SCHOOL OF ARTIFICIAL INTELLIGENCE

BIG DATA AND DATABASE MANAGENMENT SYSTEM 21AIE304

HOTEL RESERVATION MANAGEMENT SYSTEM

B.TECH
CSE-AI(Semester-5)

Under the supervision of:

Dr.Sanjanasri.J.P

SUBMITTED BY:
BATCH A
GROUP-8
SYKAM SUMANJALI(CB.EN.U4AIE21068)
SUDA HARI PRIYA(CB.EN.U4AIE21067)
R HEMA RADHIKA(CB.EN.U4AIE21050)
SOUVIK GORAIN(CB.EN.U4AIE21065)

Acknowledgement:

We would like to express our sincere gratitude to our esteemed faculty Dr.Sanjanasri.J.P Mam for their invaluable guidance and support throughout our Big Database Management System project on the Hotel Reservation System. Your expertise and mentorship were instrumental in shaping our project's success. Your dedication to our learning and development has been a source of inspiration. Thank you for your unwavering commitment to our academic growth. We deeply appreciate your contributions to our team's journey.

TABLE OF CONTENTS:

Abstract:

The hotel reservation system analysis project was conceived in the context of the hospitality industry, where efficient management of reservations is critical to both customer satisfaction and business success. With the increasing volume of data generated by hotel bookings, there is a pressing need to harness this information to optimize occupancy rates, improve customer experience, and enhance operational efficiency. This project, therefore, aimed to analyze hotel reservation data to unearth patterns and insights that could inform strategic decision-making. By employing advanced data processing and visualization techniques using MySQL, Apache Spark, and Power BI, the project sought to transform raw booking data into actionable intelligence. Such an analysis is not only pivotal in understanding customer preferences and booking trends but also vital in forecasting future demand, shaping marketing strategies, and streamlining hotel operations, thereby playing a crucial role in the competitive landscape of the hospitality sector.

INTRODUCTION:

Objectives:

1. Data-Driven Insight Generation: To extract and analyze key patterns and trends from the hotel reservation dataset, providing a deep understanding of customer booking behavior.
2. Operational Optimization: To identify areas for operational improvements in hotel booking management, such as peak booking times, most preferred room types, and cancellation trends.
3. Enhanced Customer Experience: To understand customer preferences and requirements better, enabling personalized guest experiences and targeted marketing strategies.
4. Forecasting and Strategy Development: To use the insights gained from the analysis for forecasting future booking trends and aiding in strategic decision-making for hotel management.
5. Technical Proficiency: To demonstrate the effective use of technology in handling big data, specifically the integration of MySQL, Apache Spark for data processing, and Power BI for data visualization.

Scope:

Project Covers:

- Analysis of a specific hotel reservation dataset from Kaggle.
- Use of MySQL for data storage and normalization.
- Data processing and analytical computations using Apache Spark.
- Visualization and dashboard creation in Power BI.

MOTIVATION:

In the realm of hotel reservations, our project isn't just a collection of data and analyses; it's a gateway to a more intuitive, efficient, and guest-centric future. By seamlessly integrating technologies like Apache Spark, MySQL, and Power BI, we're not merely decoding numbers; we're decoding the keys to exceptional guest experiences, operational excellence, and a competitive edge in the hospitality industry. As we delve into the intricacies of reservations, customer preferences, and market trends, we're not just creating a system; we're shaping the future of hospitality, one reservation at a time. Welcome to a project that goes beyond expectations, paving the way for a new era in data-driven hotel management.

METHODOLOGY:

Dataset Overview:

We got the dataset “Hotel_Reservations.csv” from Kaggle.
Source: Hotel_Reservations.csv

Data Storage and Connectivity:

The normalized tables were stored in MySQL for efficient data retrieval and management. JDBC (Java Database Connectivity) was used to establish a connection between MySQL and Apache Spark, allowing seamless data transfer for analysis.

Data Analysis with Apache Spark:

Apache Spark was employed for in-depth analysis of the hotel reservation data. Various queries and transformations were applied to gain insights into customer behaviour, booking patterns, and overall trends. Spark SQL was utilized for querying the data stored in MySQL.

Data Visualization with PowerBI:

PowerBI was chosen for its powerful visualization capabilities. The connectivity between PowerBI and MySQL was established using MySQL Connector/NET. Visualizations included: Trend Analysis, Bar graphs.

ANALYSIS OF HOTEL RESERVATIONS :

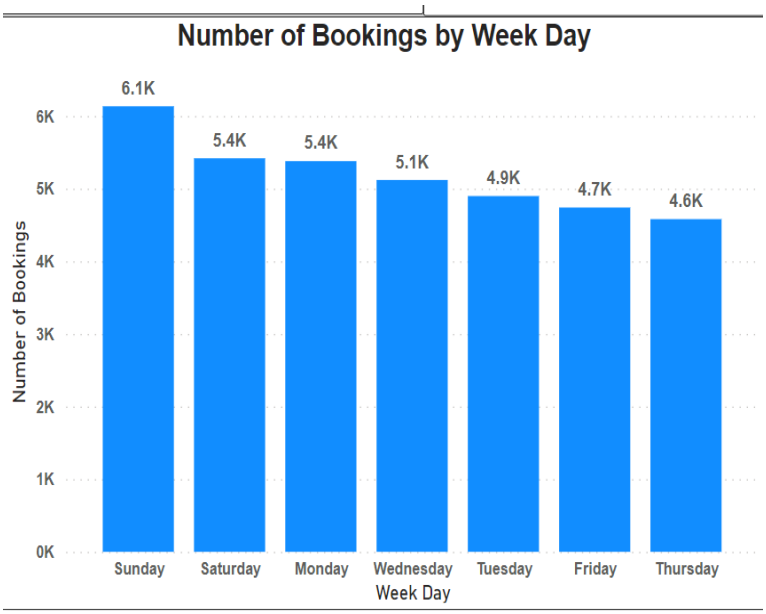
Analysis-1: DAY OF THE WEEK WITH MOST BOOKINGS

Methodology:

In this analysis, we aim to determine the day of the week that experiences the highest number of hotel bookings. To achieve this, we utilized the Booking_Details table in Apache Spark, specifically focusing on the arrival_year, arrival_month, and arrival_date columns to derive a full arrival date. Subsequently, we extracted the day of the week using the date_format function.

RESULTS:

```
scala> mostBookingsDayOfWeek.show()
+------------------+-----+
|arrival_day_of_week|count|
+------------------+-----+
|            Sunday| 6135|
+------------------+-----+
```



INFERENCE:

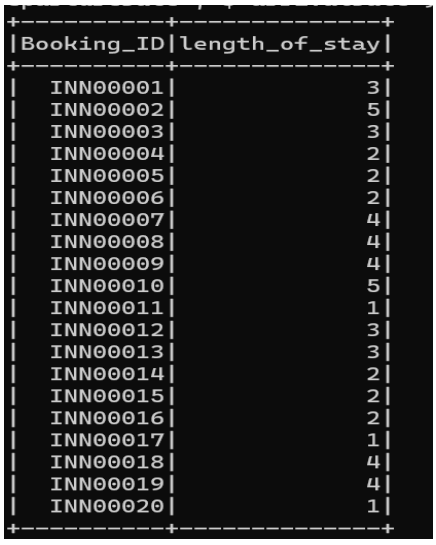
The output of the analysis provides valuable insights into the booking patterns of hotel guests. The day of the week with the highest count represents the peak period for hotel reservations. This information can be utilized by hotel management to optimize resource allocation, staffing, and marketing strategies to cater to the increased demand on the most booked day of the week.

Analysis -2: The Length of Stay for Each Booking

Methodology:

In this analysis, we aim to calculate the length of stay for hotel bookings by considering both the number of weekend nights (no_of_weekend_nights) and week nights (no_of_week_nights). The calculation involves determining the arrival date, adding the total number of nights, and subsequently calculating the departure date.

RESULTS:



Booking_ID	length_of_stay
INN00001	3
INN00002	5
INN00003	3
INN00004	2
INN00005	2
INN00006	2
INN00007	4
INN00008	4
INN00009	4
INN00010	5
INN00011	1
INN00012	3
INN00013	3
INN00014	2
INN00015	2
INN00016	2
INN00017	1
INN00018	4
INN00019	4
INN00020	1

INFERENCE:

The output of this analysis provides valuable information about the length of stay for each hotel booking. Hotel management can use this data to understand the average duration of guest stays, which can inform pricing strategies, housekeeping schedules, and resource planning. This insight contributes to the overall optimization of hotel operations and enhances the guest experience.

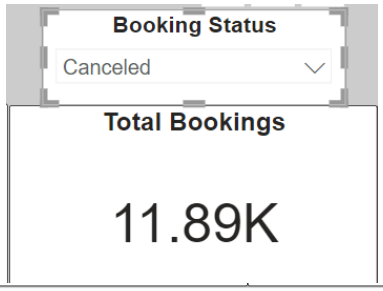
Analysis -3: Cancellation Analysis

Methodology:

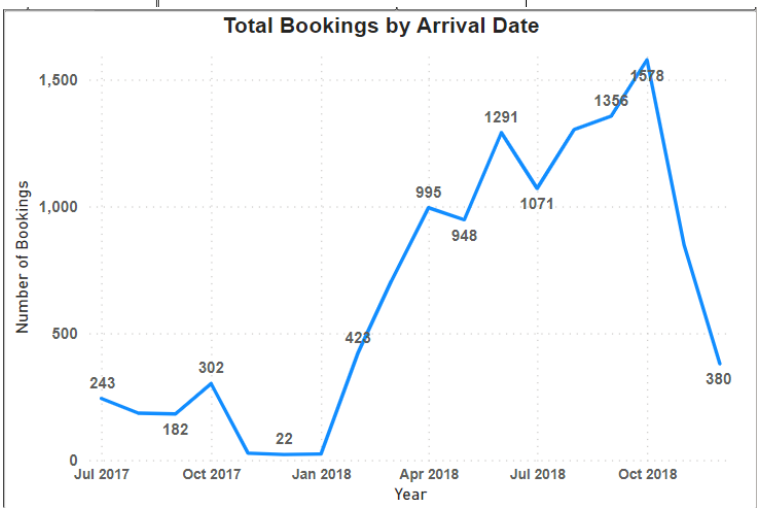
This analysis aims to provide insights into the booking status of hotel reservations, focusing on the total number of bookings, the count of cancellations, and the percentage of cancellations relative to the total bookings. Utilizing the Customer_Details table in Apache Spark, we extracted and calculated relevant metrics to understand the overall booking dynamics.

RESULTS:

```
cancellationPercentage: Double = 32.76361130254997
```




```
scala> statusCountsDF.show()
+--------------+-----+
|booking_status|count|
+--------------+-----+
|    Canceled\r|11885|
|Not_Canceled\r|24390|
+--------------+-----+
```



INFERENCE:

Understanding the booking status metrics is crucial for hotel management to optimize strategies for customer retention, pricing adjustments, and resource allocation. This analysis Offers insights into the number of bookings canceled, aiding in understanding customer behavior.

Analysis -4: AVG LEAD TIME FOR EACH MONTH

Methodology:

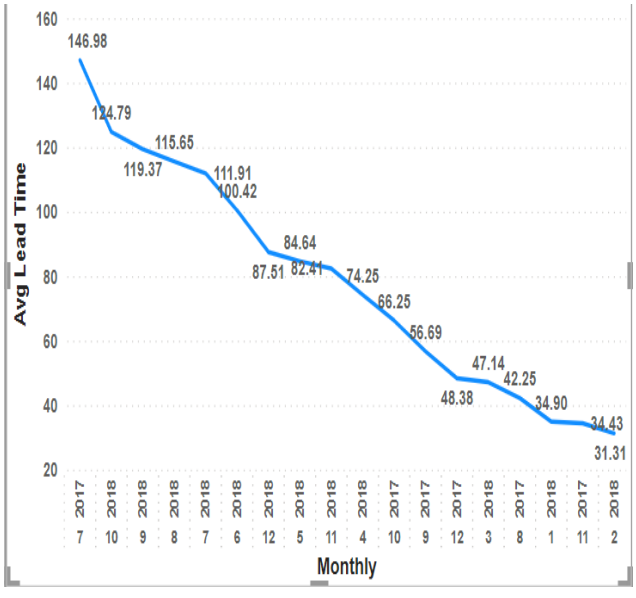
This analysis focuses on understanding the average lead time for hotel bookings, considering variations across different months. From the Booking_Details table, we employed a window function to calculate the average lead time for each combination of arrival year and month. The results provide insights into the booking patterns over time.

INFERENCE:

The output of this analysis presents the average lead time for hotel bookings, categorized by both arrival year and month. Understanding these trends can be valuable for hotel management in anticipating demand patterns, optimizing marketing strategies, and adjusting pricing based on historical lead time data. Variations in average lead time across months may also inform operational planning and resource allocation.

RESULTS:

```
scala> result.show()
+------------+-------------+------------------+
|arrival_year|arrival_month|     avg_lead_time|
+------------+-------------+------------------+
|        2018|            1| 34.89842209072978|
|        2018|            2| 31.31338028169014|
|        2018|            3| 47.13952502120441|
|        2018|            4| 74.25255847953217|
|        2018|            5| 84.63933795227098|
|        2018|            6|100.41554792382142|
|        2017|            7|146.97796143250687|
|        2018|            7| 116.2722602739726|
|        2017|            8| 42.250493096646941|
|        2018|            8| 96.13139260424862|
|        2017|            9| 56.69254093389933|
|        2018|            9| 96.95337237041856|
|        2017|           10| 66.25196027182436|
|        2018|           10|103.72823020500282|
|        2017|           11|34.425038639876355|
|        2018|           11| 71.99328859060402|
|        2017|           12| 48.37823275862069|
|        2018|           12| 75.48924197285668|
+------------+-------------+------------------+
```



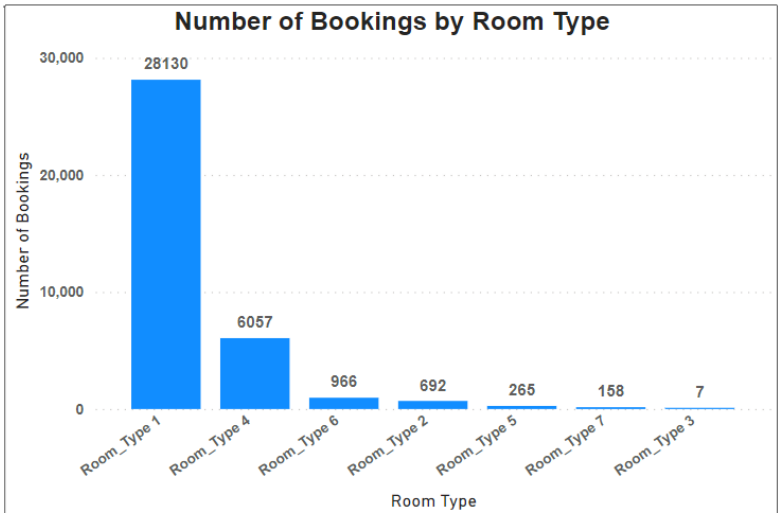
Analysis -5: ROOM UTILIZATION RATE BY ROOM TYPE

Methodology:

This analysis aims to assess the room utilization rate for different room types in a hotel. Utilizing the Booking_Details table in Apache Spark, we calculated the utilization rate by dividing the count of bookings by the count of distinct arrival dates for each room type. The results provide insights into the efficiency of room utilization across various room categories.

RESULTS:

room_type_reserved	utilization_rate
Room_Type 7	5.266666666666667
Room_Type 2	22.322580645161292
Room_Type 3	1.1666666666666667
Room_Type 1	907.4193548387096
Room_Type 5	8.548387096774194
Room_Type 6	31.161290322580644
Room_Type 4	195.38709677419354



INFERENCE:

The output of this analysis provides valuable information on how efficiently different room types are being utilized in the hotel. The utilization rate reflects the average number of bookings per available room on distinct arrival dates. This metric is essential for optimizing room allocation strategies, identifying high-demand room types, and ensuring efficient use of hotel resources.

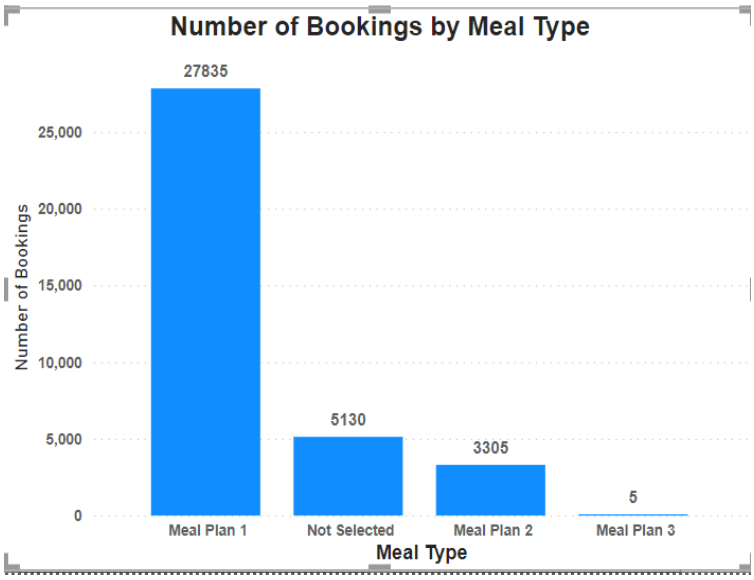
Analysis -6: BOOKINGS BY MEAL PLANS

Methodology:

This analysis focuses on understanding the distribution of hotel bookings based on different types of meal plans. Using the Booking_Details table in Apache Spark, we identified distinct meal plans and calculated the total number of bookings for each plan. The results provide insights into guest preferences regarding meal plans, allowing hotel management to tailor services and optimize offerings.

RESULTS:

```
scala> bookingsCountByMealPlan.show()
+-----------------+--------------+
|type_of_meal_plan|total_bookings|
+-----------------+--------------+
|      Meal Plan 1|         27835|
|     Not Selected|          5130|
|      Meal Plan 3|             5|
|      Meal Plan 2|          3305|
+-----------------+--------------+
```



INFERENCE:

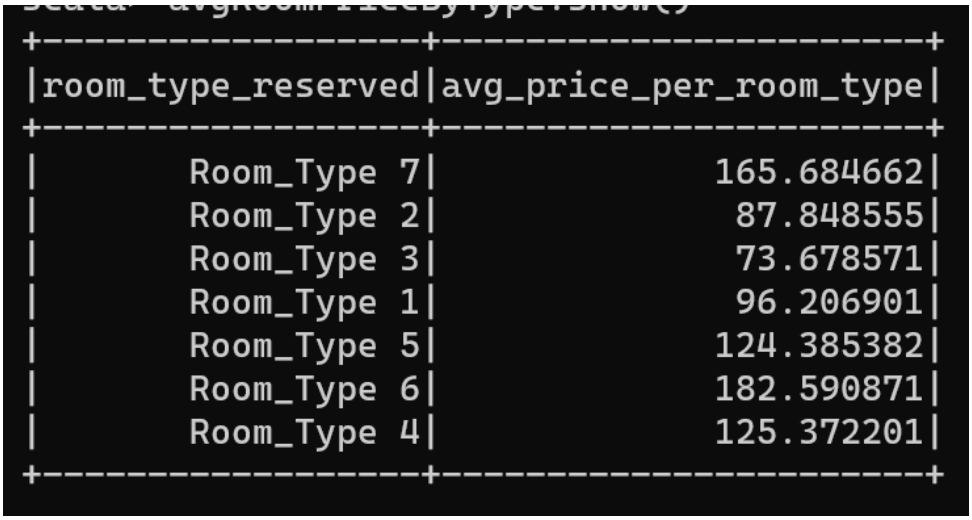
The output of this analysis provides a breakdown of the total number of bookings associated with each meal plan. This information is crucial for understanding guest preferences and allows hotel management to optimize meal plan offerings based on demand. Additionally, it aids in resource planning, staff allocation, and pricing strategies tailored to different meal plan categories.

Analysis -7: HIGH VALUE CUSTOMERS AND THEIR AVG ROOM PRICE BY ROOM TYPE

Methodology:

This analysis aims to identify and analyze high-value customers within a hotel reservation system. High-value customers are defined as repeated guests with zero previous cancellations. Using Apache Spark, we filtered and identified these high-value customers in the Customer_Details table. Subsequently, we joined this subset with the Booking_Details table to gather additional booking details. The analysis focuses on calculating the average room price for each room type booked by these high-value customers.

RESULTS:



room_type_reserved	avg_price_per_room_type
Room_Type 7	165.684662
Room_Type 2	87.848555
Room_Type 3	73.678571
Room_Type 1	96.206901
Room_Type 5	124.385382
Room_Type 6	182.590871
Room_Type 4	125.372201

INFERENCE:

The output of this analysis provides the average room price for each room type booked by high-value customers. This information is valuable for understanding the spending patterns of customers who contribute significantly to the hotel's revenue. It enables hotel management to tailor marketing strategies, personalize services, and optimize pricing for different room types to enhance the overall customer experience.

Analysis -8: PRICE CHANGE ANALYSIS

Methodology:

This analysis focuses on understanding the price changes in hotel bookings over time. We utilized Apache Spark to join the Customer_Details and Booking_Details tables based on the common column "Booking_ID." Subsequently, a window function was applied to calculate the price change for each booking by comparing the average room price with the previous day's price using the lag Function.

RESULTS:

Booking_ID	arrival_year	arrival_month	arrival_date	avg_price_per_room	price_change
INN32763	2017	7	1	101.50	NULL
INN27897	2017	7	1	80.00	-21.50
INN00584	2017	7	1	80.00	0.00
INN34752	2017	7	1	101.50	21.50
INN07285	2017	7	1	101.50	0.00
INN16947	2017	7	1	101.50	0.00
INN06955	2017	7	1	101.50	0.00
INN16843	2017	7	1	101.50	0.00
INN14890	2017	7	1	101.50	0.00
INN02487	2017	7	1	101.50	0.00
INN03899	2017	7	1	101.50	0.00
INN17341	2017	7	1	101.50	0.00
INN21337	2017	7	1	80.00	-21.50
INN02006	2017	7	1	101.50	21.50
INN07680	2017	7	1	101.50	0.00
INN30226	2017	7	1	101.50	0.00
INN05885	2017	7	1	80.00	-21.50
INN13132	2017	7	1	101.50	21.50
INN09122	2017	7	1	101.50	0.00
INN24533	2017	7	1	101.50	0.00

INFERENCE:

The output of this analysis provides a detailed view of the average room price changes for each booking over time. The "price_change" column represents the daily variation in room prices. This information is valuable for identifying trends, understanding price fluctuations, and potentially optimizing pricing strategies based on demand patterns.

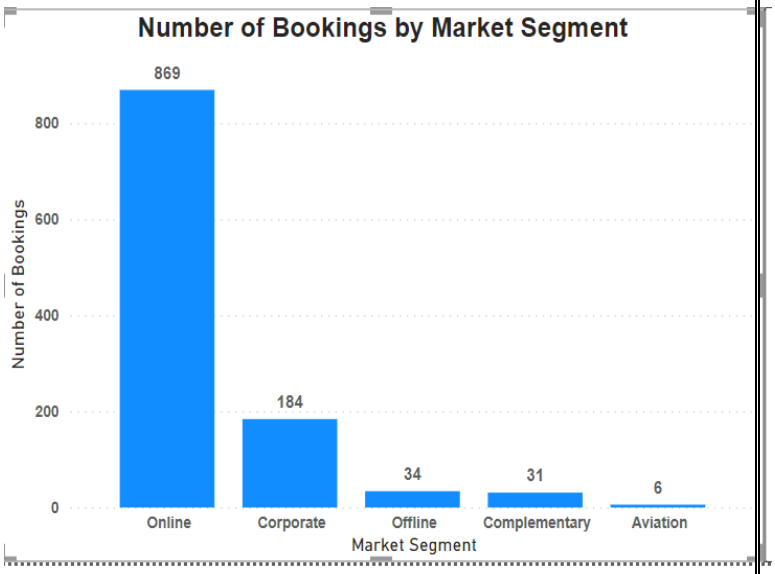
Analysis -9: PERCENTAGE OF BOOKINGS WITH CAR PARKING SPACE BY MARKET SEGMENT

Methodology:

This analysis aims to understand the utilization of car parking spaces based on different market segments. Using Apache Spark, we grouped the data in the Booking_Details table by market segment type, calculated the count of bookings with car parking requirements, and determined the overall percentage of bookings with car parking needs within each market segment.

RESULTS:

```
scala> parkingPercentageBySegment.show()
+-------------------+-------------------+
|market_segment_type|parking_percentage|
+-------------------+-------------------+
|      Complementary| 7.928388746803069|
|           Aviation|                4.8|
|          Corporate| 9.122459097669806|
|             Online|3.7434306883777033|
|            Offline|0.3229483282674772|
+-------------------+-------------------+
```



INFERENCE:

The output of this analysis reveals the percentage of bookings with car parking requirements within each market segment. This information is valuable for understanding the parking demand patterns across different customer segments, allowing hotel management to allocate resources effectively and tailor services based on the needs of each market segment.

Analysis -10: SPECIAL REQUESTS PERCENTAGE

Methodology:

This analysis focuses on understanding the prevalence of special requests in hotel bookings. Using Apache Spark, we created a new column indicating whether each booking has special requests based on the "no_of_special_requests" column. Subsequently, we calculated the percentage of bookings with special requests.

RESULTS:

```
scala> percentageWithSpecialRequests.show()
+-------------------------------+
|percentage_with_special_requests|
+-------------------------------+
|              45.480358373535495|
+-------------------------------+
```

INFERENCE:

The output of this analysis provides the percentage of bookings that include special requests. This information is valuable for hotel management to understand the prevalence of guest preferences for additional services or accommodations. It enables the hotel to anticipate and fulfill specific guest needs, enhancing the overall guest experience.

Analysis -11: REPEATED GUESTS BY ROOM TYPE

Methodology:

This analysis focuses on understanding the distribution of repeated guests across different room types within a hotel. Using Apache Spark, we joined the Customer_Details and Booking_Details tables based on the common column "Booking_ID." Subsequently, we filtered the data to include only repeated guests (repeated_guest equals 1) and grouped the results by room type. The outcome provides insights into the popularity of specific room types among guests who have visited the hotel more than once.

RESULTS:

```
scala> totalRepeatedGuestsByRoomType.show()
+------------------+---------------------+
|room_type_reserved|total_repeated_guests|
+------------------+---------------------+
|       Room_Type 7|                   21|
|       Room_Type 2|                    7|
|       Room_Type 1|                  808|
|       Room_Type 5|                   17|
|       Room_Type 6|                   10|
|       Room_Type 4|                   67|
+------------------+---------------------+
```

INFERENCE:

The output of this analysis provides insights into the distribution of repeated guests across different room types. By understanding which room types are more popular among guests who choose to return, hotel management can tailor marketing strategies, allocate resources efficiently.

CONCLUSION:

The insights gathered from these analyses collectively contribute to informed decision-making, enabling the hotel to:

- Optimize pricing strategies based on historical trends.
- Tailor services to meet the specific preferences of different customer segments.
- Enhance operational efficiency by allocating resources effectively.
- Foster customer loyalty by understanding and catering to repeat guests.
- Improve overall guest satisfaction through personalized service offerings.

This project demonstrates the power of combining database management systems, big data analytics, and visualization tools to extract actionable insights from a hotel reservation system. The data-driven approach adopted in this project empowers the hotel management to make informed decisions, optimize operations, and ultimately enhance the guest experience.