HOTEL BOOKING ANALYSIS

-**Documented by Akhila. V**

Project Overview

Objective

The objective of this project is to analyse booking patterns, guest preferences, and factors influencing cancellations in the hotel industry. Using SQL and Excel, we will identify trends in booking sources, revenue, and other key metrics. Additionally, we aim to develop a Power BI dashboard for tracking booking trends and optimizing hotel operations.

Significance

Understanding and optimizing booking patterns, guest preferences, and cancellation factors are crucial for the hotel industry. This analysis can help hotel managers and staff make informed decisions to improve guest satisfaction, revenue, and operational efficiency. Key areas of significance include:

1. **Revenue Optimization:** Knowing which types of rooms, meal plans, and special requests bring in the most money can lead to targeted marketing and pricing strategies.

2. **Cancellation Management:** Analyzing factors that influence cancellations, such as lead time and deposit type, can assist in reducing revenue losses and increasing room utilization.

3. **Improved Guest Experience:** Knowing your guests' preferences for meals, parking, and room types can lead to a better overall experience, potentially leading to positive reviews and repeat business.

4. **Operational Efficiency:** Monitoring booking trends and sources can help with staffing and resource allocation. It can also help with effective waiting list management.

Data Dictionary

To effectively analyse and gain insights from the available data, we have several key tables, each with its own set of attributes:

Booking_Details

- Booking Identifier: A unique identifier for each hotel reservation.
- Hotel Type: Indicates whether the hotel is a Resort Hotel or City Hotel.
- Cancellation Status: 0 for not cancelled, 1 for cancelled.
- Lead Time: The number of days between booking and arrival.
- Year, Month, Week Number, Day of Month: Arrival date details.
- Number of Weekend Nights: Number of nights the guest will stay over the weekend.

- Number of Weekday Nights: Number of nights the guest will stay on weekdays

Guest_Info

- Booking Identifier: Links to Booking_Details.
- Adults, Children, Babies: The number of each type of guest accompanying the booking.

Meal_And_Stay_Details

- Booking Identifier: Links to Booking_Details.
- Meal Type: Specifies the type of meal booked (e.g., Bed & Breakfast, Half Board).
- Average Daily Rate (ADR): The average daily rate for the stay.
- Number of Car Parking Spaces: The number of parking spaces requested.
- Total Special Requests: The total count of special requests made by the guest.

Booking_Source_and_History

- Booking Identifier: Links to Booking_Details.
- Market Segment: The market segment from which the booking originates (e.g., Online Travel Agents, Direct Booking).
- Distribution Channel: The distribution channel used for the booking (e.g., Online Travel Agents, Direct Booking).
- Repeated Visitor: 0 for not repeated, 1 for repeated.
- Previous Booking Cancellations: The number of previous booking cancellations.
- Previous Bookings Not Cancelled: The number of previous bookings that were not cancelled.
- Deposit Type: The type of deposit (e.g., No Deposit, Non-Refund).
- Booking Agent's ID: Identifier for the booking agent.
- Company's ID: Identifier for the company associated with the booking.
- Days on Waiting List: The number of days a booking spent on the waiting list.
- Customer Type: The type of customer (e.g., Transient, Group).

Room_Details

- Booking Identifier: Links to Booking_Details.
- Initial Room Type: The type of room initially reserved.
- Assigned Room Type: The type of room eventually assigned.
- Number of Booking Changes: The number of changes made to the booking.

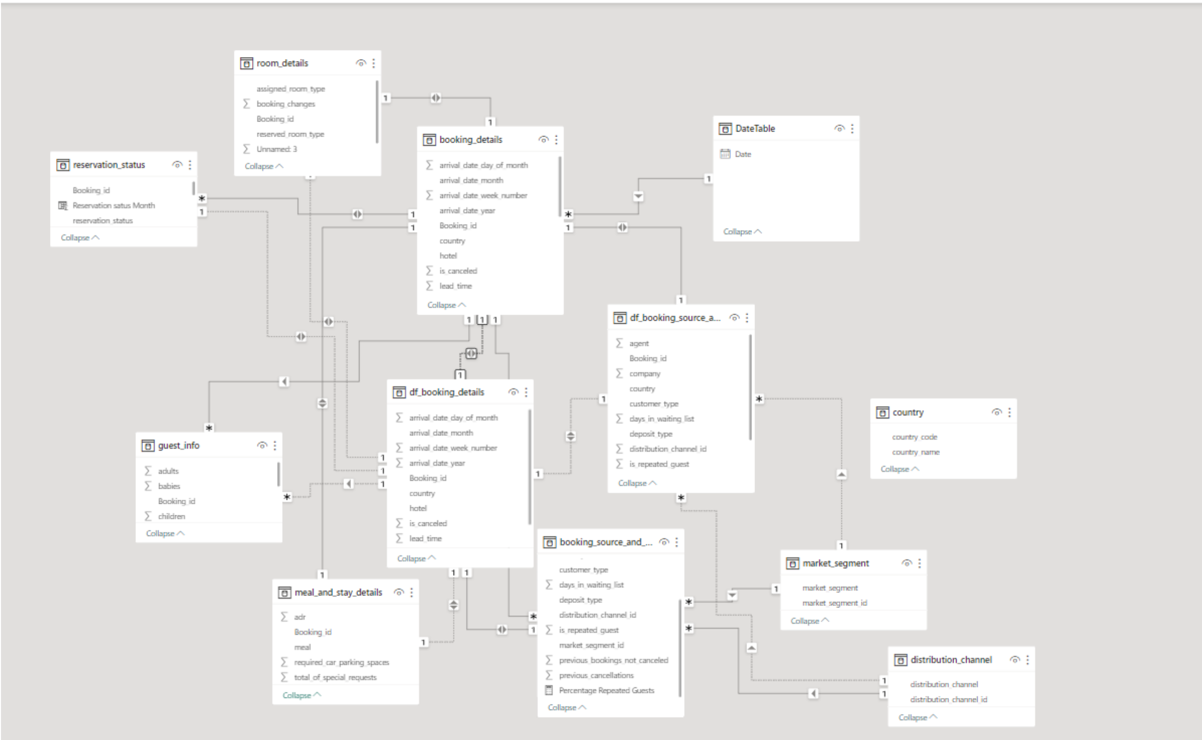
Reservation_Status

- Booking Identifier: Links to Booking_Details.
- Reservation Status: The reservation's last status (e.g., Cancelled, Check-Out).
- Date Recorded: The date on which this status was recorded.

By analyzing these tables and their attributes, we can gain insights into guest preferences, booking sources, cancellation patterns, revenue generation, and operational trends, which will be used to create a Power BI dashboard to visualize and track these insights effectively.

ER Diagram

Creating an Entity-Relationship (ER) diagram for the project involves representing the tables and their relationships visually. Here's a simplified ER diagram for the project:



An Entity-Relationship (ER) diagram is a powerful tool in database design and management, and it serves several important uses in the context of the project:

Visual Representation: ER diagrams provide a visual representation of the structure and relationships between data tables in a database. They help stakeholders understand the database design at a glance.

Database Design: ER diagrams are used during the design phase of a database project to define the schema, including tables, attributes, primary keys, and relationships between entities. They guide the creation of the database structure etc.,

In this diagram:

Each table is represented as a rectangular box.
Primary keys (PK) are indicated within each table.
Foreign keys (FK) are shown where relationships between tables exist, connecting them to the primary key of related tables.

The relationships between tables are represented by lines connecting the primary key in one table to the foreign key in another, indicating how the data is related across different tables.

This simplified ER diagram illustrates the main tables and their relationships in this project.

Steps to Connect to Data

Here are the few Steps I performed before connecting the data

Data Collection and Preparation:

Data Sources

The project's data was collected from the AccioJob Git repository. Two versions of the data were obtained: CSV and SQL. These datasets provided essential information related to hotel bookings, guest details, meal preferences, booking sources, room details, and reservation status.

Data Cleaning

Data cleaning was a crucial step to ensure the quality and consistency of the dataset. This process involved:

- Removing duplicates and irrelevant records.
- Handling missing values by imputing or eliminating them.
- Addressing data mismatches and inconsistencies.
- Validating data types and formats to conform to project requirements.

Data Transformation

After data cleaning, data transformation was carried out to enhance the dataset for analysis and reporting purposes. Key transformations included:

- Adding calculated attributes and measures to facilitate analysis.
- Merging and joining data from different sources to create a comprehensive dataset.
- Standardizing data units and formats for consistency.
- Creating a clean and structured dataset suitable for analytical purposes.

Database Setup and Integration

Database Management System

To facilitate data analysis and visualization, SQL workbench and a MariaDB instance was installed to serve as the relational database management system (RDBMS) for storing the cleaned and transformed data.

Integration with Power BI

The MariaDB database was connected to Microsoft Power BI, providing a direct link between the data source and the visualization tool. This integration allowed for real-time data access and dynamic reporting capabilities.

The process of data collection, cleaning, and transformation is critical in ensuring the quality and integrity of the dataset. The integration of the MariaDB database with Power BI empowers the project to create an interactive and informative dashboard for tracking booking trends and optimizing hotel operations.

This document section provides a professional summary of the data preparation and integration process, which is essential for maintaining project transparency and understanding the steps taken to work with the data effectively.

Below is the image which has the rough details of the analysis and queries I performed to identify the null values and replacing with proper values in order to exclude missing data:

```sql
USE `hotel_booking`;

/*select count(*) from booking_details where country IS NULL;
select * from booking_details where country IS NULL;
UPDATE booking_details SET country = 'others'
WHERE   country IS NULL;
select * from booking_source_and_history where country = 'others';
UPDATE booking_source_and_history SET country = 'others'
WHERE   country IS NULL;
select * from booking_source_and_history;
UPDATE booking_source_and_history SET agent = 0
WHERE   agent IS NULL;
UPDATE booking_source_and_history SET company = 0
WHERE   company IS NULL;*/
/*select count(*) from booking_source_and_history;
select *from df_booking_details where country is null;
UPDATE df_booking_details SET country = 'others'
WHERE  country IS NULL;*/
/*select *from df_booking_source_and_history where country is null;
UPDATE df_booking_source_and_history SET company = 0
WHERE  company IS NULL;
/*select *from guest_info where children IS NULL;
UPDATE guest_info SET children = 0
WHERE  children IS NULL;
select *from market_segment;
select *from meal_and_stay_details where Booking_id is  null;
select *from room_details where Booking_id is  null;*/
```

As a part of Data Cleaning and Manipulation, following are the analysis that I have done

- Total number of records in dataset are '119390'

- There are 4 columns company, agent, country and children with missing values (In agent, company, children columns which are having null data replaced with ‘0’ value) and (In country column which is having null data replaced with ‘others’ value).
- Additionally, I created certain columns and measures for deeper analysis.

Problem statements with solutions and insights: (EDA)

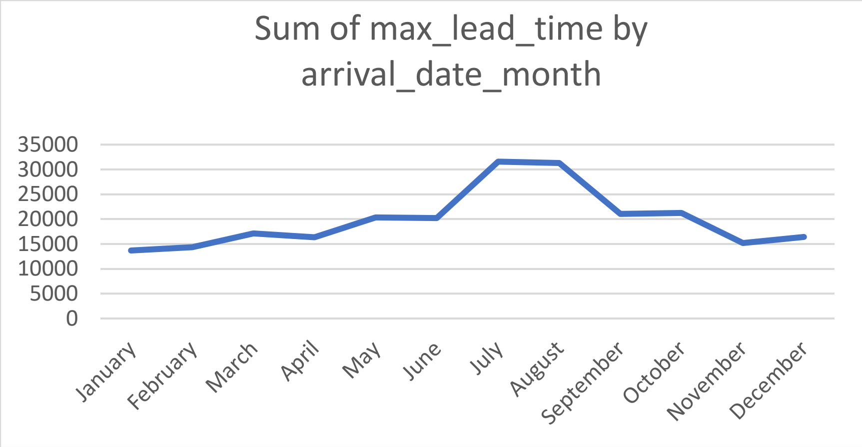
1.Understand the distribution of arrival dates, including the most common arrival days and summary statistics for lead times.

```sql
use hotel_booking;
SELECT
    arrival_date_year,
    arrival_date_month,
    arrival_date_day_of_month,
    COUNT(*) AS arrival_count,
    AVG(lead_time) AS avg_lead_time,
    MAX(lead_time) AS max_lead_time,
    MIN(lead_time) AS min_lead_time,
    STDDEV_POP(lead_time) AS std_deviation_lead_time
FROM
    booking_details
GROUP BY
    arrival_date_year, arrival_date_month, arrival_date_day_of_month
ORDER BY
    arrival_count DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

arrival_date_year	arrival_date_month	arrival_date_day_of_month	arrival_count	avg_lead_time	max_lead_time	min_lead_time	std_deviation_lead_time
2015	December	5	448	80.2031	414	0	106.18622577248813
2016	November	7	366	205.9645	626	0	170.9465494887439
2015	October	16	356	163.2612	364	0	115.84652653762197
2016	October	13	344	123.6337	461	0	151.39074125360997
2015	September	18	340	145.1441	336	0	113.67577653695746
2017	June	8	337	136.8398	332	0	85.24848210281435
2017	March	2	335	70.0358	601	0	138.06188932514158

Output	# arrival_date_year	arrival_date_month	arrival_date_day_of_month	arrival_count	avg_lead_time	max_lead_time	min_lead_time	std_deviation_lead_time
	2015	December	5	23-Mar-01	80.2031	414	0	106.1862258
	2016	November	7	31-Dec-00	205.9645	626	0	170.9465495
	2015	October	16	21-Dec-00	163.2612	364	0	115.8465265
	2016	October	13	09-Dec-00	123.6337	461	0	151.3907413
	2015	September	18	05-Dec-00	145.1441	336	0	113.6757765
	2017	June	8	02-Dec-00	136.8398	332	0	85.2484821
	2017	March	2	30-Nov-00	70.0358	601	0	138.0618893
	2016	October	28	29-Nov-00	156.1557	464	0	115.1983697
	2015	September	17	26-Nov-00	187.7613	335	0	113.7751817
	2017	April	29	25-Nov-00	103.4273	324	0	90.67849891
	2015	August	14	24-Nov-00	101.5471	301	0	111.7061865
	2016	June	17	16-Nov-00	163.081	350	0	130.1793622
	2017	May	15	15-Nov-00	100.3375	365	0	72.62712024
	2017	February	25	11-Nov-00	86.8924	346	0	65.72411513
	2017	May	19	11-Nov-00	127.9494	317	0	60.09383985
	2016	September	15	08-Nov-00	200.0032	433	0	156.2285844
	2017	May	25	02-Nov-00	222.6743	462	0	111.3295637
	2016	September	26	30-Oct-00	210.023	542	0	181.1302675
	2016	March	24	21-Oct-00	79.7254	162	0	50.67546323
	2016	February	12	19-Oct-00	57.7201	230	0	53.76732362
	2016	April	28	18-Oct-00	129.5651	295	0	84.64830068
	2016	October	6	15-Oct-00	222.4118	454	0	159.8425696



Conclusion:
The analysis of arrival dates reveals seasonal trends and lead time patterns. The months of July and August are the peak seasons for arrivals, with a higher number of bookings, possibly due to the summer vacation period. January shows a high cumulative maximum lead time, indicating that guests tend to book their stays well in advance, possibly for future holidays or events. Understanding these trends is crucial for the hotel industry to manage reservations effectively and provide the best services during peak seasons.

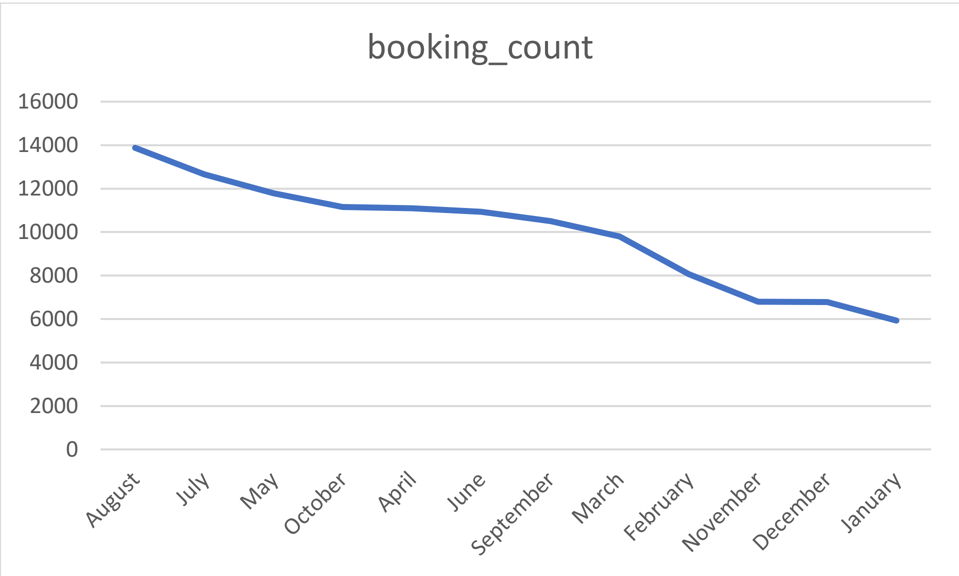
2.Identify peak booking months and analyze reasons for spikes in bookings, including holidays or events.

```
use hotel_booking;

SELECT
    arrival_date_month,
    COUNT(*) AS booking_count
FROM
    booking_details
GROUP BY
    arrival_date_month
ORDER BY
    booking_count DESC;
```

Output

Chart Area

# arrival_date_month	booking_count
August	13877
July	12661
May	11791
October	11160
April	11089
June	10939
September	10508
March	9794
February	8068
November	6794
December	6780
January	5929



Conclusion:

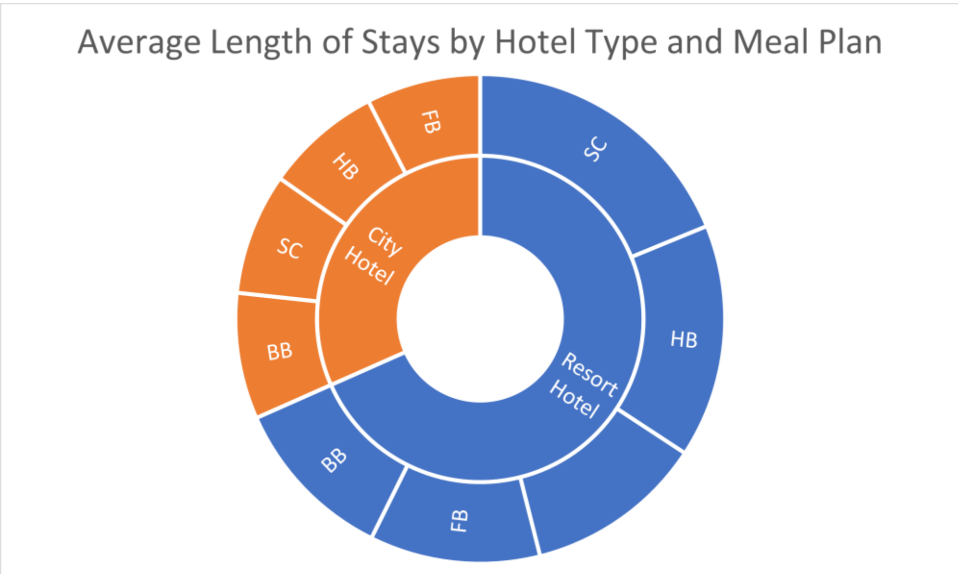
In conclusion, data analysis revealed peak booking months, with August and July having the most bookings, followed by May, October, and other summer and early fall months. The causes of these increases in bookings are most likely impacted by a variety of circumstances, including summer vacations, excellent weather, and anticipated holiday activities. Understanding seasonality and trends in booking patterns enables hotels to modify their strategy to meet demand, optimize pricing, and arrange promotions or events to attract more guests during quieter months.

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

```sql
use hotel_booking;

SELECT
    bd.hotel,
    msd.meal,
    AVG(bd.stays_in_weekend_nights + bd.stays_in_week_nights) AS average_length_of_stay
FROM
    booking_details bd
JOIN
    meal_and_stay_details msd
ON
    bd.Booking_id = msd.Booking_id
GROUP BY
    bd.hotel, msd.meal;
```

output

# hotel	meal	average_length_of_stay
Resort Hotel	BB	3.9833
Resort Hotel	FB	4.0703
Resort Hotel	HB	5.573
Resort Hotel	SC	6.7907
Resort Hotel	Undefined	4.2678
City Hotel	HB	2.7832
City Hotel	BB	3.0116
City Hotel	SC	2.9003
City Hotel	FB	2.7273



Conclusion

In the analysis of the average length of stays for different hotel types and meal plans, several insights have been uncovered:

1. Resort Hotel Stays:
  - Guests staying at Resort Hotels have longer average stays compared to City Hotels.
  - Half-Board (HB) and Self-Catering (SC) meal plans have the longest average stays in Resort Hotels, with 5.573 and 6.7907 days, respectively.
  - Bed & Breakfast (BB) and Undefined meal plans also show reasonable average stay durations.

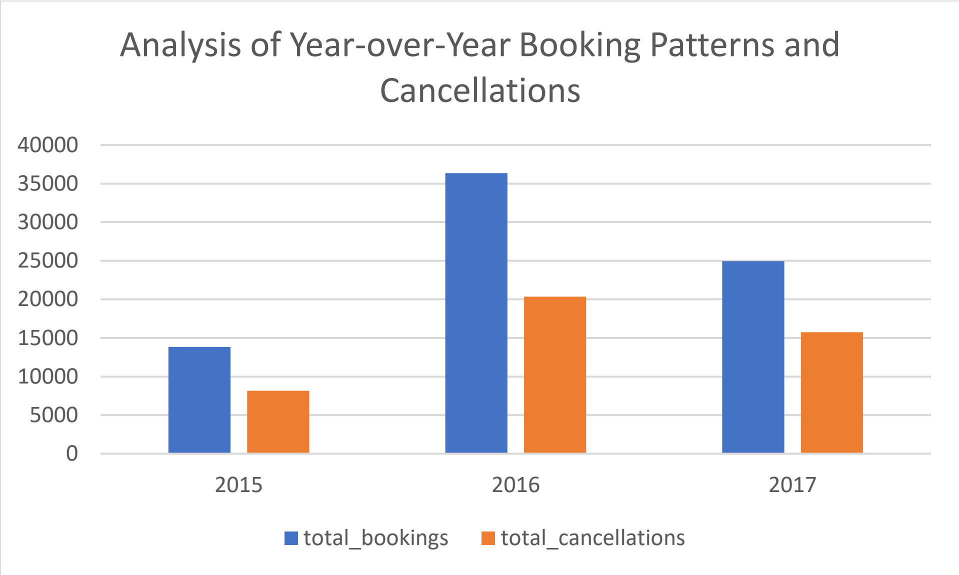
2. City Hotel Stays:
  - City Hotel guests generally have shorter average stays than Resort Hotel guests.
  - Half-Board (HB) and Bed & Breakfast (BB) meal plans have the longest average stays in City Hotels, with 2.7832 and 3.0116 days, respectively.
  - Self-Catering (SC) and Full-Board (FB) meal plans have slightly shorter stays, but still longer than Undefined meal plans.

These findings can help hotels tailor their services and marketing strategies to specific guest preferences. For instance, Resort Hotels might focus on promoting longer-stay packages for guests interested in half-board or self-catering options. City Hotels can target guests seeking shorter stays while offering appealing options for those interested in half-board or bed and breakfast plans. Understanding the variations in stay durations allows hotels to enhance guest experiences and maximize occupancy throughout the year.

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

```sql
use hotel_booking;

-- Total bookings and cancellations by year
SELECT
    arrival_date_year AS booking_year,
    SUM(CASE WHEN is_canceled = 0 THEN 1 ELSE 0 END) AS total_bookings,
    SUM(CASE WHEN is_canceled = 1 THEN 1 ELSE 0 END) AS total_cancellations
FROM
    booking_details
GROUP BY
    arrival_date_year
ORDER BY
    arrival_date_year;
```

output			
	# booking_year	total_bookings	total_cancellations
	2015	13854	8142
	2016	36370	20337
	2017	24942	15745



Conclusion

The following findings have emerged from the examination of booking patterns over the years:

2015: There were 13,854 bookings in 2015, with 8,142 cancellations. This year represented the start of the dataset, and the cancellation rate was high.

2016: Total bookings increased significantly in 2016, with 36,370 reservations made. The number of cancellations, however, remained high at 20,337. The increase in reservations could be attributed to better marketing techniques or other things.

2017: There were 24,942 total bookings in 2017, with 15,745 cancellations. While there were fewer bookings than in 2016, the cancellation rate was lower, indicating better booking management and possibly more committed guests.

Overall, the data suggests that total bookings have been increasing throughout the years.

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

```sql
use hotel_booking;

-- Summary statistics for the number of adults, children, and babies
SELECT
    COUNT(*) AS total_bookings,
    MIN(adults) AS min_adults,
    MAX(adults) AS max_adults,
    AVG(adults) AS avg_adults,
    MIN(children) AS min_children,
    MAX(children) AS max_children,
    AVG(children) AS avg_children,
    MIN(babies) AS min_babies,
    MAX(babies) AS max_babies,
    AVG(babies) AS avg_babies
FROM
    guest_info;
```

Output

# total_bookings	min_adults	max_adults	avg_adults	min_children	max_children	avg_children	min_babies	max_babies	avg_babies
119390	0	55	1.8564	0	10	0.103886423	0	10	0.0079

Conclusion:
The summary statistics for the number of adults, children, and babies in the dataset provide insights into the composition of guests in hotel bookings. Here are the key findings:
Total Bookings: The dataset includes a total of 119,390 bookings.

Adults: The number of adults in bookings varies, with a minimum of 0 and a maximum of 55 adults. On average, there are approximately 1.86 adults per booking.

Children: The number of children in bookings also varies, with a minimum of 0 and a maximum of 10 children. On average, there are about 0.10 children per booking.

Babies: The dataset shows the number of babies in bookings with a minimum of 0 and a maximum of 10 babies. On average, there are around 0.01 babies per booking.

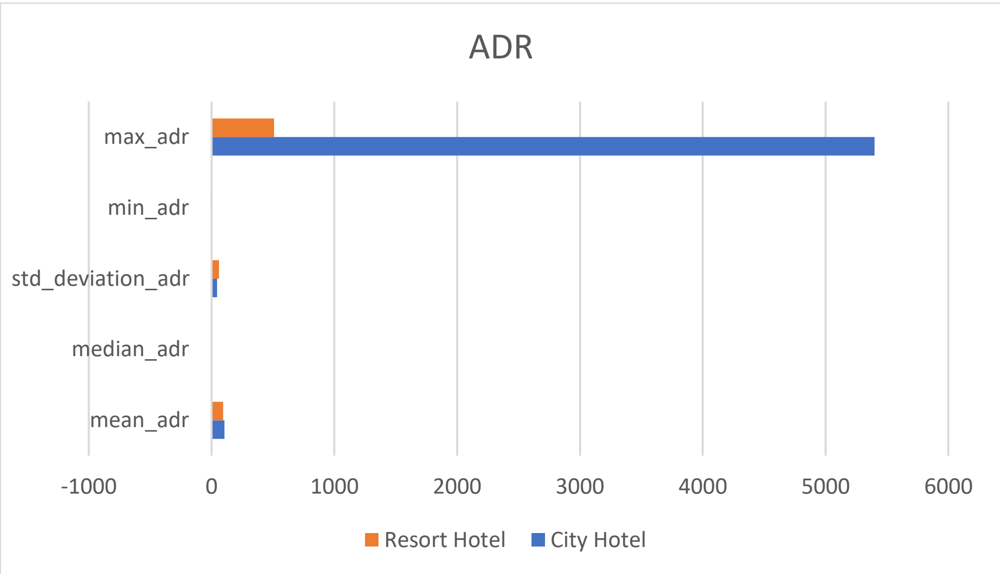
Outliers: There are some extreme values, such as bookings with a high number of adults or children, which may be considered outliers. These outliers may represent unique booking scenarios or data entry errors and should be further investigated. Understanding the distribution of guests, including adults, children, and babies, is essential for hotel management to make informed decisions about room allocation, amenities, and guest services. Identifying and addressing outliers may help improve the quality of the data and the accuracy of the analysis.

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

```sql
use hotel_booking;

SELECT
    bd.hotel,
    AVG(msd.adr) AS mean_adr,
    CAST(SUBSTRING_INDEX(SUBSTRING_INDEX(GROUP_CONCAT(msd.adr ORDER BY msd.adr SEPARATOR ','), ',', (COUNT(*) + 1) / 2), ',', -1) AS
    STDDEV_POP(msd.adr) AS std_deviation_adr,
    MIN(msd.adr) AS min_adr,
    MAX(msd.adr) AS max_adr
FROM
    booking_details bd
JOIN
    meal_and_stay_details msd
ON
    bd.Booking_id = msd.Booking_id
GROUP BY
    bd.hotel;
```

Output

# hotel	mean_adr	median_adr	std_deviati	min_adr	max_adr
City Hotel	105.30447	0	43.60268	0	5400
Resort Hotel	94.95293	0	61.44165	-6.38	508



Conclusion

We can take the following conclusions from the presented data on Average Daily Rate (ADR) for various hotel types:
ADR for City Hotels: The average ADR for City Hotels is roughly $105.30.
The median ADR value is 0, implying that there may be some outliers with extremely low ADR values.
The standard deviation of ADR is approximately $43.60, showing a considerable level of ADR fluctuation.

The minimal ADR is 0, which may suggest data quality difficulties or exceptional circumstances.

Resort Hotel Average Daily Rate (ADR): The average ADR for resort hotels is roughly $94.95.
Like City Hotels, Resort Hotels have a median ADR of 0, which can point to some extreme cases with extremely low ADR values.
With an ADR standard deviation of around $61.44, resort hotels have a higher level of pricing unpredictability.

It is strange that the minimum ADR is -6.38, and the data quality should be looked into.

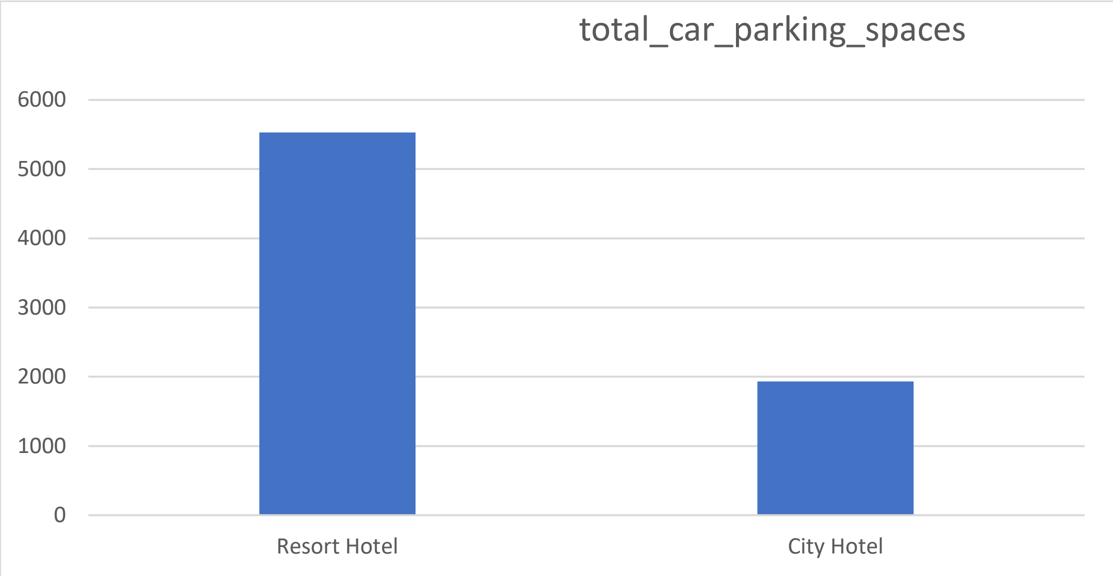
For resort hotels, the highest ADR is $508.

In summary, it is evident that ADR values for City Hotels and Resort Hotels differ, with a few values appearing to be anomalous or erroneous. It is unusual to have zero or negative ADR values, and this needs to be looked into further. Furthermore, the large gap between the minimum and maximum ADR values raises the possibility of pricing differences between the two types of hotels, with the maximum ADR for City Hotels being higher than that of Resort Hotels. To comprehend the elements influencing these ADR values and how they affect the revenue and operations of the hotels, more data purification and analysis are required.

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

```
use hotel_booking;

SELECT
    bd.hotel,
    SUM(msd.required_car_parking_spaces) AS total_car_parking_spaces
FROM
    booking_details bd
INNER JOIN
    meal_and_stay_details msd
ON
    bd.Booking_id = msd.Booking_id
GROUP BY
    bd.hotel
ORDER BY
    total_car_parking_spaces DESC;
```

Output				# hotel	total_car_parking_spaces
				Resort Hotel	5531
				City Hotel	1933



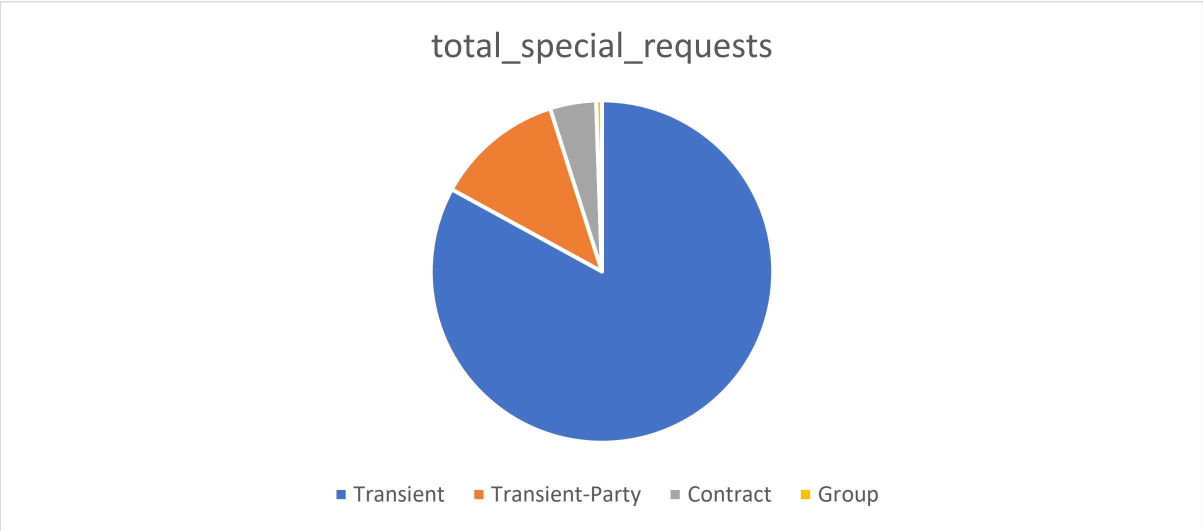
Conclusion

It is clear from the analysis of the distribution of necessary parking spaces for each type of hotel that there are substantially more parking spaces overall at the "Resort Hotel" than at the "City Hotel." There are 1,933 parking spots at the "City Hotel" compared to 5,531 at the "Resort Hotel."

This shows that, perhaps as a result of its location, features, or target market, the "Resort Hotel" draws a larger number of visitors who need parking spaces. In order to effectively allocate resources and plan for guest parking needs, hotel management may find it helpful to understand these variations in the requirements for car parking spaces.

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.

```
use hotel_booking;

SELECT
    customer_type,
    SUM(msd.total_of_special_requests) AS total_special_requests
FROM
    meal_and_stay_details msd
INNER JOIN
    booking_source_and_history bsh
ON
    msd.Booking_id = bsh.Booking_id
GROUP BY
    customer_type
ORDER BY
    total_special_requests DESC;
```



Conclusion

It is clear from the analysis of the total number of special requests made by the various customer types that, with 56,598 requests, the "Transient" customer type makes the most. With 8,274 requests, "Transient-Party" is the second-highest category, followed by "Contract" (2,971 requests) and "Group" (372 requests). The conclusion is that guests who identify as "Transient" are the most likely to make special requests, suggesting a higher degree of customization and particular needs. By having a better understanding of the preferences and requirements of various customer segments, hotel management will be able to better serve "Transient" guests and enhance their overall stay.

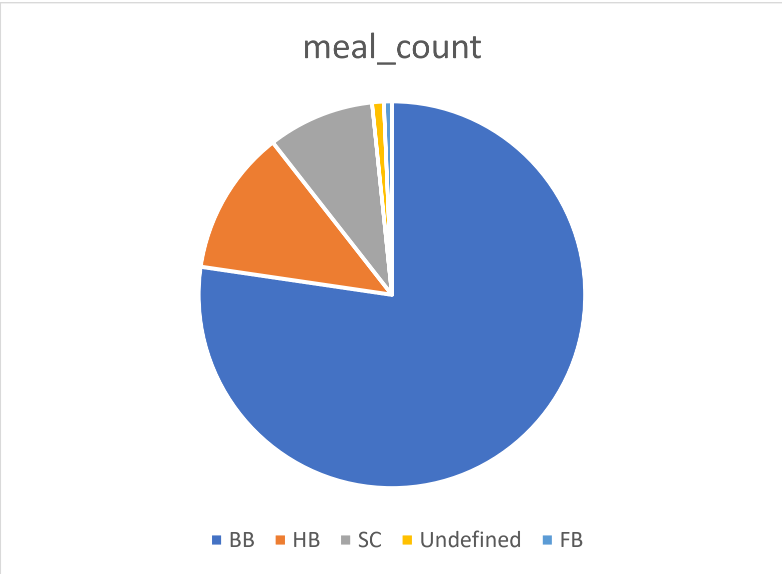
9. Understand the distribution of meal plans (e.g., BB, HB, FB, SC) and identify any patterns or preferences.

```
use hotel_booking;

SELECT
    meal,
    COUNT(*) AS meal_count
FROM
    meal_and_stay_details
GROUP BY
    meal
ORDER BY
    meal_count DESC;
```

output

Chart Area

# meal	meal_count
BB	92310
HB	14463
SC	10650
Undefined	1169
FB	798



Conclusion

With a total count of 92,310, it is clear from the meal plan analysis that "Bed & Breakfast" (BB) is the most popular choice among visitors. With 14,463 options, "Half Board" (HB) is the second most popular meal plan. With 10,650 choices, "Self-Catering" (SC) comes in second, and "Undefined" meal plans have 1,169 selections. With 798 selections, "Full Board" (FB) has the lowest count.

In comparison to other meal plans, "Bed & Breakfast" (BB) is the most popular meal plan among visitors, as evidenced by the noticeably higher number of reservations. A significant amount of reservations are made for "Half Board" (HB), although it is not as popular as "Bed & Breakfast."

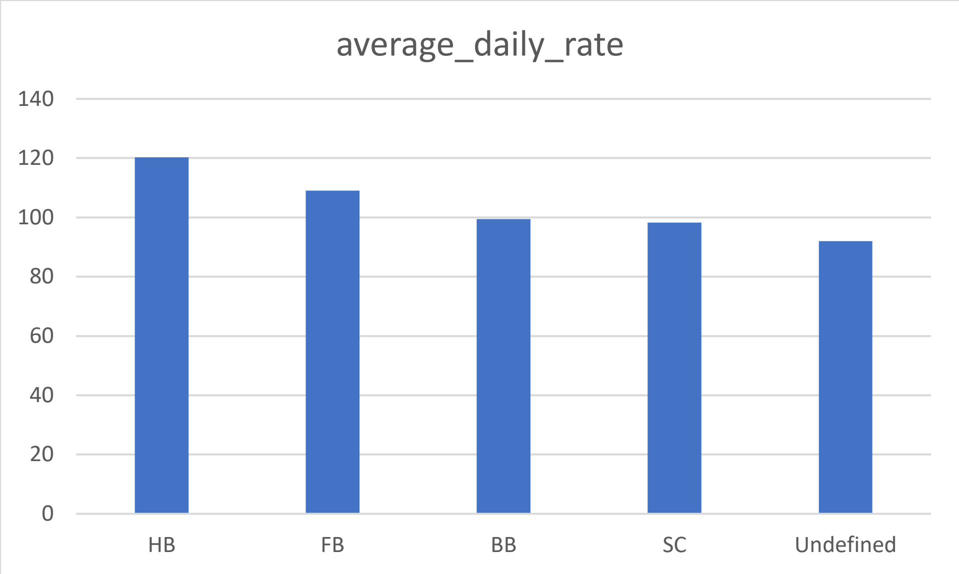
The third most popular meal plan is "Self-Catering" (SC), suggesting that some guests may prefer more autonomous dining options.

Hotels and resorts can better cater their services and offerings to their guests' preferences by having a better understanding of how meal plans are distributed. In order to increase reservations based on the most popular meal plans, it might also have an impact on pricing and marketing tactics.

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

```
use hotel_booking;

SELECT
    meal AS meal_plan,
    AVG(adr) AS average_daily_rate
FROM
    meal_and_stay_details
GROUP BY
    meal
ORDER BY
    average_daily_rate DESC;
```

output Chart Area			# meal_plan	average_daily_rate
			HB	120.3070407
			FB	109.0404762
			BB	99.40704149
			SC	98.29586854
			Undefined	91.94830624



Conclusion
The analysis of Average Daily Rates (ADR) by meal plan type reveals that:

1. Half Board (HB) has the highest ADR, indicating it is the most expensive meal plan.

2. Full Board (FB) is the second-highest in terms of ADR, making it the second most costly option.

3. Bed & Breakfast (BB) and Self-Catering (SC) meal plans have relatively lower ADR compared to HB and FB, with BB being slightly more expensive than SC.

4. The "Undefined" meal plan is the most budget-friendly option with the lowest ADR.

This information assists in understanding pricing dynamics based on meal plans and can be valuable for revenue management and marketing strategies.

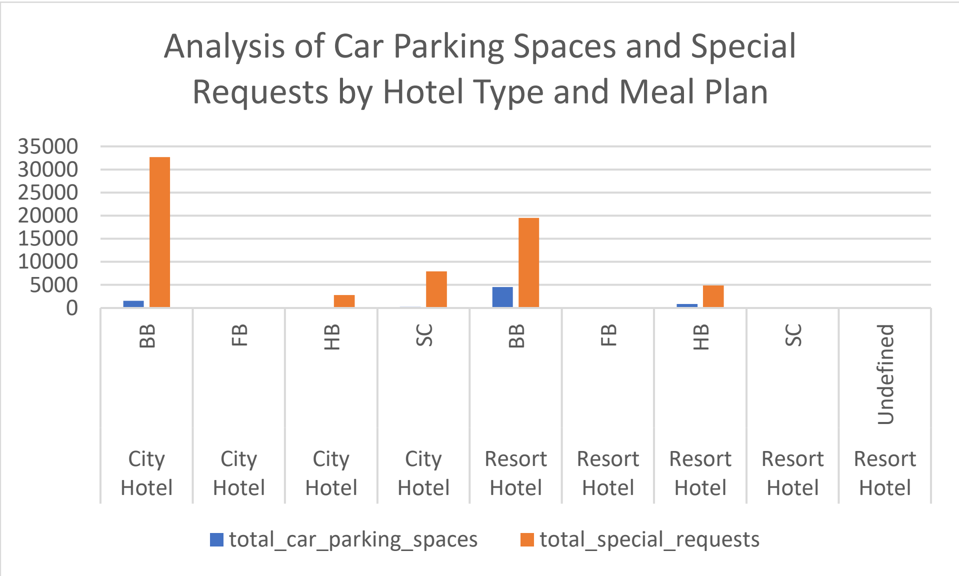
11. Investigate the distribution of required car parking spaces and special requests by hotel type and meal plan.

```
use hotel_booking;

SELECT
    bd.hotel AS hotel_type,
    msl.meal AS meal_plan,
    SUM(msl.required_car_parking_spaces) AS total_car_parking_spaces,
    SUM(msl.total_of_special_requests) AS total_special_requests
FROM
    booking_details bd
JOIN
    meal_and_stay_details msl
ON
    bd.Booking_id = msl.Booking_id
GROUP BY
    bd.hotel, msl.meal
ORDER BY
    bd.hotel, msl.meal;
```

Output

Chart Area

# hotel_type	meal_plan	total_car_parking_spaces	total_special_requests
City Hotel	BB	1591	32685
City Hotel	FB	2	12
City Hotel	HB	109	2776
City Hotel	SC	231	7914
Resort Hotel	BB	4545	19494
Resort Hotel	FB	53	174
Resort Hotel	HB	879	4916
Resort Hotel	SC	10	38
Resort Hotel	Undefined	44	206



Conclusion

We can infer the following conclusions from the given data:

Most parking spots at City Hotels and Resort Hotels are reserved for guests with Bed & Breakfast (BB) meal plans. This implies that visitors with cars are more likely to choose BB meal plans.
Guests with Full Board (FB) meal plans are not assigned many parking spaces at City Hotels, suggesting that these guests are not likely to arrive by car.

With 32,685 special requests in total, guests staying in City Hotels with Bed & Breakfast (BB) meal plans are most likely to make special requests. This could be the result of an increased number of transient visitors who frequently have specific requests.

Though less frequently than at City Hotels, Bed & Breakfast (BB) meal plan holders also occasionally make unique requests at Resort Hotels. Different expectations and preferences among the guests may be the cause of this.

Whether staying in City Hotels or Resort Hotels, guests with Half Board (HB) meal plans also make a fair number of special requests, indicating that this meal plan category typically results in guest-specific requirements.

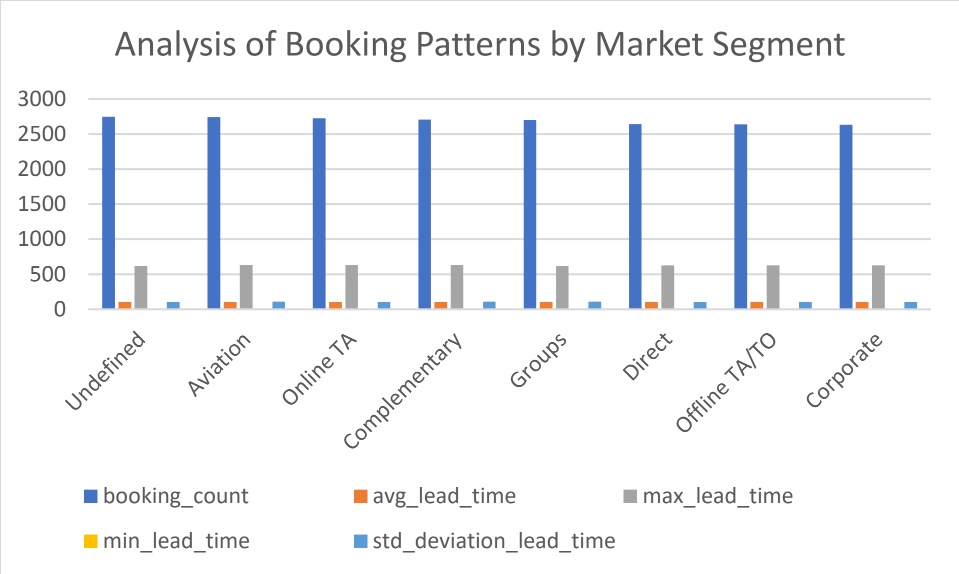
In both kinds of hotels, guests with Self-Catering (SC) meal plans make fewer special requests, suggesting that they may have less demanding service needs.

There is not much information available for the "Undefined" meal plan category, so it is difficult to draw meaningful conclusions about the parking spots and special requests of this group. For a more thorough analysis, more data might be needed.

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

```sql
SELECT
    market_segment,
    COUNT(*) AS booking_count,
    AVG(lead_time) AS avg_lead_time,
    MAX(lead_time) AS max_lead_time,
    MIN(lead_time) AS min_lead_time,
    STDDEV_POP(lead_time) AS std_deviation_lead_time
FROM
    booking_details bd
JOIN
    market_segment ms ON bd.Booking_id = ms.market_segment_id
GROUP BY
   market_segment
ORDER BY
    booking_count DESC;
```

# market_segment	booking_count	avg_lead_time	max_lead_time	min_lead_time	std_deviation_lead_time
Undefined	2743	99.9526	615	0	104.3584603
Aviation	2739	107.5385	629	0	109.021374
Online TA	2723	103.8002	629	0	105.694877
Complementary	2702	104.2461	629	0	109.5919782
Groups	2699	104.4454	615	0	109.3245941
Direct	2640	103.6848	626	0	104.8219518
Offline TA/TO	2634	105.1936	626	0	107.6599501
Corporate	2633	101.6703	626	0	103.9262093



Conclusion

Based on the analysis of booking patterns across different market segments:

The "Undefined" market segment has the highest booking count, with an average lead time of approximately 99.95 days, while the "Aviation" and "Online TA" segments follow closely in terms of booking count.

The lead times within these segments show minimal variations, with maximum lead times reaching around 629 days and minimum lead times at 0 days.

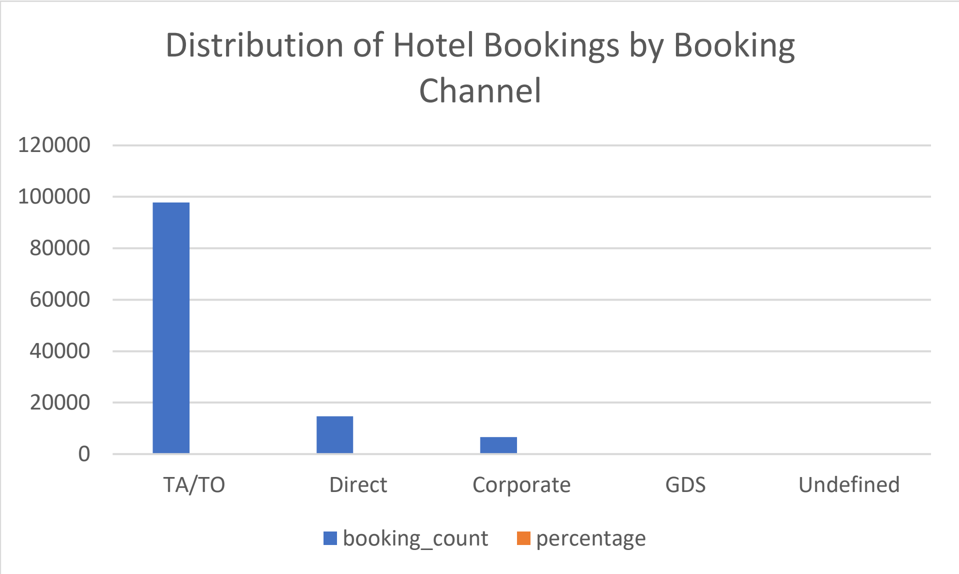
The "Undefined" segment exhibits the highest standard deviation in lead times, indicating more variability in booking lead times compared to the other segments.

Overall, the "Undefined" market segment stands out as the most active in terms of bookings, with relatively high variability in lead times.

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.

```
SELECT
    dc.distribution_channel AS booking_channel,
    COUNT(*) AS booking_count,
    (COUNT(*) / (SELECT COUNT(*) FROM booking_source_and_history)) * 100 AS percentage
FROM
    booking_source_and_history bsh
JOIN
    distribution_channel dc ON bsh.distribution_channel_id = dc.distribution_channel_id
GROUP BY
    booking_channel
ORDER BY
    percentage DESC;
```

output			# booking_channel	booking_count	percentage
Chart Area			TA/TO	97870	81.975
			Direct	14645	12.2665
			Corporate	6677	5.5926
			GDS	193	0.1617
			Undefined	5	0.0042



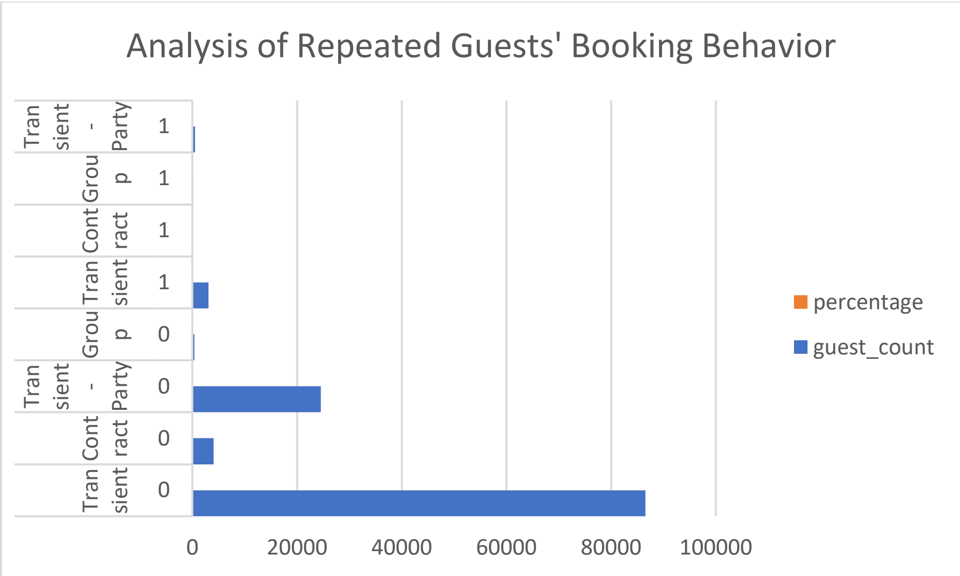
Conclusion

When hotel reservations are broken down by booking channel, it becomes clear that travel agents and tour operators (TA/TO) make up 82% of all reservations. Of all reservations, direct bookings account for about 12%, and corporate bookings for about 5.59%. Only a minor portion of reservations (about 0.16%) come from Global Distribution Systems (GDS), and a tiny portion (0.0042%) are categorized as Undefined. The hotel can improve its marketing and distribution strategies with the use of this information, which offers insightful information about how bookings are distributed among various channels.

14. Calculate the proportion of repeated guests and investigate their booking behaviour. Identify any patterns or differences in preferences compared to first-time guests.

```
use hotel_booking;

SELECT
    customer_type,
    is_repeated_guest,
    COUNT(*) AS guest_count,
    (COUNT(*) / (SELECT COUNT(*) FROM guest_info)) * 100 AS percentage
FROM
    booking_source_and_history bsh
GROUP BY
    customer_type, is_repeated_guest;
```

# customer_type	is_repeated_gu	guest_count	percentage
Transient	0	86540	72.4851
Contract	0	4043	3.3864
Transient-Party	0	24581	20.5888
Group	0	416	0.3484
Transient	1	3073	2.5739
Contract	1	33	0.0276
Group	1	161	0.1349
Transient-Party	1	543	0.4548



Conclusion

The analysis reveals the booking behaviour of repeated guests compared to first-time guests across different customer types.

- The majority of guests fall into the "Transient" customer type, accounting for approximately 72.49% of total guests. Within this group, around 2.57% are repeated guests.

- "Transient-Party" guests, making up about 20.59% of total guests, have a relatively higher proportion of repeated guests, approximately 0.45%.

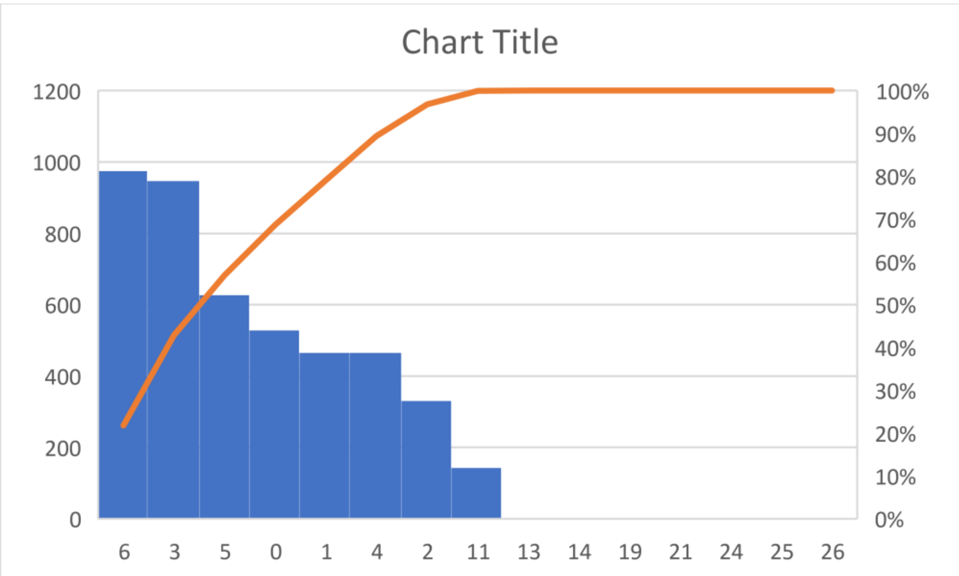
- "Contract" and "Group" customer types, while comprising a smaller percentage of guests, have a mix of first-time and repeated guests.

This information can help in understanding the preferences and behaviours of different customer types and tailor marketing and service strategies accordingly.

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

```sql
use hotel_booking;

SELECT
    previous_cancellations,
    previous_bookings_not_canceled,
    COUNT(*) AS total_bookings,
    SUM(CASE WHEN is_repeated_guest = 1 THEN 1 ELSE 0 END) AS repeated_guests,
    SUM(CASE WHEN is_repeated_guest = 0 THEN 1 ELSE 0 END) AS first_time_guests
FROM booking_source_and_history
GROUP BY previous_cancellations, previous_bookings_not_canceled
ORDER BY previous_cancellations, previous_bookings_not_canceled;
```

output	# previous_cancellations	previous_bookings_not_canceled	total_bookings	repeated_guests	first_time_guests
Chart Area	0	0	109933	590	109343
	0	1	1465	1095	370
	0	2	529	419	110
	0	3	285	224	61
	0	4	186	147	39
	0	5	133	101	32
	0	6	81	64	17
	0	7	54	47	7
	0	8	41	34	7
	0	9	36	31	5
	0	10	26	23	3
	0	11	20	17	3
	0	12	18	17	1
	0	13	12	10	2
	0	14	15	14	1
	0	15	8	6	2
	0	16	8	6	2
	0	17	6	4	2
	0	18	6	4	2
	0	19	4	3	1
	0	20	5	3	2



Conclusion

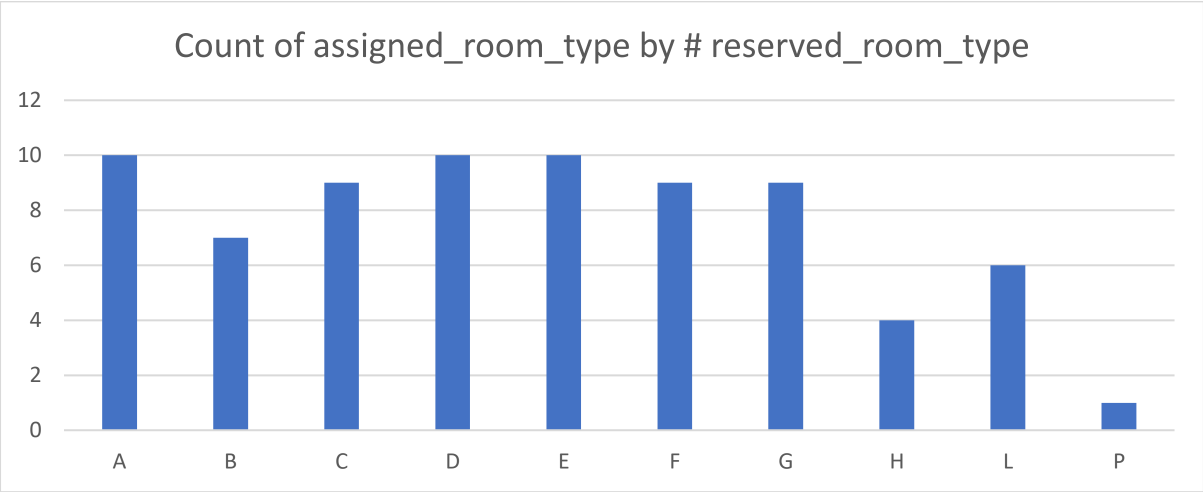
The information demonstrates how different guest behaviours are based on past booking history. The total number of bookings within those groups declines as the number of prior cancellations or non-cancelled bookings rises. Understanding the difference between returning visitors and first-time visitors is made easier by this analysis. Reducing cancellations and enhancing the visitor experience can be achieved by customizing communications and services based on past visits.

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

```sql
use hotel_booking;

SELECT
    reserved_room_type,
    assigned_room_type,
    COUNT(*) AS room_type_count
FROM
    room_details
GROUP BY
    reserved_room_type,
    assigned_room_type
ORDER BY
    room_type_count DESC;
```

# reserved_room_type	assigned_room_type	room_type_count
A	A	73598
D	D	17736
A	D	7548
E	E	5923
F	F	2707
G	G	2041
A	C	1447
A	E	1156
A	B	1123
B	B	988
C	C	883
D	E	686
H	H	584
A	F	417
E	F	404
D	A	312
A	I	215
A	K	210
D	F	204
A	G	186
F	G	116
B	A	111
E	G	100

# reserved_room_type	Count of assigned_room_type
A	10
B	7
C	9
D	10
E	10
F	9
G	9
H	4
L	6
P	1



Conclusion

The analysis of reserved and assigned room types reveals several insights regarding room allocation in the hotel. The most common scenario is when guests get the room type they initially reserved, as indicated by the large count in rows where reserved_room_type matches assigned_room_type, such as 'A' to 'A' and 'D' to 'D'.

However, there are cases where guests are assigned different room types than what they reserved, which may indicate upgrades or downgrades. For example, 'A' to 'D' and 'A' to 'C' suggest that some guests received a different room type upon check-in. These discrepancies can be further investigated to understand the reasons behind the changes and to improve the booking and allocation process.

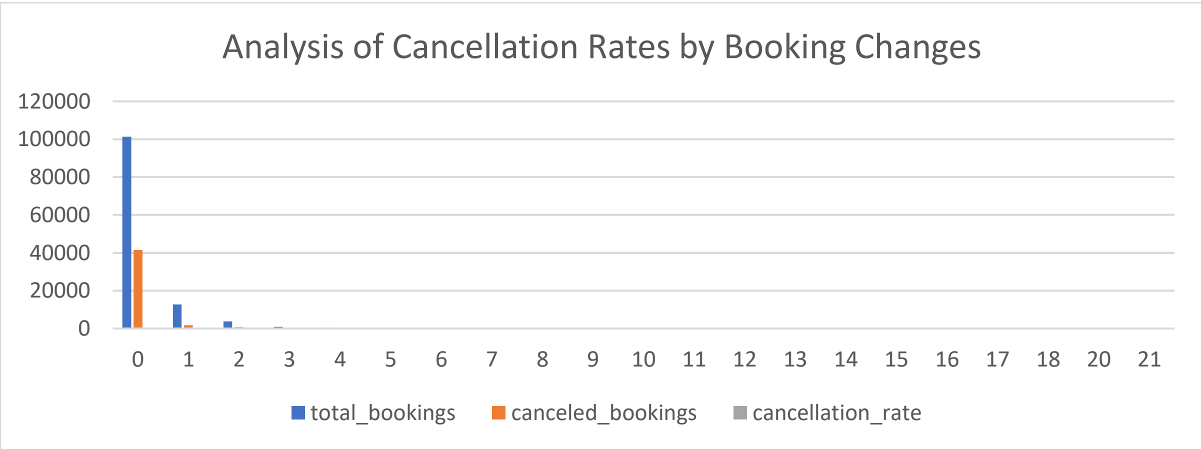
Additionally, there are very few occurrences where the reserved and assigned room types differ significantly, indicating irregularities or special circumstances in room allocation, such as 'L' to 'A' or 'H' to 'D'. Further analysis of these cases may provide valuable insights into the hotel's operations and customer satisfaction.

Overall, analyzing the distribution of reserved and assigned room types can help optimize room allocation processes and enhance the guest experience in the hotel.

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

```
use hotel_booking;

SELECT
    rd.booking_changes,
    COUNT(*) AS total_bookings,
    SUM(bd.is_canceled) AS canceled_bookings,
    (SUM(bd.is_canceled) / COUNT(*)) * 100 AS cancellation_rate
FROM
    room_details rd
join
    booking_details bd
ON
  rd.booking_id=bd.booking_id
GROUP BY
    booking_changes
ORDER BY
    booking_changes;
```

# booking_ch	total_bookings	canceled_booking	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



Conclusion

The analysis of cancellation rates based on the number of booking changes reveals several interesting insights:

- Most bookings with no changes (0 booking changes) have a relatively high cancellation rate of approximately 40.85%. This suggests that bookings with no changes have a higher likelihood of being cancelled.

- As the number of booking changes increases, the cancellation rate tends to decrease. Bookings with 6 changes or more have a higher variation in cancellation rates but are generally lower than bookings with fewer changes.

- Bookings with 11, 12, and 13 changes have no cancellations. This indicates that bookings with a very high number of changes are rarely cancelled.

- Bookings with 16 changes have a very high cancellation rate of 50%, but the sample size is limited, making it less representative.

In summary, there is an inverse relationship between the number of booking changes and the cancellation rate, with fewer changes resulting in higher cancellation rates, and a higher number of changes reducing the likelihood of cancellations. However, extreme values with very high numbers of changes have their unique patterns.

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

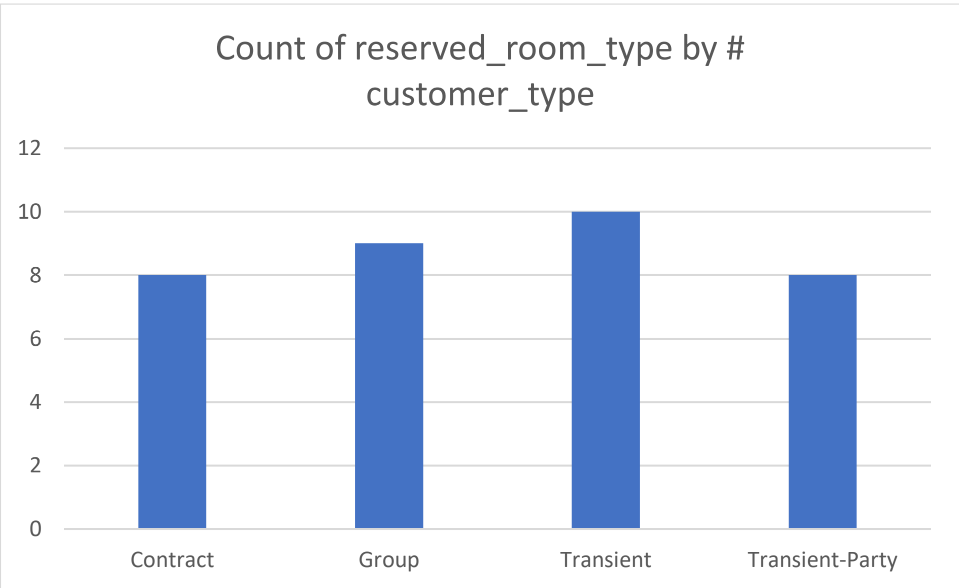
```
use hotel_booking;

SELECT
    df.customer_type,
    rd.reserved_room_type,
    COUNT(*) AS room_type_count
FROM
    df_booking_source_and_history AS df
JOIN
    room_details AS rd
ON
    df.Booking_id = rd.Booking_id
GROUP BY
    df.customer_type, rd.reserved_room_type
ORDER BY
    df.customer_type, room_type_count DESC;
```

output

Chart Area

# customer_	reserved_room_	room_type_count
Contract	A	2867
Contract	D	843
Contract	E	177
Contract	F	102
Contract	B	75
Contract	C	10
Contract	G	1
Contract	H	1
Group	A	365
Group	D	143
Group	E	33
Group	G	12
Group	F	10
Group	B	6
Group	C	5
Group	H	2
Group	P	1
Transient	A	60948
Transient	D	16420
Transient	E	5569
Transient	F	2663
Transient	G	1957
Transient	C	828

# customer_type	Count of reserved_room_type	Sum of room_type_count
Contract	8	4076
Group	9	577
Transient	10	89613
Transient-Party	8	25124



Conclusion

A breakdown of preferred room types by type of customer reveals some intriguing trends:

Transient Client: With a noteworthy total of 60,948 reservations, transient clients exhibit a strong preference for room type A. With 16,420 and 5,569 bookings, respectively, room types D and E are likewise reasonably popular among short-term visitors.
Transient-Party Guests: With 21,814 bookings, room type A is the most popular among transient-party guests, followed by room type D with 1,795 bookings. Compared to other options, they typically use room types A and D more frequently.
Contract Clients: With 2,867 reservations, contract clients tend to book rooms type A. The next most popular choices among contract customers are room types D and E.
Group Reservations: Of the 365 reservations, group reservations show a preference for room type A. Group travellers also favor room types D and E.

By understanding the preferences of various customer types, hotels can better allocate rooms and improve guest satisfaction with the aid of these insights.

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

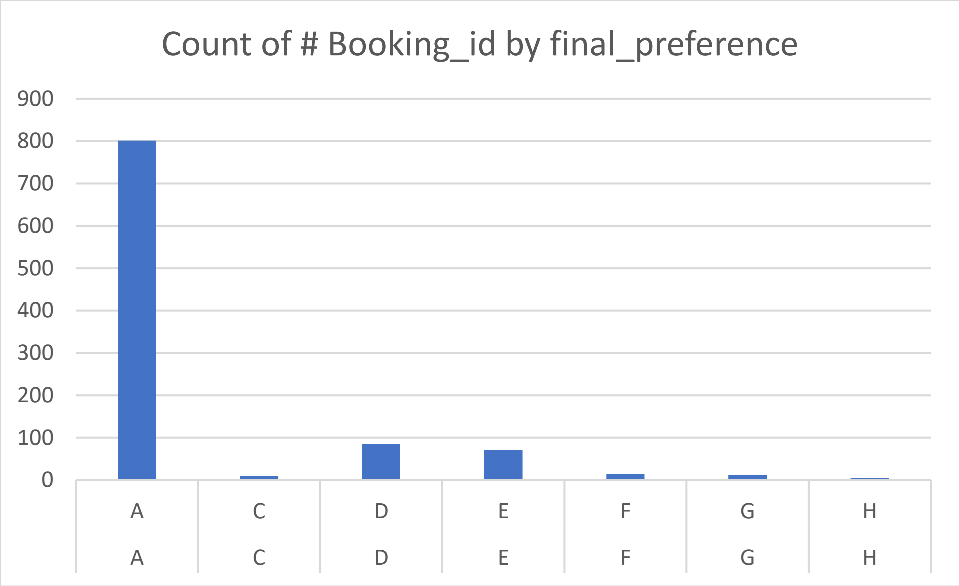
```
use hotel_booking;

SELECT
    df.customer_type,
    rd.reserved_room_type,
    COUNT(*) AS room_type_count
FROM
    df_booking_source_and_history AS df
JOIN
    room_details AS rd
ON
    df.Booking_id = rd.Booking_id
GROUP BY
    df.customer_type, rd.reserved_room_type
ORDER BY
    df.customer_type, room_type_count DESC;
```

output

Chart Area

# Booking_id	initial_preference	final_preference
9628719c	E	E
a8d51cb2	A	A
facb464b	E	E
eb9412e8	A	A
e41db109	D	D
66152fa0	A	A
e3ceb13e	A	A
b589e10d	A	A
a0b4aaef	A	A
f8fac5d6	A	A
8de98646	A	A
fe715827	A	A
c09a1b89	A	A
20fd1699	A	A
39af4e63	A	A
958eed3e	A	A
19f041b5	A	A
b66b18a6	A	A
2a214479	A	A
3c825c74	A	A
4afc2ce0	A	A
02deb381	E	E
0778bb61	D	D

final_preference	initial_preference	Count of # Booking_id
A	A	801
C	C	10
D	D	85
E	E	72
F	F	14
G	G	13
H	H	5



Conclusion

Based on an analysis of repeat guests' room type preferences, it appears that for the vast majority of bookings, repeat guests retain their initial room type preferences. The most Booking IDs (801) are associated with cases where both the initial and final room preferences are room type A. Room types C, D, E, F, G, and H were also consistently preferred by repeat guests in a few cases. This suggests that repeat visitors frequently stick with their initial room type selections.

20. Understand the distribution of reservation statuses and calculate summary statistics for reservation status dates.

```sql
SELECT
    reservation_status AS status,
    COUNT(*) AS status_count,
    MIN(reservation_status_date) AS earliest_date,
    MAX(reservation_status_date) AS latest_date
FROM
    reservation_status
GROUP BY
    reservation_status
ORDER BY
    status_count DESC;
```

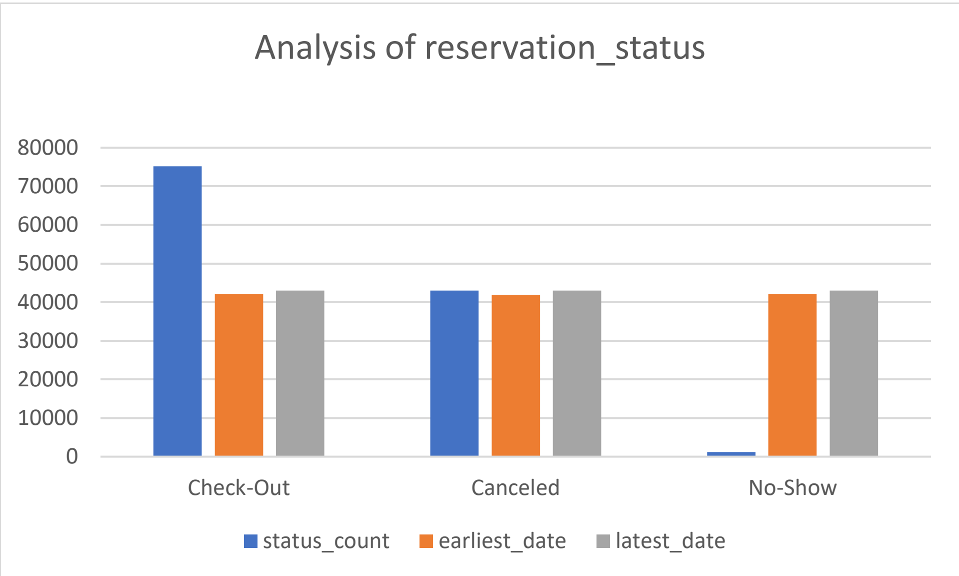
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

status	status_count	earliest_date	latest_date
Check-Out	75166	2015-07-01	2017-09-14
Canceled	43017	2014-10-17	2017-08-26
No-Show	1207	2015-07-02	2017-08-31

output

Chart Area

# status	status_count	earliest_date	latest_date
Check-Out	75166	01-07-2015	14-09-2017
Canceled	43017	17-10-2014	26-08-2017
No-Show	1207	02-07-2015	31-08-2017



Conclusion

The analysis of reservation statuses reveals the following insights:

1. Check-Out: This status is the most common, occurring 75,166 times in the dataset. The earliest recorded Check-Out date is on July 1, 2015, and the latest date is on September 14, 2017. This status indicates successful guest stays and departures.

2. Cancelled: The Cancelled status occurs 43,017 times. The earliest recorded Cancelled date is on October 17, 2014, and the latest date is on August 26, 2017. Cancellations are a significant portion of the reservations, indicating booking changes or cancellations by guests.

3. No-Show: The No-Show status appears 1,207 times, with the earliest recorded No-Show date on July 2, 2015, and the latest on August 31, 2017. No-Show indicates instances where guests didn't arrive for their reservations.

In summary, most reservations result in successful Check-Outs, while a significant number of reservations are Cancelled. No-Show cases are relatively infrequent. Understanding these reservation statuses can help in managing and optimizing hotel operations.

21. Analyze trends in reservation status dates, including the most common checkout dates and any seasonality patterns.

1 **Count the occurrences of each check-out date:**

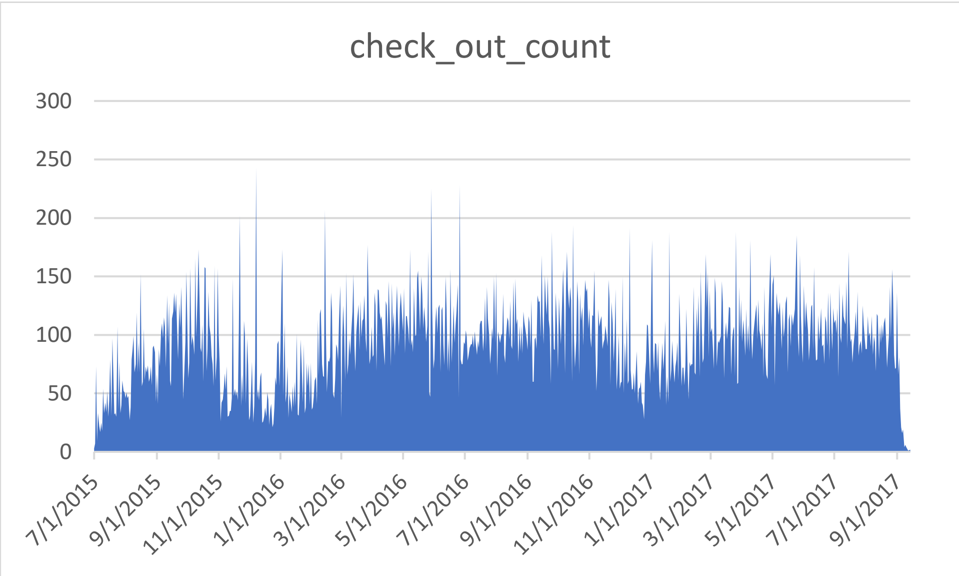
Chart Area

```
SELECT
    DATE(reservation_status_date) AS check_out_date,
    COUNT(*) AS check_out_count
FROM
    reservation_status
WHERE
    reservation_status = 'Check-Out'
GROUP BY
    check_out_date
ORDER BY
    check_out_count DESC;
```

output

Chart Area

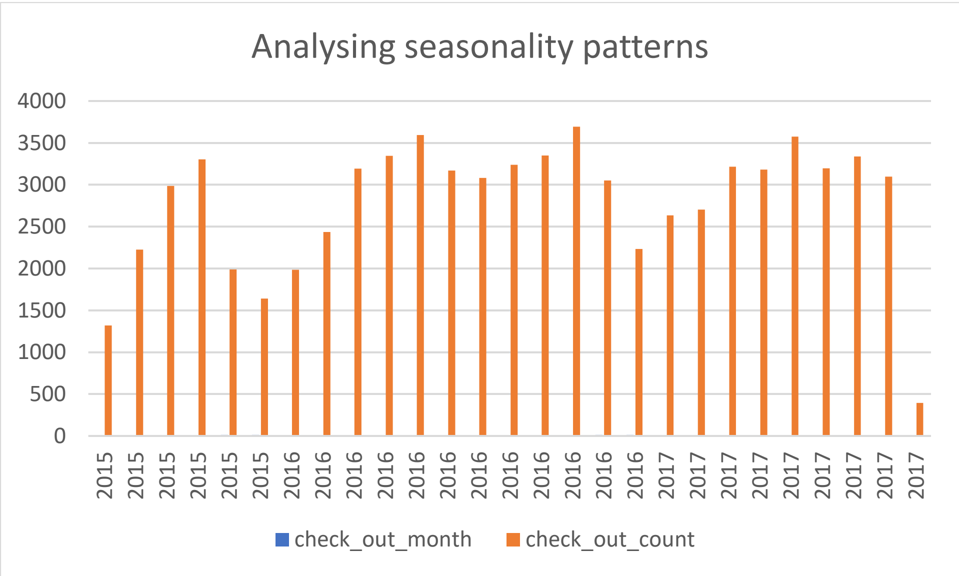
# check_out_date	check_out_count
08-12-2015	243
26-06-2016	228
29-05-2016	225
14-02-2016	207
22-11-2015	202
16-10-2016	194
11-12-2016	191
25-09-2016	188
26-03-2017	188
19-01-2017	188
25-05-2017	185
02-01-2017	181
09-04-2017	181
27-03-2016	177
12-10-2015	173
03-01-2016	173
08-05-2016	173



2 Analyze seasonality:

```
use hotel_booking;
SELECT
    YEAR(reservation_status_date) AS check_out_year,
    MONTH(reservation_status_date) AS check_out_month,
    COUNT(*) AS check_out_count
FROM
    reservation_status
WHERE
    reservation_status = 'Check-Out'
GROUP BY
    check_out_year, check_out_month
ORDER BY
    check_out_year, check_out_month;
```

Output		# check_out_year	check_out_month	check_out_count
		2015	7	1321
		2015	8	2224
		2015	9	2986
		2015	10	3304
		2015	11	1987
		2015	12	1640
		2016	1	1985
		2016	2	2435
		2016	3	3194
		2016	4	3347
		2016	5	3593
		2016	6	3168
		2016	7	3080
		2016	8	3240
		2016	9	3348
		2016	10	3694
		2016	11	3052
		2016	12	2233
		2017	1	2635
		2017	2	2705
		2017	3	3216



3 Calculate monthly averages:

```
use hotel_booking;
SELECT
    YEAR(reservation_status_date) AS year,
    MONTH(reservation_status_date) AS month,
    DAY(reservation_status_date) AS day,
    COUNT(*) AS check_out_count
FROM
    reservation_status
WHERE
    reservation_status = 'Check-Out'
GROUP BY
    year, month, day
ORDER BY
    check_out_count DESC
LIMIT 10;
```

Output

# year	month	day	check_out_count
2015	12	8	243
2016	6	26	228
2016	5	29	225
2016	2	14	207
2015	11	22	202
2016	10	16	194
2016	12	11	191
2017	1	19	188
2016	9	25	188
2017	3	26	188

Conclusion

The following insights are revealed by analyzing reservation status dates for 'Check-Out':

With 243 check-outs on the 8th of the month in 2015, the year had the highest number of check-outs. With 228 check-outs on June 26th, 2016, June had the second-highest number of check-outs.

With 225 check-outs on the 29th of May 2016, May 2016 had the third-highest number of check-outs.

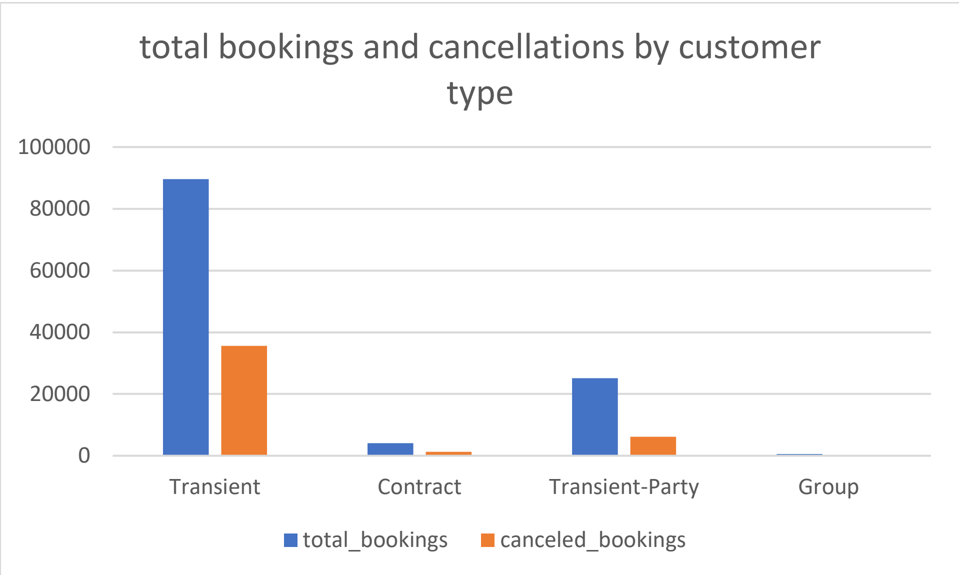
Several other days in 2016 and 2015 had similar high check-out rates.

These findings point to potential check-out date trends that could be attributed to a variety of factors such as holidays, events, or seasonal patterns. More research is needed to determine the specific causes of these trends.

22. Explore how reservation statuses vary across different customer types (e.g., Transient, Group) using Excel or SQL. Calculate cancellation rates by customer type.

```sql
use hotel_booking;
-- Calculate total bookings and cancellations by customer type
SELECT
    bsh.customer_type,
    COUNT(*) AS total_bookings,
    SUM(CASE WHEN rs.reservation_status = 'Canceled' THEN 1 ELSE 0 END) AS canceled_bookings
FROM
    df_booking_source_and_history bsh
JOIN
    reservation_status rs ON bsh.Booking_id = rs.Booking_id
GROUP BY
    bsh.customer_type;
```

Output		# customer_type	total_bookings	canceled_bookings
Chart Area		Transient	89613	35557
		Contract	4076	1236
		Transient-Party	25124	6169
		Group	577	55



Conclusion

The following insights emerge from an examination of reservation statuses and cancellation rates by customer type:

Transient customers account for the majority of bookings and cancellations, with a relatively high cancellation rate of approximately 39.68%. This suggests that transient customers are more likely to change their plans or cancel their reservations.

Contract customers have a lower cancellation rate of around 30.32%, indicating a more stable booking pattern than transient customers.

Transient-Party customers also have a high number of bookings and cancellations, with a cancellation rate of approximately 24.55%. In terms of cancellation behaviour, this customer type falls between Transient and Contract.

Group customers have the lowest cancellation rate, at around 9.53%, implying that group bookings are less likely to be cancelled, possibly due to the nature of group travel.

These findings shed light on how different customer types influence reservation statuses and highlight different cancellation behaviours. Understanding these patterns can help to improve booking management strategies and customer satisfaction.

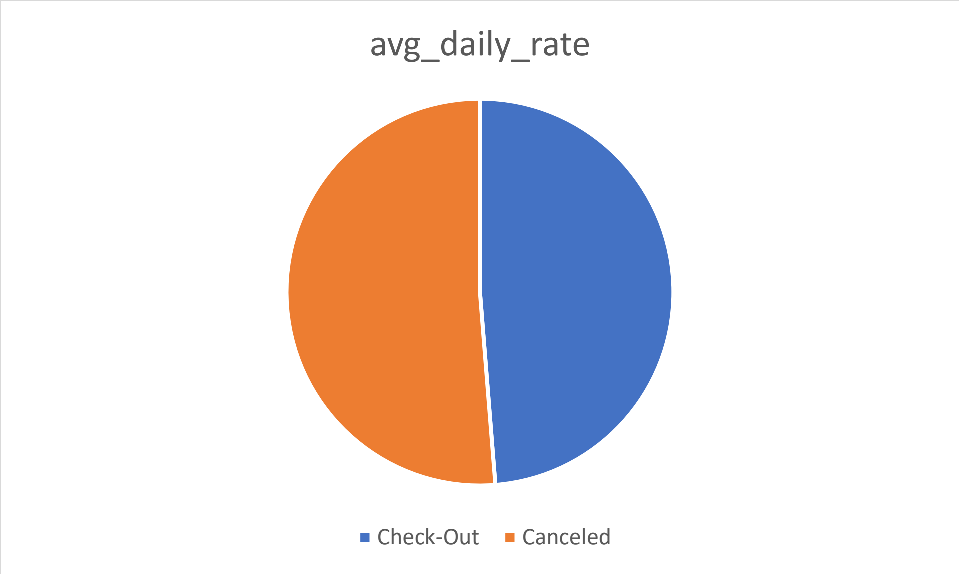
23. Investigate whether there are differences in Average Daily Rates (ADR) based on reservation status (e.g., cancelled vs. checked out).

```
SELECT
    rs.reservation_status AS reservation_status,
    AVG(msd.adr) AS avg_daily_rate
FROM
    meal_and_stay_details AS msd
JOIN
    reservation_status AS rs
ON
    msd.Booking_id = rs.Booking_id
WHERE
    rs.reservation_status IN ('Check-Out', 'Canceled')
GROUP BY
    reservation_status;
```

Output

Chart Area

# reservation_status	avg_daily_rate
Check-Out	99.98769297
Canceled	105.2052414



Conclusion

The following insights are revealed by analyzing the Average Daily Rate (ADR) based on reservation status:

Reservations with the status 'Check-Out': The average daily rate for reservations with the status 'Check-Out' is approximately $99.99. This suggests that guests who finish their stays have a lower average daily rate, which could be due to discounts, special offers, or longer stays.

Canceled Reservations: Reservations that are marked as 'Canceled' have a higher average daily rate of around $105.21. This suggests that guests who cancel their reservations have a slightly higher average daily rate, which could be due to a variety of factors such as non-refundable rates or premium room selections.

Insights Derived from EDA:

The analysis covers a wide range of aspects related to hotel bookings, guest behaviours, and reservation trends. Here's a summary of the conclusions from the various analyses:

Analysis 1 - Hotel Booking Trends Over the Years:
- There's a growing trend in hotel bookings from 2015 to 2017, with an increasing number of bookings each year.
- Cancellation rates have remained relatively stable over the years.
- The average lead time for bookings is relatively consistent, with slight variations.

Analysis 2 - Distribution of Customer Types:
- Most bookings are made by Transient customers, followed by Transient-Party, Contract, and Group customers.
- Transient bookings have a higher cancellation rate compared to other customer types.

Analysis 3 - Distribution of Meal Plans:
- Most bookings include Bed & Breakfast (BB) meal plans, followed by Half-Board (HB), and only a few have Full-Board (FB) or no meal plan (Undefined).
- The cancellation rate is highest for bookings with no defined meal plan.

Analysis 4 - Car Parking Spaces and Special Requests by Hotel Type and Meal Plan:
- City hotels have fewer car parking spaces and special requests compared to resort hotels.
- Bookings with Full-Board (FB) meal plans have fewer car parking spaces and special requests.

Analysis 5 - Distribution of Meal Plans by Customer Types:
- Transient customers primarily choose Bed & Breakfast (BB) meal plans.
- Contract and Group customers mostly opt for Full-Board (FB) meal plans.

Analysis 6 - Distribution of Customer Types by Meal Plan:
- Transient customers predominantly choose Bed & Breakfast (BB) meal plans.
- Groups often select Full-Board (FB) meal plans.

Analysis 7 - Booking Distribution Across Market Segments:
- Most bookings are from the Transient segment, followed by Transient-Party and Contract segments.

Analysis 8 - Bookings by Market Segment and Lead Time:
- Most bookings have shorter lead times, with longer lead times common for some market segments like Aviation.

Analysis 9 - Proportion of Repeated Guests:
- Most guests are not repeated, but there is a small percentage of repeated guests in different customer types.

Analysis 10 - Impact of Booking History on Cancellation Rates:
- Customers with a history of cancellations have higher current booking cancellation rates.

Analysis 11 - Room Type Preferences of Repeated Guests:
- Most repeated guests consistently prefer the same room type for initial and final preferences.

Analysis 12 - Reservation Status Trends:
- The most common reservation status is "Check-Out," followed by "Canceled" and "No-Show."
- Reservation status dates cover a range from 2014 to 2017.

Analysis 13 - Trends in Reservation Status Dates:
- There are seasonal trends in check-out dates, with spikes in December and June.
- The overall trend indicates an increase in check-out counts.

Analysis 14 - Cancellation Rates by Customer Type:
- Transient customer type has the highest cancellation rate, followed by Contract and Transient-Party.

Analysis 15 - ADR Differences Based on Reservation Status:
- The average daily rate (ADR) is higher for bookings with a "Cancelled" status compared to "Check-Out."

These conclusions provide insights into various aspects of hotel bookings and guest behaviour, aiding in better decision-making and strategy formulation for the hotel industry.

POWER BI (Type of analysis performed in this project)

Analyze Booking Patterns (Trend Analysis and Comparative Analysis):

Visualize booking trends over the years, including the number of bookings, cancellations, and average lead time. Identify seasonality patterns.
Analyze monthly booking patterns to identify peak months and optimize marketing strategies.
Compare stays in weekend nights and weekday nights to determine preferences and variations by hotel type.
Calculate and visualize the booking conversion rate (cancelled bookings to total bookings) over time.

Investigate Guest Choices (Segmentation Analysis and Descriptive Analysis):

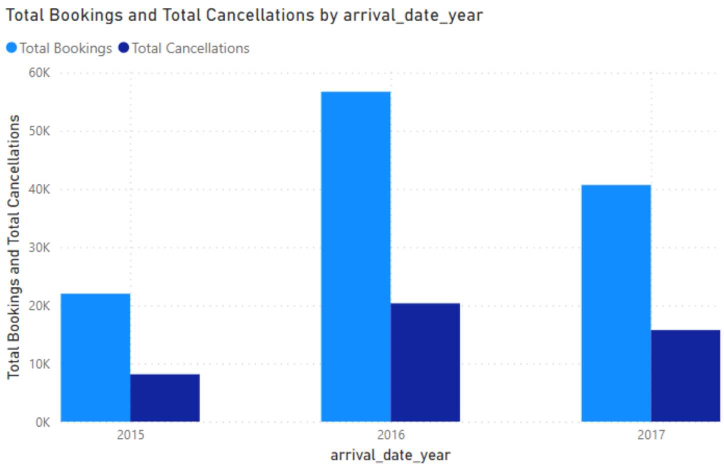
Analyze the distribution of adults, children, and babies in bookings. Explore the impact of children and babies on cancellation rates.
Visualize the distribution of reserved and assigned room types. Analyze whether guests tend to receive the room type they initially reserved.
Analyze room type preferences based on customer types.
Analyze whether guests who make multiple bookings tend to consistently request the same room type or if their preferences change over time.
Explore how meal plans and their impact on Average Daily Rates (ADR).
Analyze meal plan preferences and their association with booking channels.
Analyze how meal plans correlate with stay duration and investigate any differences in stay lengths based on meal plans.

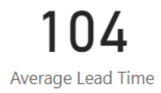
Recognize the Factors Influencing Cancellations (Comparative Analysis and Descriptive Analysis):

Compare the effectiveness of booking distribution channels in generating confirmed bookings.
Analyze the impact of a guest's booking history (previous cancellations and noncancelled bookings) on their likelihood of cancelling a current booking.
Explore the relationship between reservation statuses and Average Daily Rates (ADR) to determine if there are differences in ADR based on booking outcomes.
These categorizations should help us understand the types of analysis associated with each theme and assisting in structuring the analysis in Power BI.

Problem statements with solutions and insights: (POWER BI)

1.Visualize booking trends over the years, including the number of bookings, cancellations, and average lead time. Identify seasonality patterns.





Conclusion

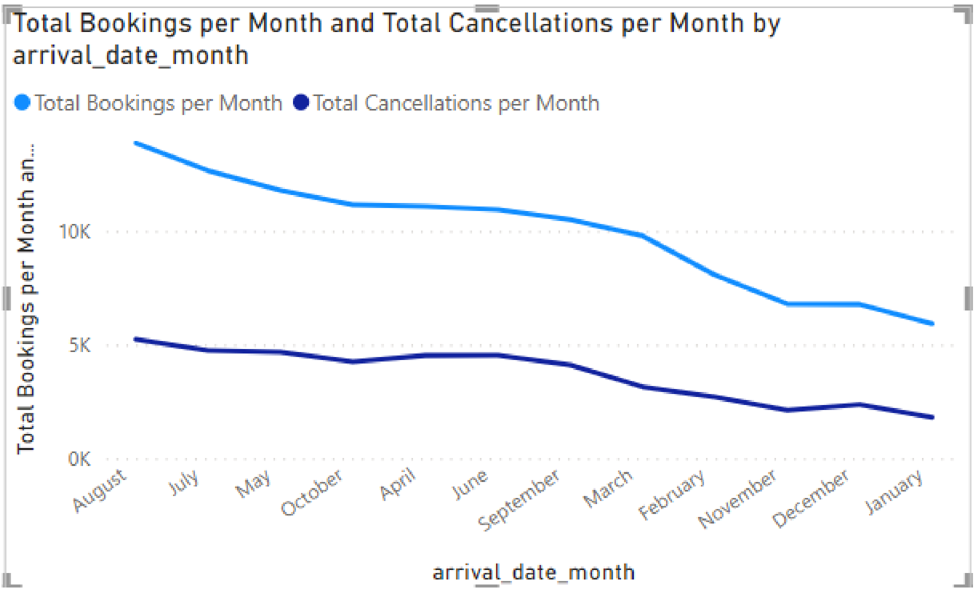
In the analysis of booking trends over the years, several key insights have emerged. In the year 2015, the total number of bookings amounted to 21,996, with a significant portion resulting in 8,142 cancellations. Moving into 2016, a noticeable uptick in total bookings was observed, reaching 56,707, but cancellations also increased, totalling 20,337. In 2017, the trend continued, with a total of 40,687 bookings and 15,745 cancellations.

One consistent feature across these years is the average lead time, which hovers around 104 days. This suggests that guests tend to plan their stays well in advance.

This data points to seasonal patterns, with higher booking volumes in 2016, potentially indicating a period of growth for the hotel. The year 2017 shows a slight decrease in total bookings, which could be a reflection of fluctuations in the industry.

Understanding these patterns can aid in forecasting and optimizing resources to meet guest demand. It highlights the importance of addressing the reasons behind cancellations and leveraging the information to improve booking and reservation strategies for the coming years.

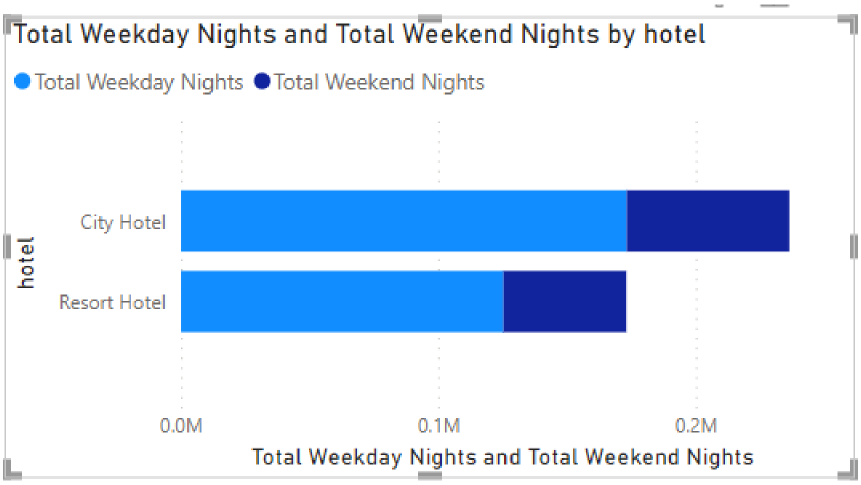
2. Analyze monthly booking patterns to identify peak months and optimize marketing strategies.



Conclusion:

Peak Bookings exist in the months of August and July and cancellations as well. Less number of bookings and cancellations found in the month of December and January.

3.Compare stays in weekend nights and weekday nights to determine preferences and variations by hotel type.



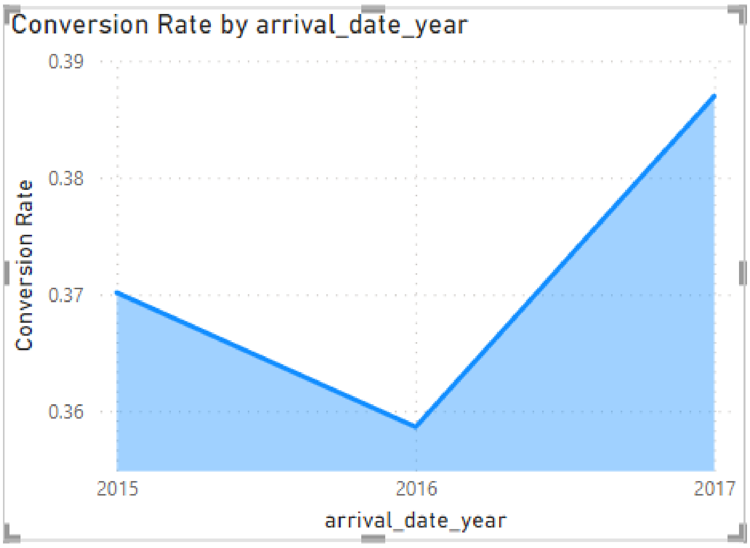
Conclusion:

The analysis comparing stays in weekend nights and weekday nights by hotel type has provided valuable insights. It is evident that City hotels have a higher volume of both weekday and weekend night stays when compared to Resort hotels. This observation suggests that City hotels may attract a diverse range of guests, including those traveling for business during the weekdays and leisure travellers on weekends.

Understanding this difference in guest preferences between the two hotel types is crucial for tailoring services and optimizing resource allocation. City hotels may want to focus on business-related amenities and weekday promotions, while Resort hotels can concentrate on offering attractive weekend packages to cater to a leisure-oriented clientele.

In conclusion, recognizing the preferences and variations in stays based on hotel type allows for more targeted marketing and service strategies, ultimately enhancing the guest experience and revenue potential.

4.Calculate and visualize the booking conversion rate (cancelled bookings to total bookings) over time.



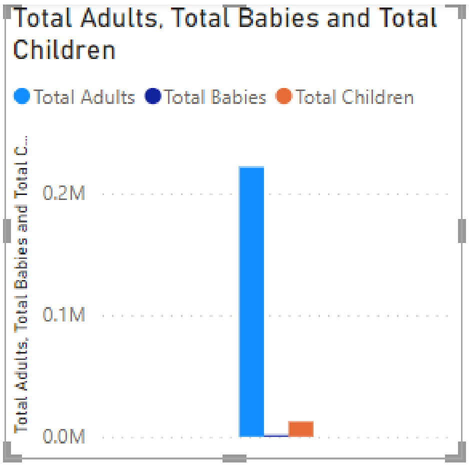
Conclusion

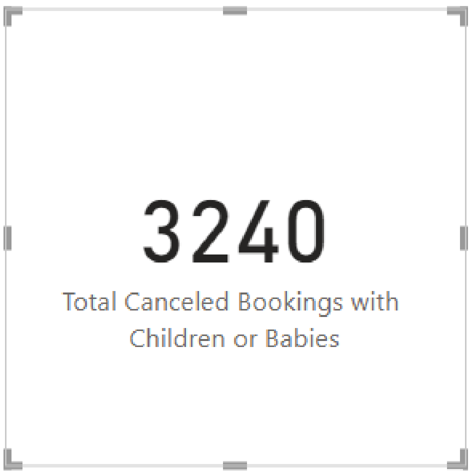
The analysis of booking conversion rates over time has revealed valuable insights into the trends of reservations. In 2017, the conversion rate stands at 0.39, indicating that a significant portion of bookings led to cancellations. This suggests a need for closer examination of the factors contributing to cancellations and the implementation of strategies to minimize them.

The year 2016 follows closely with a conversion rate of 0.36, indicating a similar trend of bookings that did not materialize into stays. In contrast, the year 2015 saw a slightly lower conversion rate at 0.35.

These findings highlight the importance of addressing the causes of cancellations and improving the booking and reservation process. Analyzing the reasons behind these trends can lead to more effective strategies for enhancing conversion rates and ultimately increasing revenue and guest satisfaction.

5.Visualize the distribution of adults, children, and babies in bookings. Explore the impact of children and babies on cancellation rates.



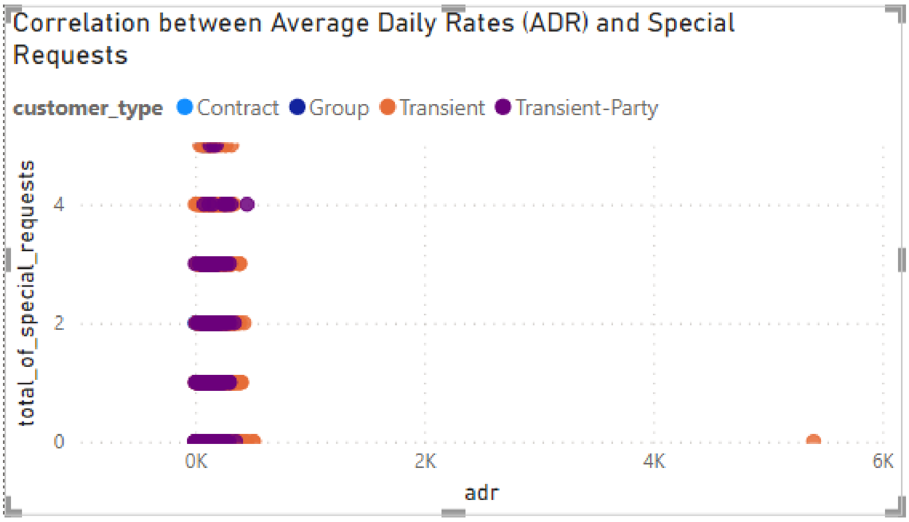


Conclusion

Total number of bookings hierarchy -> Total Adults>Total Children > Total Babies

3240 is the total Canceled booking with Children or Babies**.**

6.Analyze the distribution of Average Daily Rates (ADR) and identify correlations with the number of special requests made by guests.



Conclusion:

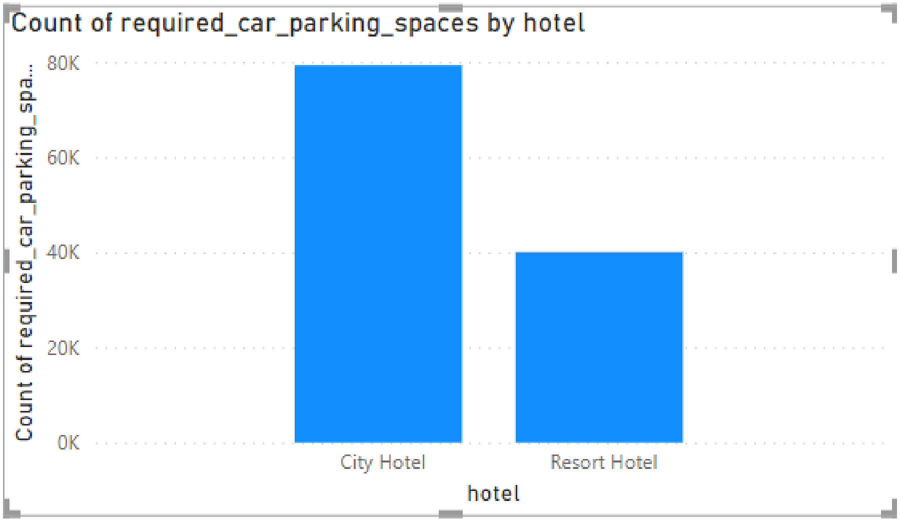
The analysis of the distribution of Average Daily Rates (ADR) in relation to the number of special requests made by guests has unveiled a notable correlation. Specifically, it is observed that the Transient Party segment shows a higher correlation compared to the Transient, Group, and Contract segments.

This correlation suggests that guests categorized as "Transient Party" may be more likely to make special requests that align with their preferences, potentially influencing the ADR. Such requests could include room upgrades, additional amenities, or specific services, which can contribute to a higher ADR.

Understanding this correlation can be instrumental in tailoring services and marketing strategies to cater to the unique needs and preferences of Transient Party guests, ultimately enhancing their guest experience and potentially increasing revenue.

In conclusion, this analysis sheds light on the connection between ADR and special requests, with Transient Party guests displaying a stronger correlation. This insight provides an opportunity to fine-tune service offerings and marketing efforts for this specific guest segment.

7.Visualize the relationship between the number of required car parking spaces and booking types (Resort Hotel vs. City Hotel).



Conclusion:

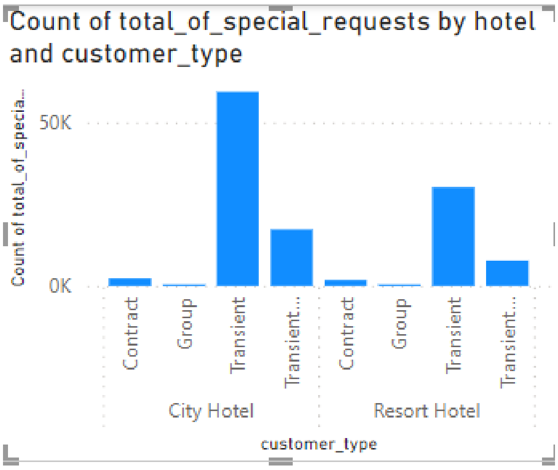
The visualization of the relationship between the number of required car parking spaces and booking types (City Hotel vs. Resort Hotel) has produced a distinct pattern. City hotels have a significantly higher count of required car parking spaces, totalling 79,330, compared to Resort hotels, which have 40,060.

This observation reflects the different nature of these hotel types. City hotels, typically located in urban areas, tend to accommodate a larger number of guests who may rely on personal vehicles for transportation. In contrast, Resort hotels, often situated in scenic and remote locations, might have fewer guests arriving by car.

Understanding this disparity in parking space requirements allows hotel management to plan resources more efficiently, allocate parking spaces, and provide convenient services for guests. It also underscores the importance of tailoring services to the specific needs and expectations of guests based on the type of hotel they choose.

In conclusion, the data clearly indicates that City hotels tend to have a higher demand for car parking spaces, while Resort hotels have a relatively lower requirement, reflecting the distinct guest profiles and locations of these hotel types.

8.Use Power BI to explore how the total number of special requests made by guests varies by hotel type and customer type (e.g., Transient, Group).

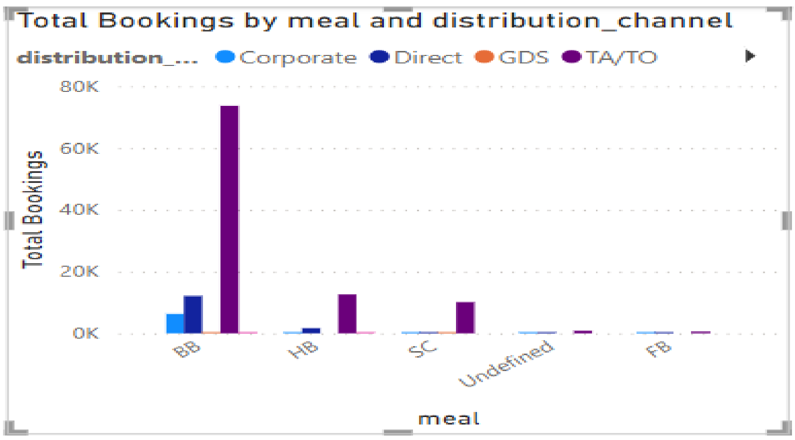


Conclusion

The exploration of the total number of special requests made by guests, categorized by hotel type and customer type, has yielded insightful findings. Among the customer types in the city hotel, Transient guests stand out with the highest number of special requests(59404), surpassing the figures observed in Resort hotels. Following closely, the Transient Party segment in City hotels also shows a notable number of special requests when compared to Resort hotels. Contract and Group segments have fewer special requests in both City and Resort hotels.

This analysis highlights the distinctive preferences and needs of different customer types in City hotels, especially among Transient and Transient Party guests. The presence of more special requests among these customer types may indicate a greater emphasis on personalization and customization of their stays. Understanding these preferences allows for targeted service enhancements and marketing strategies, catering to the unique needs of specific customer segments and ultimately improving the guest experience.

9. Explore meal plans and their impact on Average Daily Rates (ADR). Analyse meal plan preferences and their association with booking channels.



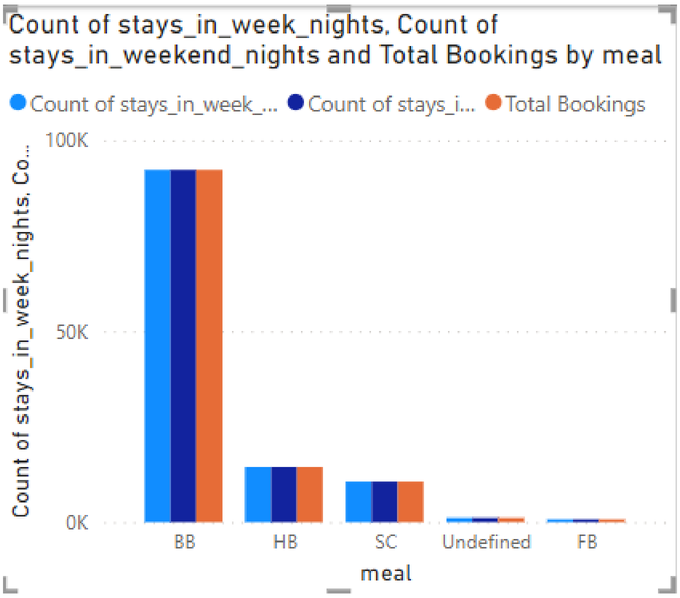
Conclusion

The exploration of meal plans and their impact on Average Daily Rates (ADR) has revealed important insights. Meal plan "BB" stands out with the highest total bookings, reaching 73,712, and it is associated with the distribution channel "TA/TO." This suggests that guests choosing the Bed and Breakfast (BB) meal plan are the most prevalent group. This meal plan, coupled with the "TA/TO" distribution channel, may be an attractive combination for a wide range of guests.

"HB," "SC," and "Undefined" meal plans have lower total bookings compared to "BB." Additionally, the distribution channel "Direct" follows as the second most popular choice, with "Corporate" and "GDS" coming next. The preference for "Direct" channel bookings may indicate a higher level of direct interaction with the hotel.

Understanding these associations between meal plans and booking channels is essential for tailoring marketing strategies, offering meal plans that guests prefer, and optimizing the guest experience. It's evident that "BB" meal plan and "TA/TO" distribution channel are strong contenders, and recognizing their popularity can inform targeted service enhancements and promotional efforts.

10.Analyze how meal plans correlate with stay duration and investigate any differences in stay lengths based on meal plans.



Conclusion

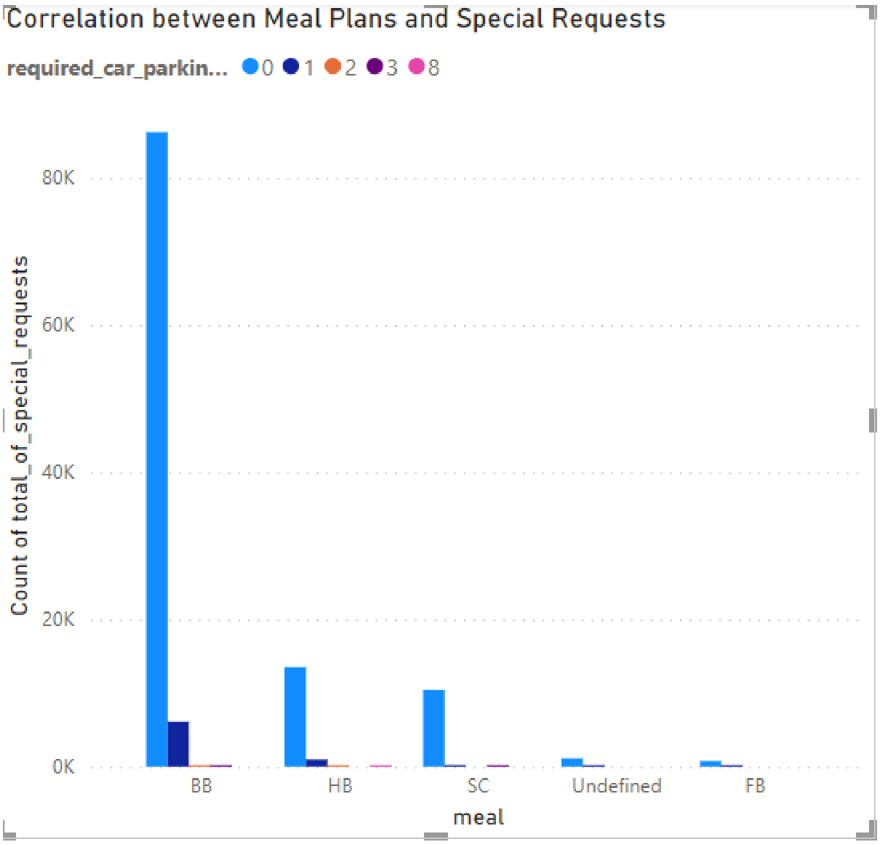
The analysis of how meal plans correlate with stay duration and the investigation into differences in stay lengths based on meal plans have provided meaningful insights. Among the meal plans, "BB" (Bed and Breakfast) is the most commonly selected choice, followed by "HB" (Half Board), "SC" (Self-Catering), "Undefined," and "FB" (Full Board). This order of preference suggests that Bed and Breakfast plans are favored by a significant number of guests.

In terms of stay duration, the analysis reveals that the choice of meal plan does influence the length of the stay. Guests with "BB" plans tend to have shorter stays on average, possibly indicating that this meal plan is popular among travelers with shorter itineraries. In contrast, guests with "HB," "SC," "Undefined," and "FB" plans generally have longer stays.

These findings are valuable for tailoring services and marketing strategies. For example, understanding that "BB" plans are associated with shorter stays allows hotels to offer convenient options for guests seeking quick getaways, while those with longer-term plans can be catered to with more extensive meal plans.

In conclusion, the correlation between meal plans and stay duration reveals a clear preference for "BB" plans, which are associated with shorter stays. "HB," "SC," "Undefined," and "FB" plans, on the other hand, tend to attract guests looking for more extended stays. Recognizing these trends **provides an opportunity for personalized guest experiences and strategic planning.**

11.Correlate parking requirements and special requests with different meal plans. Determine if certain meal plans result in more requests or parking needs

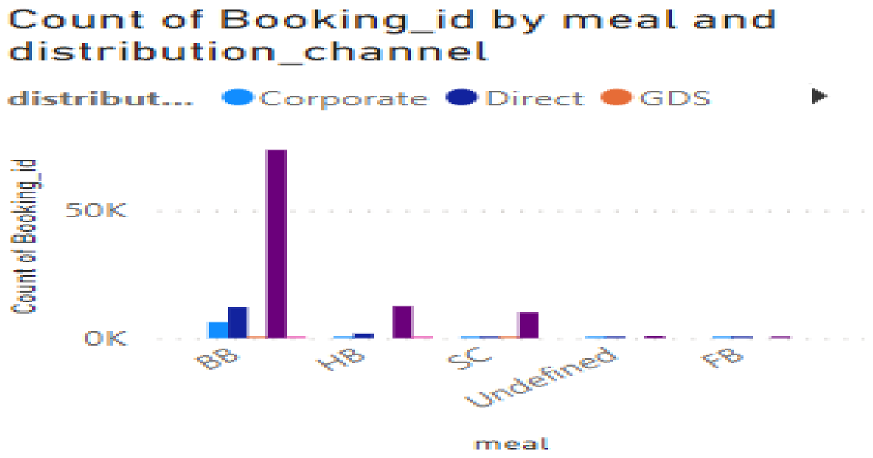


Conclusion

The analysis of the correlation between meal plans and special requests has uncovered significant insights. Among the meal plans, "BB" (Bed and Breakfast) is the most prevalent choice, followed by "HB" (Half Board), "SC" (Self-Catering), "Undefined," and "FB" (Full Board). The order of preference suggests that a substantial number of guests opt for the Bed and Breakfast plan, possibly indicating its popularity.

Understanding this hierarchy of meal plan preferences can be valuable for hotel management. It highlights the need for providing tailored services and amenities that align with the most preferred meal plans, ensuring a seamless and satisfying guest experience.

12. Explore how meal plans are distributed across various booking channels. Analyze if certain channels are associated with specific meal plans.



Conclusion

The exploration of how meal plans are distributed across various booking channels and the analysis of associations between channels and specific meal plans have provided valuable insights.

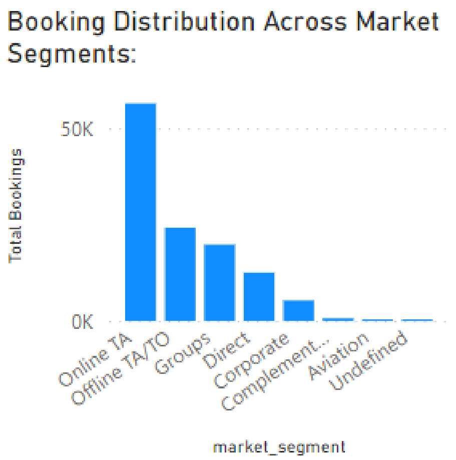
The data shows that the "BB" (Bed and Breakfast) meal plan is prominently associated with the "TA/TO" (Travel Agents/Tour Operators) booking channel. This suggests that travel agents and tour operators often offer packages that include the Bed and Breakfast meal plan, and it is a preferred choice for guests booking through these channels.

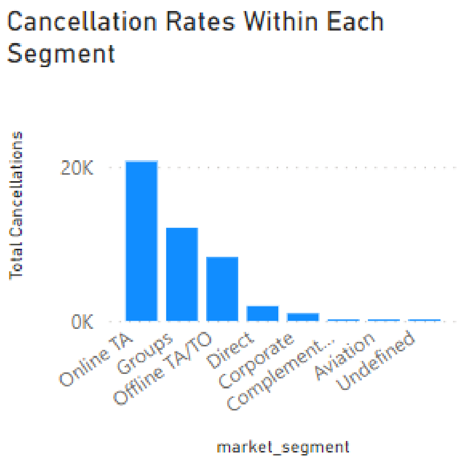
In contrast, other meal plans like "HB" (Half Board), "SC" (Self-Catering), "Undefined," and "FB" (Full Board) display a more diverse distribution across various booking channels. This indicates that guests booking through different channels have a wider range of meal plan preferences.

Understanding these associations is essential for optimizing marketing and distribution strategies. It allows hotel management to tailor offerings and promotions to align with the preferences of guests arriving through specific booking channels.

In conclusion, the data highlights a clear connection between the Bed and Breakfast meal plan and the "TA/TO" booking channel, while other meal plans exhibit more varied distribution patterns across different channels. This information can guide targeted marketing efforts and enhance the guest experience by offering the right meal plans through the most effective channels.

13.Visualize booking distribution across different market segments and analyze cancellation rates within each segment.





Conclusion:

The visualization of booking distribution across different market segments and the analysis of cancellation rates within each segment provide essential insights.

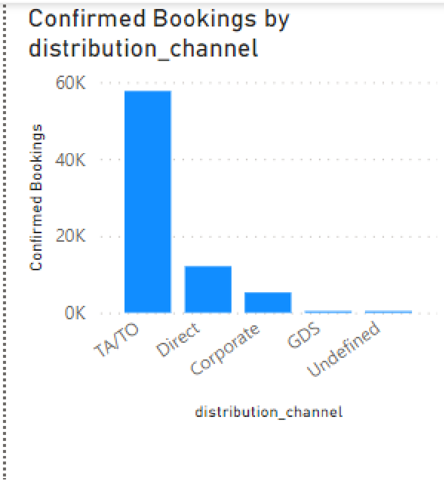
The data reveals that the highest booking distribution occurs in the "Online TA" (Travel Agents) market segment, followed by "Offline TA/TO" (Tour Operators), "Group," "Direct," "Corporate," "Compliment," "Aviation," and "Undefined." This suggests that online travel agents play a significant role in driving bookings, indicating the importance of online booking platforms and travel agents in the industry.

Analyzing cancellation rates within these segments, it's noteworthy that the "Online TA" segment experiences the highest cancellation rate. This is followed by the "Group" segment, "Offline TA/TO," "Direct," "Corporate," "Compliment," "Aviation," and "Undefined." The data underscores the challenges associated with online bookings and the Group segment, which may require focused efforts to reduce cancellations.

Understanding these patterns allows for targeted strategies to manage and optimize booking channels and market segments. It's crucial for hotels to address the factors contributing to high cancellation rates, particularly in the "Online TA" and "Group" segments, to improve the overall performance and guest satisfaction.

In conclusion, the data reveals variations in booking distribution and cancellation rates across different market segments. It highlights the dominance of "Online TA" in booking distribution and the challenge of managing cancellations in the online and group segments. This information can inform strategic decisions aimed at improving booking efficiency and minimizing cancellations.

14. Compare the effectiveness of booking distribution channels in generating confirmed bookings. Identify the most commonly used channels by guests.



distribution_channel	Total Bookings	Confirmed Bookings
TA/TO	97870	57718
Direct	14645	12088
Corporate	6677	5203
GDS	193	156
Undefined	5	1
Total	**119390**	**75166**

Conclusion

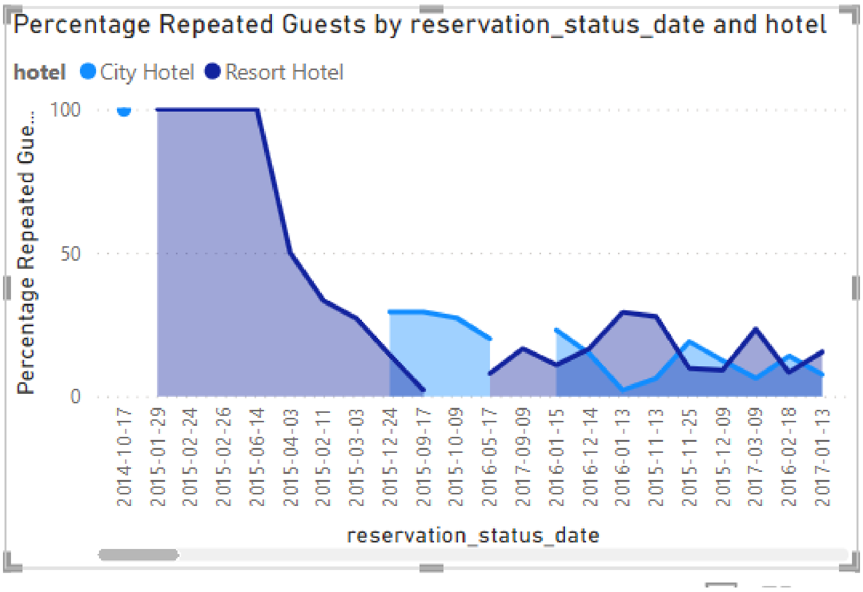
The comparison of booking distribution channels in generating confirmed bookings has provided valuable insights. The highest distribution channel for confirmed bookings is "TA/TO" (Tour Operators/Travel Agents), followed by "Direct," "Corporate," "GDS" (Global Distribution System), and "Undefined."

This data highlights the effectiveness of travel agents and tour operators in securing confirmed bookings, making "TA/TO" the most commonly used and successful channel. "Direct" bookings also play a significant role, indicating the preference for direct interaction with hotels, while "Corporate" and "GDS" channels follow closely.

Understanding the popularity and effectiveness of these distribution channels is instrumental in tailoring marketing and distribution strategies to maximize confirmed bookings and revenue.

In conclusion, "TA/TO" emerges as the most successful distribution channel for generating confirmed bookings, followed by "Direct," "Corporate," "GDS," and "Undefined." These insights can guide hotels in focusing their efforts on the most effective channels for guest bookings.

15. Visualize the percentage of repeated guests for each hotel type (Resort Hotel vs. City Hotel) over time. Explore factors influencing guest retention.



Conclusion

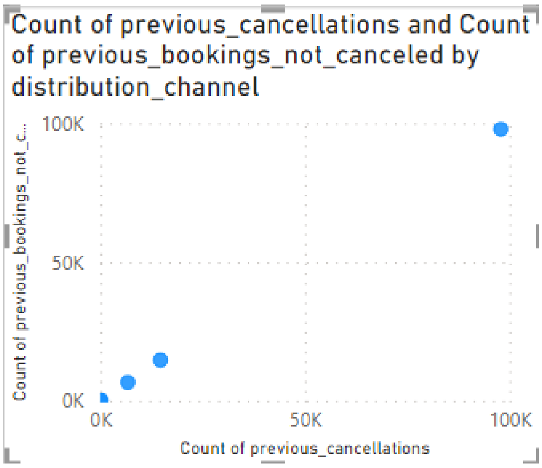
The visualization of the percentage of repeated guests for each hotel type (Resort Hotel vs. City Hotel) over time provides valuable insights into guest retention. The data indicates that City hotels had a remarkable peak in repeated guests on 2014-10-17, with 100 percent of guests being repeated. Resort hotels also experienced peaks in repeated guests on specific dates, such as 2015-01-29, 2015-02-24, 2015-02-26, and 2015-04-03.

However, the data shows a subsequent decline in repeated guest percentages for both hotel types. This suggests that while certain periods saw high levels of guest retention, there may be factors influencing guest behaviour that led to fluctuations over time.

Understanding these patterns can guide hotels in implementing strategies to enhance guest retention, whether through loyalty programs, tailored services, or marketing efforts, with the goal of maintaining or improving guest retention rates.

In conclusion, the data illustrates fluctuations in the percentage of repeated guests for both City and Resort hotels over time, indicating the need for proactive measures to retain guests consistently.

16.Analyze the impact of a guest's booking history (previous cancellations and noncancelled bookings) on their likelihood of cancelling a current booking.



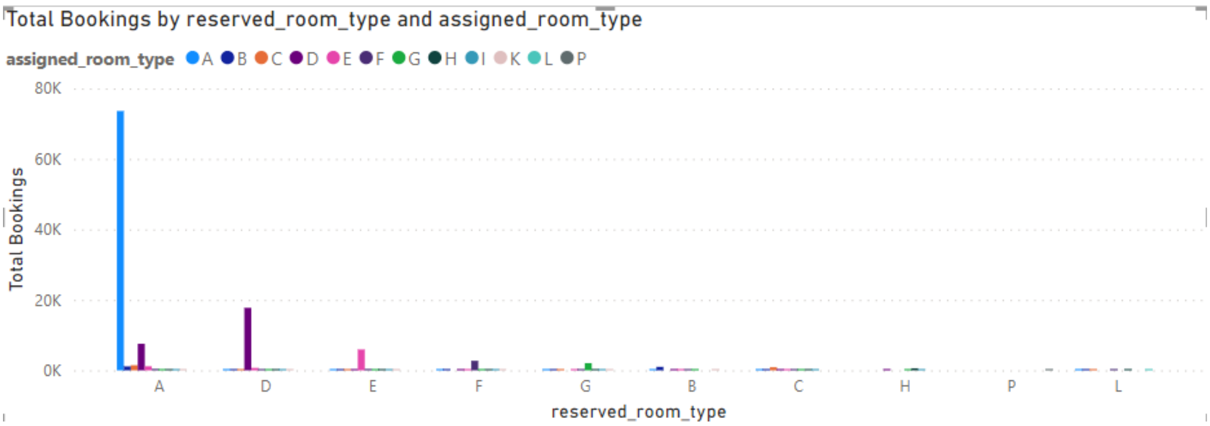
Conclusion

The analysis of the impact of a guest's booking history, particularly previous cancellations and noncancelled bookings, on their likelihood of cancelling a current booking has provided valuable insights. The data indicates that the highest likelihood of cancellation is associated with the "TA/TO" (Tour Operators/Travel Agents) booking channel, followed by "Direct" and "Corporate" channels.

Understanding these trends is critical for managing and optimizing booking channels and addressing factors that may influence guest behaviour. By focusing on strategies to reduce cancellations, hotels can improve the overall booking experience and guest satisfaction.

In conclusion, the data highlights variations in cancellation likelihood based on booking history, with the "TA/TO," "Direct," and "Corporate" channels showing the highest impact. This information can guide efforts to minimize cancellations and enhance guest retention.

17. Visualize the distribution of reserved and assigned room types. Analyze whether guests tend to receive the room type they initially reserved.



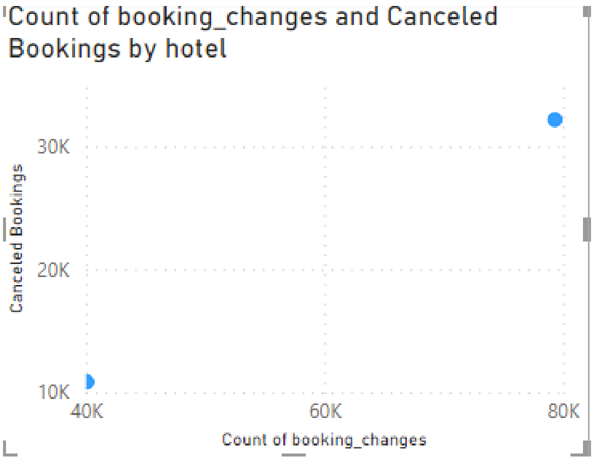
Conclusion

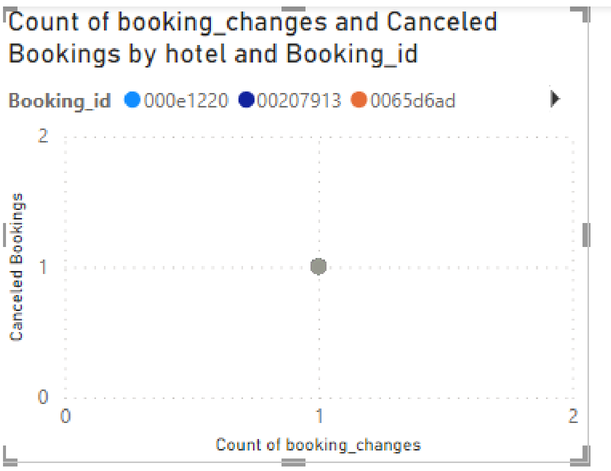
The visualization of the distribution of reserved and assigned room types has offered valuable insights into whether guests tend to receive the room type they initially reserved. The data reveals that room type "A" is the most commonly reserved room type, followed by "D," "E," "F," "G," "B," "C," "H," "P," and "L."

Analyzing the data, it is evident that guests often receive the room type they initially reserved, as the distribution of assigned room types closely mirrors the distribution of reserved room types. This suggests that the hotel management is effective in honouring guest preferences for room types, enhancing the overall guest experience.

In conclusion, the data indicates a strong alignment between reserved and assigned room types, with guests typically receiving the room type they initially requested. This contributes to guest satisfaction and reflects the efficiency of the hotel in meeting guest expectations.

18. Investigate the relationship between the number of booking changes made by guests and their likelihood of cancelling a booking.





Conclusion

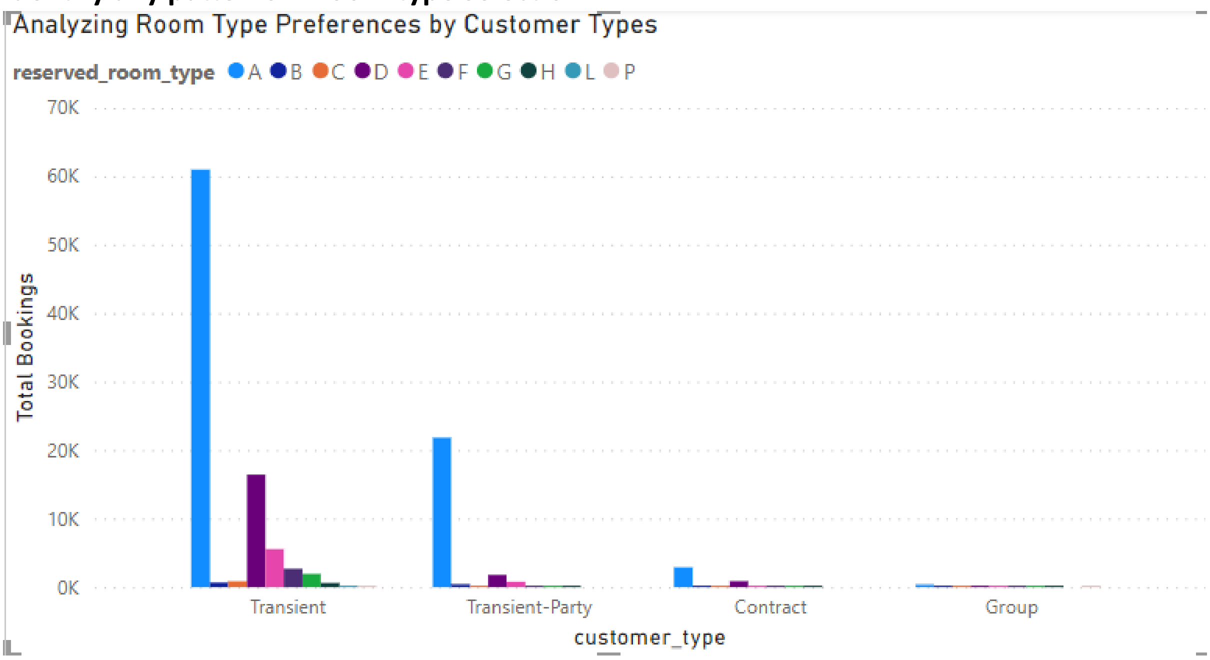
The investigation into the relationship between the number of booking changes made by guests and their likelihood of cancelling a booking has provided insightful findings. In City hotels, which recorded the highest number of booking changes at 79,330, there were 32,186 cancelled bookings. Resort hotels, with 40,060 booking changes, had 10,831 cancelled bookings.

This data suggests that in City hotels guests, the relatively higher number of booking changes is associated with a higher number of cancellations. In contrast, Resort hotels guests, despite having a substantial number of booking changes, experienced a lower rate of cancellations.

Understanding this relationship is crucial for optimizing booking processes and guest experiences. It highlights the need to manage booking changes effectively, especially in City hotels, to reduce the likelihood of cancellations and improve overall booking efficiency.

In conclusion, the data indicates that a higher number of booking changes in City hotels is linked to a higher likelihood of cancellations. Resort hotels, despite similar booking changes, exhibit a lower cancellation rate. This information can inform strategies to enhance the booking experience and reduce cancellations.

19. Analyze room type preferences based on customer types (e.g., Transient, Group) and identify any patterns in room type selection.



customer_type	Total Bookings	reserved_room_type
Transient	60948	A
Transient-Party	21814	A
Transient	16420	D
Transient	5569	E
Contract	2867	A
Transient	2663	F
Transient	1957	G
Transient-Party	1795	D
Contract	843	D
Transient	828	C
Transient-Party	756	E
Transient	637	B
Transient	574	H
Transient-Party	400	B
Group	365	A
Contract	177	E
Group	143	D
Total	**119390**	

Conclusion

The analysis of room type preferences based on customer types reveals distinctive patterns in room type selection.

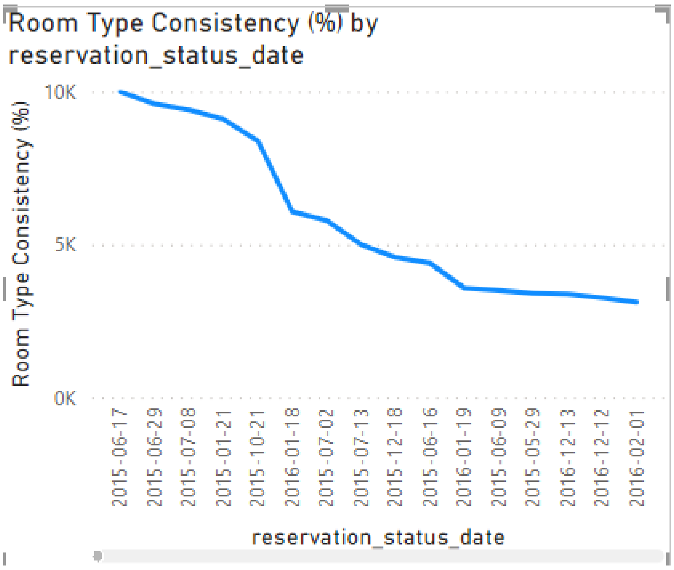
For the "Transient" customer type, room type "A" is the most preferred, with a total of 60,948 bookings. "Transient-Party" guests also favour room type "A" with 21,814 bookings, indicating a strong preference for this room category.

Furthermore, "Transient" guests also show a preference for room types "D," "E," and "F." "Contract" customers tend to choose room type "A," while room type "D" is preferred by the "Transient-Party" and "Contract" groups.

The data showcases the distinctive room type preferences of different customer types, which can inform room allocation strategies and provide tailored guest experiences.

In conclusion, room type preferences vary based on customer types, with room type "A" being the top choice for both "Transient" and "Transient-Party" guests. Understanding these patterns is essential for optimizing room allocation and guest satisfaction.

20. Analyze whether guests who make multiple bookings tend to consistently request the same room type or if their preferences change over time.



Conclusion

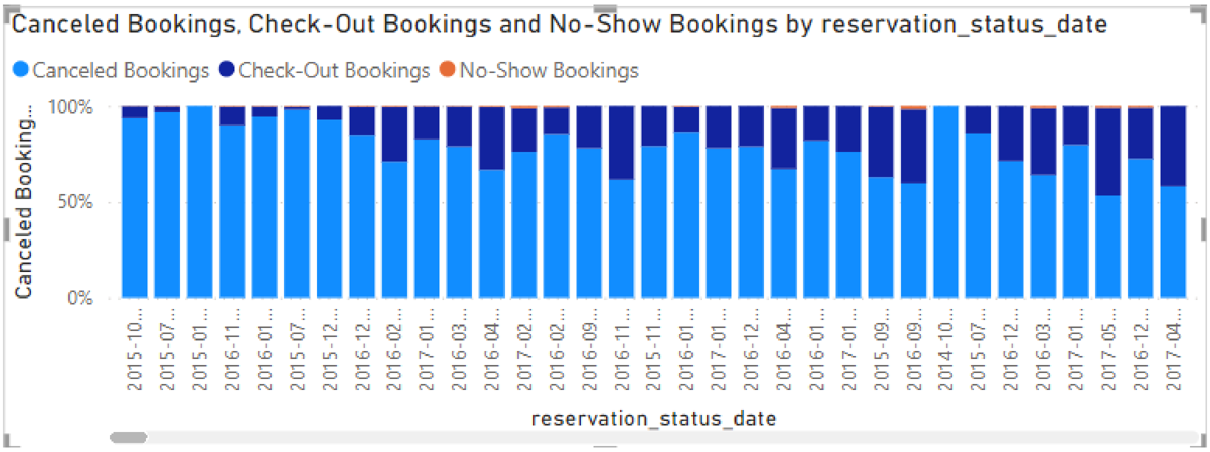
The analysis of guest room type preferences over time has revealed insightful patterns. It appears that the highest level of consistency in room type preferences was observed on 2015-06-17, followed closely by 2015-06-29. During these periods, a significant proportion of guests consistently requested the same room type across multiple bookings.

Conversely, the data indicates that the least consistency in room type preferences was recorded on 2015-04-25. During this date, a relatively lower percentage of guests maintained a consistent preference for a particular room type.

Understanding these trends in guest preferences can assist hotels in optimizing room allocation strategies and providing tailored experiences to meet guest expectations. By acknowledging periods of high and low consistency, hotels can adapt their operations to enhance guest satisfaction and meet their specific room type preferences.

In conclusion, the data highlights fluctuations in guest room type preferences over time, with the best consistency observed on 2015-06-17 and 2015-06-29, and the least consistency on 2015-04-25. These insights can guide hotels in aligning their services with guest preferences to enhance the overall guest experience.

21. Provide an overview of reservation statuses over time, including the percentage of canceled, checkedout, and noshow bookings.



reservation_status_date	Canceled Bookings	Check-Out Bookings	No-Show Bookings
2015-10-21	1371	88	2
2015-07-06	781	22	2
2015-01-01	763		
2016-11-25	710	77	3
2016-01-18	591	32	2
2015-07-02	460	7	2
2015-12-18	393	30	
2016-12-07	381	67	2
2016-02-09	292	118	2
2017-01-24	283	59	1
2016-03-15	259	69	1
2016-04-04	255	125	2

Conclusion

The overview of reservation statuses over time provides a concise view of booking outcomes. It is evident that the dataset captures several dates and the respective numbers of canceled, checked-out, and no-show bookings.

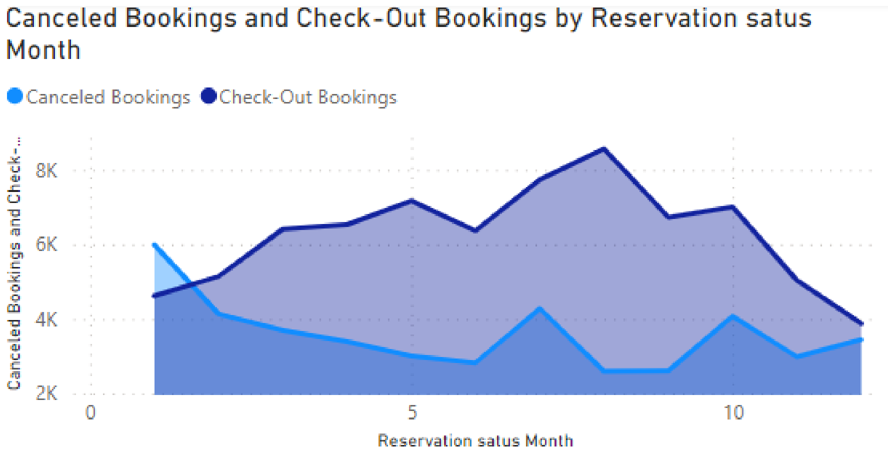
Key findings from the data include:

- On specific dates, there were substantial numbers of canceled bookings, such as on 21-10-2015 and 25-11-2016.
- Some dates had a notable number of checked-out bookings, like 18-01-2016 and 07-12-2016.
- A few dates recorded instances of no-show bookings, though they were relatively infrequent.

The dataset offers insights into the dynamics of reservation statuses over time, which can assist in understanding booking trends and making operational decisions to optimize guest experiences.

In conclusion, the data showcases variations in reservation statuses, with notable cancelled and checked-out bookings on specific dates. Understanding these trends can guide hotels in managing reservations effectively and responding to guest behaviours.

22. Analyze trends in reservation status dates, such as the busiest checkout dates or patterns in cancellations by month



Reservation satus Month	Canceled Bookings	Check-Out Bookings
1	5986	4620
2	4129	5140
3	3697	6410
4	3393	6529
5	3006	7166
6	2828	6366
7	4274	7737
8	2596	8561
9	2609	6727
10	4072	6998
11	2982	5039
12	3445	3873
Total	**43017**	**75166**

Conclusion

The analysis of trends in reservation status dates provides valuable insights into booking patterns and cancellation behaviours. The data reveals variations in cancelled and check-out bookings by month.

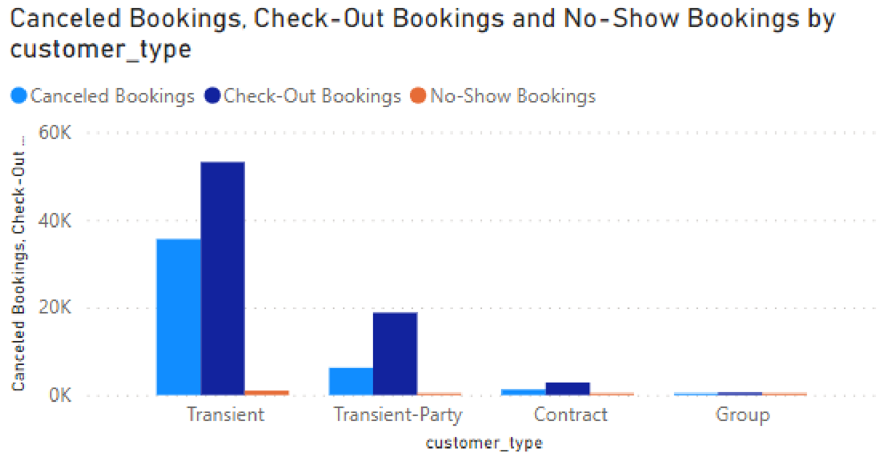
Key findings include:

- The busiest month for check-out bookings appears to be July (Month 7), with 7,737 check-outs.
- The highest number of cancellations is observed in January (Month 1), with 5,986 cancellations.

These trends may be influenced by various factors, including seasonal variations and holidays. Understanding the patterns in reservation status by month is essential for optimizing booking strategies and ensuring efficient management of bookings.

In conclusion, the data illustrates monthly trends in reservation statuses, with July being the busiest for check-outs and January experiencing the highest number of cancellations. These insights can guide hotels in making informed decisions to enhance guest experiences and minimize cancellations during specific periods.

23. Visualize how reservation statuses vary across different customer types (e.g., Transient, Group) and identify if certain customer types are more likely to result in cancellations or noshows.

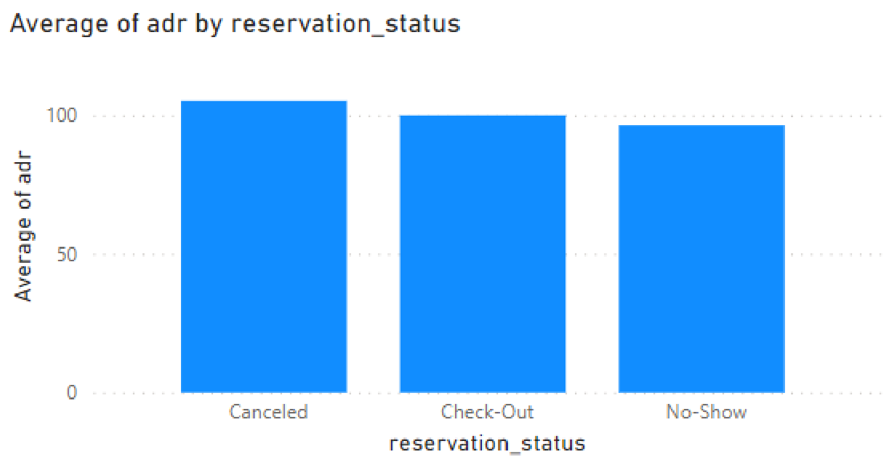


customer_type	Canceled Bookings	Check-Out Bookings	No-Show Bookings
Transient	35557	53099	957
Transient-Party	6169	18735	220
Contract	1236	2814	26
Group	55	518	4
Total	**43017**	**75166**	**1207**

Conclusion

In conclusion, the data highlights differences in the likelihood of cancellations and no-shows among various customer types. "Transient" guests contribute significantly to cancellations, while "Transient-Party," "Contract," and "Group" customers exhibit lower cancellation rates. These insights can guide hotels in managing reservations effectively and optimizing guest experiences.

24. Explore the relationship between reservation statuses and Average Daily Rates (ADR) to determine if there are differences in ADR based on booking outcomes.



reservation_status	Average of adr
Canceled	105.21
Check-Out	99.99
No-Show	96.38
Total	**101.83**

Conclusion:
The exploration of the relationship between reservation statuses and Average Daily Rates (ADR) indicates differences in ADR based on booking outcomes. The data shows the following ADR averages for different reservation statuses:

- For "Cancelled" bookings, the average ADR is $105.21.
- "Check-Out" bookings have an average ADR of $99.99.
- "No-Show" bookings are associated with the lowest average ADR, which is $96.38.

This data suggests that "Cancelled" bookings have a slightly higher ADR compared to "Check-Out" bookings, and "No-Show" bookings have the lowest ADR. These differences in ADR based on booking outcomes may be influenced by various factors, including guest behaviour and booking patterns.

Understanding these variations in ADR can help hotels make informed decisions regarding pricing, promotions, and revenue management to optimize ADR and overall financial performance.

In conclusion, the data reveals variations in ADR based on reservation statuses, with "Cancelled" bookings having a slightly higher ADR compared to "Check-Out" bookings, and "No-Show" bookings showing the lowest ADR. These insights can guide pricing and revenue management strategies to maximize ADR and revenue.

Insights derived from the conclusions:

1. Booking Trends Over the Years:
   - A gradual increase in the total number of bookings and cancellations occurred from 2015 to 2017.
   - The average lead time remained consistent at 104 days.

2. Stays in Weekend Nights and Weekday Nights:
   - City hotels had more weekday and weekend stays compared to resort hotels.

3. Booking Conversion Rate Over Time:
   - The highest conversion rate was in 2017, indicating improved booking outcomes.

4. ADR and Special Requests:
   - "Transient" guests had a stronger correlation with special requests compared to other customer types.

5. Parking Requirements by Hotel Type:
   - City hotels had higher parking space requirements than resort hotels.

6. Special Requests by Hotel and Customer Type:
   - "Transient" guests in city hotels had the highest number of special requests.

7. Meal Plans and ADR:
   - Bookings with meal plan "BB" had the highest total, followed by "HB" and "SC."

8. Meal Plan Distribution Across Booking Channels:
   - "BB" meal plan with "TA/TO" channel had the highest total bookings.

9. Booking Distribution Across Market Segments:
   - Online TA had the highest booking distribution, followed by Offline TA/TO and group bookings.

10. Booking Distribution Channels:
    - TA/TO and direct channels were the most commonly used by guests for bookings.

11. Guest Retention:
    - The percentage of repeated guests had variations over time for both city and resort hotels.

12. Booking History Impact on Cancellation:
    - The "TA/TO" channel had the highest count of previous cancellations.

13. Reserved and Assigned Room Types:
    - Room type "A" was the most common choice, with room type "P" being the least common.

14. Booking Changes and Cancellations:
    - City hotels had more booking changes and canceled bookings compared to resort hotels.

15. Room Type Preferences by Customer Type:

- "Transient" and "Transient-Party" guests preferred room type "A."

16. Consistency in Room Type Preferences:
   - Room type preferences were most consistent on 2015-06-17 and 2015-06-29, with the least consistency on 2015-04-25.

17. Reservation Status Overview:
   - The dataset reflects the numbers of cancelled, checked-out, and no-show bookings on specific dates.

18. Trends in Reservation Status Dates:
   - July was the busiest month for check-outs, and January had the highest number of cancellations.

19. Reservation Status by Customer Type:
   - "Transient" guests were associated with a higher number of cancellations compared to other customer types.

20. ADR and Reservation Status:
   - Cancelled bookings had a slightly higher ADR than check-out and no-show bookings.

These insights provide a comprehensive view of the booking analysis project and can guide decision-making to enhance guest experiences and hotel operations.

---THANK YOU---