



# *DATA ANALYSIS*

# Exploratory Data Analysis

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypotheses and to check assumptions with the help of summary statistics and graphical representations.

In statistics, exploratory data analysis (EDA) is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling and thereby contrasts traditional hypothesis testing. Exploratory data analysis has been promoted by John Tukey since 1970 to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments. EDA is different from initial data analysis (IDA), which focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA.



## Steps involved in EDA :-

- **Data Analysis (Data Pre-Processing, Cleaning and Manipulation).**
- **Data Visualisation (Visualize relationships in data using different types of plots).**

# HOTEL BOOKING

This is an exploratory analysis of The Superstore Sales Dataset from JESSE MOSTIPAK on Kaggle

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

All personally identifying information has been removed from the data.

The data set that is going to be used is the 'Hotel Booking Dataset'. It contains Hotel Booking data, we are going to clean the data if required and perform Exploratory Data Analysis(EDA). The tools that are going to be used for the EDA would be NumPy, Pandas, Matplotlib and Seaborn.

# Dataset Description

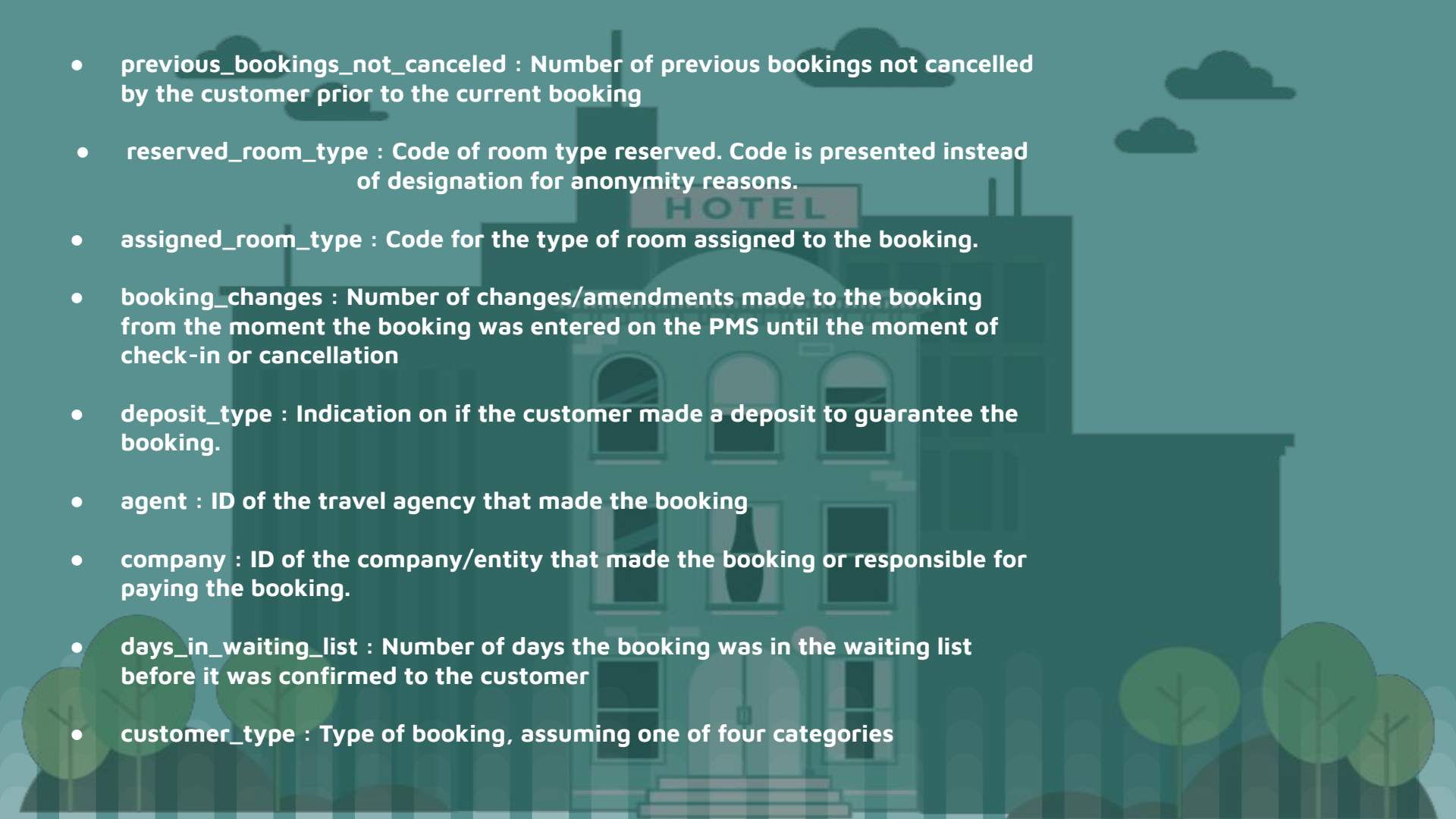
- **hotel** : Hotel(There is only two categories one is resort hotel and other is city hotel)
- **is\_canceled** : The value represents 1 if the booking is canceled or it represents 0 .
- **lead\_time** :It is the number of days between the time a guest books their room and the time they are scheduled to arrive at the hotel.
- **arrival\_date\_year** : It indicates the year in which the guest will arrive to the hotel.
- **arrival\_date\_month** : It indicates the month in which guest will arrive to the hotel.
- **arrival\_date\_week\_number** : This indicates the week number in which the guest will arrive to the hotel
- **arrival\_date\_day\_of\_month** : This indicates the day of the month in which the guest will arrive to the hotel.
- **stays\_in\_weekend\_nights** : Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
- **stays\_in\_week\_nights** : Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel



- 
- **adults** : Number of adults
  - **children** : Number of children
  - **babies** : Number of babies
  - **meal** : Type of meal booked . Categories are presented in standard hospitality meal packages:

(BB:"Bed & Breakfast"; HB:"Bed,Breakfast & Dinner ; FB: "Bed & All meals(Breakfast,Lunch & Dinner)

- **country** : Country of origin.
- **market\_segment** : Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"
- **distribution\_channel** : Booking distribution channel. The term "TA" means "Travel Agents" and "TO" means "Tour Operators"
- **is\_repeated\_guest** : The value represents 1 if the booking name was from a repeated guest or it represents 0 .
- **previous\_cancellations** : Number of previous bookings that were cancelled by the customer prior to the current booking

- 
- **previous\_bookings\_not\_canceled** : Number of previous bookings not cancelled by the customer prior to the current booking
  - **reserved\_room\_type** : Code of room type reserved. Code is presented instead of designation for anonymity reasons.
  - **assigned\_room\_type** : Code for the type of room assigned to the booking.
  - **booking\_changes** : Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation
  - **deposit\_type** : Indication on if the customer made a deposit to guarantee the booking.
  - **agent** : ID of the travel agency that made the booking
  - **company** : ID of the company/entity that made the booking or responsible for paying the booking.
  - **days\_in\_waiting\_list** : Number of days the booking was in the waiting list before it was confirmed to the customer
  - **customer\_type** : Type of booking, assuming one of four categories

- **adr** : Average Daily Rate (ADR) measures the average rental revenue earned for an occupied room per day .

"The operating performance of a hotel or other lodging business can be determined by using the ADR."

- **required\_car\_parking\_spaces** : Number of car parking spaces required by the customer
- **total\_of\_special\_requests**: Number of special requests made by the customer (e.g. twin bed or high floor)
- **reservation\_status** : Reservation last status, assuming one of three categories

Canceled – booking was canceled by the customer

Check-Out – customer has checked in but already departed

No-Show – customer did not check-in and did inform the hotel of the reason why

- **reservation\_status\_date** : Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel



# ***PERFORMING EDA AND TRIES TO ANSWERING FOLLOWING QUESTIONS :-***

- Which type of Hotel does the guest prefer the most ?
- What is the percentage of canceled or not canceled of hotel bookings ?
- Top 10 Countries Earned most Revenue from Hotel Bookings ?
- In which year hotels earned more revenue ?
- In which month the more hotel booked ?
- In which week the more hotel booked ?
- How many types of market segment of hotel booking present ?

- Different types of meals preferred by customers ?
- Different types of customer who booked hotel ?
- How many types of distribution channel of hotel booking ?
- Top 10 Agents who Booked most of the hotels ?
- Different types of Reservation in hotel ?
- How many times the rooms are reserved by customer ?

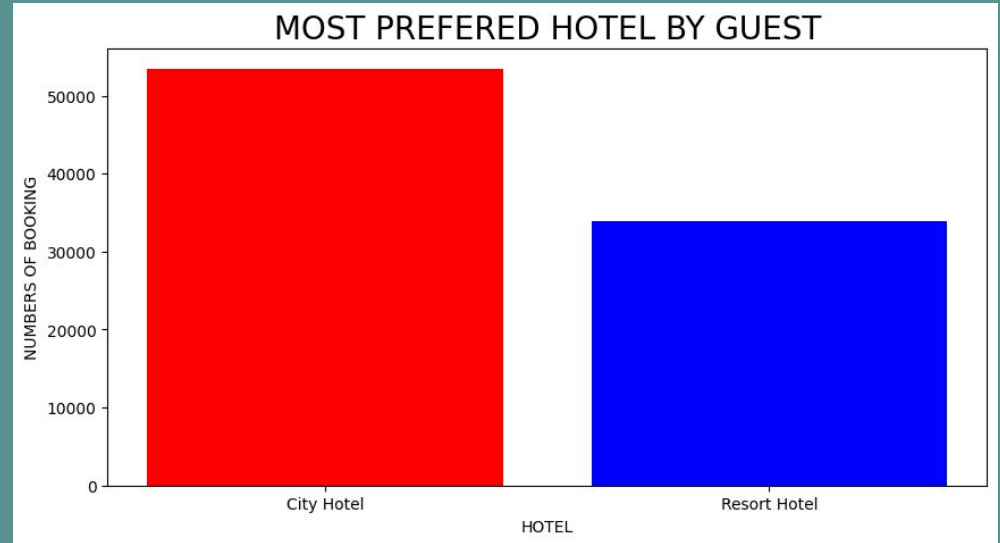
- What is the percentage of repeated guest in hotel ?
- What is the percentage of guest who required car parking spaces ?
- How many times types of room are assigned by hotels ?
- What is the Percentage distribution of deposit type?
- What is the distribution of booking changes made by guests?

The background is a solid teal color. Several light bulbs are hanging from the top by thin black lines. Most bulbs are dark grey with a small horizontal line inside, representing an unlit bulb. One bulb at the bottom center is glowing, indicated by a yellow lightning bolt inside and several short lines radiating outwards. At the bottom of the image, there is a row of vertical bars of varying heights, resembling a bar chart or a data visualization.

# INSIGHTS

# Which type of Hotel does the guest prefer the most ?

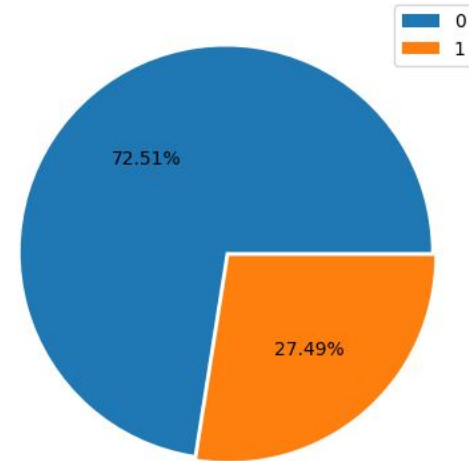
- There are 53428 times City Hotel preferred by guest.
- There are 33968 times Resort Hotel preferred by guest.



# What is the percentage of canceled or not canceled of hotel bookings ?

- The Hotel Booking were not canceled 63371 times. The Percentage of Not Canceled Hotel booking is 72.51%
- The Hotel Booking were canceled 24025 times. The Percentage of Canceled Hotel booking is 27.49%

PERCENTAGE OF CANCELED/NOT CANCELED OF HOTEL BOOKING

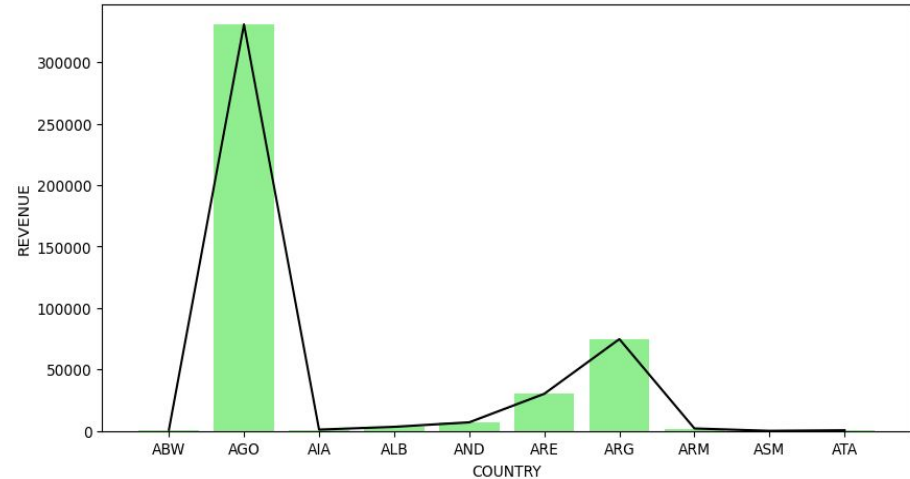




# Top 10 Countries Earned most Revenue from Hotel Bookings ?

- AGO is the one country who earned most of revenue from hotel bookings.
- The Top 10 Countries who earned most of revenue are AGO, ARG, ARE, AND, ALB, ARM, AIA, ABW, ASM and ATA.

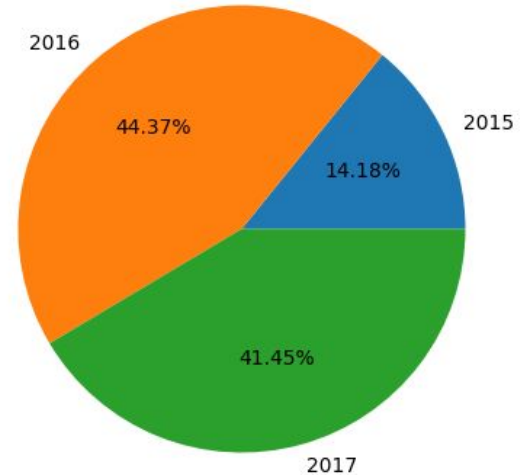
TOP 10 COUNTRIES EARNED MOST REVENUE FROM HOTEL BOOKINGS



# In which year hotels earned more revenue ?

- The hotel earned more revenue in 2016 year. The revenue of year 2016 was 15291176 rupees.
- The hotel earned average revenue in 2017 year. The revenue of year 2016 was 14284246 rupees.
- The hotel earned less revenue in 2015 year. The revenue of year 2016 was 4885517 rupees.

REVENUE EARNED BY HOTELS IN DIFFERENT YEAR



# In which month the more hotel booked ?

- The hotel had more booking in August Month.
- The hotel had average booking in October Month.
- The hotel had less booking in January Month.



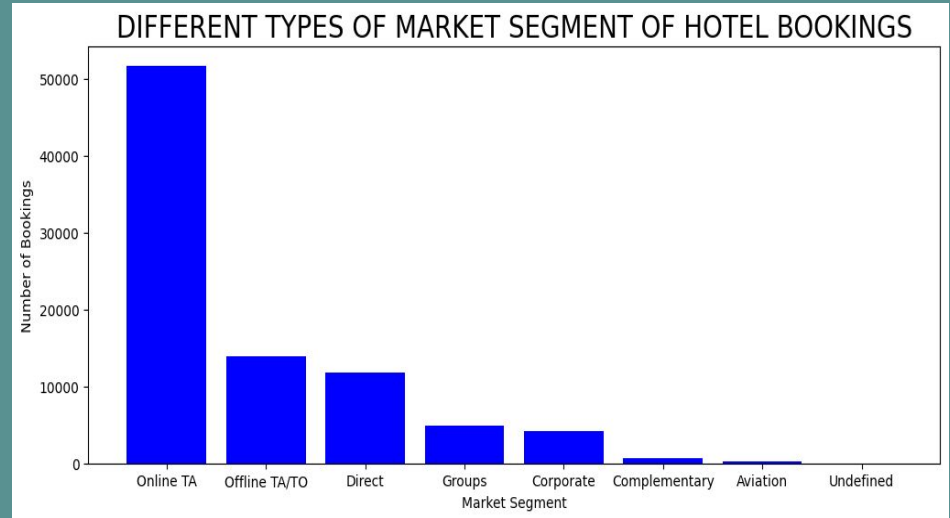
# In which week the more hotel booked ?

- The hotel had more booking in 33 Week.
- The hotel had average booking in 51 Week.
- The hotel had less booking in 25 Week.



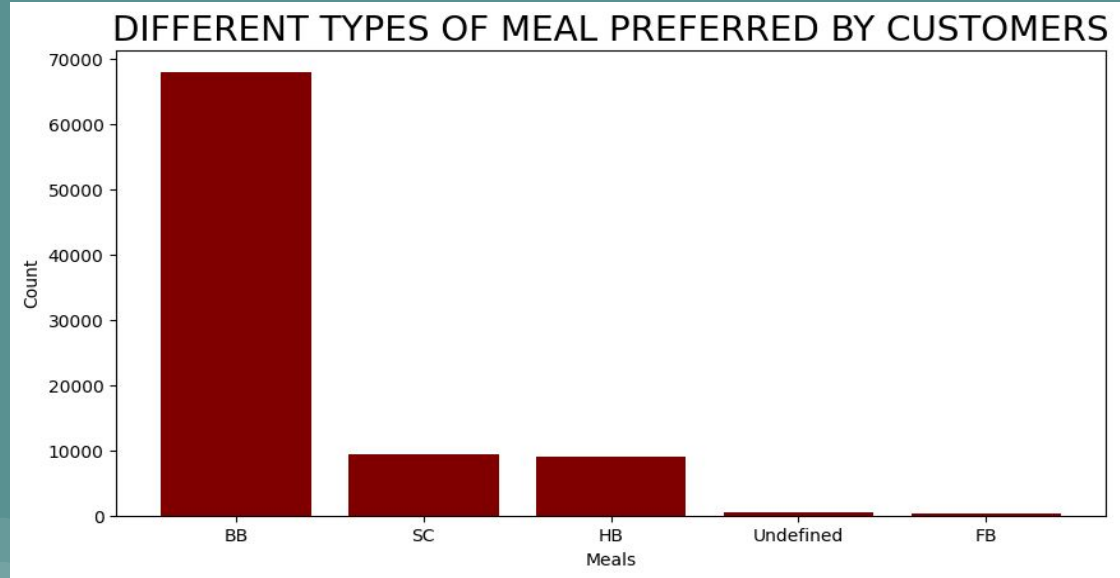
# How many types of market segment of hotel booking present ?

- Online TA is a most famous type of market segment for Hotel Booking.
- The Three most famous type of market segment for Hotel Booking are Online TA, Offline TA/To, and Direct.
- Undefined and Aviation are least famous type of market segment for Hotel Booking.



# Different types of meals preferred by customers ?

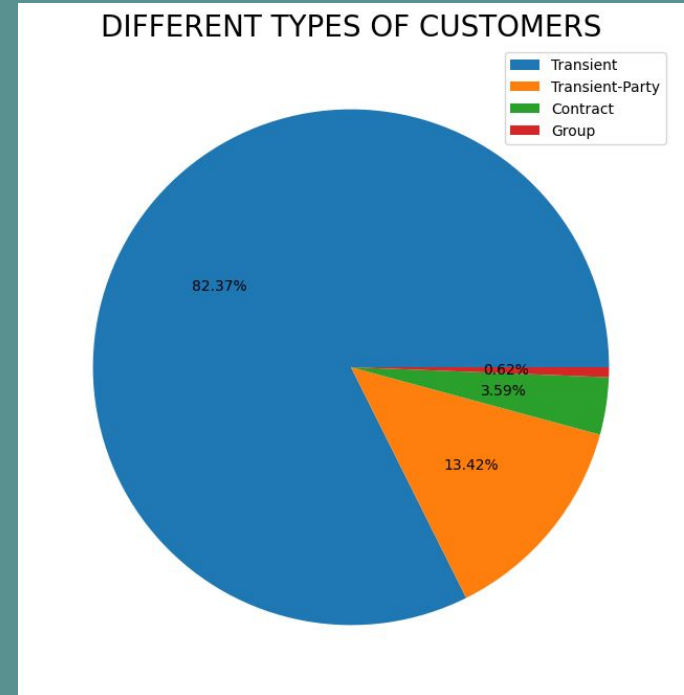
- Most preferred meal type is BB.
- Least preferred meal type is FB





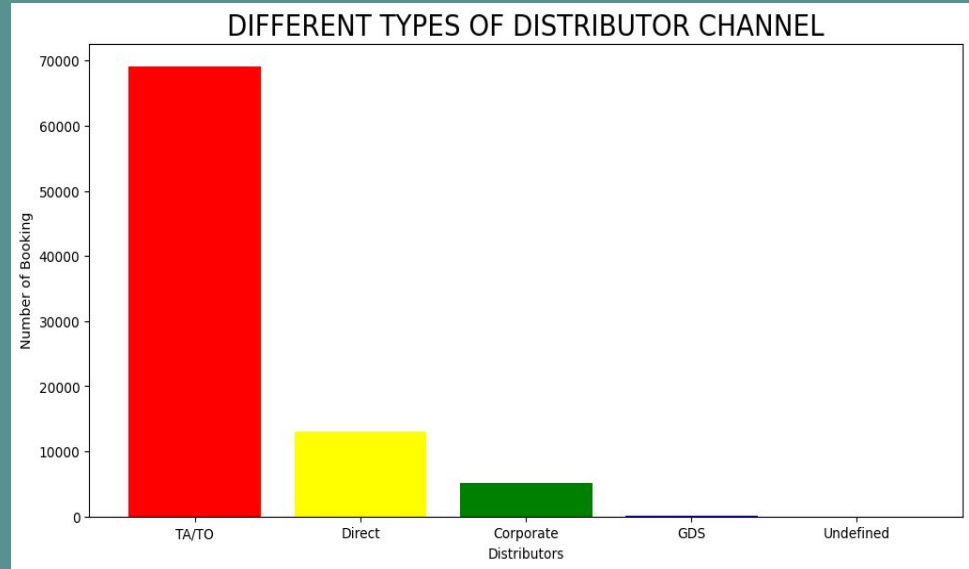
# Different types of customer who booked hotel ?

- Transient Customer have more stake in booking of hotels. Transient Customer have 82.37% stake between all.
- Group Customer have least stake in booking of hotels. Group Customer have 0.62% stake between all.



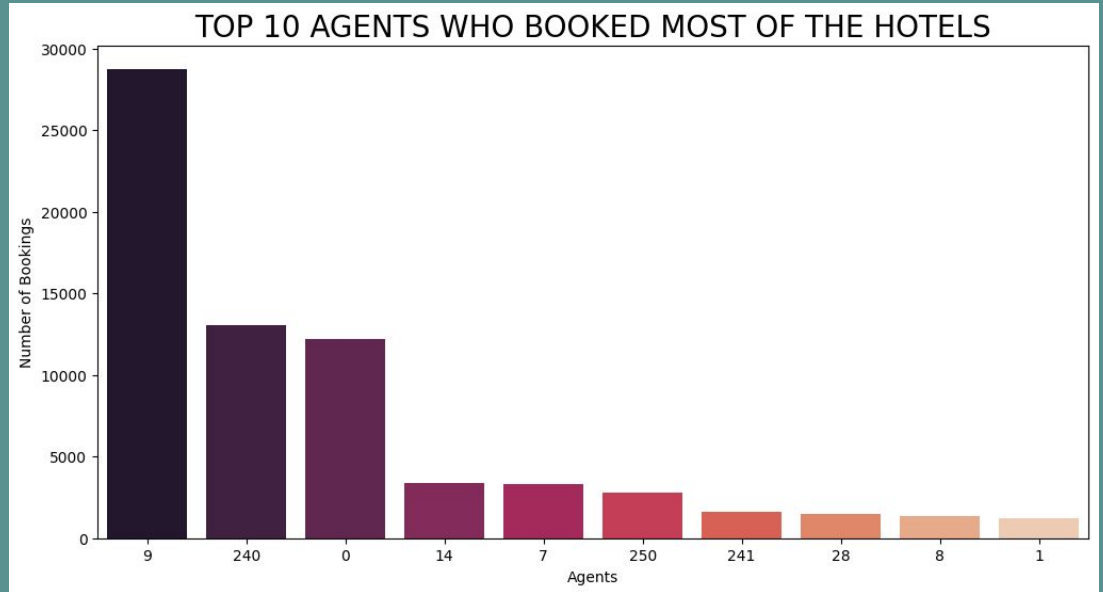
# How many types of distribution channel of hotel booking ?

- TA/TO is a most famous distributor channel for Hotel Booking.
- The Three most famous type of market segment for Hotel Booking are Online TA/TO, Direct, and Corporate.
- Undefined and GDS are least famous type of market segment for Hotel Booking.



# Top 10 Agents who Booked most of the hotels ?

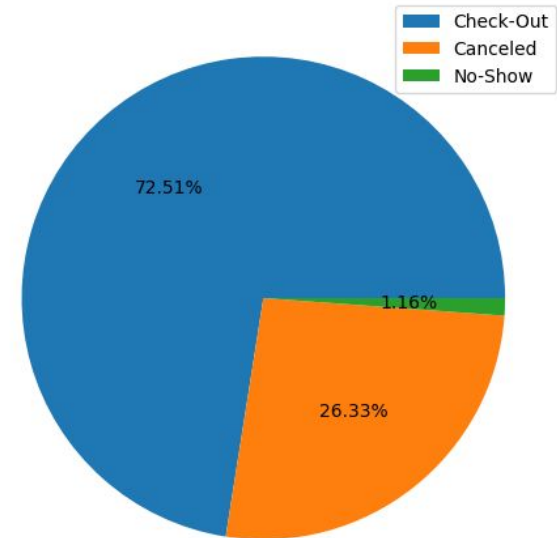
- 9 is the one agent who booked most of hotel.
- The Top Three Agent who booked most of hotel are 9, 240, and 0.



# Different types of Reservation in hotel ?

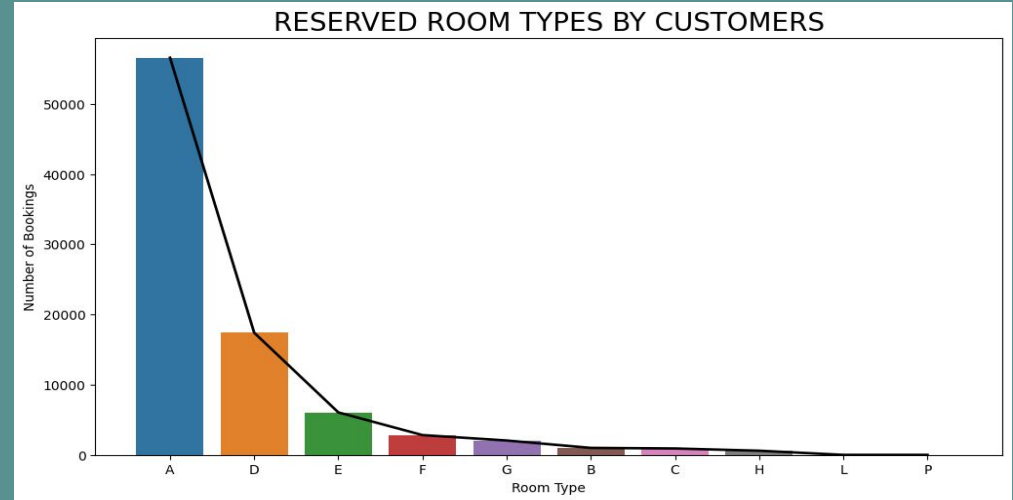
- Check-Out have more stake in reservation status of hotels. Transient Customer have 72.51% stake between all.
- No-Show have less stake in booking of hotels. No-Show have 1.16% stake between all.

DISTRIBUTION OF RESERVATION STATUS



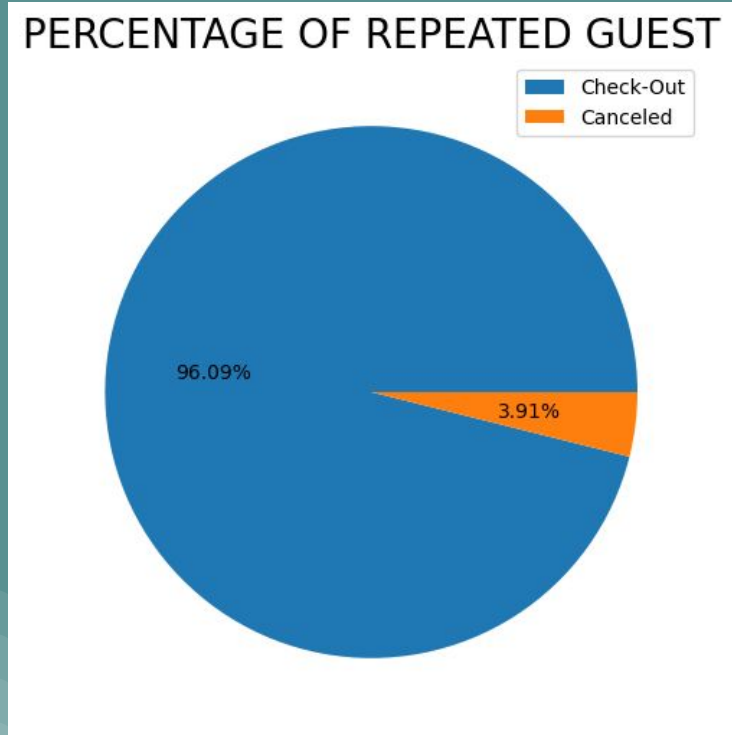
# How many times the rooms are reserved by customer ?

- Room A is the most reserved room type between them.
- Room L and P are the least reserved room type between them.



# What is the percentage of repeated guest in hotel ?

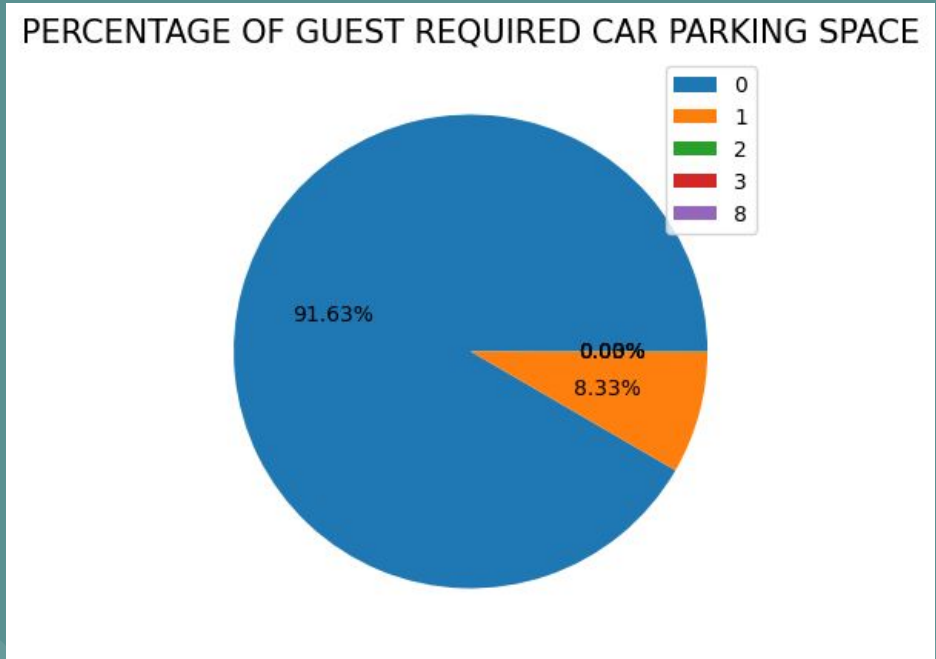
- There is a 3.91% are repeated guest.
- There is a 96.09% are not repeated guest.





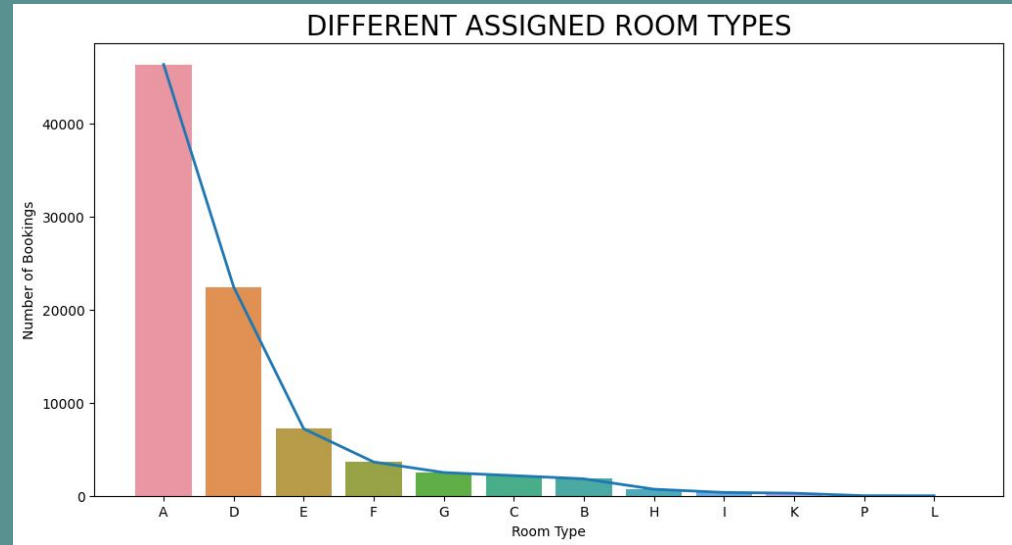
# What is the percentage of guest who required car parking spaces ?

- The percentage of guest whose didn't required car parking space is 91.63%.
- The percentage of guest whose required car parking space is 8.33%.



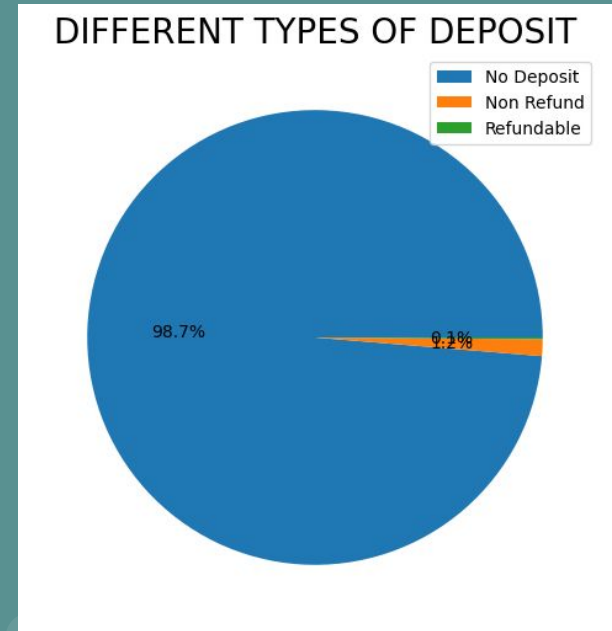
# How many times types of room are assigned by hotels ?

- Room A is the most assigned room type between them.
- Room P and L are the least assigned room type between them.



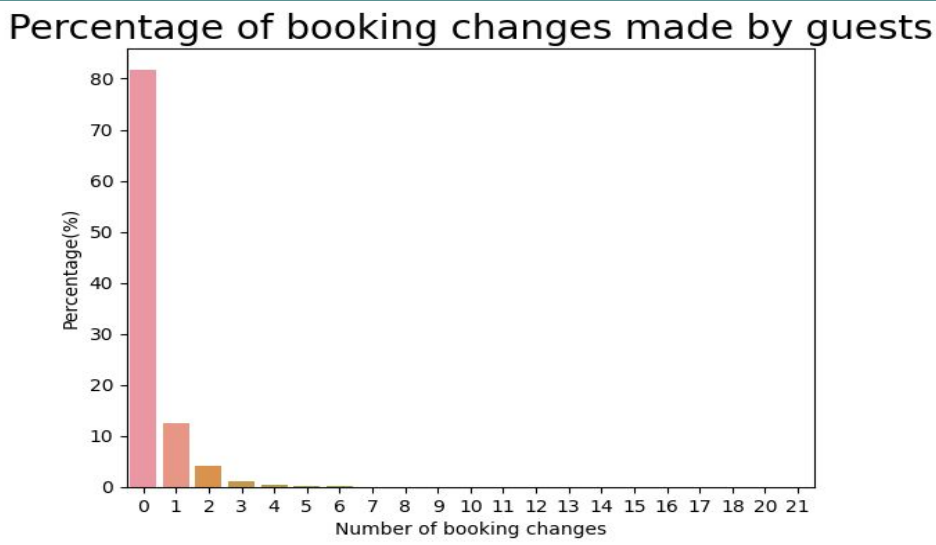
# What is the Percentage distribution of deposit type?

- Around 98.7% booking required no deposit.
- Around 1.2 % has no refund and only 0.1% has refundable.

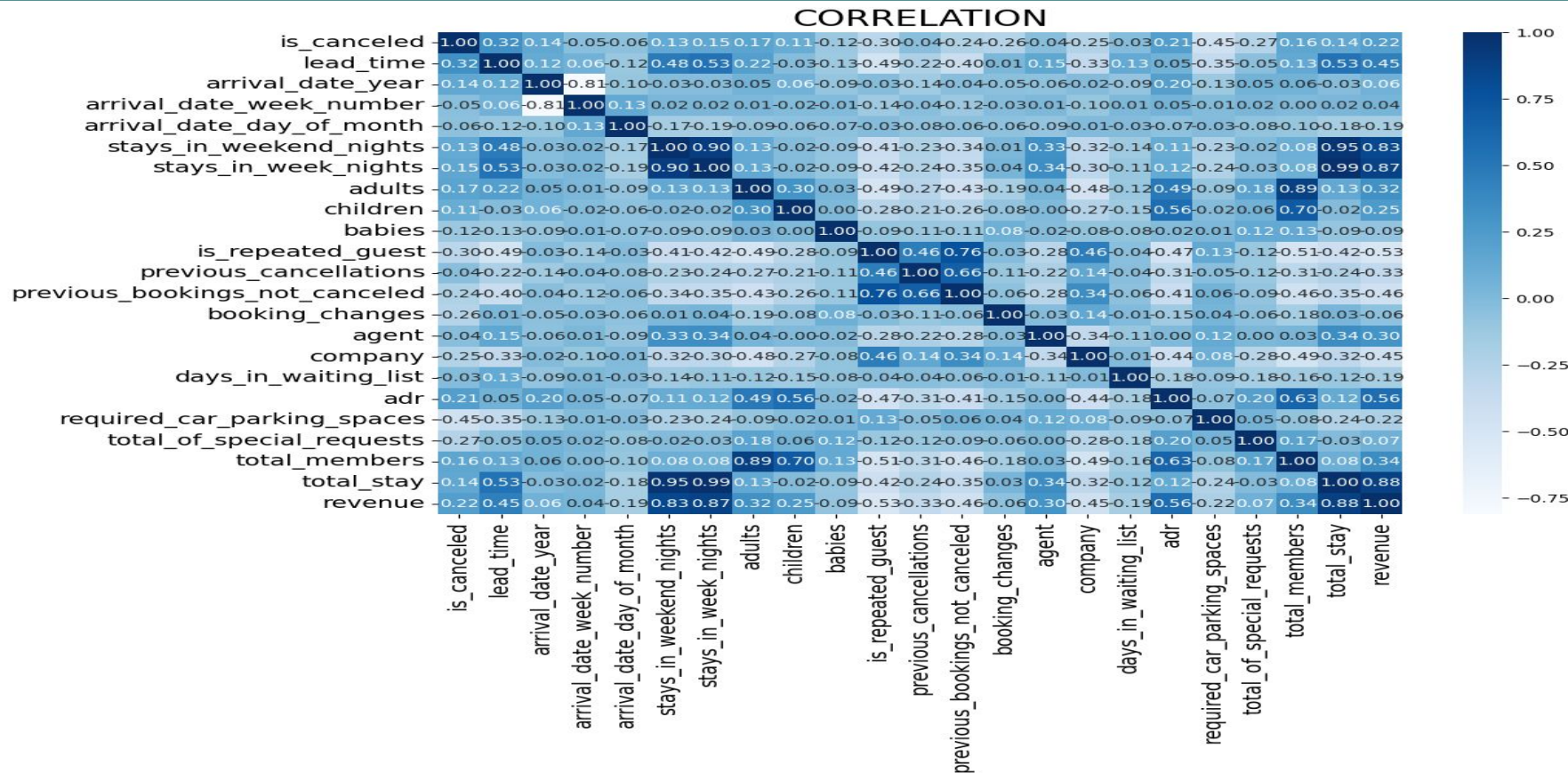


# What is the distribution of booking changes made by guests?

- More than 82% of bookings were not changed.
- Around 12% of guest changed 1 booking and less than 5% has made 2.



# CORRELATION



- Total\_stay and lead time have slight correlation it. This might means that customer plan reservation before their actual arrival.
- Adr(Average Daily Rate) is slightly correlated with total people, which makes sense as more number of people means more revenue.
- Previous booking not cancelled and repeated guest have high correlation as repeated guest are most likely to be who has not cancelled their previous bookings.
- We can also see that some columns have high correlation between them for example total people with adult and children it is because as those column are dervied from them by addition.



# *Data Analysis Conclusion of Hotel Booking Dataset*

- Which type of Hotel does the guest prefer the most ?

There are 53428 times City Hotel preferred by guest.  
There are 33968 times Resort Hotel preferred by guest.

- What is the percentage of canceled or not canceled of hotel bookings ?

The Hotel Booking were not canceled 63371 times. The Percentage of Not Canceled Hotel booking is 72.51%  
The Hotel Booking were canceled 24025 times. The Percentage of Canceled Hotel booking is 27.49%

- Top 10 Countries Earned most Revenue from Hotel Bookings ?

AGO is the one country who earned most of revenue from hotel bookings.  
The Top 10 Countries who earned most of revenue are AGO, ARG, ARE, AND, ALB, ARM, AIA, ABW, ASM and ATA.

- In which year hotels earned more revenue ?

The hotel earned more revenue in 2016 year. The revenue of year 2016 was 15291176 rupees.  
The hotel earned average revenue in 2017 year. The revenue of year 2016 was 14284246 rupees.  
The hotel earned less revenue in 2015 year. The revenue of year 2016 was 4885517 rupees.

- In which month the more hotel booked ?

The hotel had more booking in August Month.  
The hotel had average booking in October Month.  
The hotel had less booking in January Month.

- In which week the more hotel booked ?

The hotel had more booking in 33 Week.  
The hotel had average booking in 51 Week.  
The hotel had less booking in 25 Week.

- How many types of market segment of hotel booking present ?

Online TA is a most famous type of market segment for Hotel Booking.  
The Three most famous type of market segment for Hotel Booking are Online TA, Offline TA/To, and Direct.  
Undefined and Aviation are least famous type of market segment for Hotel Booking.

- Different types of meals preferred by customers ?

Most preferred meal type is BB.  
Least preferred meal type is FB.

- Different types of customer who booked hotel ?

Transient Customer have more stake in booking of hotels. Transient Customer have 82.37% stake between all. Group Customer have least stake in booking of hotels. Group Customer have 0.62% stake between all.

- How many types of distribution channel of hotel booking ?

TA/TO is a most famous distributor channel for Hotel Booking.

The Three most famous type of market segment for Hotel Booking are Online TA/TO, Direct, and Corporate. Undefined and GDS are least famous type of market segment for Hotel Booking.

- Top 10 Agents who Booked most of the hotels ?

9 is the one agent who booked most of hotel.

The Top Three Agent who booked most of hotel are 9, 240, and 0.

- Different types of Reservation in hotel ?

Check-Out have more stake in reservation status of hotels. Transient Customer have 72.51% stake between all. No-Show have less stake in booking of hotels. No-Show have 1.16% stake between all.

- How many times the rooms are reserved by customer ?

Room A is the most reserved room type between them.

Room L and P are the least reserved room type between them.

- What is the percentage of repeated guest in hotel ?

There is a 3.91% are repeated guest.

There is a 96.09% are not repeated guest.

- What is the percentage of guest who required car parking spaces ?

The percentage of guest whose didn't required car parking space is 91.63%.

The percentage of guest whose required car parking space is 8.33%.

- How many times types of room are assigned by hotels ?

Room A is the most assigned room type between them.

Room P and L are the least assigned room type between them.

- What is the Percentage distribution of deposit type?

Around 98.7% booking required no deposit and

Around 1.2 % has no refund and only 0.1% has refundable.

- What is the distribution of booking changes made by guests?

More than 82% of bookings were not changed.

Around 12% of guest changed 1 booking and less than 5% has made 2.

- CORRELATION

Total\_stay and lead time have slight correlation it. This might mean that customer plan reservation before their actual arrival.

Adr(Average Daily Rate) is slightly correlated with total people, which makes sense as more number of people means more revenue.

Previous booking not cancelled and repeated guest have high correlation as repeated guest are most likely to be who has not cancelled their previous bookings.

We can also see that some columns have high correlation between them for example total people with adult and children it is because as those column are derived from them by addition.

THANK  
YOU!