Hotel Bookings Exploratory Data Analysis

Technical Documentation:

By Nisha Kumari

Abstract:

Hotel industry is a very volatile industry and the booking depends on many factors.The main object of hotel analy sis is to explore and analyze data to discover important factors that govern the bookings and give insights to hotel management, which can performs various campaigns to boost the business and performance.

A.)Exploring Problem Parameters:

1. Examining the guest distribution.
2. Exploring the bookings.
3. Inspecting the hotel traffic.
4. Scanning special requests & amenities.
5. Study on average daily rates.

B.) Problem Statement:

We are provided with data of hotel bookings we had analyze data on following questions.

1. Types of customers with respect to the hotels.
2. Ratio of new customers & repeated customers.
3. Customers from different countries.
4. Change in bookings.
5. Total number of canceled Bookings by Hotel type.
6. Bookings ratio between Resort Hotel & City Hotel.
7. Total number of cancellation through market segments.
8. Total number of bookings through distribution Channels.
9. Most Busy year of Hotel Bookings.
10. Most Busiest Month of Hotel Bookings.
11. Guest coming from different Countries in resort Hotel.
12. Guest coming from different Countries in City Hotel.
13. Number of Stays in week nights.
14. Number of stays in weekend nights.
15. Relation between total of the special request is canceled or not.
16. Relations between total of special request with respect to customer.
17. Types of meal preferred by the customer.
18. Required parking space with respect to customer.
19. Study on Average Daily Rates (ADR).

20. Yearly average of ADR with respect to customer.

21. Monthly Average Daily rates ADR.

22. ADR Vary Over the Year by Hotel Type.

23. Relationship between Average Daily Rate(ADR) and Arrival Month by Booking Cancellation Status.

C.) Data description:

First of all, we will need to understand what every feature in data means. The data table consists of 119,390 rows and 32 columns. We will starts with defining each columns mentioned below:

1. hotel:	Hotel type (City hotels, Resort hotels)
2. is_canceled:	value indicate booking is canceled or not.
3. lead_time:	How long in advance the booking was made
4. arrival_date_year:	Customer arrival year
5. arrival_date_month:	In which month of the year customer visited hotel.
6. arrival_date_week_number:	In which week of the year customer arrived.
7. arrival_date_day_of_month:	Date of the month customer visited hotel.
8. stays_in_weekend_nights:	Customer booked to stay in hotel during weekend nights.
9. stays_in_week_nights:	Customer stayed in hotel during week nights.
10. adults:	Number of adults.
11. children:	Number of children.
12. babies:	Number of babies.
13. meal:	Type of meal booked.
14. country:	Country of origin of customer.
15. market_segment:	Booking sources .
16. distribution_channel:	Booking Mediums.
17. is_repeated_guest:	if the booking was from a repeated guest or not
18. previous_cancellations:	Number of previous bookings were canceled by the customers.
19. previous_bookings_not_canceled:	Number of previous bookings that were canceled by the customers.
20. reserved_room_type:	type of room Booked
21. assigned_room_type:	Room assigned to guests.
22. booking_changes:	Number of changes made to the bookings.
23. deposit_type:	Indication on if the customer made a deposit to guarantee the booking.
24. agent:	ID of the travel agency that made the bookings
25. Company:	ID of the company that made the booking or responsible for paying the booking.
26. days_in_waiting_list:	Number of days the booking was in the waiting list.
27. customer_type:	Type of booking, assuming one of four categories.
28. adr:	Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights.
29. required_car_parking_spaces:	Number of car parking space required by the customer.
30. total_of_special_requests:	Number of special requests made by the customer(e.g. twin bed or high floor).
31. reservation_status:	Reservation last status, assuming one of three categories:
32. reservation_status_date:	Reservation on which date status changed to reserved.

D) Introduction:

This project data set consist of the booking details of two hotels, viz. City Hotel & Resort Hotel from year 2015 to 2017. The data set consists of 32 columns and 119,390 observations. Each observation represents a hotel booking. The data sets comprehend hotel bookings to arrive between the year 2015 and the 2017, including bookings that effectively arrived and bookings that were canceled. The main objective of this analysis is to do the Exploratory Data Analysis from hotel point of view so that the executives of the hotels could plan and amend their policies accordingly for both hotels and guests in future.

E) Data Processing:

Exploratory Data Analysis

Exploratory Data Analysis is a data analytics process to understand the data in deep and learn the different data characteristics by using different models, graphs and plots. All of this can be done with Exploratory Data Analysis. It helps you gather insights and make better sense of the data, and removes unnecessary values from data. EDA can be done through different process are:

a.) Data Collection:

First of all, we collect data for measuring and analysing different types of information. The main objective of data collection is to gather information and analyse them to make critical business decisions.

b.) Data Loading:

We find and load data using different libraries in our system in python. Using pandas we can manipulate data easily.

c.) Data Cleaning:

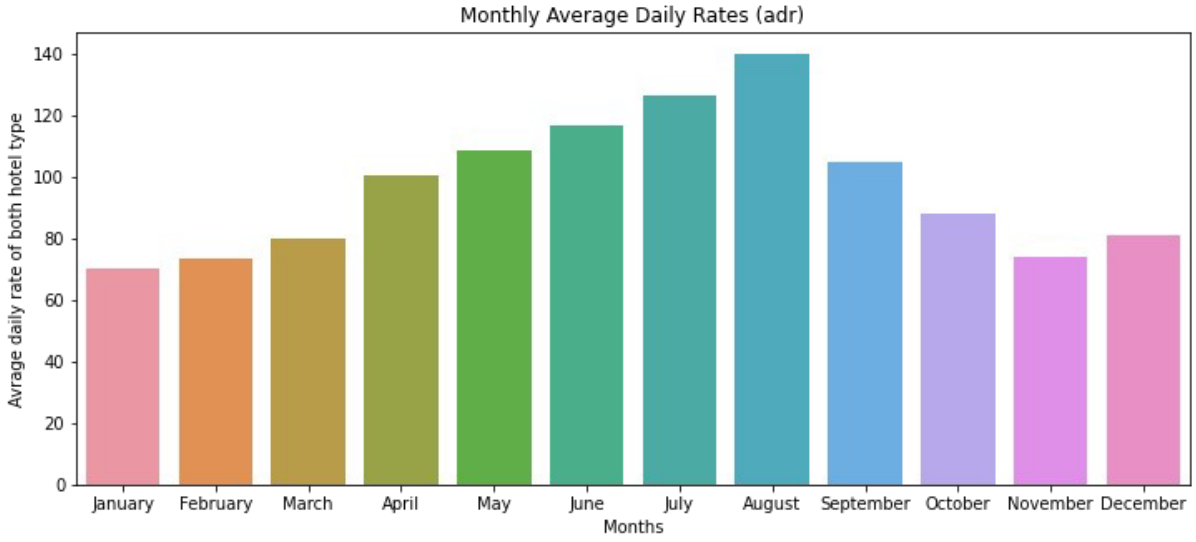
This refers to removing unwanted variables and values from your dataset and getting rid of any irregularities in it such as:Handling missing values, Removing duplicates values, Formatting data into proper data type and Adding or removing columns required for analysis.

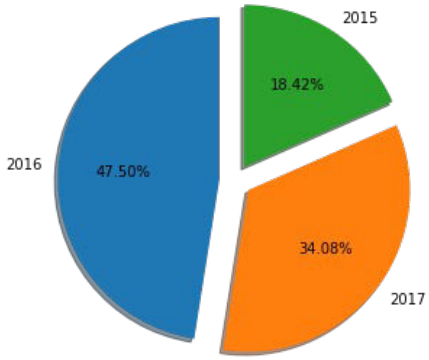
F.) Visualization:

Data visualization is the representation of data through use of common graphics, such as charts.plots and etc. These visual displays of information communicate in a way that is easy to understand.

Types of data visualizations we used:

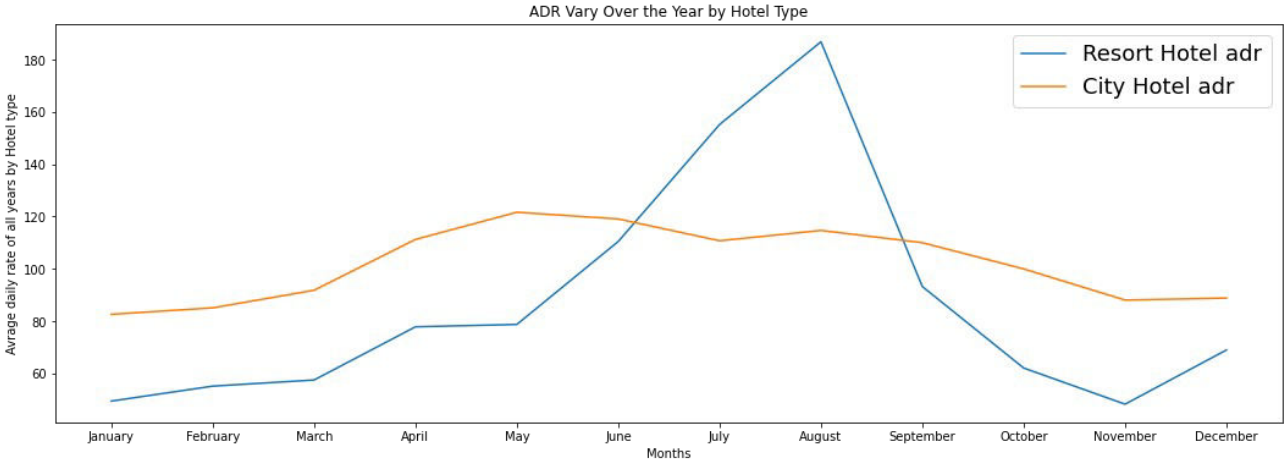
a.) Bar plot: A bar chart is a chart that represents data with rectangular bars. It can be plotted vertically or horizontally such as:





b.) Pie charts: These graphs are divided into sections that represent parts of a whole. They provide a simple way to organize data and compare the size of each component to one other.

c.) Line plot: These visuals show change in one or more quantities by plotting a series of data points over time. Line graphs utilize lines to demonstrate these changes while area charts connect data points with line segments, stacking variables on top of one another and using colour to distinguish between variables.



G.) Conclusion

In the end, as a conclusion we observe that:

- The City Hotel & Resort Hotel are attracting 'Transient' type of customers the most, we conclude that 'Transient' type of customers are most repeated in both the hotels, The hotels are attending most guests from PTR followed by GBR & FRA also'Transient' type customers are making the most changes in bookings.
- We observe that the number of Bookings and cancellations are more in City Hotel than resort hotel, the highest segment from where the booking and cancellation don e are Online TA, the highest distribution channels where bookings done are TA/TO and we saw more than 60% of the people booked from city hotel.
- We saw most busy year is 2016,most busy month is august, most guest are coming from PRT country in both hotels and maximum number of stays in weekend nights are in resort hotel than city hotel.
- In both City Hotel and Resort Hotel type 1 special requests are the most cancelled, whereas typ e 5 are the least canceled,most special requests are coming from transient and transient party t ype of customers,BB meal is preferred most by the guests followed by SC & HB and Transient type of customers are requiring car parking the most.
- The average daily rate has increased in both hotels.The City Hotel is generating more adr than the Resort Hotel, but in 2015 the resort hotel is slightly ahead compared to the city hotel, the highest average daily rate (ADR) has occurred in August,rooms are more demanding during

July & August in resort hotel and for the City Hotel,rooms are demanding during April,May & June.