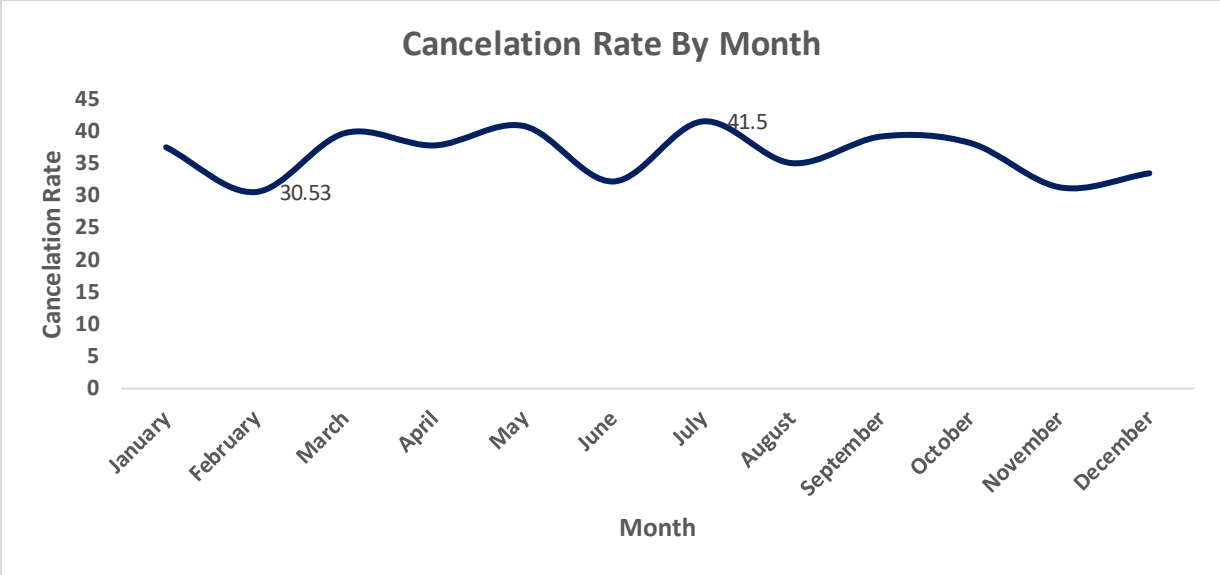


Figure 3 Average Length of Stay Chart

This chart compares the average duration of guest stays between two hotel types: Resort Hotel and City Hotel. The data reveals that guests tend to stay significantly longer at Resort Hotels (4.32 nights on average) compared to City Hotels (2.97 nights). This insight suggests that Resort Hotels attract more leisure-oriented travelers seeking extended vacations, while City Hotels cater to shorter stays, likely driven by business or transit purposes. Understanding these patterns can help tailor services, pricing strategies, and promotional packages to better suit each hotel's guest profile

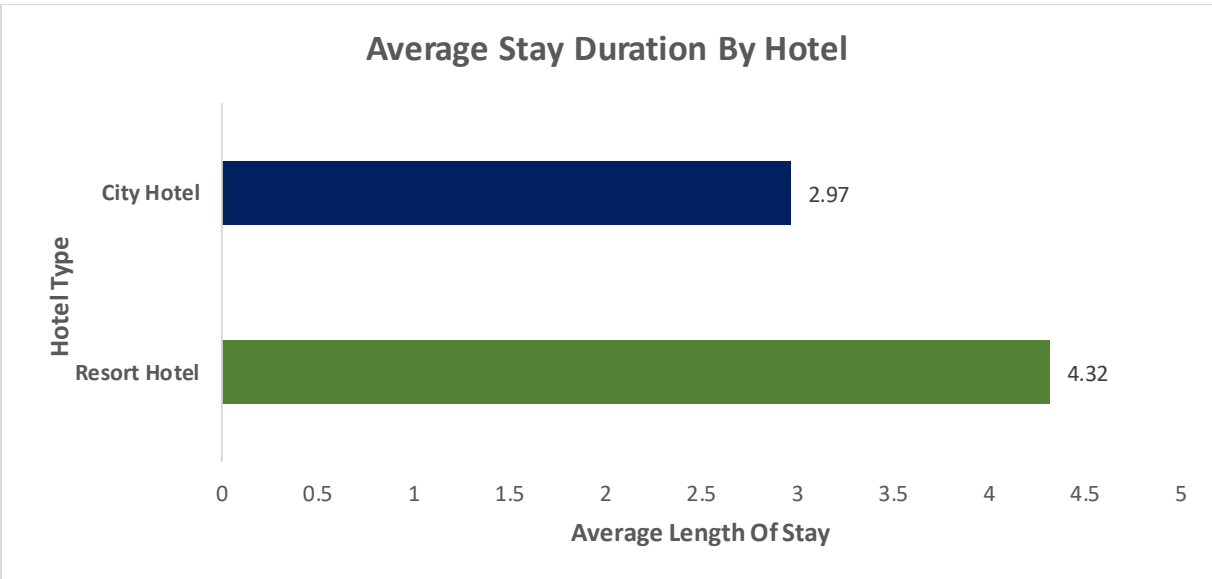


Figure 4 Top Countries by Bookings

This chart highlights the top ten countries by total hotel bookings, offering a clear view of international demand. Portugal (PRT) leads by a wide margin with 48,436 bookings, followed by the United Kingdom (GBR) and France (FRA). The data underscores Portugal’s strong domestic market presence and reveals key international contributors such as Spain, Germany, and Italy. These insights are valuable for targeting marketing efforts, customizing guest experiences, and identifying growth opportunities in high-performing regions.

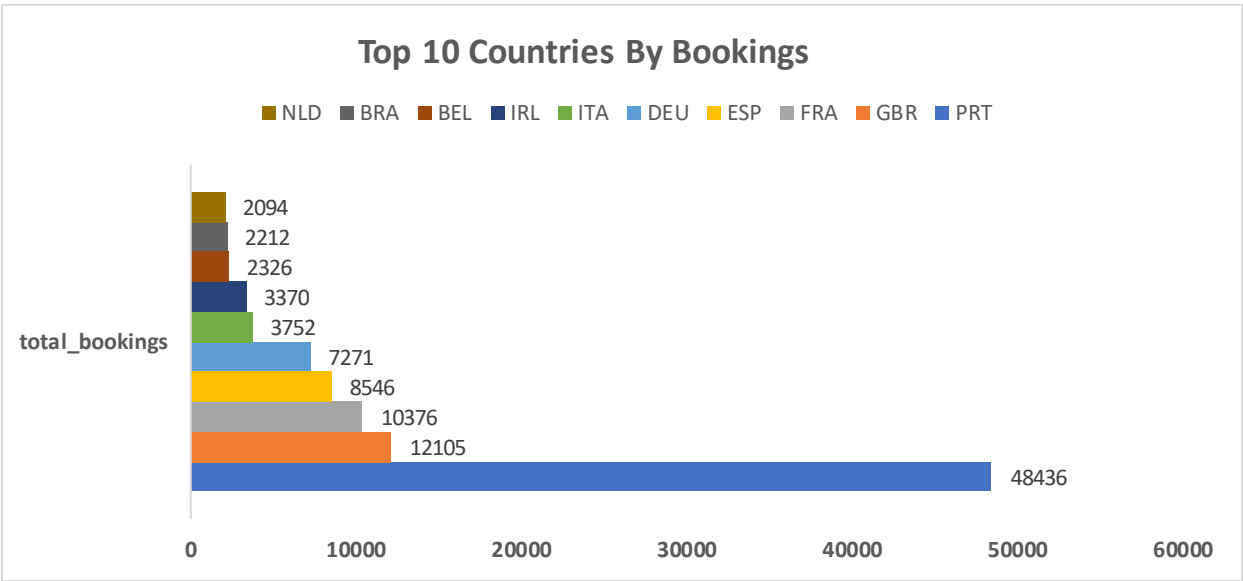


Figure 5 Impact of Lead Time on Cancellations

This chart shows how booking lead time affects cancellations. Short lead times (0–7 days) have the lowest cancellation rate, while long lead times (90+ days) show the highest, with over 25,000 cancellations from nearly 51,000 bookings. Mid-range categories vary, but notably, the 31–90 day group has a high cancellation count despite fewer bookings. The trend suggests that longer planning windows increase the risk of cancellations.

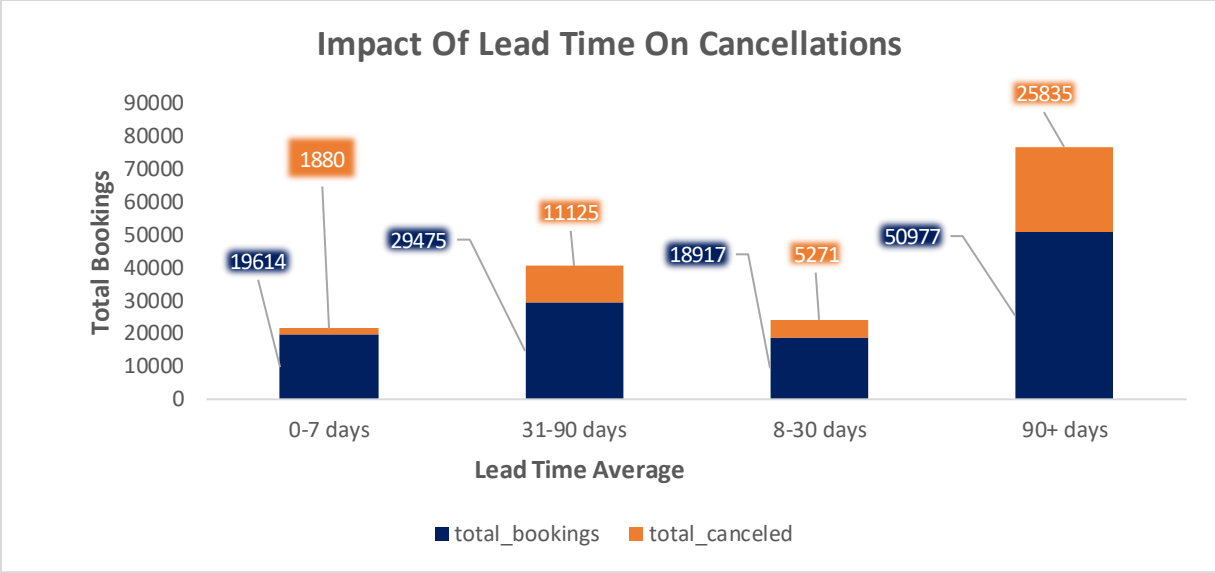


Figure 6 Top Countries by Booking Rank (Horizontal View)

This horizontal bar chart ranks countries by hotel bookings, with the United Kingdom (GBR) at the top, followed by Spain (ESP), Italy (ITA), and Belgium (BEL). The ranking progresses down to Norway (NOR), which holds the 20th position. Displaying ranks on the x-axis and countries on the y-axis offers a clear visual hierarchy, making it easy to compare performance across regions.

The chart highlights strong demand from Western European countries, while also including others like China (CHN) and Israel (ISR), indicating a mix of regional and international interest. This format is especially useful for identifying top markets and evaluating where strategic efforts should be focused to boost bookings

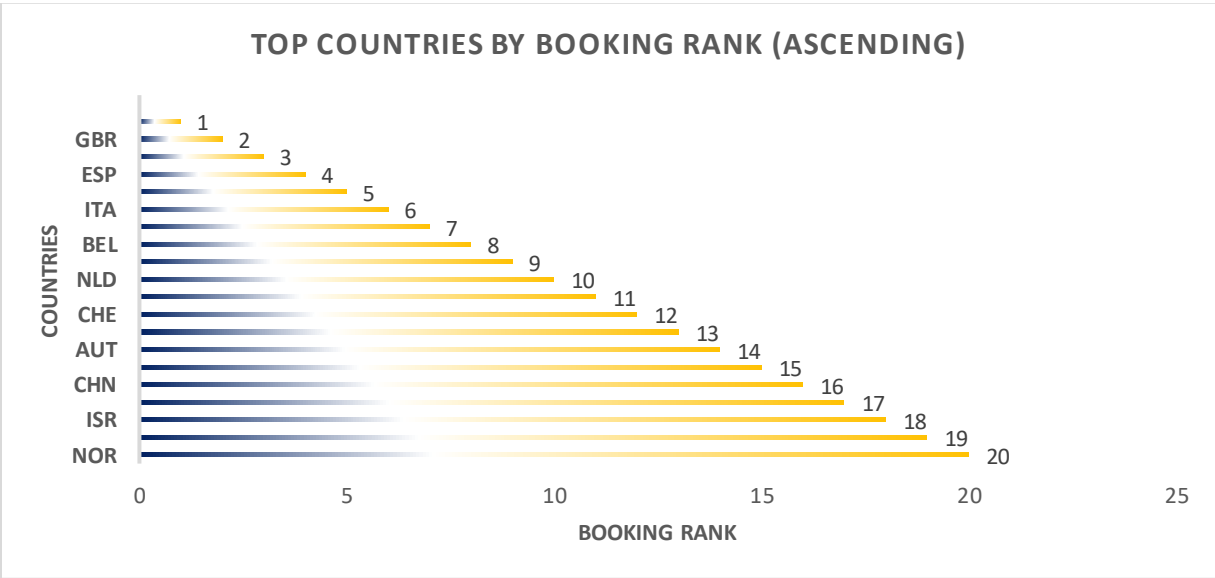


Figure 7 Month-over-Month Booking Growth (Aug 2015 – Aug 2017)

This line graph tracks the percentage change in bookings over a two-year span, from August 2015 to August 2017. The y-axis ranges from -60% to +80%, capturing both sharp declines and strong surges in booking activity. The x-axis marks each month within the period, providing a detailed timeline of fluctuations.

Key observations:

- **Peak Growth:** The most significant spike occurs around **January 2016**, suggesting a seasonal surge—possibly due to post-holiday travel or winter promotions.
- **Volatility:** The graph shows multiple ups and downs, indicating a dynamic market with frequent changes in consumer behavior or external influences.
- **Trend Line:** The blue line connecting data points helps visualize the overall momentum, making it easier to spot patterns or anomalies.

This visualization is ideal for identifying seasonal trends, evaluating marketing impact, or forecasting future booking behavior.

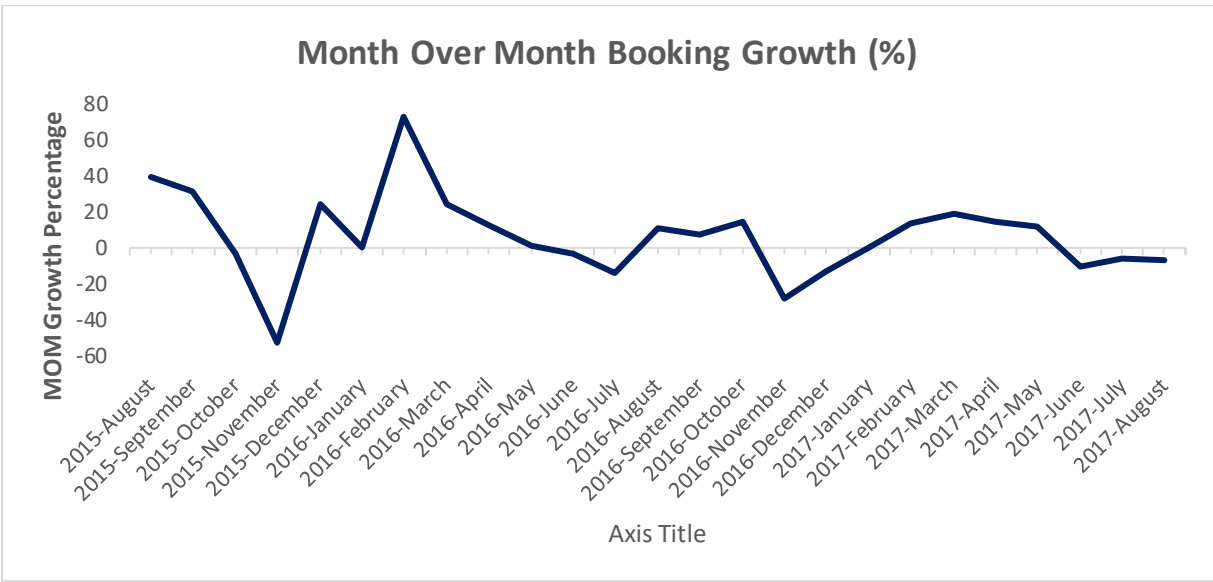


Figure 8 Cumulative Hotel Bookings (July 2015 – August 2017)

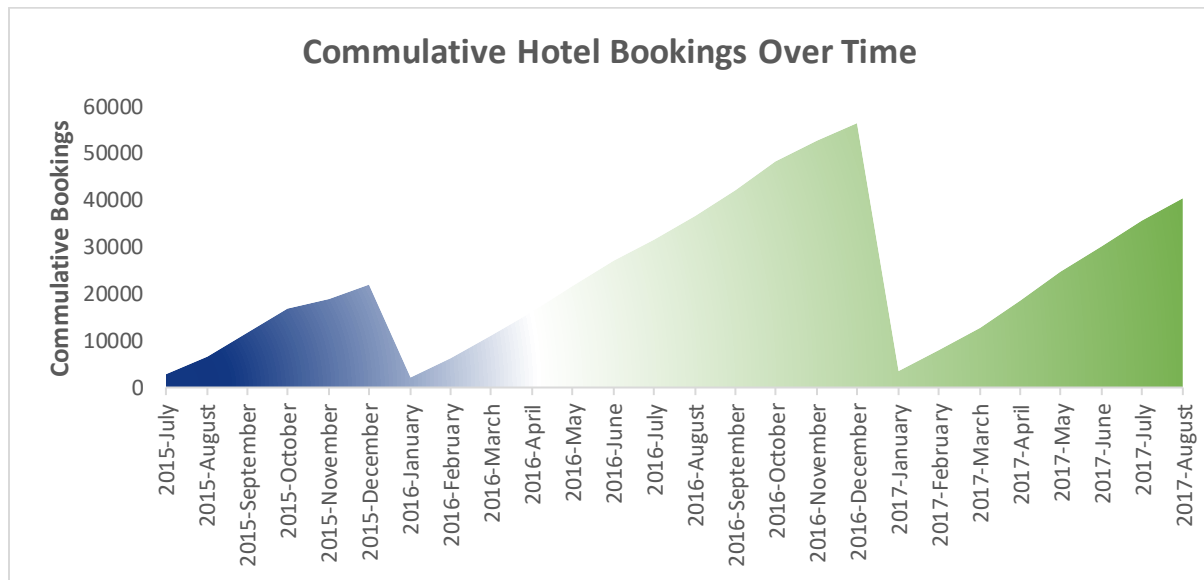
This graph illustrates the steady accumulation of hotel bookings over a two-year period, from **July 2015 to August 2017**. The x-axis tracks time month-by-month, while the y-axis measures the total number of bookings, climbing from **0 to 60,000**.

Key features:

- **Upward Trajectory:** The curve shows a consistent rise, reflecting sustained growth in hotel bookings over time.

- **Seasonal Surges:** Certain months exhibit sharper increases, hinting at peak travel seasons or successful promotional campaigns.
- **Gradient Color Scheme:** The use of darker tones for earlier dates and lighter tones for recent ones adds visual depth, helping viewers intuitively follow the timeline.

This visualization is ideal for understanding long-term booking momentum and identifying periods of accelerated growth. It can also serve as a foundation for forecasting future demand or evaluating the impact of strategic initiatives.



7. Key Insights

Key Insights from Hotel Booking Data Analysis (2015–2017)

This analysis combines three visualizations—Top Countries by Booking Rank, Month-over-Month Booking Growth, and Cumulative Hotel Bookings Over Time to uncover patterns and strategic insights in hotel booking behavior across regions and time.

1. Geographic Demand Concentration

Chart: Top Countries by Booking Rank (Ascending)

- **Western Europe dominates** hotel booking activity, with the United Kingdom (GBR), Spain (ESP), and Italy (ITA) leading the rankings.
- **Emerging markets** like China (CHN) and Israel (ISR) also show notable booking activity, suggesting potential for international expansion.

- This geographic distribution highlights where marketing efforts and partnerships may yield the highest returns.

2. Booking Volatility & Seasonality

Chart: Month-over-Month Booking Growth (%)

- Booking growth is highly volatile, with sharp peaks and troughs throughout the two-year period.
- The largest spike occurs around January 2016, likely driven by seasonal travel or promotional campaigns.
- These fluctuations underscore the importance of seasonal planning, dynamic pricing, and targeted marketing during high-growth periods.

3. Sustained Long-Term Growth

Chart: Cumulative Hotel Bookings Over Time

- The cumulative bookings show a consistent upward trend, reaching nearly 60,000 bookings by August 2017.
- Periods of accelerated growth align with peaks in the MoM growth chart, reinforcing the impact of seasonal or strategic factors.
- This long-term trajectory confirms a healthy demand curve, validating continued investment in booking infrastructure and customer acquisition.

Strategic Takeaways

- **Focus on top-performing regions** (e.g., GBR, ESP, ITA) for immediate ROI, while nurturing growth in emerging markets.
- **Leverage seasonal trends** to optimize campaigns—especially around January and other peak months.
- **Maintain momentum** by analyzing what drives cumulative growth and replicating successful strategies.

8. Conclusion

This analysis provides a comprehensive view of hotel booking dynamics between July 2015 and August 2017, revealing both short-term fluctuations and long-term growth patterns. By examining geographic booking ranks, month-over-month growth rates, and cumulative booking trends, we've identified key drivers of demand and strategic opportunities for expansion.

The data underscores the importance of focusing on high-performing regions like Western Europe, while also recognizing the potential in emerging markets. Seasonal volatility in booking growth highlights the need for agile marketing and pricing strategies, especially during peak travel periods. Meanwhile, the steady rise in cumulative bookings confirms a robust and growing market, validating continued investment in customer acquisition and operational scalability.

Overall, these insights equip stakeholders with actionable intelligence to optimize resource allocation, enhance campaign timing, and strengthen market positioning in an increasingly competitive hospitality landscape.