# Hotel Bookings Exploratory Data Analysis:

(by Reena verma )

**Objective:**

**Businesses which generate tremendous quantities of data have to not only leverage it to recognize their modern-day performance (i.e. enterprise intelligence) however additionally – and most importantly – to generate prescriptive analytics to orient their strategy towards Data science, which can also additionally at the beginning appear overly ‘techy’ and ‘expensive’ is genuinely pretty feasible. The data analytics performs a pivotal position in hotel industry as it is key to advertising strategy, facilitates in building customer loyalty, and all in all amplify their customer base. The 5 approaches wherein data analytics makes an effective effect at the hotel industry are Customer Data Analysis & Market Segmentation, Real-Time Data and Hotel Pricing Strategies, Managing Hotel Booking Channels, Inventory Management and Demand Forecasting. This is best a partial synopsis of the feasible strategic makes use of data science. However, the primary message holds actual: do not be frightened of facts. Data is everywhere and can enhance the competitiveness of your enterprise. Data also can guide product development, advertising and operations whilst additionally helping in the transformation towards “smart” businesses. This is true for any single expert within the travel-value chain, from hotelier and destination manager to service provider.**

**The project gives various insight of hotel bookings which can be leveraged to improve business performance and customer service**

**Keywords: Exploratory data analysis, pandas, numpy, seaborn, matplotlib,**

**Python, Google Collaboratory**

Dataset:

**We are been given with the hotel bookings data. This data contains booking information for a city hotel and a resort hotel. It contains the following features.**

**- hotel: the Name of hotel ( City or Resort)**

**- is\_canceled: anal. Whether the booking is canceled or not (0 for no canceled and 1 for canceled)**

**- lead\_time: time (in days) between booking transaction and actual arrival.**

**- arrival\_date\_year: deta frame show the Year of arrival**

**- arrival\_date\_month: showing every month of arrival**

**- arrival\_date\_week\_number: showing every week number of arrival date.**

**- arrival\_date\_day\_of\_month: which of the Day of month of arrival date**

**- stays\_in\_weekend\_nights: total number No. of weekend nights spent in a hotel**

**- stays\_in\_week\_nights: total number No. of weeknights spent in a hotel**

**- adults:total number showing the No. of adults in single booking record.**

**- children:what is the total No. of children in single booking record.**

**- babies: booking record of No. of babies in single booking record.**

**- meal: what Type of meal chosen**

**- country: mention Country of origin of customers (as mentioned by them)**

**- market\_segment: What segment via booking was made and for what purpose.**

**- distribution\_channel: total Via which medium booking was made.**

**- is\_repeated\_guest: most of the Whether the customer has made any booking before(0 for No and 1 for Yes)**

**- previous\_cancellations: total No. of previous canceled bookings.**

**- previous\_bookings\_not\_canceled: total No. of previous non-canceled bookings.**

**- reserved\_room\_type: what is Room type reserved by a customer.**

**- assigned\_room\_type: which of Room type assigned to the customer.**

**- booking\_changes: No. of booking changes done by customers**

**- deposit\_type: Type of deposit at the time of making a booking (No deposit/ Refundable/ No refund)**

**- agent: what is Id of agent for booking**

**- company: what is Id of the company making a booking**

**- days\_in\_waiting\_list: total No. of days on waiting list.**

**- customer\_type: what is the Type of customer(Transient, Group, etc.)**

**- adr: Average Daily rate.**

**- required\_car\_parking\_spaces: total No. of car parking asked in booking**

**- total\_of\_special\_requests: total no. of special request.**

**- reservation\_status: Whether a customer has checked out or canceled,or not showed**

**- reservation\_status\_date: what is Date of making reservation status.**

## Data Cleaning :

### (1) Removing Duplicate rows

**All duplicate rows were removed and dropped.**

### (2) Handling null values

* **Null values in columns childrens, company and agent were replaced by 0.**
* **Null values in column country were replaced by 'others'.**

**(3) We have Converted the columns to appropriate data types so that we would be able to extract usefull insights**

* **Changed data type of children, company, agent are converted to integer type.**
* **Changed data type of reservation\_status\_date to date type.**

### (4) Removing outliers

* **One outlier was found in the adr column. Simply dropped it.**

### (5) Creating new columns

* **Created new column total\_stay by adding stays\_in\_weekend\_nights+stays\_in\_week\_nights**
* **Created new column total\_people by adding adults+children+babies.**

## Exploratory Data Analysis:

## Performed Exploratory Data Analysis and tried answering the following questions:

Q1Which meal type is most preffered meal of customers?

**Ans. BB is the most preferred meal of customers as represented into the data approximately 78%peoples preffered for it and HB and HC places on the same instance and in the last FB and UNDIFINED are the least.**

**Q2 Which agent makes most no. of bookings?**

**Ans. Agent no. 9 has made most no. of bookings then comes 240 and 14 then 7 , 250 , 241 , 28, 8, 1, 6 in the decreasing order of the sales**

**Q3 Which room\_type is in most demand and which room\_type generates highest adr?**

**Ans. A types of rooms are most demanded then comes D,E,F,G, etc.**

**Q4 What is Booking Percentage in each hotel?**

**Ans. Around 60% bookings are for City hotel and 40% bookings are for Resort hotel, therefore City Hotel is busier than Resort hotel.**

**Q5  Which amoung the hotel has higher which has lower lead time?**

**Ans. City hotel has slightly higher median lead time. Also median lead time is significantly higher for both hotels, this means customers generally plan their hotel visits way early.**

**Q6  which hotel seems to make more revenue?**

**Ans. Overall adr of City hotel is slightly higher than Resort hotel and no. of bookings of City hotel is also higher than Resort hotel. Hence, City hotel is makes more revenue.**

**Q7  Which hotel has longer waiting time and which has less waiting time ?**

**Ans. City hotel has higher waiting time and resort hotel has less waiting time as the no of bookings in the city hotel is more compared with the resort hotels**

**Q8)Arrival of booking from diff segments?**

**Ans. Booking from the different segment type.**

1. **looking forward from different segments between hotel and market segments and number of booking from online TA is the most benefit sector for city hotels as compare to resort hotel**
2. **According to the graph the green colour show the percentage of booking for city hotel and blue colour y show the percentage of resort hotels the booking in the graph higher booking from online TA**

**Q9)Deposite whether it is refundable nonrefundable or no depoist?**

**Ans. percentage of each type of deposit**

**we can represent our date from different methods and one of the most important method is pie chart**

1. **In the chart out of 100 percent 87.6% are no deposits**
2. **Out of 100 percent 12.2% are non refundable**
3. **the rest 0.1% is refundable which is less than among of no refund and no deposits**

**Q10)No of passengers by cancellation by the types of hotels?**

**Ans. Almost 30 % of City Hotel bookings got canceled.**

**Q11)Meals type by no of pessengers that were preffered to eat by the pessengers?**

**Ans. Most popular meal type is BB(Bed and Breakfast).**

**Mainly performed using Matplotlib and Seaborn library and the following graph and plots had been used:**

* **Heatmap.—Correlation Matrix**
* **Scatter Plot.--- Scatter Plot Data Analysis**
* **Bar Chart --- Bar plot**
* **Pie Chart. --- Pie plot**
* **Line Plot.**

### Univariate Analysis:

**Performed univariate analysis and made following conclusions:**

**1.) Agent no. 9 has made most no. of bookings.**

**2.) Most demanded room type is A, but better adr generating rooms H, G and C. Hotels should increase the no. of room types A and H to maximise revenue.**

**3.) Around 60% bookings are for City hotel and 40% bookings are for Resort hotel, therefore City Hotel is busier than Resort hotel.**

**4.) Guests use different channels for making bookings out of which most preferred way is TA/TO.**

**5.) Most of the guests came from european countries, with highest number of guests from Portugal.**

### Bivariate Analysis :

**We tried to answer following questions**

**1.) Overall adr of City hotel is slightly higher than Resort hotel and no. of bookings of City hotel is also higher than Resort hotel. Hence, City hotel is makes more revenue.**

**2.) City hotel has slightly higher median lead time. Also median lead time is significantly higher for both hotels, this means customers generally plan their hotel visits way early.**

**3.) Almost 30 % of City Hotel bookings got canceled.**

**4.) Both hotels have very small percentage that customer will repeat, but Resort hotel has slightly higher repeat % than City Hotel.**

**5.) TA/TO is mostly used for planning Hotel visits well ahead of time.**

**6.) While booking via TA/TO one may have to wait a little longer to confirm booking of rooms.**

**7.) GDS channel brings higher revenue generating deals for City hotel, in contrast to that most bookings come via TA/TO. City Hotel can work to increase outreach on GDS channels to get more higher revenue generating deals.**

**8.) TA/TO has highest booking cancellation %. Therefore, a booking via TA/TO is 30% likely to get cancelled.**

**9.) Longer lead time has no affect on cancellation of bookings.**

**10.) Not getting same room as demanded is not the case of cancellation of rooms. A significant percentage of bookings are not cancelled even after getting different room as demanded.**

**11.) Not getting same room do affects the adr, people who didn't got same room have paid a little lower adr.**

**12.) Arrivals in hotels increases at weekends and also the avg adr tends to go up as month ends.**

**13.)Moslty bookings are done by couples(bookings have two adults.)**

## Conclusion:

**(1) Around 60% bookings are for City hotel and 40% bookings are for Resort hotel, therefore City Hotel is busier than Resort hotel. Also the overall adr of City hotel is slightly higher than Resort hotel.**

**(2) Mostly guests stay for less than 5 days in hotel and for longer stays Resort hotel is preferred.**

**(3) Both hotels have significantly higher booking cancellation rates and very few guests less than 3 % return for another booking in City hotel. 5% guests return for stay in Resort hotel.**

**(4) Most of the guests came from european countries, with most of guests coming from Portugal.**

**(5) Guests use different channels for making bookings out of which most preferred way is TA/TO.**

**(6) For hotels higher adr deals come via GDS channel, so hotels should increase their popularity on this channel.**

**(7) Almost 30% of bookings via TA/TO are cancelled.**

**(8) Not getting same room as reserved, longer lead time and waiting time do not affect cancellation of bookings. Although different room allotment do lowers the adr.**

**(9) July- August are the most busier and profitable months for both of hotels.**

**(10) Within a month, adr gradually increases as month ends, with small sudden rise on weekends.**

**(11) Couples are the most common guests for hotels, hence hotels can plan services according to couples needs to increase revenue.**

**(12) More number of people in guests results in more number of special requests.**

**(13) Bookings made via complementary market segment and adults have on average high no. of special request.**

**(14) For customers, generally the longer stays (more than 15 days) can result in better deals in terms of low adr.**

**And many more conclusions.**

## Challenges:

**(1) There was a lot of duplicate data.**

**(2) Data was present in wrong datatype format.**

**(3) Choosing appropriate visualization techniques to use was difficult.**

**(4) A lot of null values were there in the dataset.**

**Summary:**

**analysing the data of booking hotel/resort rooms.This data set contains booking information for a city hotel and a resort hotel.**

**So firstly, we import the python libraries for analyze and visualise our data.**

**After that we cleaned and preprocessed the data.**

**Then we performed the exploratory data analysis to extract information from the data to answer some questions and we did null values treatment, encoding of categorical column and then model bulding.**

**Then we performed feature selection and feature engineering, and then made the predictive model using the Decision Tree for prediction.**