Hotel Booking Analysis - Capstone Project

Submitted By *Archana Vijayan*

Overview

This hotel booking data analytics project aims to optimize operations and guest satisfaction by analyzing factors like seasonal trends, customer preferences, and operational efficiency. The project's goals include refining booking strategies, enhancing customer experiences, and providing stakeholders with a powerful tool – a Power BI dashboard – to make informed decisions in the competitive hospitality landscape. Ultimately, the project seeks to translate complex data insights into actionable strategies for improved business performance.

The Process

1. Data Acquisition from GitHub:Initiate the hotel booking data analytics project by acquiring the necessary datasets from a designated GitHub repository. Utilize Git commands or GitHub's interface to clone or download the relevant data files.
2. Connecting SQL Database:Establish a connection between the SQL database and Excel for seamless data transfer. Utilize Excel as a data analysis tool to further explore and preprocess the hotel booking data.
3. Connecting with Tools:To facilitate seamless data integration and processing, the dataset is connected with various analytical tools. Interfaces with Power BI, Excel, and MySQL Workbench are established, enabling efficient and effective handling of the dataset within these tools. This connectivity lays the groundwork for subsequent analysis and exploration.
4. Data Transformation and Cleansing in SQL:Execute SQL queries for data transformation and cleansing. Address missing values, standardize formats, and resolve discrepancies within the SQL database. Ensure data quality and consistency.
5. Exploratory Data Analysis (EDA) in SQL:Employ SQL queries to perform exploratory data analysis (EDA) on the hotel booking dataset. Craft queries to understand the distribution, summary statistics, and relationships within the data. Use aggregate functions, grouping, and filtering to extract valuable insights directly from the SQL database.
6. Excel Analysis for Data Exploration:Utilize Excel's data analysis features to explore trends, calculate metrics, and perform initial data exploration. This can include pivot tables, charts, and other Excel functionalities.
7. Problem Statement Solution in Power BI and EXCEL:The core analysis phase utilizes Power BI, a powerful analytical tool. Here, the specified problem statements are addressed using Power BI's robust features for data visualization, exploration, and analysis. The tool's capabilities are harnessed to derive valuable insights and solutions from the university rankings data. This step ensures a comprehensive and visually intuitive understanding of the dataset, aiding in informed decision-making.

Objective

In the dynamic realm of hospitality, understanding and harnessing the power of data is paramount for strategic decision-making and operational excellence. This data analytics project delves into the intricacies of hotel booking data, aiming to unravel patterns and insights that can shape the industry landscape. By employing rigorous Exploratory Data Analysis (EDA) and leveraging the visualization capabilities of Power BI, the project strives to offer a holistic perspective on crucial aspects such as booking trends, guest demographics, and pricing dynamics. Through a meticulous exploration of arrival dates, booking patterns, and customer preferences, the project not only seeks to unveil the nuances within the dataset but also endeavors to provide actionable insights that can propel the hospitality sector toward enhanced efficiency, superior customer experiences, and sustained competitiveness.

The project will involve the following tasks.

1. Temporal Dynamics:Analyze the distribution of arrival dates to identify common arrival days and discern patterns in lead times, facilitating a comprehensive understanding of temporal dynamics.
2. Seasonal Influences:Investigate peak booking months and scrutinize reasons for spikes, including the impact of holidays and events, to inform strategic marketing and resource planning.
3. Guest Behavior and Preferences:Calculate average lengths of stays for different hotel types and explore variations based on meal plans, unraveling insights into guest behavior and preferences.
4. Temporal Trends:Conduct a nuanced analysis of booking patterns over the years, examining year-over-year changes in bookings and cancellations to uncover temporal trends.
5. Demographic Distribution:Understand the distribution of guests by age groups (adults, children, babies) and identify outliers, providing insights into the demographic composition of clientele.
6. Pricing Strategies:Calculate summary statistics for Average Daily Rates (ADR) and explore differences between Resort Hotel and City Hotel bookings, aiding in the formulation of effective pricing strategies.
7. Parking Preferences:Analyze the distribution of required car parking spaces for each hotel type, determining if specific types attract more guests with cars and informing resource allocation.
8. Customer Preferences:Compare total special requests made by different customer types, unveiling trends in customer preferences and informing personalized service strategies.
9. Meal Plan Insights:Understand the distribution of meal plans and their variations, offering insights into guest preferences and influencing marketing and catering strategies.
10. Revenue Optimization:Analyze Average Daily Rates (ADR) by meal plan type, identifying variations in pricing and contributing to revenue optimization strategies.
11. Comprehensive Impact Analysis:Investigate the combined impact of parking spaces, special requests, and meal plans by analyzing their distribution based on hotel types, offering a holistic understanding.

12. Guest Retention Strategies:Explore the proportion of repeated guests and their booking behavior, unveiling insights that can shape guest retention and loyalty strategies.
13. Historical Booking Impact:Examine the impact of a guest's booking history on their likelihood of canceling a current booking, providing valuable insights into guest behavior and risk mitigation.
14. Room Allocation Consistency:Understand the distribution of reserved and assigned room types, calculating summary statistics for consistency, aiding in optimizing room allocation processes.
15. Adaptability and Continuous Improvement:Leverage the project findings to foster a culture of continuous improvement, encouraging adaptability to changing market trends and maintaining a competitive edge.

Significance:

The significance of this data analytics project lies in its potential to inform decision-making across various facets of the hospitality industry. The detailed EDA enables stakeholders to make informed decisions by providing insights into peak booking periods, guest demographics, and pricing dynamics. This informed decision-making is crucial for strategic planning and resource optimization, enhancing overall operational efficiency.

Furthermore, the project contributes to improving the customer experience by tailoring services based on guest preferences identified through the analysis of meal plans, room types, and other factors. This personalized approach fosters customer satisfaction and loyalty, critical components for success in the competitive hospitality landscape.

Operational efficiency is another key area where the project adds value. By analyzing booking patterns, cancellation rates, and room type preferences, the project aids in optimizing resource allocation, staffing levels, and overall infrastructure. This optimization not only improves day-to-day operations but also informs long-term planning and investment decisions.

Strategic marketing and revenue optimization are facilitated through insights into peak booking months, reasons for spikes, and the impact of different booking channels. Hotels can tailor marketing efforts, optimize pricing strategies, and capitalize on preferred booking channels to maximize revenue.

Moreover, the project contributes to guest retention and loyalty by understanding the impact of booking history on cancellation likelihood and analyzing the behavior of repeated guests. This knowledge allows hotels to tailor offerings to specific guest needs, fostering loyalty and encouraging repeat business.

The analysis of the distribution of required car parking spaces contributes to resource planning and infrastructure investment, ensuring that hotels can meet guest needs effectively. Transparent reporting and communication of insights, facilitated by Power BI's visual dashboards, contribute to clear and comprehensible presentations of complex data, aiding effective communication among stakeholders.

Finally, the project promotes a culture of continuous improvement by encouraging regular analysis and adaptation to changing trends. This adaptability is crucial for staying competitive and relevant in a dynamic hospitality industry. In summary, the objectives and significance of this project intertwine to create a comprehensive approach to leveraging data for informed decision-making and strategic positioning in the hospitality sector.

Data Dictionary

Tables

1. Booking_details
2. booking_source_and_history
3. country
4. df_booking_details
5. df_booking_source_and_history
6. distribution_channel
7. guest_info
8. market_segment
9. meal_and_stay_details
10. reservation_status
11. room_details

Variables

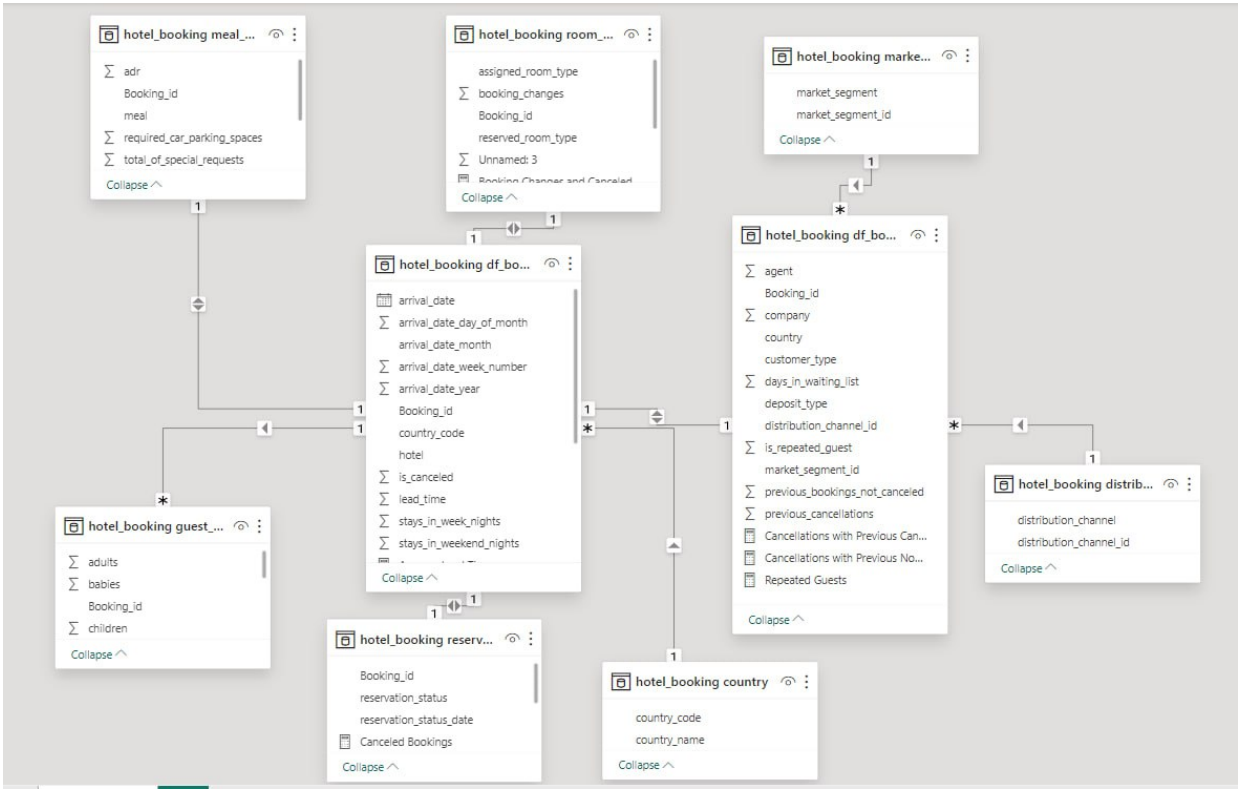
Variable	Type	Description
hotel	factor	Hotel (City Hotel or Resort Hotel)
is_canceled	integer	Value indicating if the booking was canceled (1) or not (0)
lead_time	integer	Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
arrival_date_year	integer	Year of arrival date
arrival_date_month	factor	Month of arrival date
arrival_date_week_number	integer	Week number of year for arrival date
arrival_date_day_of_month	integer	Day of arrival date
stays_in_weekend_nights	integer	Number of weekend nights (Saturday or Sunday) the guest

		stayed or booked to stay at the hotel
stays_in_week_nights	integer	Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
adults	integer	Number of adults
children	integer	Number of children
babies	integer	Number of babies
meal	factor	Type of meal booked. Undefined/SC - no meal package; BB - Bed & Breakfast; HB - Half board (breakfast and one other meal – usually dinner); FB - Full board (breakfast, lunch and dinner)
country	factor	Country of origin. Categories are represented in the ISO 3155–3:2013 format
market_segment	factor	Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”
distribution_channel	factor	Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”
is_repeated_guest	integer	Value indicating if the booking name was from a repeated guest (1) or not (0)
previous_cancellations	integer	Number of previous bookings that were cancelled by the customer prior to the current booking
previous_bookings_not_canceled	integer	Number of previous bookings not cancelled by the customer prior to the current booking
reserved_room_type	factor	Code of room type reserved. Code is presented instead of designation for anonymity reasons

assigned_room_type	factor	Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons
booking_changes	integer	Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation
deposit_type	factor	Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay
agent	factor	ID of the travel agency that made the booking
company	factor	ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons
days_in_waiting_list	integer	Number of days the booking was in the waiting list before it was confirmed to the customer

customer_type	factor	Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking
adr	numeric	Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
required_car_parking_spaces	integer	Number of car parking spaces required by the customer
total_of_special_requests	integer	Number of special requests made by the customer (e.g. twin bed or high floor)
reservation_status	factor	Reservation last status, assuming one of three categories: Canceled – booking was canceled by the customer; Check-Out – customer has checked in but already departed; No-Show – customer did not check-in and did inform the hotel of the reason why
reservation_status_date	factor	Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel

ER Diagram

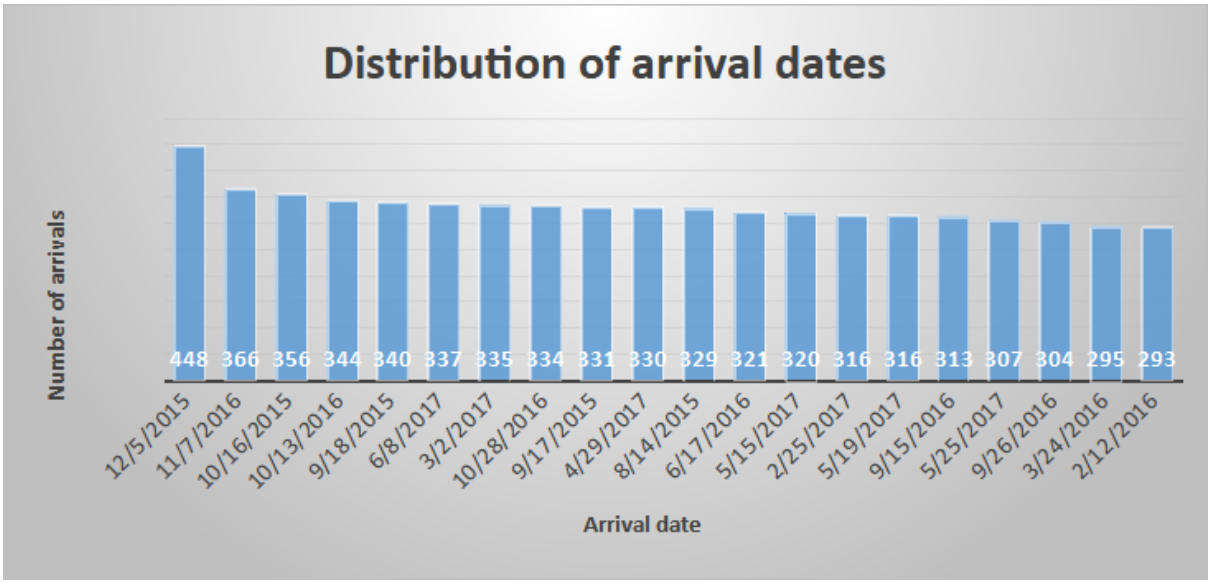

hotel_booking meal_...
adr
Booking_id
meal
required_car_parking_spaces
total_of_special_requests
Collapse
hotel_booking room_...
assigned_room_type
booking_changes
Booking_id
reserved_room_type
Unnamed: 3
Collapse
hotel_booking marke...
market_segment
market_segment_id
Collapse
hotel_booking df_bo...
arrival_date
arrival_date_day_of_month
arrival_date_month
arrival_date_week_number
arrival_date_year
Booking_id
country_code
hotel
is_canceled
lead_time
stays_in_week_nights
stays_in_weekend_nights
Collapse
hotel_booking df_bo...
agent
Booking_id
company
country
customer_type
days_in_waiting_list
deposit_type
distribution_channel_id
is_repeated_guest
market_segment_id
previous_bookings_not_canceled
previous_cancellations
Cancellations with Previous Can...
Cancellations with Previous No...
Repeated Guests
Collapse
hotel_booking distrib...
distribution_channel
distribution_channel_id
Collapse
hotel_booking guest_...
adults
babies
Booking_id
children
Collapse
hotel_booking reserv...
Booking_id
reservation_status
reservation_status_date
Canceled Bookings
Collapse
hotel_booking country
country_code
country_name
Collapse

EDA Questions

1.Understand the distribution of arrival dates, including the most common arrival days and summary statistics for lead times.

arrival_date	date_count
12/5/2015	448
11/7/2016	366
10/16/2015	356
10/13/2016	344
9/18/2015	340
6/8/2017	337
3/2/2017	335
10/28/2016	334
9/17/2015	331
4/29/2017	330

Top 10 most common arrival days



Monthly summary statistics for lead times

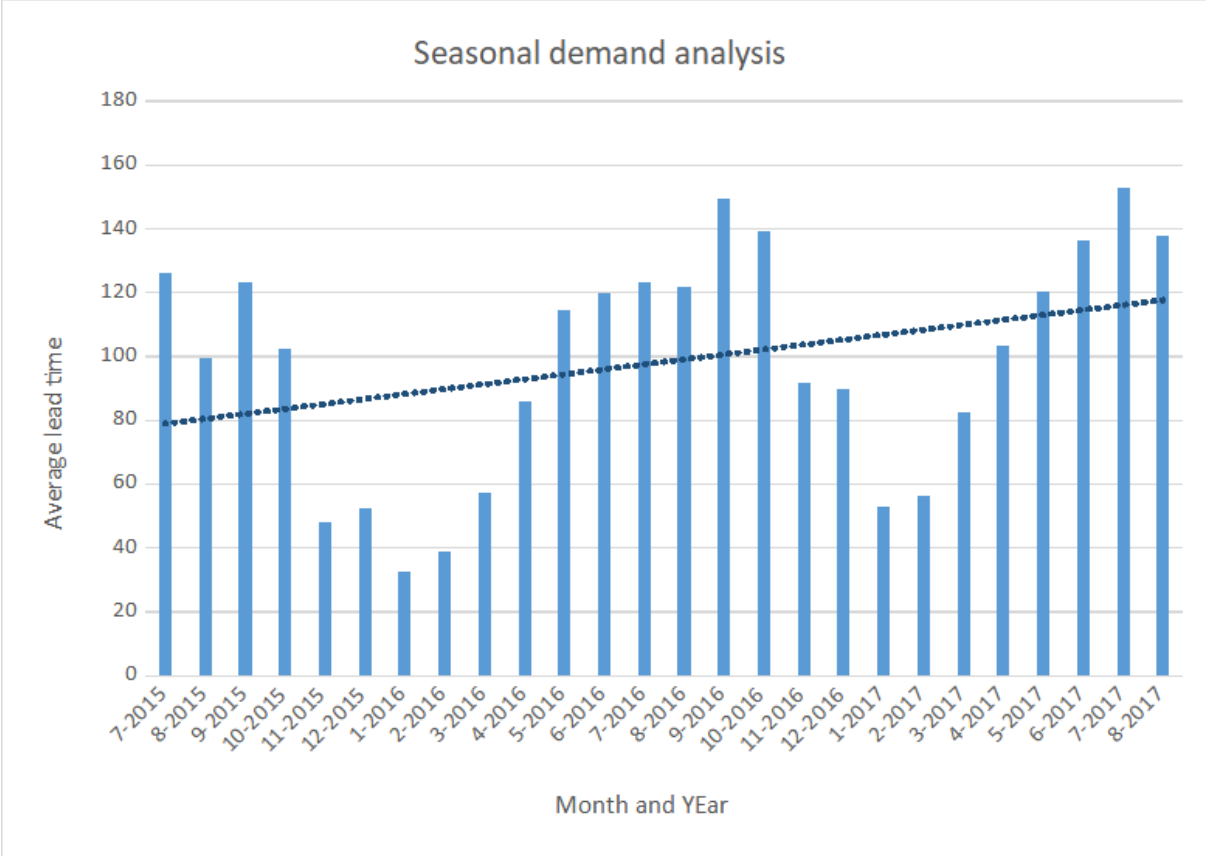
year	month	average_lead_time	min_lead_time	max_lead_time	stddev_lead_time	Month and year
2015	7	125.9679	0	737	101.228986410525	7-2015
2015	8	99.3924	0	460	107.649817523454	8-2015
2015	9	123.1382	0	382	116.008112384314	9-2015
2015	10	102.394	0	532	113.102335508728	10-2015
2015	11	48.0919	0	305	58.5167345694327	11-2015
2015	12	52.3863	0	414	63.6174575744578	12-2015
2016	1	32.5231	0	284	46.2936568848551	1-2016
2016	2	38.8412	0	709	43.4128716025588	2-2016
2016	3	57.3261	0	395	66.7010205008983	3-2016
2016	4	85.8473	0	443	76.5880802983176	4-2016
2016	5	114.6743	0	468	105.460508096818	5-2016
2016	6	119.8235	0	366	96.72143959586	6-2016
2016	7	123.2594	0	424	95.632455616674	7-2016
2016	8	121.634	0	454	98.6966163875811	8-2016

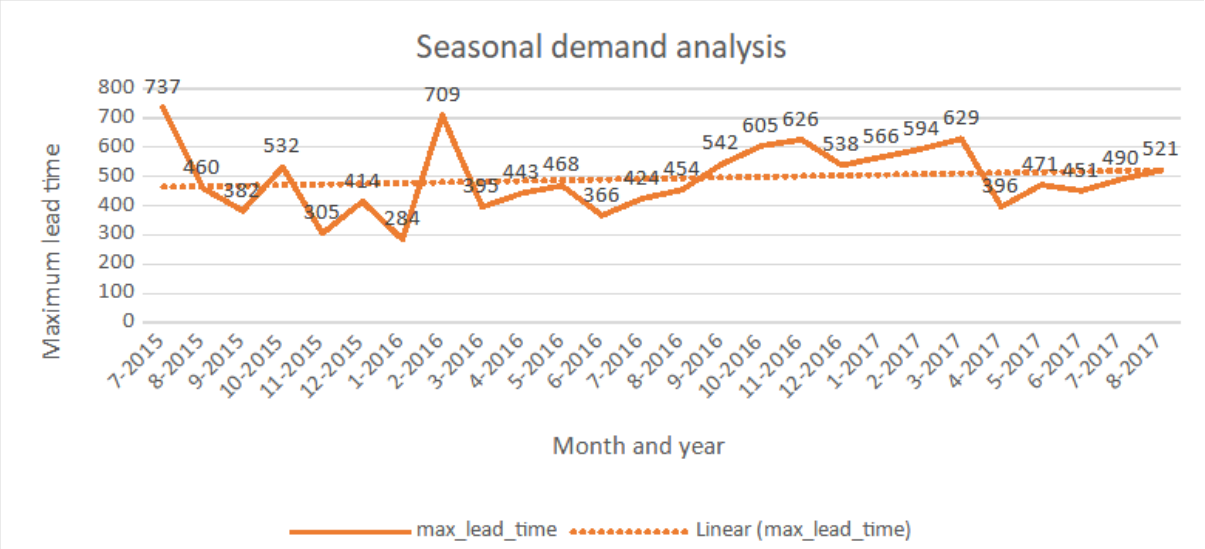
2016	9	149.5095	0	542	132.460300415238	9-2016
2016	10	139.4572	0	605	132.199517621932	10-2016
2016	11	91.548	0	626	114.82606594062	11-2016
2016	12	89.5798	0	538	113.199894752503	12-2016
2017	1	53.119	0	566	89.532177649881	1-2017
2017	2	56.2605	0	594	89.3191020336203	2-2017
2017	3	82.5366	0	629	101.232348075223	3-2017
2017	4	103.4291	0	396	83.3556692632014	4-2017
2017	5	120.2249	0	471	102.372838675416	5-2017
2017	6	136.1415	0	451	108.675182856909	6-2017
2017	7	152.974	0	490	121.985025304611	7-2017
2017	8	137.7986	0	521	113.058522368786	8-2017

a. Peak Booking Days:
    i. December 5, 2015, has the highest count of bookings (448), indicating it was a peak day for hotel reservations.
    ii. November 7, 2016, and October 16, 2015, also have substantial booking counts, suggesting these dates might correspond to popular travel periods.
b. Seasonal Patterns:
    i. There might be seasonal trends in hotel bookings, as seen in the variation of counts across different months and years.
    ii. For instance, the high counts in September 2015 and October 2016 might suggest a particular seasonal demand during these periods.

c. Consistent Popular Dates:Some dates, like November 7, 2016, October 16, 2015, and October 13, 2016, consistently appear among the top 10, indicating their popularity across different years.
d. Yearly Variation:Analyzing the top dates across different years can reveal whether certain periods consistently attract more bookings or if there are variations from year to year.
e. Potential Promotions or Special Events:Consider exploring external factors, such as promotions or local events, that might have influenced the high booking counts on these specific dates.
f. Strategic Planning:Hotel management can use these insights for strategic planning, allocating resources, and optimizing services during periods of high demand.
g. Demand Forecasting:Understanding the popularity of these dates can aid in forecasting future demand, helping the hotel make informed decisions about pricing, staffing, and inventory.

2.Identify peak booking months and analyze reasons for spikes in bookings, including holidays or events.



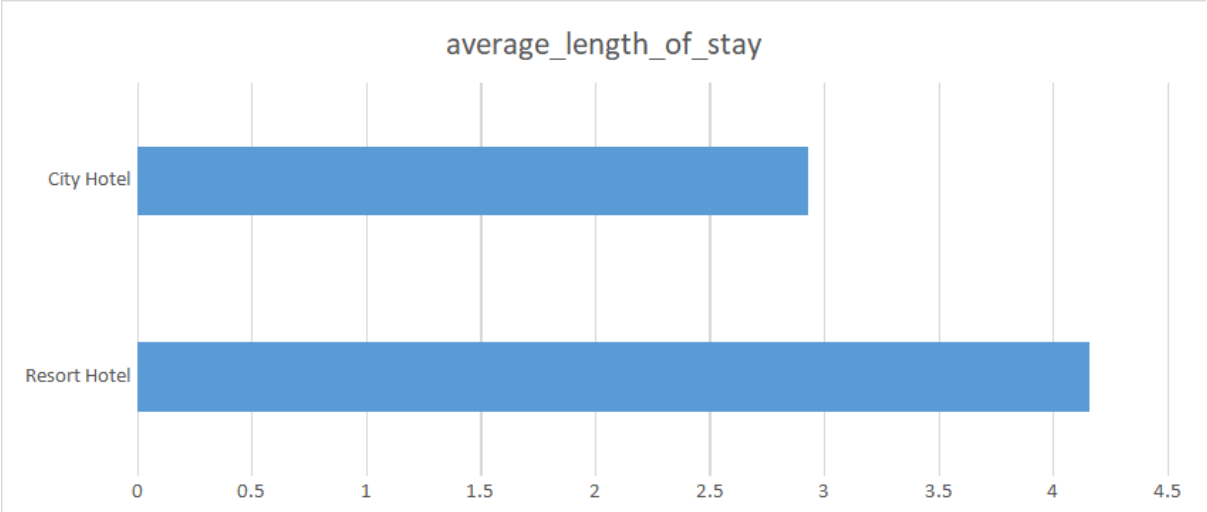


1. Seasonal Variations in Lead Times:Lead times display noticeable seasonality, with peaks observed in the summer months (July and August) across the three years. This suggests that customers tend to plan and book their stays well in advance during peak vacation seasons.
2. Yearly Increasing Trend in Average Lead Time:There is a consistent upward trend in average lead times from 2015 to 2017, indicating a potential shift in booking behavior towards longer planning horizons. This trend may be attributed to changing customer preferences or external factors influencing travel planning.
3. Lead Time Variability:The standard deviation of lead times fluctuates, indicating varying degrees of uncertainty in booking behaviors. Higher standard deviations might signify periods of increased unpredictability or external influences affecting booking patterns.
4. Short-Term Booking Trends:Despite the general trend of increasing lead times, instances of zero lead time (last-minute bookings) persist throughout the years. This suggests a segment of customers who prefer spontaneous bookings, possibly taking advantage of discounts or deals.
5. Impact on Operational Planning:Longer lead times may necessitate enhanced operational preparations, particularly during peak months, to meet the anticipated demand. On the other hand, short lead times require agile operational responses to accommodate sudden surges in bookings.
6. Implications for Marketing Strategies:The observed lead time patterns provide insights for crafting effective marketing strategies. Tailoring promotions to incentivize early bookings during peak months or offering last-minute deals could capitalize on observed booking behaviors.

3.Calculate the average length of stays for different hotel types and explore variations by meal plans.

1. Average Length of Stay by Hotel Type:
    a. Resort Hotel:The average length of stay for guests in Resort Hotels is 4.1557 days. This metric represents the overall average duration of guests' stays in Resort Hotels, encompassing various meal plans.

b. City Hotel:In City Hotels, the average length of stay is notably shorter at 2.9303 days. This suggests a difference in guest behavior and stay durations compared to Resort Hotels.



hotel	average_length_of_stay
Resort Hotel	**4.1557**
City Hotel	**2.9303**

2. Variations by Meal Plans:
    a. Resort Hotel:
        i. Bed and Breakfast (BB): Guests opting for the BB meal plan have an average stay of 3.7639 days in Resort Hotels.
        ii. Full Board (FB): For guests with the FB meal plan, the average length of stay increases to 4.7588 days, indicating a longer duration compared to BB.
        iii. Half Board (HB): Guests on the HB meal plan have the longest average stay among all meal plans, with an average length of stay of 5.6152 days.
        iv. Self Catering (SC): Guests choosing the SC meal plan have a considerable average stay of 7.0602 days, signifying a preference for more extended stays.
        v. Undefined: For guests with an undefined meal plan, the average length of stay is 4.4156 days.
    b. City Hotel:
        i. Bed and Breakfast (BB): In City Hotels, guests with the BB meal plan stay, on average, for 2.9948 days.
        ii. Half Board (HB): The average length of stay for guests with the HB meal plan is slightly lower at 2.7874 days.
        iii. Self Catering (SC): Guests opting for the SC meal plan in City Hotels have an average stay of 2.6696 days.

iv. Full Board (FB): Guests with the FB meal plan in City Hotels have the shortest average stay, with a duration of 1.6667 days.

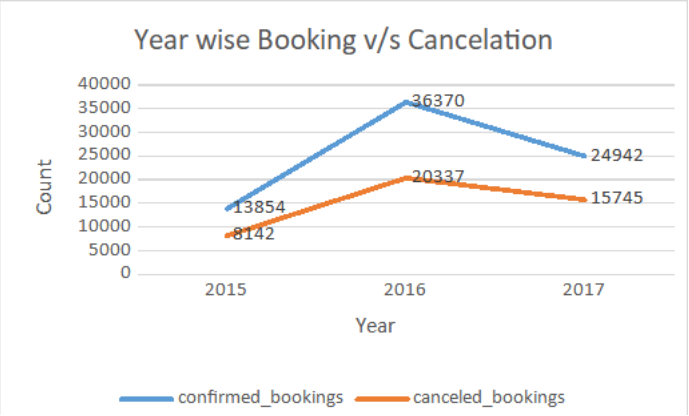
hotel	meal	average_length_of_stay
Resort Hotel	BB	3.7639
Resort Hotel	FB	4.7588
Resort Hotel	HB	5.6152
Resort Hotel	SC	7.0602
Resort Hotel	Undefined	4.4156
City Hotel	BB	2.9948
City Hotel	HB	2.7874
City Hotel	SC	2.6696
City Hotel	FB	1.6667

Inferences:

1. Guests in Resort Hotels generally have longer stays compared to City Hotels.
2. Among Resort Hotel guests, those on the SC meal plan tend to have the longest average stays, followed by HB, FB, BB, and Undefined.
3. In City Hotels, guests with the BB meal plan have the longest average stay, while those with the FB meal plan have the shortest duration.
4. Understanding these variations is crucial for tailoring marketing strategies, optimizing operational planning, and providing services that align with guest preferences for different hotel types and meal plans.

4.Analyze how booking patterns have evolved over the years, including yearoveryear changes in bookings and cancellations.

Year	confirmed	cancele
2015	13854	8142
2016	36370	20337
2017	24942	15745

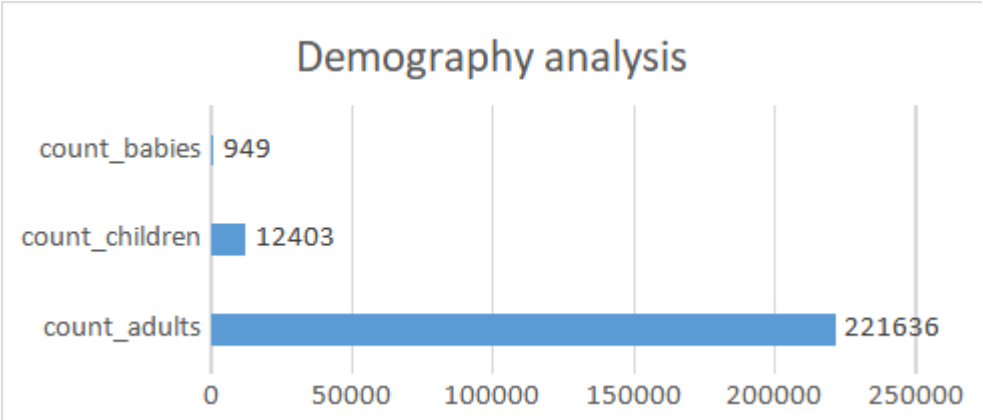


1. Total Bookings:The total number of bookings has shown a consistent upward trend over the three years, indicating a growth in overall demand for hotel accommodations. The progression is as follows:
    a. 2015: 13,854 bookings
    b. 2016: 36,370 bookings
    c. 2017: 24,942 bookings
2. Year-over-Year Changes:The year-over-year changes in total bookings highlight substantial growth from 2015 to 2016, with a nearly threefold increase. However, there is a noticeable decrease in total bookings from 2016 to 2017, indicating a shift in the booking landscape.
3. Confirmed Bookings:Confirmed bookings represent successful reservations that guests honored by checking in. The confirmed bookings for each year are as follows:
    a. 2015: 13,854 bookings
    b. 2016: 36,370 bookings
    c. 2017: 24,942 bookings
4. Canceled Bookings:Canceled bookings refer to reservations that were initially made but later canceled by guests. The number of canceled bookings for each year is as follows:
    a. 2015: 8,142 cancellations
    b. 2016: 20,337 cancellations
    c. 2017: 15,745 cancellations
5. Cancellation Rate:While the absolute number of cancellations has increased, the cancellation rate (canceled bookings as a percentage of total bookings) might provide further insights into the changing trend. The cancellation rate is calculated by dividing the number of canceled bookings by the total bookings and multiplying by 100.

Inferences:

1. The significant increase in total bookings from 2015 to 2016 suggests a period of substantial growth in demand for hotel accommodations.
2. The notable decrease in total bookings from 2016 to 2017 indicates a change in the booking landscape, possibly influenced by external factors such as economic conditions, industry trends, or changes in customer behavior.
3. The steady increase in canceled bookings highlights the importance of managing cancellation policies effectively and understanding the factors influencing guests' decisions to cancel.
4. Analyzing the cancellation rate can provide more nuanced insights into the proportion of bookings that are canceled, helping hotels refine their strategies for reducing cancellations and optimizing revenue.
5. The year-over-year changes in bookings and cancellations point to the dynamic nature of the hospitality industry, emphasizing the need for adaptability and strategic planning to navigate evolving trends and customer preferences.

5.Understand the distribution of the number of adults, children, and babies and identify any outliers.



Identified outliers are			
adults	children	babies	SUM
55	0	0	55
50	0	0	50
40	0	0	40
27	0	0	27
27	0	0	27
26	0	0	26
26	0	0	26
26	0	0	26
26	0	0	26
26	0	0	26
20	0	0	20
20	0	0	20
2	10	0	12
2	0	10	12
10	0	0	10
1	0	9	10

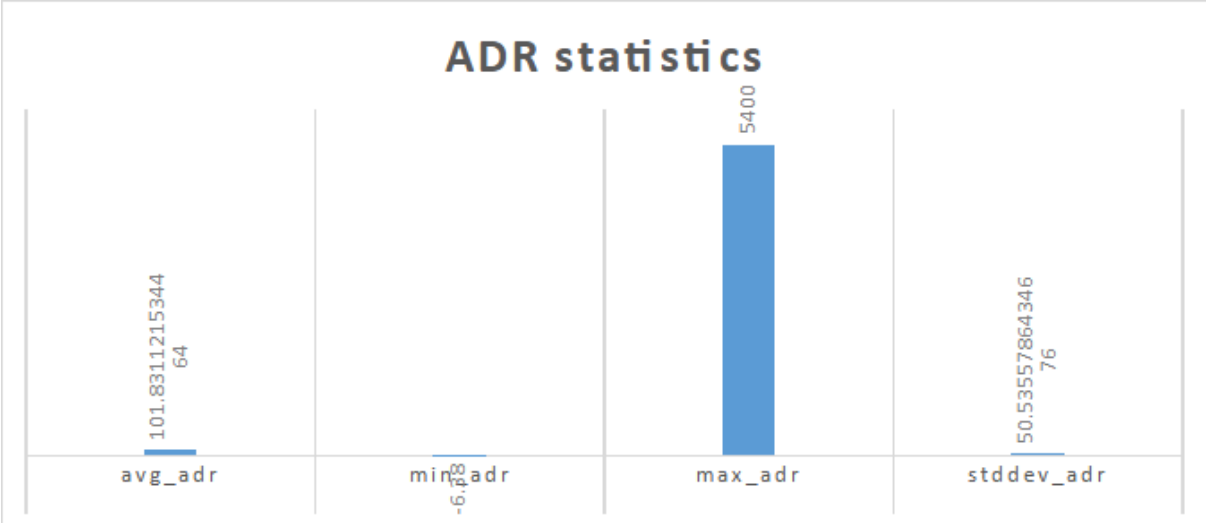
Observations:

1. The dataset contains instances where only adults are present, with counts ranging from 55 to 10.
2. There are cases where children or babies are present without adults, indicating potential data anomalies or errors.
3. Some records show the presence of both children and babies, contributing to the total counts.

Implications:

1. The identified outliers suggest potential data entry errors or anomalies that may need further investigation and correction.
2. Presence of only children or babies without adults may indicate incomplete or inaccurate guest information.
3. Addressing these outliers is crucial for maintaining data accuracy and ensuring that guest demographics are properly represented.

6.Calculate summary statistics for ADR and explore differences between Resort Hotel and City Hotel bookings.



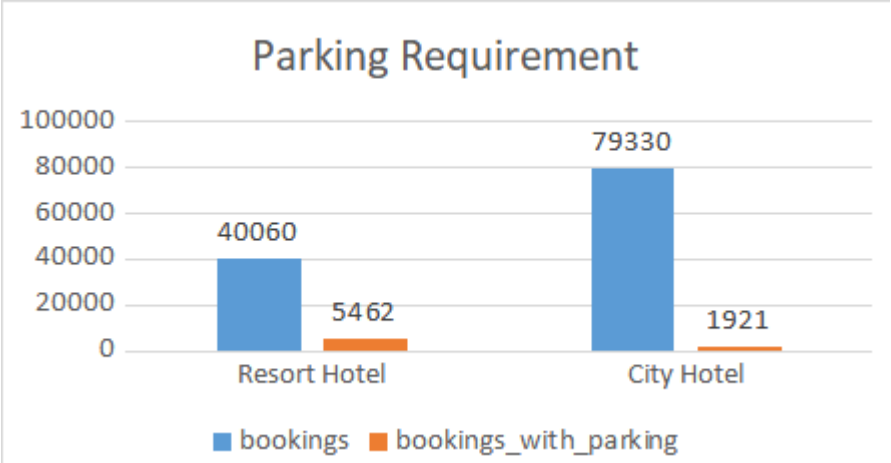
avg_adr	min_adr	max_adr	stddev_adr
101.831121534464	-6.38	5400	50.5355786434676

hotel	total_bookings	avg_adr	avg_nights	avg_guests
Resort Hotel	39760	94.953	4.3208	2.01
City Hotel	78735	105.307	2.9777	1.94

Insights:

1. The lower average ADR for Resort Hotels (94.95) compared to City Hotels (105.31) indicates potential pricing distinctions. Resort Hotels may adopt a strategy of offering more affordable accommodations, possibly targeting longer-stay guests.
2. Resort Hotels, with an average stay of 4.32 nights, cater to guests seeking extended vacations or resort-style experiences. In contrast, City Hotels, with an average stay of 2.98 nights, may attract guests looking for shorter stays, possibly related to business or city exploration.
3. Despite the higher ADR in City Hotels, Resort Hotels have a slightly higher average number of guests per booking (2.01 vs. 1.95). This suggests that Resort Hotels may focus on accommodating families or larger groups, influencing their pricing and marketing strategies.
4. The presence of negative ADR values requires careful investigation, as negative rates are typically illogical. Rectifying these anomalies is crucial for maintaining data integrity and ensuring the reliability of ADR metrics.

7.Analyze the distribution of required car parking spaces for each hotel type and determine if one type attracts more guests with cars.



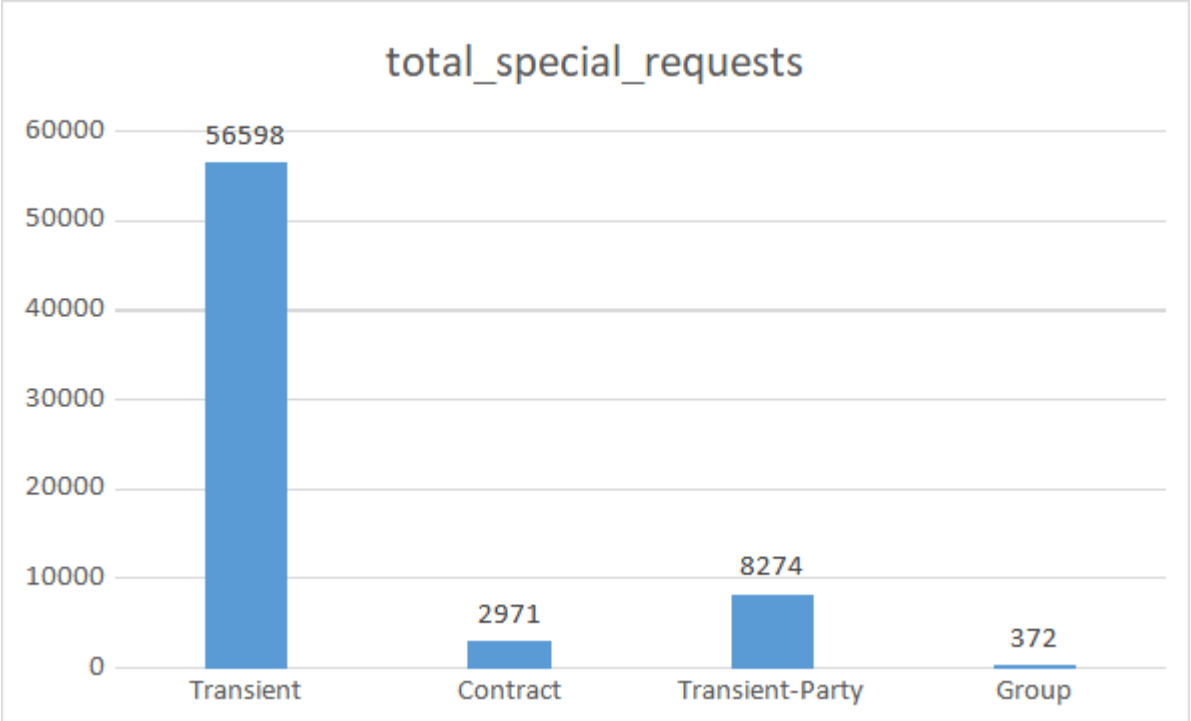
hotel	bookings	bookings_with_parking	bookings_without_parking	ratio_of_parking_required
Resort Hotel	**40060**	**5462**	**34570**	**0.1363**
City Hotel	**79330**	**1921**	**77404**	**0.0242**

1. The data reveals that both Resort and City Hotels offer car parking spaces for their guests, with varying ratios of bookings with parking.
2. Resort Hotels have a total of 40,060 bookings, of which 5,462 (13.63%) include parking spaces. The majority of guests at Resort Hotels do not require parking, possibly indicating a higher percentage of guests arriving by other means of transportation.
3. City Hotels, with 79,330 total bookings, have a much lower ratio of bookings with parking at 2.42%. This suggests that a smaller percentage of guests at City Hotels

opt for parking spaces, which could be attributed to the urban location where alternative transportation methods may be more common.

4. The significant difference in the ratio of parking-required bookings between Resort and City Hotels suggests that guests' parking preferences vary based on the type and location of the accommodation.

8.Compare the total number of special requests made by different customer types (e.g., Transient, Group) and identify which customer type makes more requests.

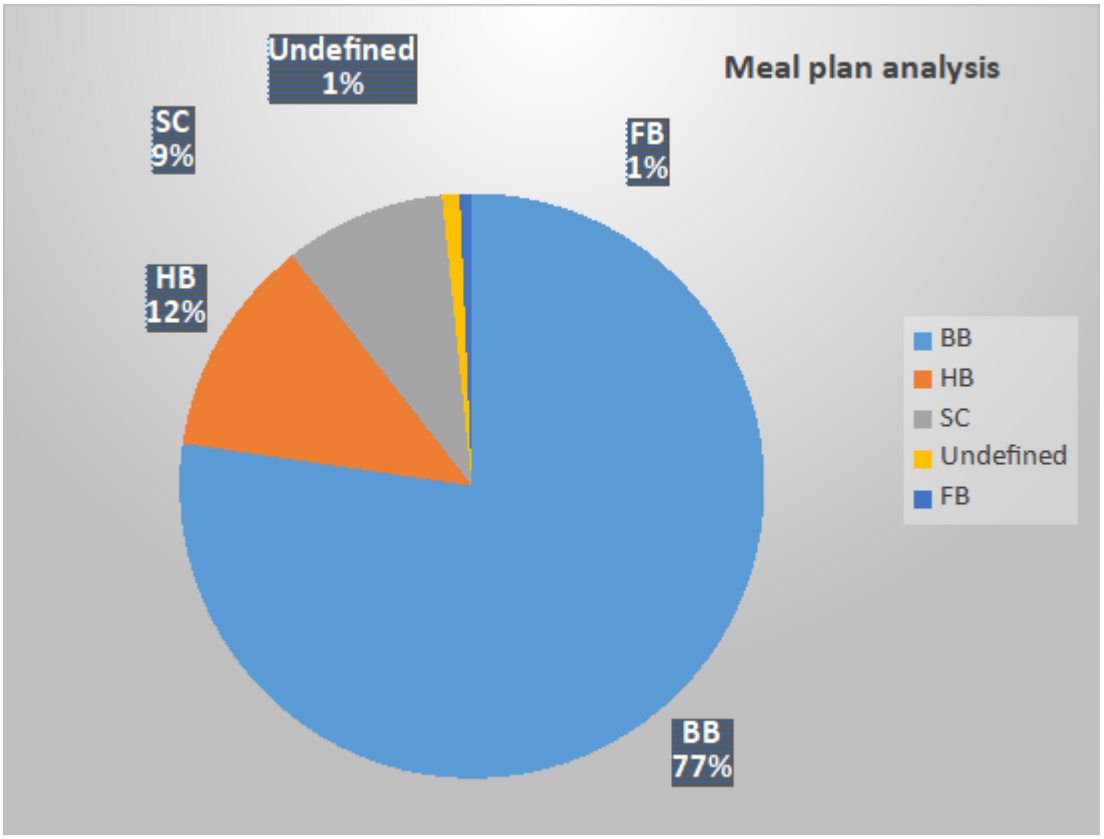


customer_type	total_special_requests
Transient	**56598**
Contract	**2971**
Transient-Party	**8274**
Group	**372**

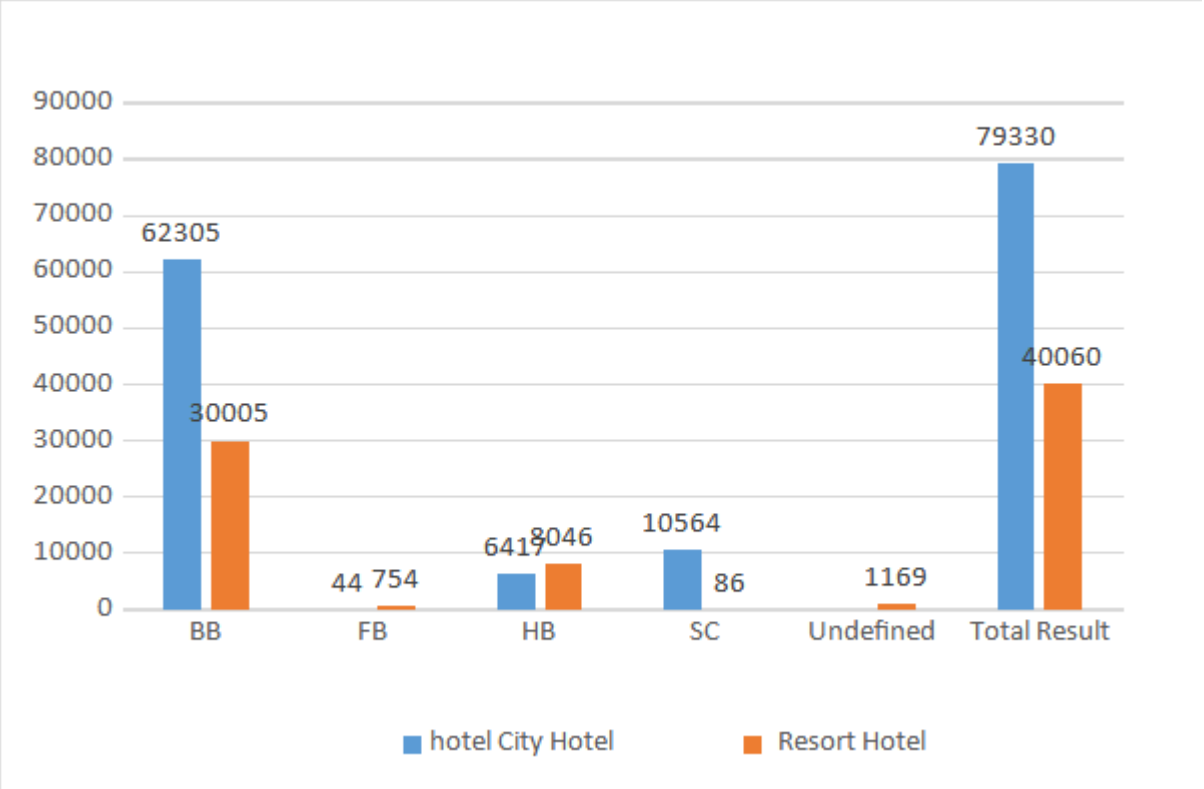
1. Transient customers stand out with the highest total special requests, indicating a substantial demand for personalized services and amenities. This customer segment may include individual travelers or small groups with diverse needs.
2. The Transient-Party segment also shows a notable number of special requests, suggesting that even when part of a larger group, these guests have specific preferences or requirements.
3. Contract customers have a comparatively lower number of special requests, potentially indicating a more standardized or predefined arrangement in their contracts, resulting in fewer individualized requests.

4. Group customers, with the smallest total special requests, may prefer a more streamlined experience or have fewer individual preferences compared to other customer types.
5. Hotel management can optimize resource allocation based on the specific needs of different customer types. This includes staffing, amenities, and services tailored to meet the expectations of each segment.

9.Understand the distribution of meal plans (e.g., BB, HB, FB, SC) and identify any patterns or preferences.



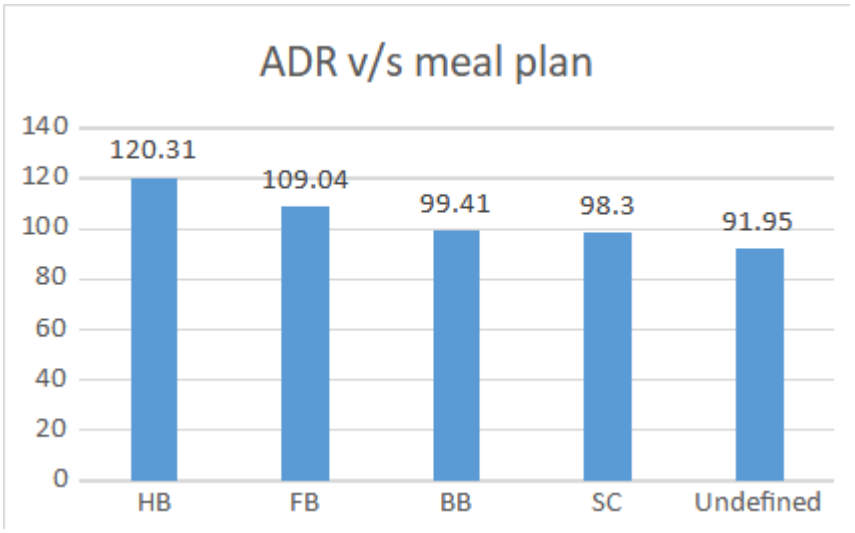
meal	total_bookings
BB	92310
HB	14463
SC	10650
Undefined	1169
FB	798

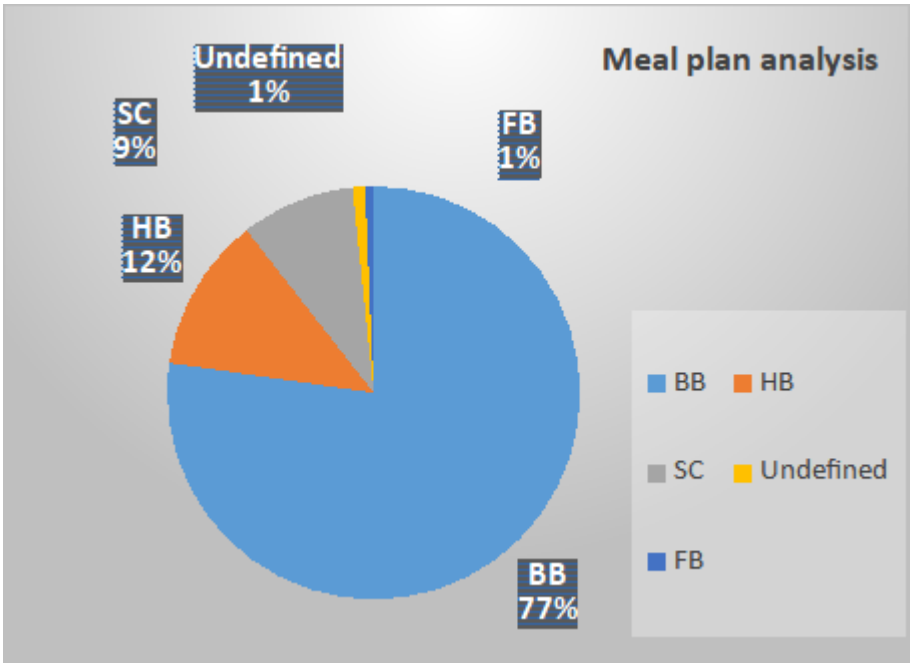


hotel	meal	total_bookings
City Hotel	SC	10564
City Hotel	HB	6417
City Hotel	FB	44
City Hotel	BB	62305
Resort Hotel	Undefined	1169
Resort Hotel	SC	86
Resort Hotel	HB	8046
Resort Hotel	FB	754
Resort Hotel	BB	30005

Insights : Very less people prefer self catering in a resort hotel. Meal plan which is less preferred in city hotel is full board. Bookings in city hotel opts at least one type of meal plan

10.Analyze Average Daily Rates (ADR) by meal plan type to identify variations in pricing.

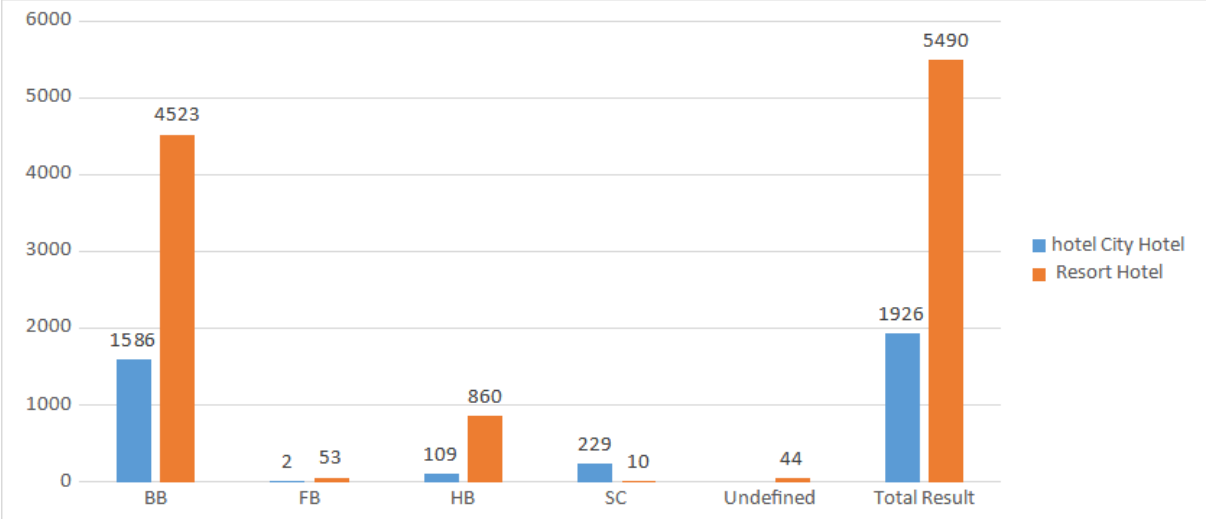




meal	avg_adr
HB	120.31
FB	109.04
BB	99.41
SC	98.3
Undefined	91.95

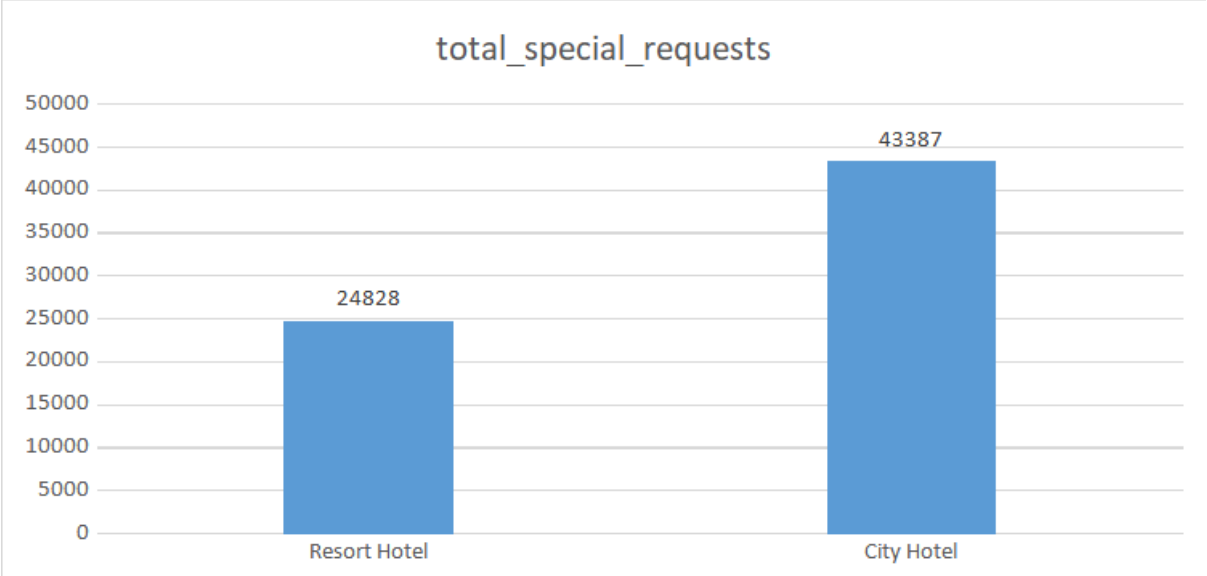
Insights: HB has more average adr, it could suggest that there may be additional services or amenities included in the Half Board package, leading to a higher overall cost for guests.

11. Investigate the distribution of required car parking spaces and special requests by hotel type and meal plan.

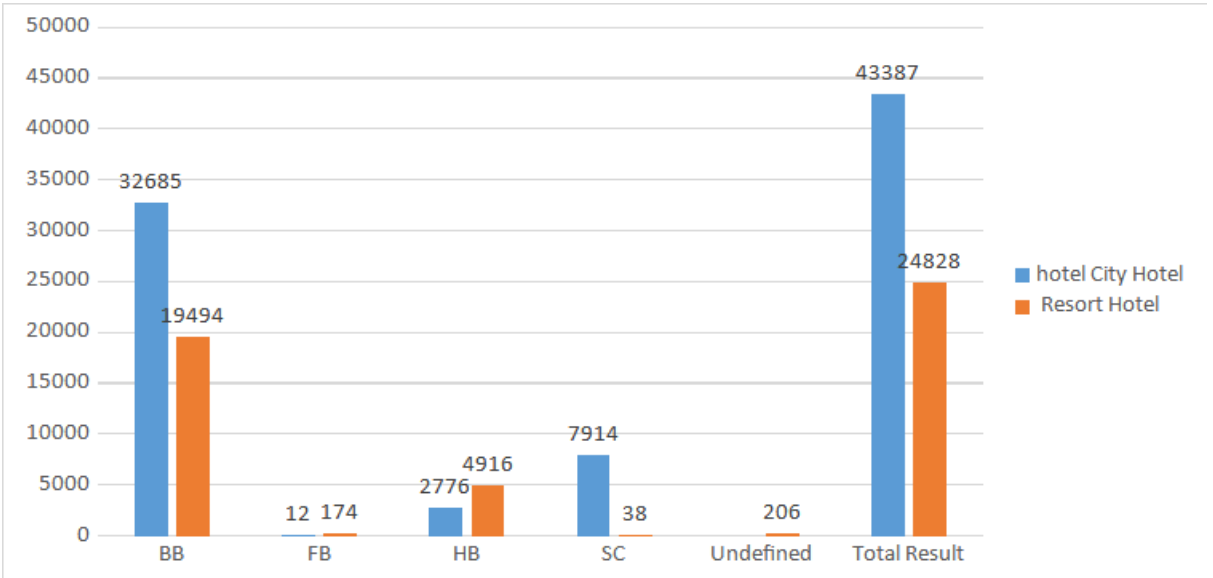


Number of car parkings each hotels based on meal plan

hotel	meal	parking	total_bookings
City Hotel	BB	1586	62305
City Hotel	FB	2	44
City Hotel	HB	109	6417
City Hotel	SC	229	10564
Resort Hotel	BB	4523	30005
Resort Hotel	FB	53	754
Resort Hotel	HB	860	8046
Resort Hotel	SC	10	86
Resort Hotel	Undefined	44	1169



hotel	total_special_requests
Resort Hotel	24828
City Hotel	43387

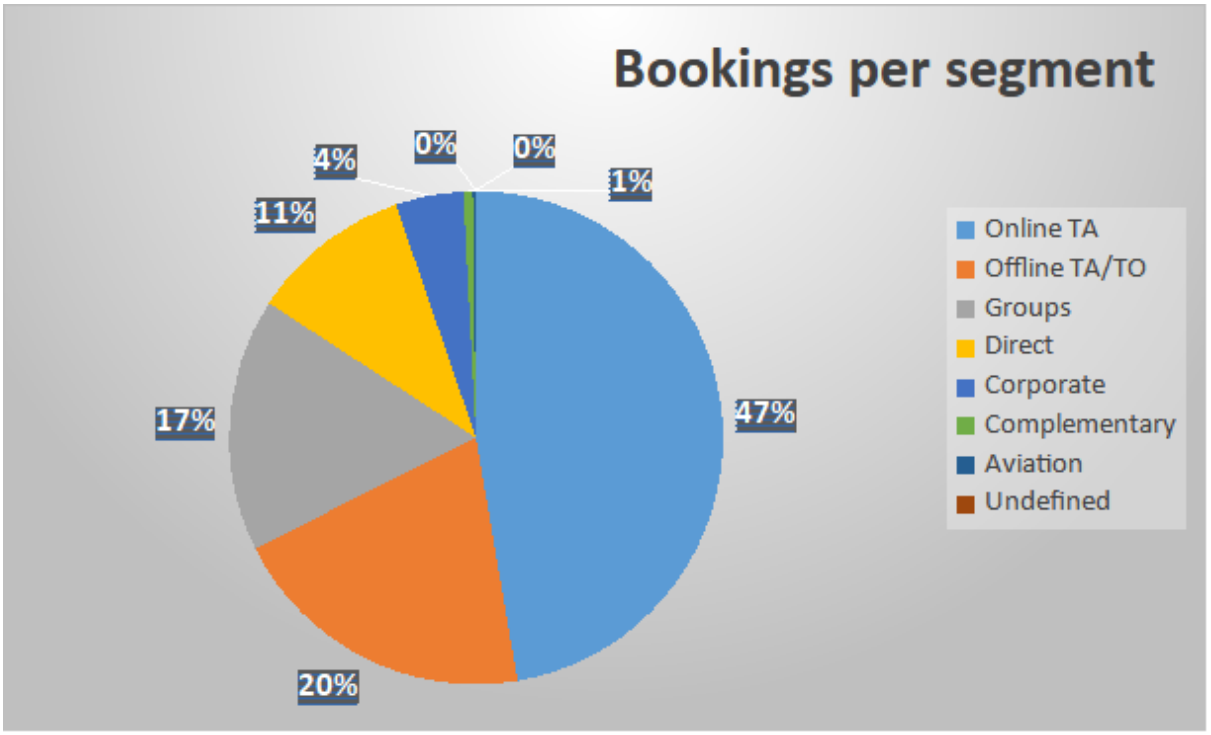


Number of special requests in each hotels based on meal plan

1. Parking Spaces Utilization:City Hotel sees higher parking utilization for Bed & Breakfast (BB) bookings, while Resort Hotel shows significant usage for both BB and Half Board (HB) guests.
2. Total Special Requests:Resort Hotel receives a substantial 24,828 special requests, indicating a high demand for personalized services.
3. City Hotel follows closely with 43,387 total special requests, showcasing a notable interest in additional services.

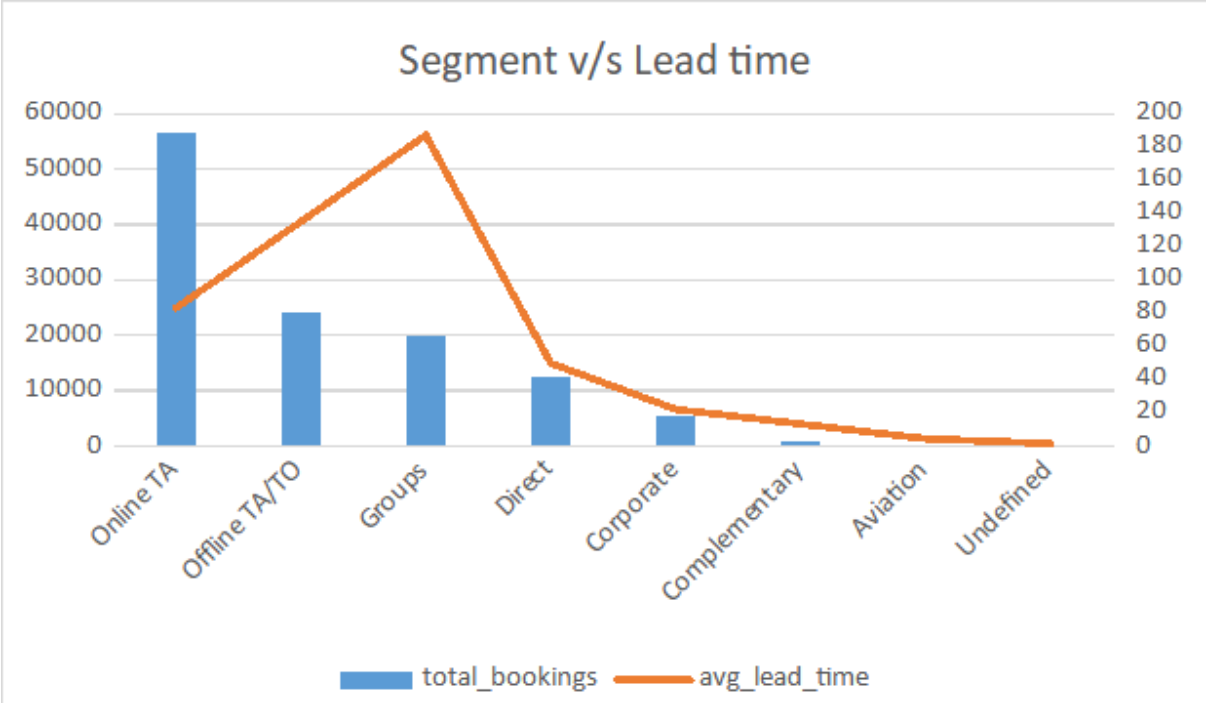
4. Special Requests by Meal Plans - Resort Hotel:BB guests contribute significantly to special requests (19,494), highlighting a preference for personalized services.HB guests also make considerable special requests (4,916), emphasizing interest in additional services beyond accommodation and meals.
5. Special Requests by Meal Plans - City Hotel:BB guests in City Hotels lead in total special requests (32,685), indicating a strong preference for personalized services. SC guests also make a notable number of special requests (7,914), showcasing diverse needs even for those opting for less inclusive meal plans.
6. Undefined Meal Plan in Resort Hotel:The "Undefined" meal plan in Resort Hotel generates special requests (206), indicating the need for customized services even when meal plans are not explicitly chosen.

12.Understand the distribution of bookings across different market segments and calculate summary statistics for lead times within each segment.



market_segment	total_bookings
Online TA	56477
Offline TA/TO	24219
Groups	19811
Direct	12606
Corporate	5295
Complementary	743

Aviation	**237**
Undefined	**2**

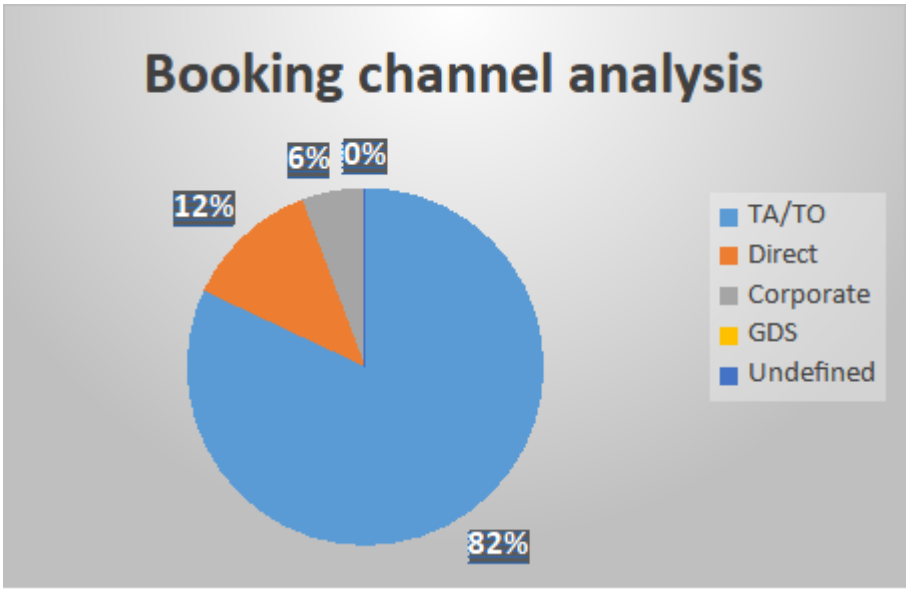


market_segment	total_bookings	avg_lead_time	min_lead_time	max_lead_time
Online TA	56477	82.9987	0	403
Offline TA/TO	24219	135.0045	0	532
Groups	19811	186.9731	0	629
Direct	12606	49.8591	0	737
Corporate	5295	22.1256	0	343
Complementary	743	13.2867	0	386
Aviation	237	4.443	0	23
Undefined	2	1.5	1	2

1. Online Travel Agencies (Online TA) contribute the highest total bookings (56,477), indicating a significant reliance on online platforms for reservations.
2. Offline Travel Agencies/Tour Operators (Offline TA/TO) follow with 24,219 total bookings, showcasing the importance of traditional offline channels.

3. The Groups segment represents 19,811 total bookings, likely associated with group reservations, emphasizing the significance of accommodating larger groups.
4. Direct bookings account for 12,606 total bookings, reflecting guests who prefer booking directly with the hotels, highlighting the importance of a seamless direct booking experience.
5. Corporate bookings, with 5,295 total bookings, indicate business-related reservations, emphasizing the need for specialized services and corporate partnership strategies.
6. The Complementary segment, with 743 total bookings, suggests complimentary stays, potentially linked to loyalty programs or partnerships.
7. Aviation-related bookings amount to 237, reflecting reservations associated with the aviation sector.
8. The Undefined segment has only 2 total bookings, requiring further investigation for clarity and categorization.
9. The average lead time for Group bookings is the longest at 186.9731 days, indicating extended planning periods for group reservations.
10. Online TA bookings have an average lead time of 82.9987 days, showcasing a moderate planning horizon for online travel agency reservations.
11. Corporate bookings have a relatively short average lead time of 22.1256 days, reflecting shorter planning cycles for business-related stays.
12. Direct bookings have an average lead time of 49.8591 days, suggesting a moderate planning horizon for guests booking directly with the hotel.
13. Complementary stays have a short average lead time of 13.2867 days, indicating shorter lead times for complimentary reservations.
14. Aviation-related bookings have the shortest average lead time at 4.443 days, reflecting the industry's need for quick and responsive planning.
15. The Undefined segment, with only 2 total bookings, has an average lead time of 1.5 days, requiring further investigation for clarity and categorization.

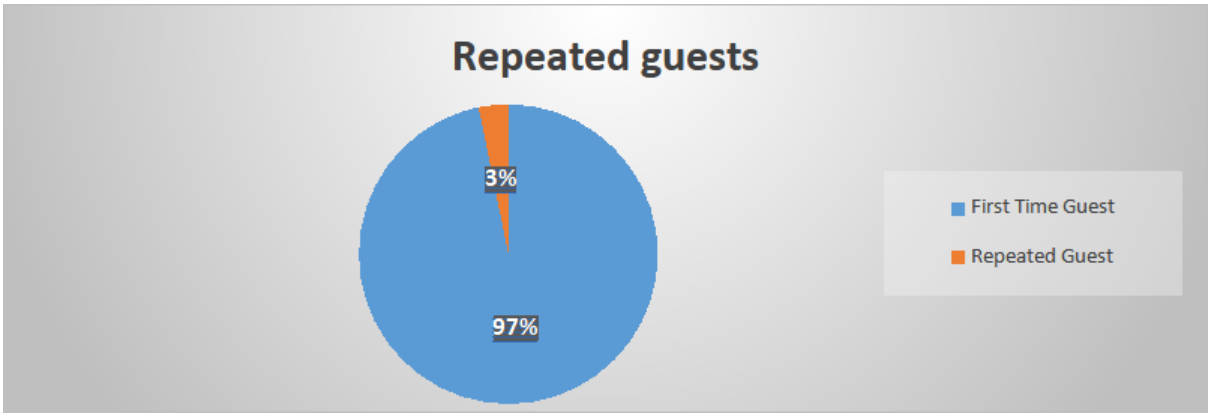
13.Analyze the distribution of bookings through different booking channels (e.g., online travel agents, direct bookings) and calculate the percentage of bookings through each channel.



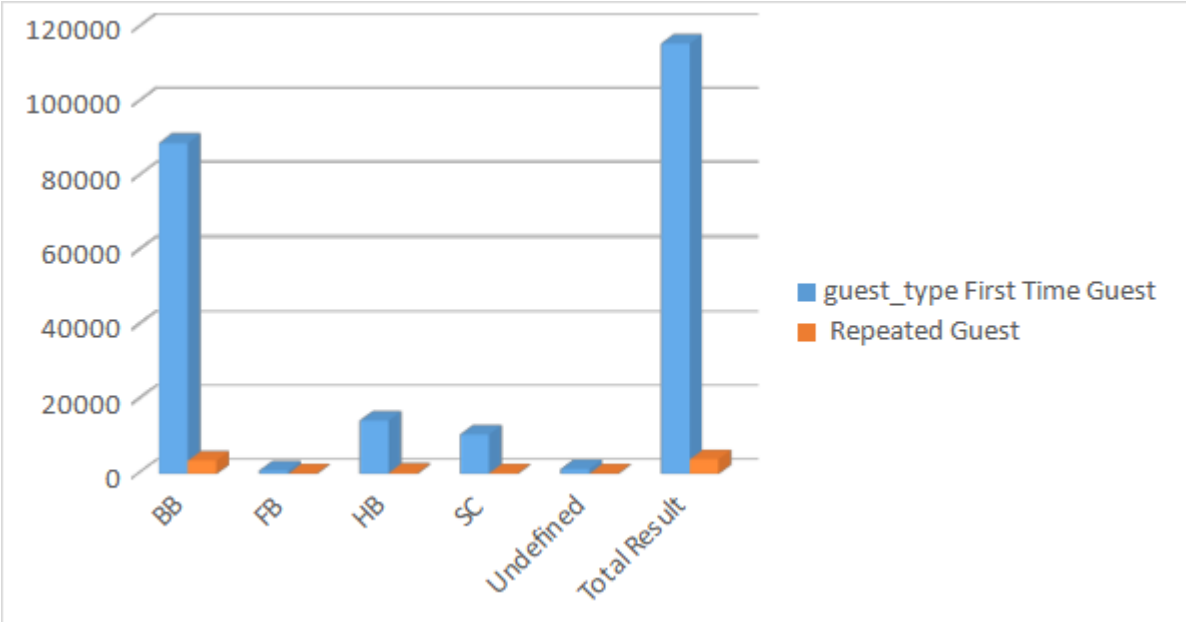
distribution_channel	total_bookings	percentage_distribution
TA/TO	97870	81.975
Direct	14645	12.2665
Corporate	6677	5.5926
GDS	193	0.1617
Undefined	5	0.0042

1. Bookings through Travel Agencies/Tour Operators (TA/TO) constitute the majority, with a total of 97,870 bookings, representing 81.975% of the total distribution.
2. Direct bookings contribute significantly, totaling 14,645, and comprise 12.2665% of the overall distribution, showcasing a notable preference for guests booking directly with the hotel.
3. Corporate bookings, with 6,677 total, make up 5.5926% of the distribution, indicating a portion of guests booking through corporate channels.
4. Bookings through Global Distribution Systems (GDS) are limited, with 193 total, constituting only 0.1617% of the distribution.
5. The Undefined channel, with only 5 total bookings, represents a minimal portion at 0.0042% of the overall distribution.
6. The distribution highlights the significance of TA/TO channels as the primary source of bookings, followed by a substantial portion of direct bookings, while corporate and GDS channels play smaller but notable roles. The Undefined channel represents a negligible fraction of the overall distribution.

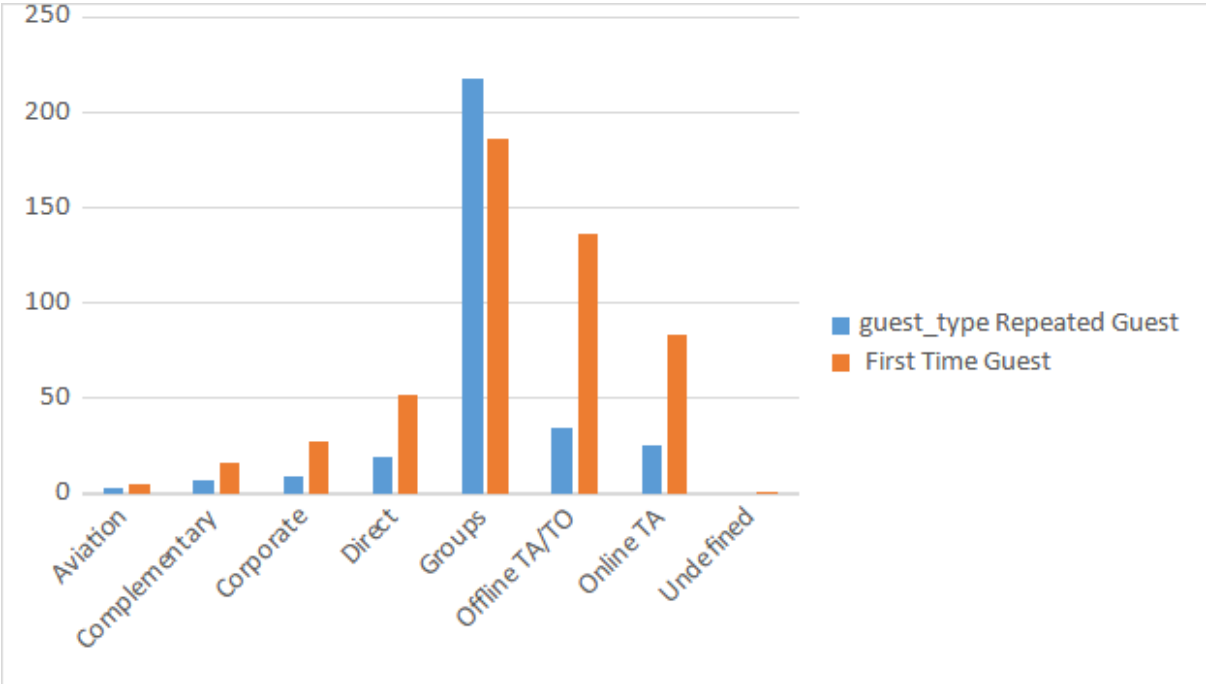
14.Calculate the proportion of repeated guests and investigate their booking behavior. Identify any patterns or differences in preferences compared to firsttime guests



guest_type	guest_count	proportion
First Time Guest	115580	96.8088
Repeated Guest	3810	3.1912



Sum of total_orders	guest_type		
meal	First Time Guest	Repeated Guest	Total Result
BB	88837	3473	92310
FB	789	9	798
HB	14277	186	14463
SC	10540	110	10650
Undefined	1137	32	1169
Total Result	115580	3810	119390



Sum of avg_lead_time	guest_type	
market_segment	Repeated Guest	First Time Guest
Aviation	3.3906	4.8324
Complementary	6.7478	16.2183
Corporate	8.9464	27.2096
Direct	18.9661	51.9413
Groups	217.8185	186.5469
Offline TA/TO	34.3166	136.6052
Online TA	25.8392	83.6094
Undefined		1.5

1. The majority of guests are first-time visitors, constituting 96.81%, while repeated guests make up 3.19%.
2. First-time guests have higher meal order volumes across all categories compared to repeated guests.
3. Repeated guests, particularly in the Undefined and Full Board categories, show lower meal order volumes.
4. Repeated guests exhibit shorter lead times, indicating more spontaneous or frequent bookings.

5. First-time guests prefer online platforms for bookings, while repeated guests show a broader distribution across various channels.
6. Groups significantly impact first-time guests, influencing longer lead times, while repeated group reservations contribute to extended planning periods.
7. Corporate bookings for repeated guests have notably shorter lead times, suggesting quicker decision-making processes for business-related stays.

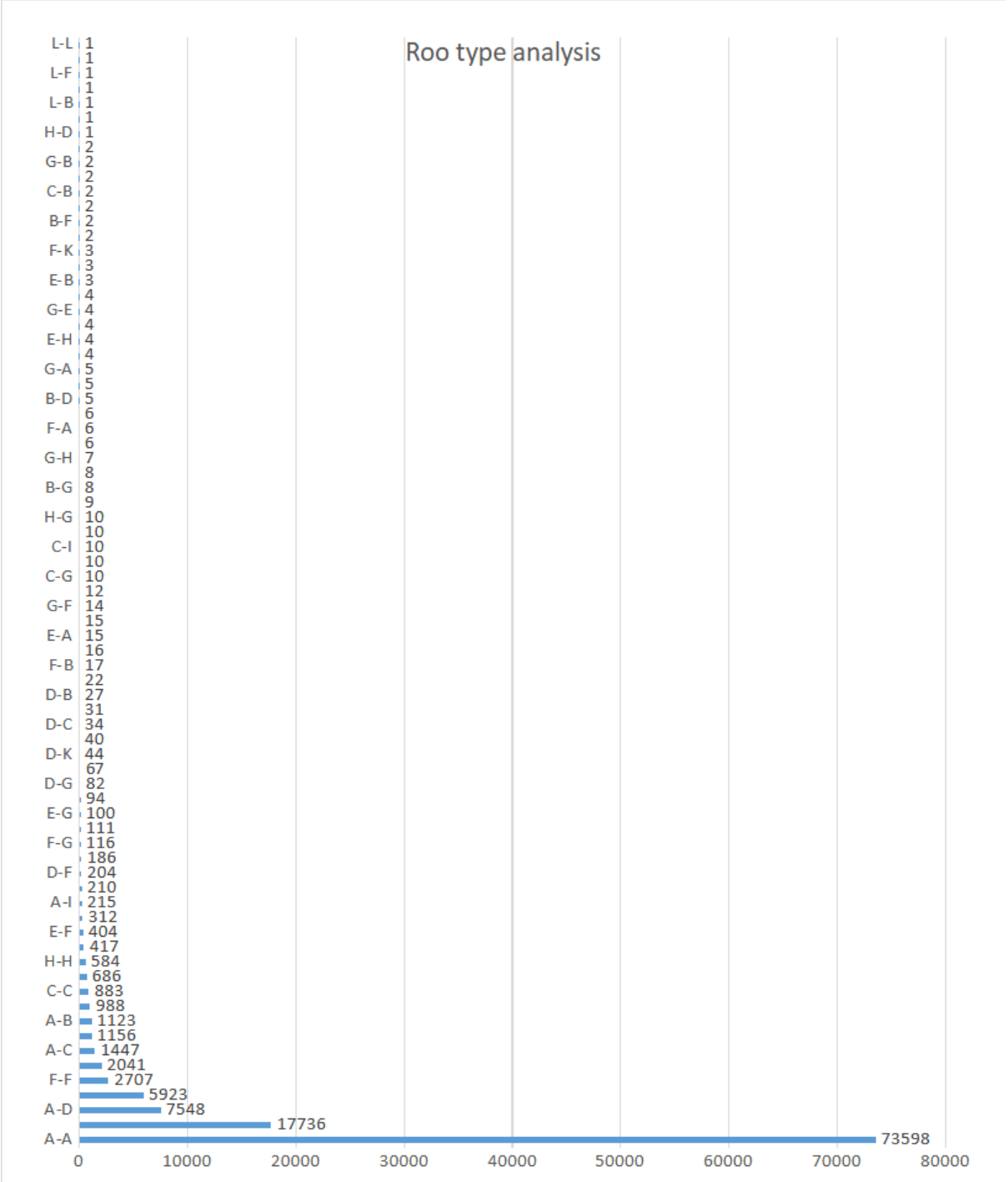
15.Explore the impact of a guest's booking history on their likelihood of canceling a current booking. Calculate cancellation rates based on previous cancellations and noncanceled bookings.

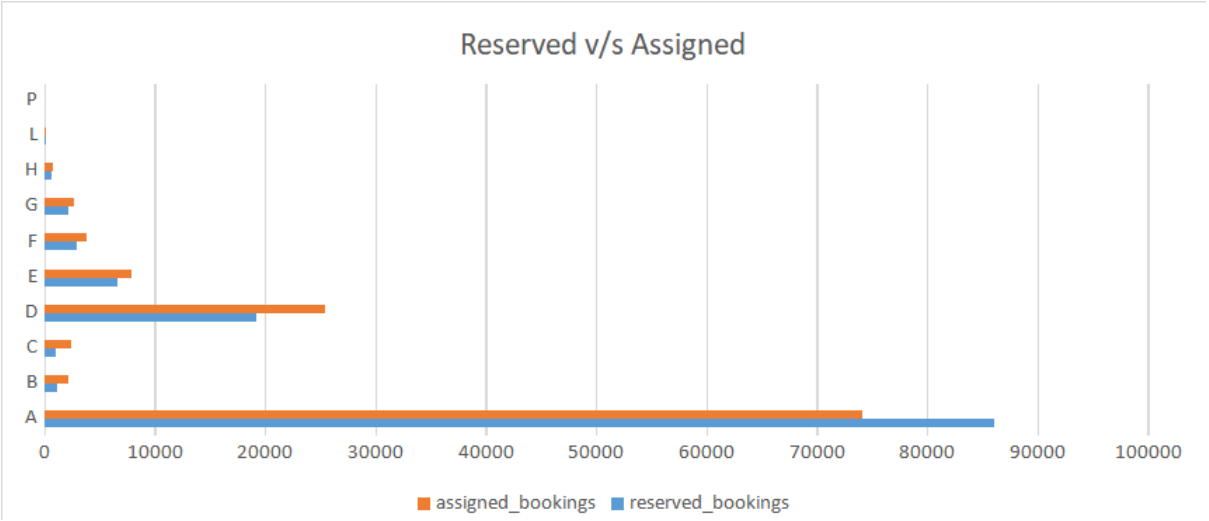
cancellations_with_previous_cancellations	cancellation_rate
5942	91.641

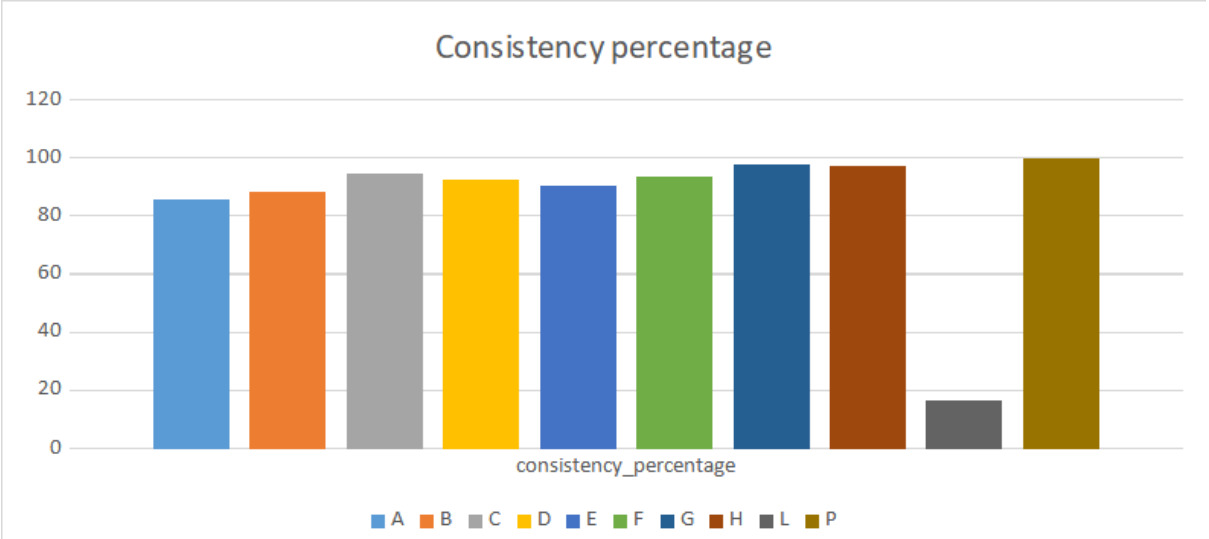
cancellations_with_previous_no_cancellations	cancellation_rate
200	5.5249

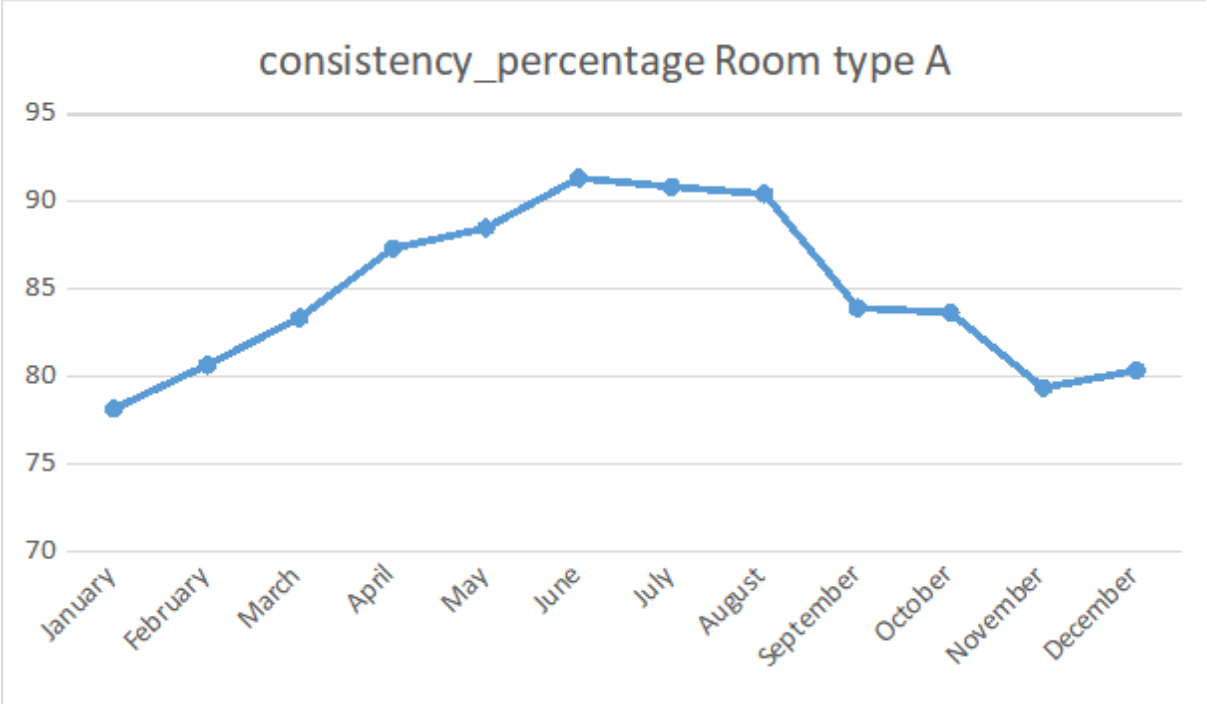
1. Bookings with previous cancellations have a higher cancellation rate of 91.641%, indicating a likelihood of guests canceling a booking if they have a history of cancellations.
2. Bookings with previous non-cancellations have a lower cancellation rate of 5.5249%, suggesting a lower likelihood of guests canceling a booking if they have a history of not canceling previous reservations.

16.Understand the distribution of reserved and assigned room types. Calculate summary statistics for the consistency between reserved and assigned room types.



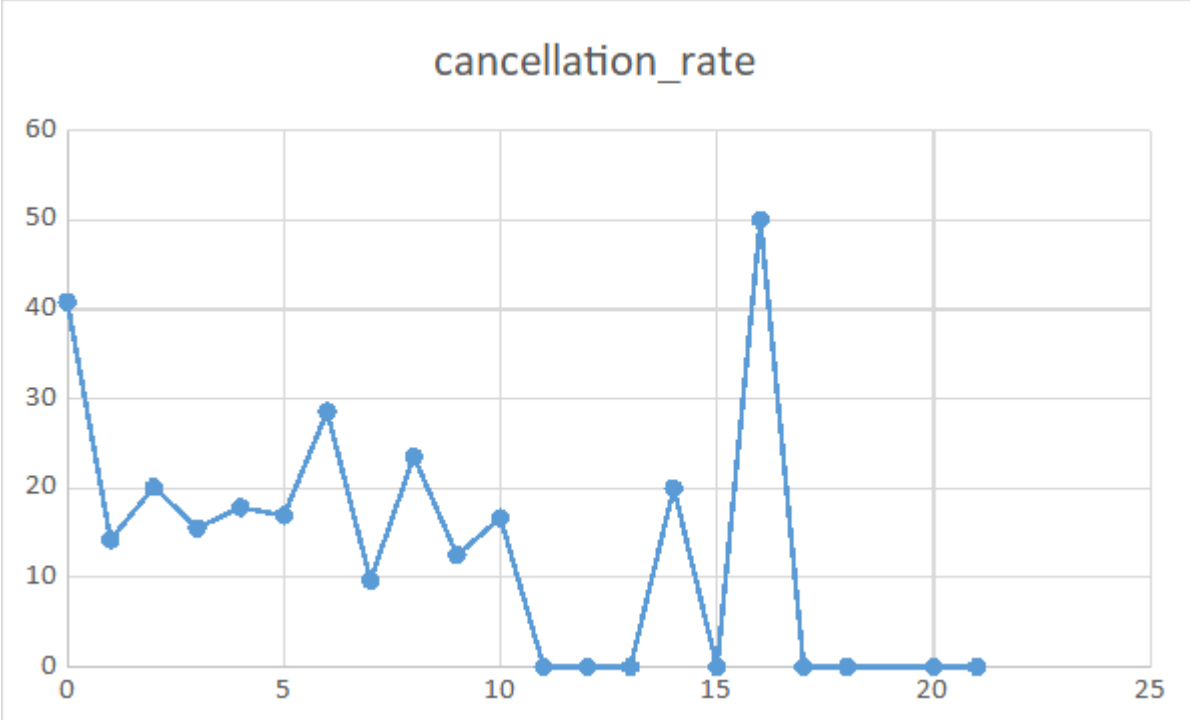

Reserved v/s Assigned
P
L
H
G
F
E
D
C
B
A
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
assigned_bookings
reserved_bookings


Consistency percentage
120
100
80
60
40
20
0
consistency_percentage
A
B
C
D
E
F
G
H
L
P



1.
2. Across all room types, there is a general consistency of 87.51% between reserved and assigned room types, indicating a relatively high match rate.
3. Room types G (97.47%) and H (97.17%) exhibit the highest consistency percentages, suggesting a strong alignment between reserved and assigned room types.
4. Room types C (94.74%) and F (93.44%) also show high consistency percentages, indicating reliable matching between reservations and assigned room types.
5. Room types A (85.59%), B (88.37%), D (92.37%), and E (90.64%) demonstrate varying consistency percentages, with A having a relatively lower rate.
6. Room types L (16.67%) and P (100%) exhibit unique scenarios; L has a notably low consistency percentage, while P shows a perfect match between reserved and assigned room types.
7. Analyzing consistency by month for room type A, January (78.17%) has the lowest percentage, while August (90.45%) and July (90.86%) have relatively higher percentages.
8. Insights:
9.
10. The overall consistency between reserved and assigned room types is high, indicating effective management of room allocations.
11. Room types G and H consistently match at a high rate, suggesting a reliable assignment process for these room types.
12. Room types C and F also exhibit strong consistency, contributing to positive guest experiences.
13. Room type A shows some variability in consistency across months, with January having the lowest and August having the highest percentages.
14. Room types L and P have unique situations: L shows a low consistency, while P demonstrates a perfect match, possibly due to specific policies or limitations for these room types.

17.Analyze the impact of booking changes on cancellation rates. Calculate cancellation rates for bookings with different numbers of changes.

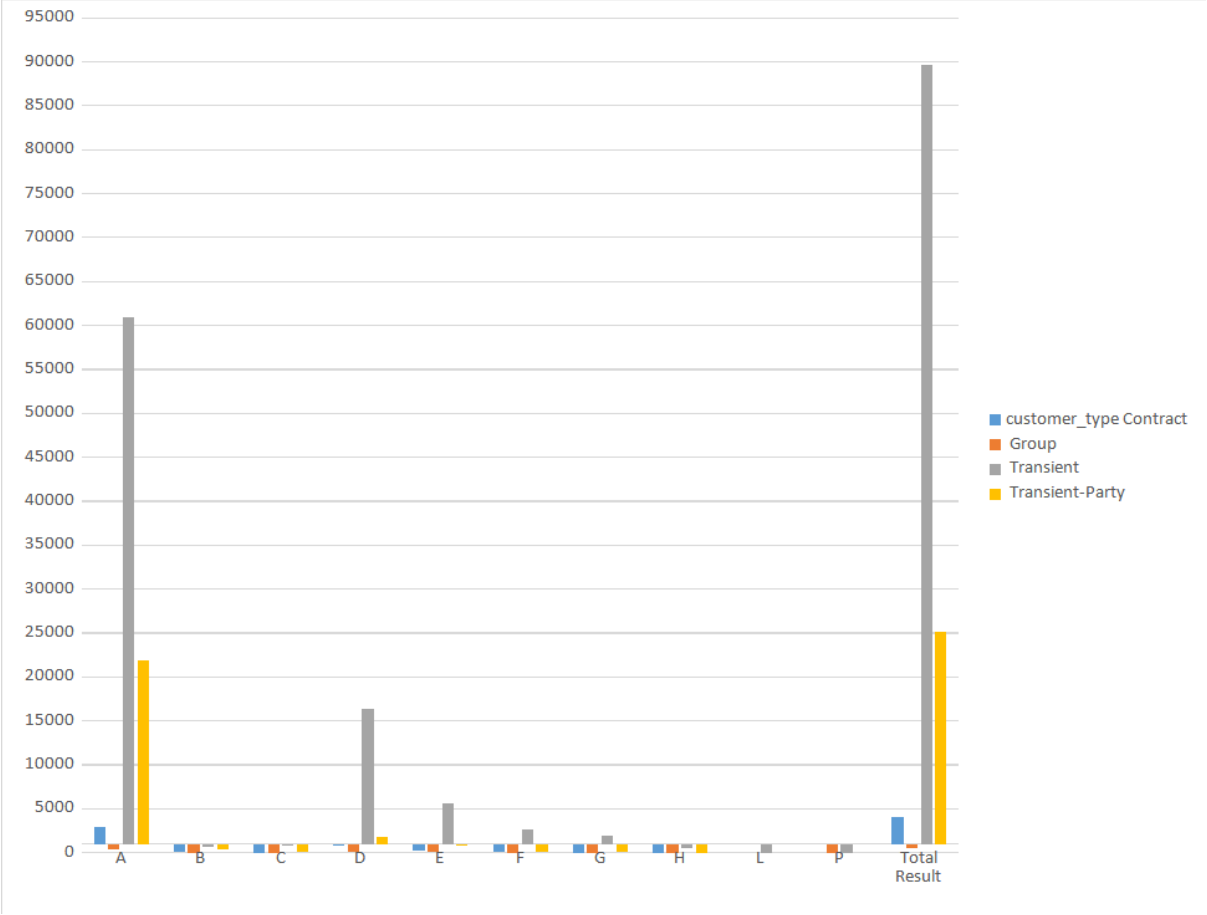


booking_changes	total_bookings	canceled_bookings	cancellation_rate
0	101314	41391	40.8542
1	12701	1807	14.2272
2	3805	766	20.1314
3	927	144	15.534
4	376	67	17.8191
5	118	20	16.9492
6	63	18	28.5714
7	31	3	9.6774
8	17	4	23.5294
9	8	1	12.5
10	6	1	16.6667
11	2	0	0

12	2	0	0
13	5	0	0
14	5	1	20
15	3	0	0
16	2	1	50
17	2	0	0
18	1	0	0
20	1	0	0
21	1	0	0

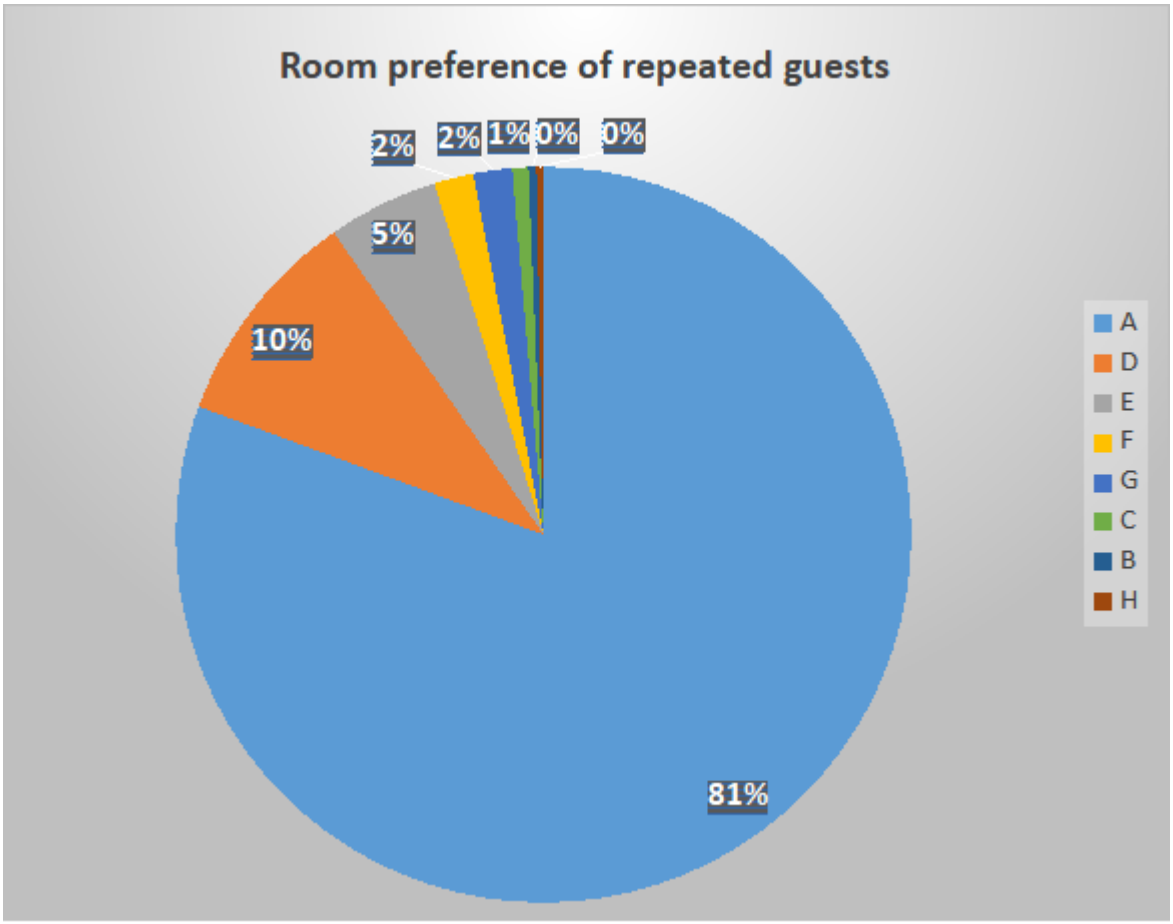
1. Bookings with 0 changes have a cancellation rate of 40.8542%, suggesting a higher likelihood of cancellation for reservations without any changes.
2. As the number of booking changes increases, the cancellation rate generally decreases. For instance, bookings with 1 change have a lower cancellation rate of 14.2272% compared to those with 0 changes.
3. The trend continues, and bookings with 7 changes or more exhibit fluctuating cancellation rates, with the highest rate at 28.5714% for 6 changes.
4. Bookings with 11, 12, 13, 15, 17, 18, 20, and 21 changes show 0% cancellation rates, indicating a potential correlation between a higher number of changes and a lower likelihood of cancellation.
5. However, it's important to note that the sample sizes for bookings with a high number of changes are relatively small, influencing the stability of the cancellation rate percentages.
6. Bookings with no changes have a higher cancellation rate, emphasizing the impact of stability in reservation details on the likelihood of cancellation.
7. As the number of changes increases, the cancellation rate generally decreases, suggesting that guests who make modifications to their bookings may be less likely to cancel.
8. Bookings with a very high number of changes (11 or more) show 0% cancellation rates, but the limited sample size for such cases requires cautious interpretation due to the smaller number of instances.

18.Explore how room type preferences vary across different customer types (e.g., Transient, Group). Identify if certain customer types have specific room preferences.



1. Transient guests show a consistent preference for various room types, especially for A, B, C, D, E, F, G, H, and P.
2. Group guests have a relatively lower impact on room type preferences, with fewer bookings across all room types.
3. The distribution of room types among different customer types provides insights into the varied preferences and needs of guests based on their booking behavior and purpose of stay.

19.Examine whether guests who make multiple bookings have consistent room type preferences or if their preferences change over time.



1. First-time guests tend to have consistent preferences, especially for room types A, D, E, F, and G.
2. Repeated guests also exhibit preferences, with room type A being the most favored, but the overall booking numbers are lower compared to first-time guests.
3. Understanding room type preferences among different guest types can help tailor accommodation offerings and enhance the overall guest experience.
4.