

# Hotel Booking Demand/Cancellation EDA

Havan Patel  
hawangpatel@lewisu.edu  
DATA-53000, Summer 2022  
Data Visualization  
Lewis University

## I. INTRODUCTION

For this project we will be analyzing hotel booking data. Booking cancellation have a substantial impact on demand management decisions in the hospitality industry. The process of online travel booking has moved firmly into the 21st century, with an estimated \$817 billion worth of online bookings by 2020. Considering that the tourism industry is worth approximately \$1.2 trillion, online bookings represent one of the largest market shares in the sector.[1] This report utilizes multiple data visualizations for exploratory analysis; the overall goal of the report is to visualize the data attributes and seek potential relationships of the data attributes with prediction on which guests were likely to cancel and adjust the overbook rate accordingly. Informative visualizations will be displayed below to educate the readers on the data and simply display the different data attributes.

## II. DATA DESCRIPTION

As described in Section I above, the dataset we are using is an Exploratory Data Analysis (EDA) to get insights from Hotel Booking Demand which consists of 116,785 instances [3]. Because the data has a high number of instances, data visualization will assist in the ability to comprehend the large amount of data. The dataset has been used for a variety of different studies in the hotel booking demand. From this dataset we will get insight as to which feature have contributed more to predicting cancellation by performing data visualization in Tableau . It always the best to practice and understand the data first and try to collect as many insights as possible from it. When analyzing the past years data, D-Edge Hospitality Solutions has found that with a global average of almost 40% cancellation rate, this trend produces a very negative impact on hotel revenue and distribution management strategies. [2] predicting reservations which might get canceled and preventing these cancellations will create a surplus revenue for both Hotels and Hotel Technology companies.

TABLE I. DATA ATTRIBUTES

Attribute	Type	Example Value	Description
Hotel	Categorical	Resort Hotel	Hotel Type
Is Canceled	Numeric	0	0 not cancelled, 1 cancelled
Lead Time	Numeric	5	Days book in advanced
Arrival Date Year	Date	2015	Arrival Year of arrival date (2015-2017)
Arrival Date Month	Date	January	Month of arrival date (Jan - Dec)
Arrival Date Week Number	Numeric	27	Week number of year for arrival date (1-53)
Arrival Date Day Of Month	Numeric	1	Day of arrival date
Stays in Weekends Nights	Numeric	0	No of weekend nights (Sat/Sun) the guest stayed or booked to stay at the hotel
Stays in Week Nights	Numeric	2	No of weeknights (Mon - Fri) the guest stayed or booked to stay at the hotel
Adults	Numeric	1	No of Adults
Children	Numeric	2	No. of Children
Babies	Numeric	0	No, Of babies
Meal	Categorical	BB	Type of meal booked.
Country	Categorical	GBR	
Market Segment	Categorical	Corporate	TA: Travel Agent, TO: Tour Operator
Distribution Channel	Categorical	TA/TO	TA: Travel Agent, TO: Tour Operator

Attribute	Type	Example Value	Description
Is Repeated Guest	Numeric	0	1: Yes 0: No
Previous Cancellation	Numeric	0	Previous bookings cancelled
Previous Bookings Not Canceled	Numeric	0	Previous booking not cancelled
Reserved Room Type	Ordinal	A	Code of room type reserved.
Assigned Room Type	Numeric	C	Code for the type of room assigned to the booking
Booking Changes	Numeric	3	Number of changes/amendments made to the booking
Deposit Type	Categorical	No Deposit	
Agent	Numeric	302	ID of the travel agency that made the booking
Company	Numeric	110	ID of the company/entity that made the booking
Days In Waiting List	Numeric	0	Number of days the booking was in the waiting list before it was confirmed to the customer
Customer Type	Categorical	Transient	Contract, Group, Transient, Transient party
Adr	Numeric	154.770	
Required Car Parking Spaces	Numeric	0	Number of car parking spaces required by the customer
Total Of Special Requests	Numeric	3	Number of special requests made by the customer
Reservation Status	Categorical	Check-out	Canceled, Check-Out, No-Show
Reservation Status Date	Date	2015	Date at which the last status was set.

### III. METHODOLOGY AND RESULTS

The data was input into Tableau, and multiple visualizations were created to conduct exploratory visualizations on the data. When booking for hotel, there are different types of hotels you can book depending on the type of travelling you do, either city hotel, or resort hotel. In Fig. 1 below, a visualization was created to compare the different type of hotels with the overall booking of the Hotel. In Fig. 1, we can clearly see, that city hotels are clearly more popular and preferred for booking at 66.45% then resort hotel. There can be many reasons as to why that is. One reason is prices vary at different type of hotels. Most of the time resort hotels tend to be more on the expensive side than city hotel, therefore people will stick with city hotels. But with the different types of hotels at different prices you also get different types of services and packages. One thing to mention is that resort hotels tend to be appropriate for large group of people. It is also important to mention, that due to potentially missing data or having null values, some results can be skewed or due to some attributes not offering information.

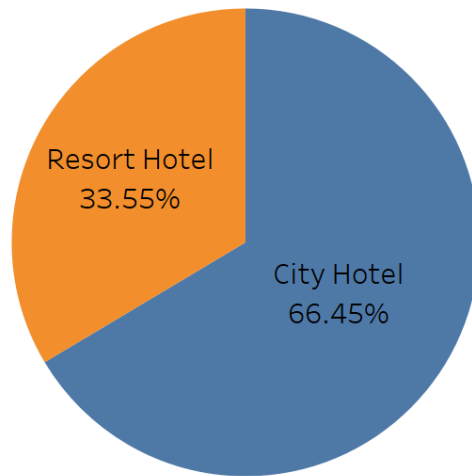
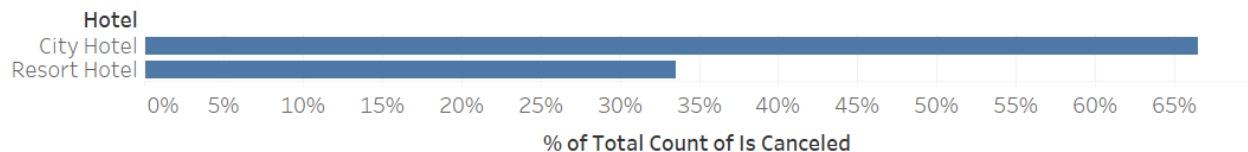


Fig. 1 Popularity of Hotel Booking

In Fig. 2, a visualization We have two graphs of cancellation between hotels and Overall Cancellation between both hotels. In the first graph we can clearly see that more than 65% of city hotels booking were cancelled and this makes sense because from previous Fig. 1, we saw there were more booking were made for city hotels. Therefore, we can say that if more booking happens to one type of hotel, then cancellation will also be occurring to that same hotel. Now if we look at the second graph in Fig 2. We see that overall, there were less cancellation between both the hotel types. In this case 0 represents not cancelled booking which is close to 65%. And 1 represent cancellation to both hotels.

### Cancellation Between Hotels



### Overall Cancellation

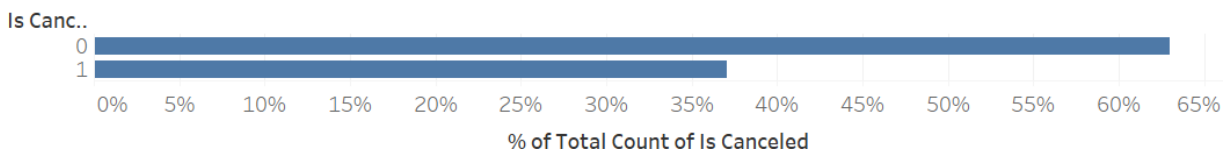


Fig. 2 Cancellation between Hotel vs Overall Cancellation

In Fig. 3, we take a different approach to the data. In the Fig. 3 we are trying to answer the question of which month has the highest number of cancellations by the hotel type. We are applying the filter on cancellation to months by the hotel type in this bar graph. From this figure we can visualize that city hotel, which has around 40% cancellation every month compared to resort hotel. Although this is consistent with amount of cancellation every month for city hotel, this shows that they need to remodel their business in a way that there are less cancellations or at least the cancellations are not consistent every month. They need to come up with a new skim. Now as per the resort hotel we see varied cancellation per months. The highest cancellation for resort hotels are happening in June, July and August and lowest in November, December and January.

## Average Cancellation Per Month

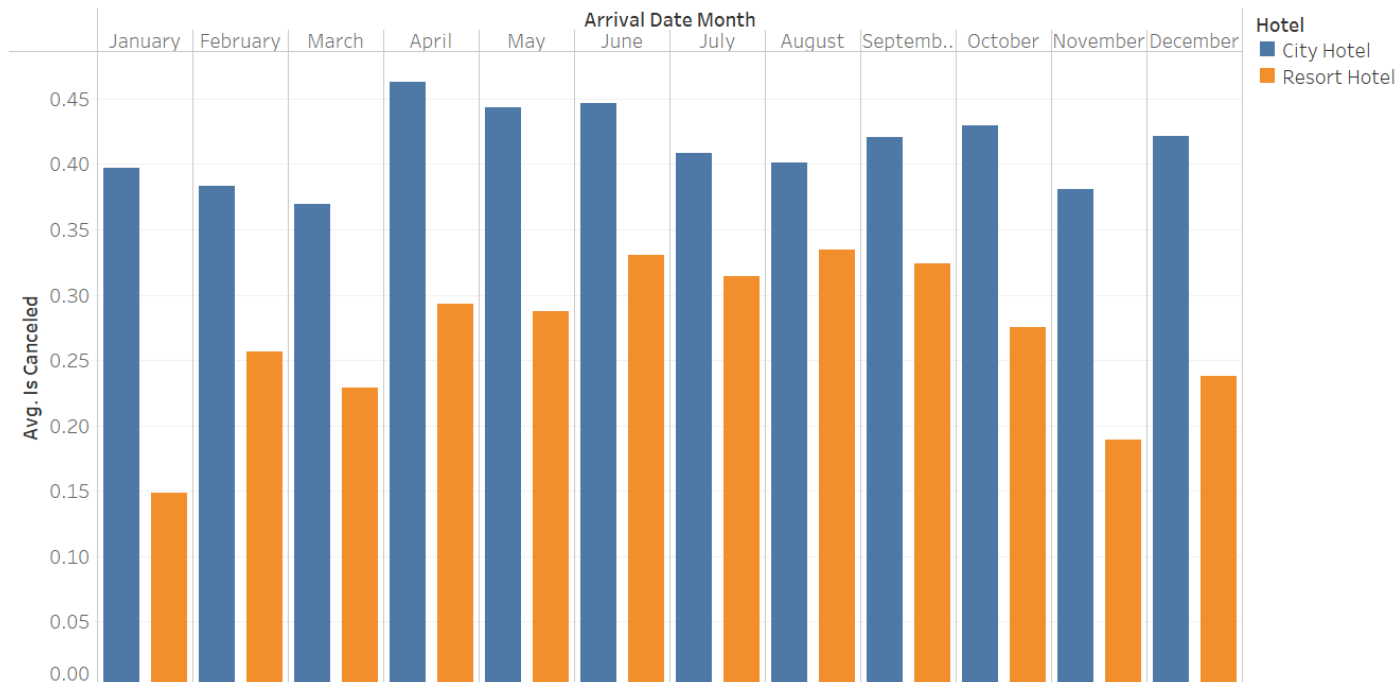


Fig. 3 Months with highest number of cancellations.

In Fig. 4, we are again continuing with cancellation status with relation to Average daily rate. From the graph we can clearly see that in August has the highest average daily rate between both the hotel types. 0 represent no cancelled and 1 represent cancellation. Because of high ADR in August, that is the reason why more of the cancellation are happening. Lower the price, the less cancellations are happening.

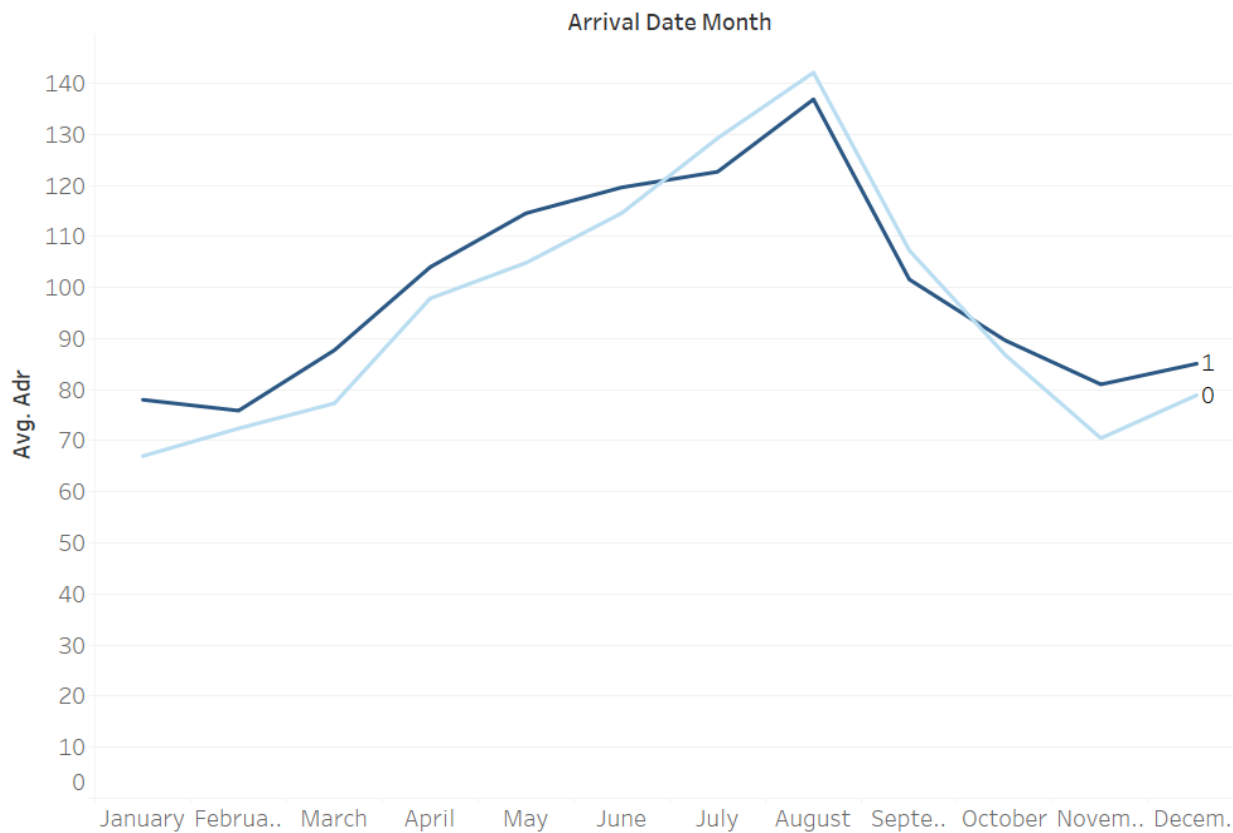


Fig. 4 Relationship between Average Daily Rate(ADR) and Arrival Month by Booking Cancellation

In Fig. 5, we are comparing the lead time vs the arrival year by cancellation status. We have a data of over the 3 year period. We can see that the booking that has a lead time of 100 days or more has more changes of getting cancelled compared to lead time with less than 100 days has fewer chance of getting cancelled. I think this happens because if you book a hotel very far in future in advance you can very easily change your mind because maybe you found a better deal, or may you won something somewhere else. This can lead to a cancellation of the booking. Now let's say your booking lead time was less than 60 days then you wouldn't change your mind as often as you would in 100 days.

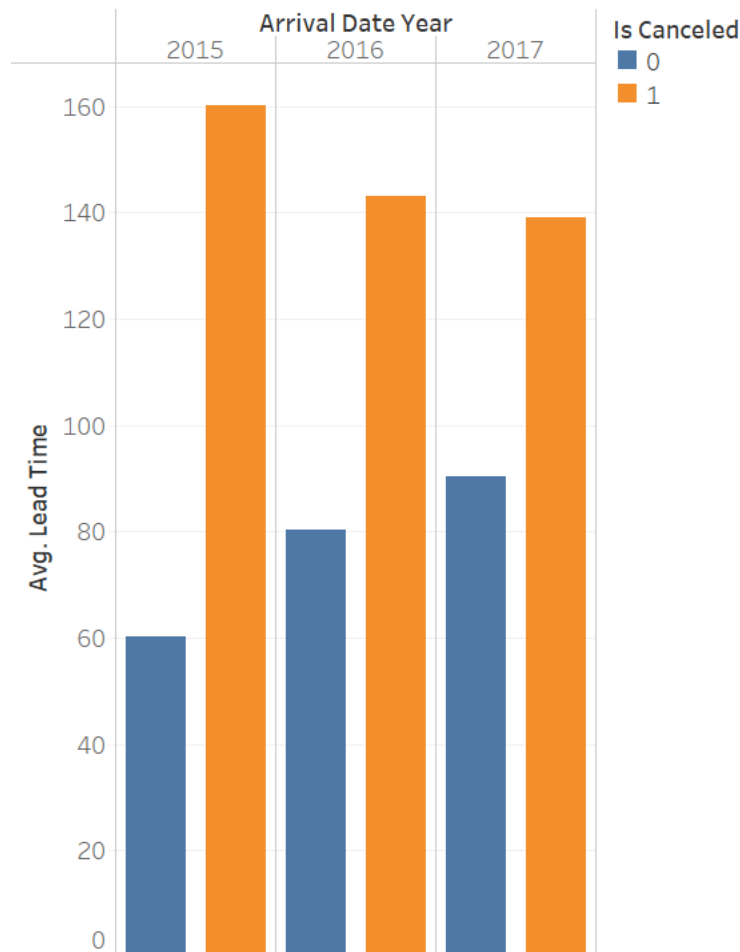


Fig. 5 Average Leadtime vs Arrival Year cancellation

In Fig. 6, let's look at which countries that books the most hotels. I have created a map chart which shows all the country with a heat map. The darker the country's color the more booking of hotels. We can clearly see that the Europe has majority of the bookings of hotels, whether that is resort or city hotels. We can also conclude that most of the visitors are coming from wester side of the Europe because they are many neighboring countries close to each other. Another interesting fact we can conclude from this visualization is that most of the data in this dataset seems to be coming from hotels bookings in Europe because that's where most of the heat map color is.

## Countries Booking The Most Hotels

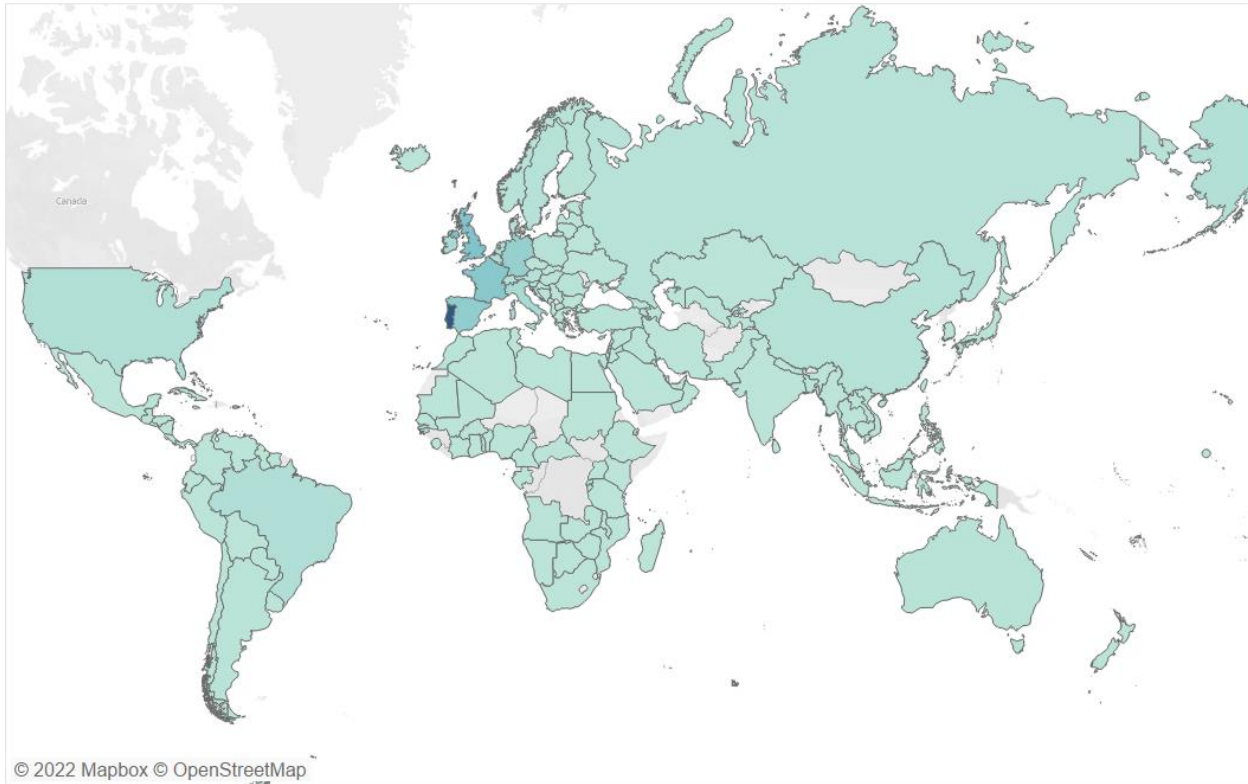


Fig. 6 Countries booking the most hotels.

### IV. DISCUSSION

In Section III, we used different types of visualizations such as pie chart, bar graphs, map, and line chart to visualize some of the attributes of our data. We created informative visualizations to provide the audience with facts about the data, and some of the correlations, not necessarily causations. We compared different attributes to one another as well as compared attributes to the overall booking and cancellation of the hotel types by visualizing the attributes, singularly or in a multivariate setting, we can start to see what attributes are contributing to the cancellation of the bookings. In this report, we learned several things about our attributes. Such as, more than 60% of population book city hotel, most bookings happen in Europe, lead time of 100 days leads to cancellation and more. From our visualization we can observe that there are few most important features in the data that helps us predict cancellation from the guests are: Lead Time, ADR, Deposit Types, Arrival Day of the Month.

Utilizing exploratory examination, permits us to begin in figuring out our information. It is the first step of completing our analysis. We were able to visually understand the different attributes and features and how they play a role in overall bookings demand of the hotels. From this hotel business can learn and change their business model to counter the cancellation of their hotel bookings.

### V. CONCLUSIONS

In this report, we used the dataset that contains data about hotel bookings and created visualizations to conduct exploratory analysis to better understand the dataset. With our data visualizations, we were able to visualize how certain attributes can affect the cancellation of the hotel bookings. For example, having lead time of 100 days in advance booking had higher chance of getting cancelled. We tried to answer the following questions 1) how many bookings were cancelled? 2) Booking ration of Resort vs City hotel 3) Cancellation of between Resort and City hotels, 4) Cancellation between months for both hotel types and more.

### REFERENCES

- [1]“Steve Deane” “Over 60 online travel booking statistics(2022)” 04-Jan-2022 [Online]. Available: <https://www.stratosjets.com/blog/online-travel-statistics/>
- [2]“How online hotel distribution is change in Europe” [Online]. Available: <https://www.d-edge.com/how-online-hotel-distribution-is-changing-in-europe/>

[3]“Jesse Mostipak“Hotel Booking Demand,” 12-Feb-2020. [Online]. Available:  
<https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand>.