Data Visualisation Assignment 1

Mina Jamshidian

15 Nov 2020

Hotel Booking Demand And Cancellation Analysis

1. Data

● Dataset description

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.
The dataset variables are described as follows:

Hotel: (H1 = Resort Hotel or H2 = City Hotel).
Is_canceled: indicating if the booking was canceled (1) or not (0)
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.
Arrival_date_month: Month of arrival date.
Is_repeated_guest: Value indicating if the booking name was from a repeated guest (1) or not (0).
Market_segment: Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”,.. .
Deposit_type: Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay.
Lead_time: Number of days that elapsed between the entering date of the booking into the PMS and the arrival date.
Country: Country of origin. Categories are represented in the ISO 3155–3:2013 format.

Link to dataset:
https://www.kaggle.com/jessemostipak/hotel-booking-demand

● Cleaning Data by Using Tableau Prep Builder

1.Rename the columns name: Renaming the columns field by appropriate name.
2.Children column: Replace the null values of the “children” column by 0.
3.Meal column: "meal" contains values "Undefined", which is equal to SC
So was replacing the “undefined values by "SC"

4.Country column: Replace the null value of the “country” column by the “Unknown” value
5.Company and Agent column: For agent column, If no agency is given, booking was most likely made without one.So null values are replaced by 0.
For company columns, If none given, it was most likely private. So null values are replaced by 0.
6.Market Segment column and Distribution Channel column: Replace the undefined value of the “Market Segment” column by the “Others” value
7.Average Daily Price column: Negative value was replaced by 0.

2. Story

Due to the increasing use of technology by people. They prefer to use online systems to book a hotel with the cancellation option which is a very common feature for those online systems.
This situation has two conditions. One of them was due to the increase in hotel reservations and the other was due to the increase in cancellations, which was a big concern for hotels owners.
Therefore, the hotels owners want to develop strategies to maximize reservation, revenue and minimize cancellation rate.
The dataset comes from booking information for two hotels that are located in Portugal which is a City Hotel in Lisbon and a Resort Hotel in the resort region of Algarve. The purpose of this work is to perform exploratory analysis on the dataset in order to understand booking seasonality demand and pricing, customer segmentation, for providing three insights to increase hotel booking and minimize cancellations.

3. Exploration

- Dashboard 1

Link to Dashboard 1 :
https://public.tableau.com/profile/mina2239#!/vizhome/DV-CA1-Mina/Dashboard1

Dashboard 1 Description:

Booking, Daily Price and length of stay Analysis

Dashboard1 gives information regarding booking, pricing and length of stay.
As you can see in Dashboard1 When the two hotels (City hotel and Resort hotel) were compared, the booking rate for City hotel with 61% is more than the Resort hotel that means the city hotel is more popular than the resort hotel in general.
The maximum booking is for summer time for both hotels especially for July and August and the minimum booking reservation is for winter time.

Average duration of stay and price for the Resort hotel during the period of summer especially for August and July is more than the City hotel.
So the City hotel owner should raise the fee in summertime for increasing their booking.

- Dashboard 2

Link to Dashboard 2 :
https://public.tableau.com/profile/mina2239#!/vizhome/DV-CA1-Mina/Dashboard2

Dashboard 2 Description:

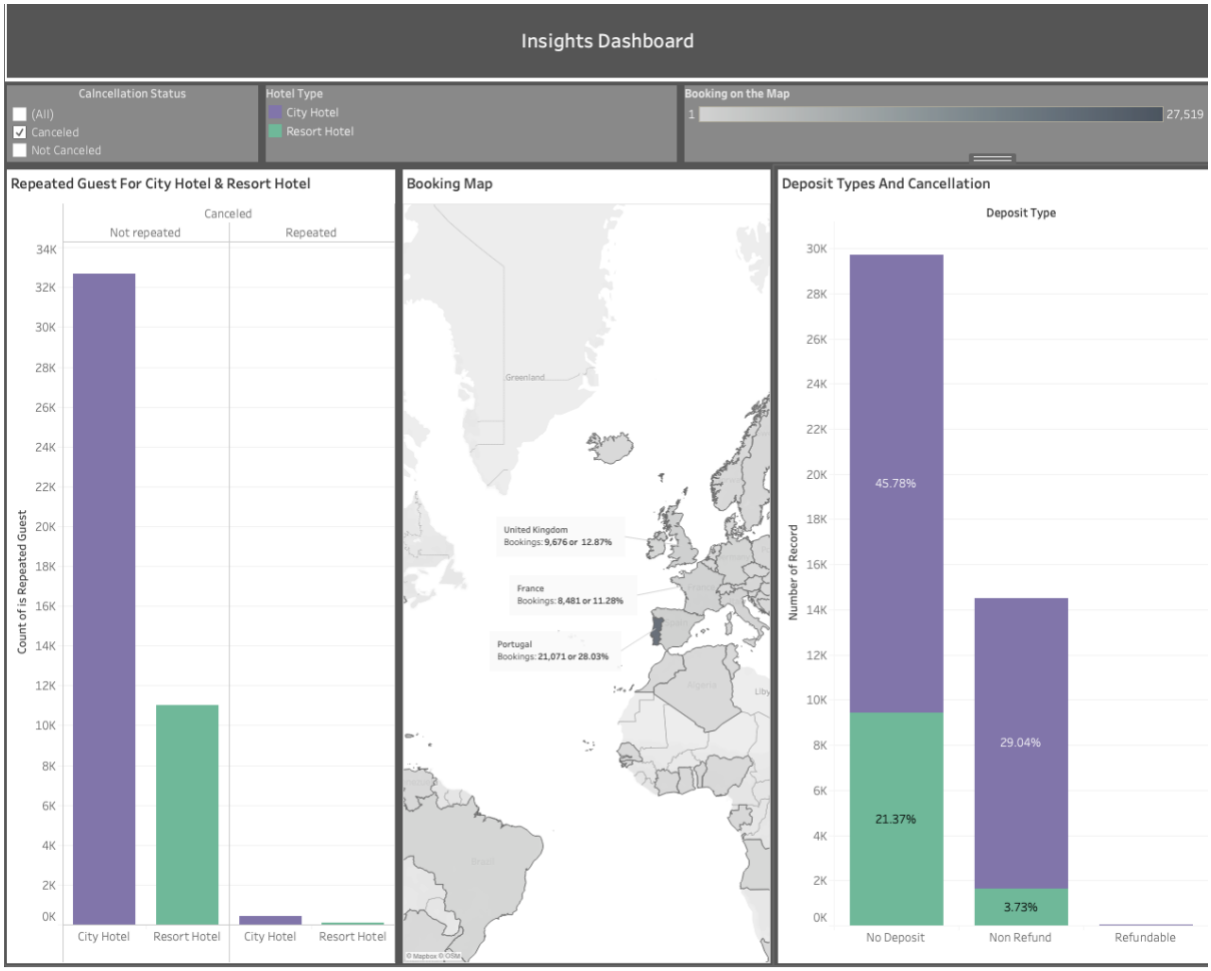
Cancellation, Lead Time And Market Segmentation Analysis

Dashboard2 gives information regarding cancellation, customer segmentation and lead time.
As you can see in Dashboard2 When the two hotels (City hotel and Resort hotel) were compared, the cancellation rate for City hotel with 42% is more than the Resort hotel in all the months that means there is a correlation between booking rate and cancellation rate as you see in dashboard1 the city hotel is more popular for booking so the rate of cancellation for this hotel is also high.

The cancellation rate for resort hotels during the summer is high because in summertime it has a higher booking rate.
After reviewing the cancellation rate, this dashboard looked at the factors that led to the cancellation of reservations.
It shows that the biggest factor that leads to the highest rate of cancellation is the number of days that elapsed between the entering date of the booking and the arrival date. Because the correlation between cancelation rate and lead time is positive. And the most 3 groups that make the highest lead time comes from “Groups”, “Online Travel Agents”, “Offline Travel Agents”.

4. Insights Dashboard

Link to Insights Dashboard :
https://public.tableau.com/profile/mina2239#!/vizhome/DV-CA1-Mina/InsightsDashboard



Insights Dashboard Description:

1. Most customers do not come back to the hotel again. So the hotels must invest in new customers and show them that their hotel is so interesting and how many options they have. Or they can also invest in previous customers by giving them a good discount because of coming back to the hotel, email them for promotion such as a special package. But in general they should focus on new customers.

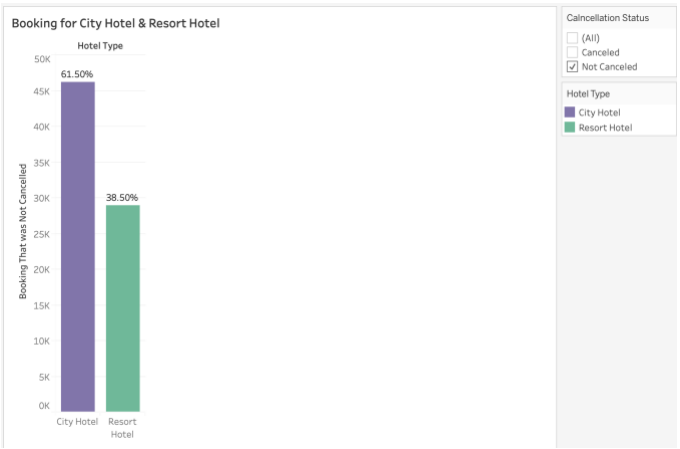
2. Mostl bookings are from Portugal, UK and France. The map shows the most customers are from those three countries more than others.So they can focus on those counties for marketing and advertising.

3. When the hotels do not get the deposit from customers, they would cancel their booking most of the time. The hotels should get deposits when the customer reserves the hotel. Therefore the hotels would not be faced by a lot of cancellation.

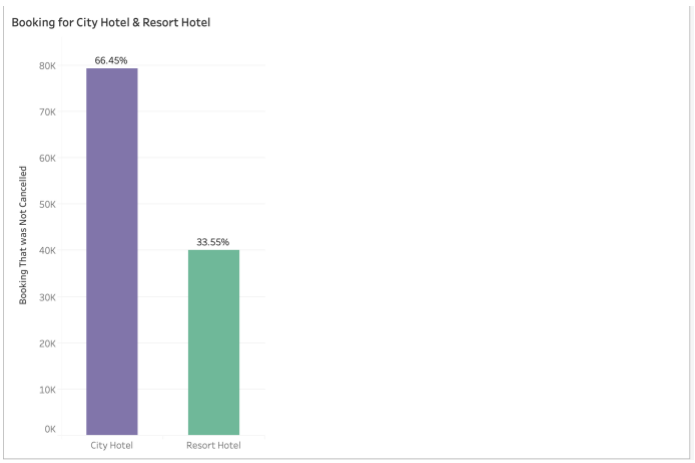
5. Iterations

Iteration 1:

Before:



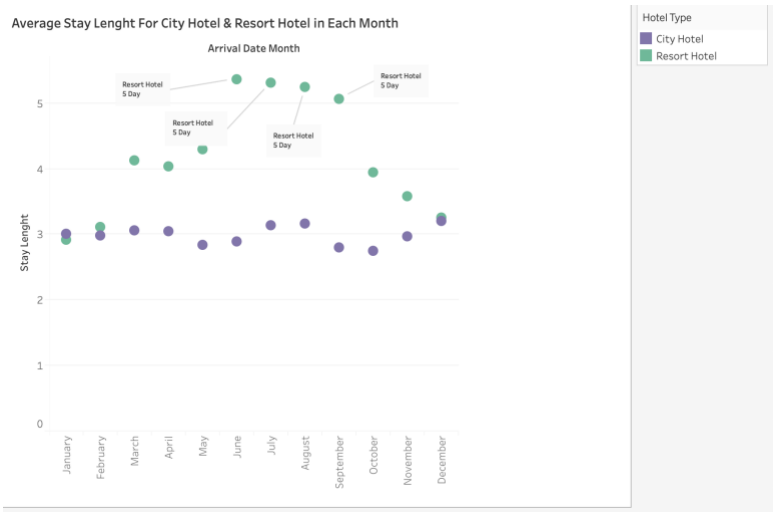
After:



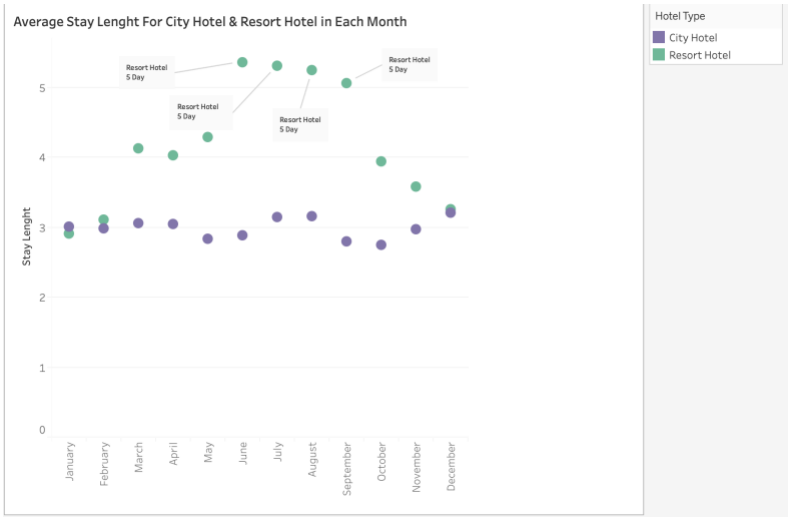
Title of x-axis (Hotel Type) that is obvious in the before graph, has been removed because the title of the graph itself shows it is about hotel type.
“Hotel type” filter box that shows the color of each hotel has been removed from the graph because x-axis labels clearly describe the sectors being represented.
“Cancellation Status” filter box has been removed from the graph because the title of y-axis shows these bookings are about not cancelled reservations.

Iteration 2:

Before:



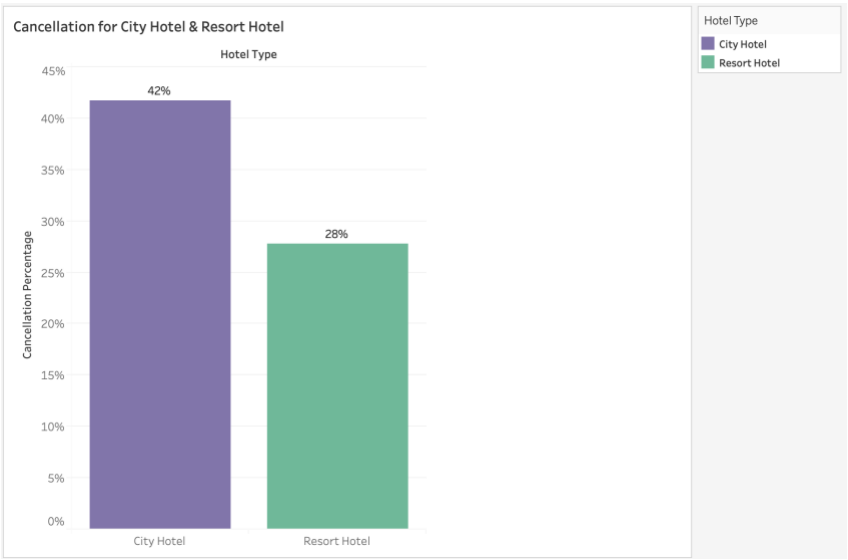
After:



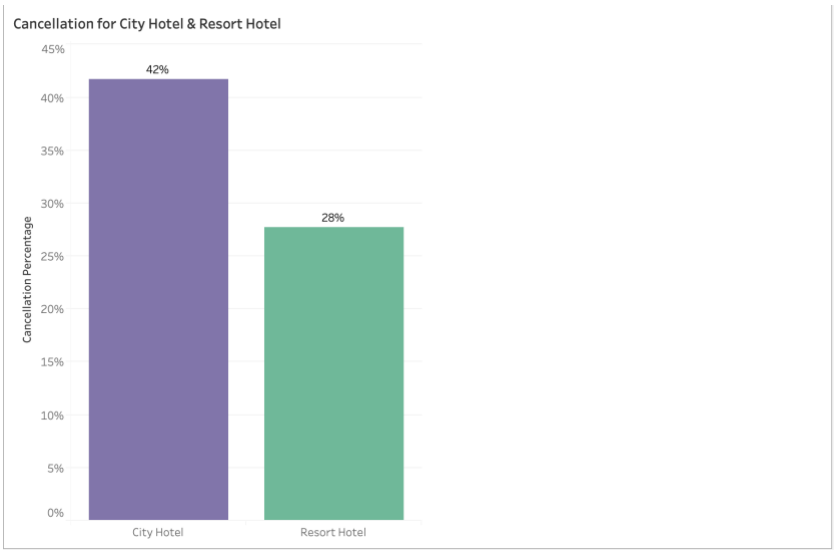
Title of x-axis (Arrival Date Month) that is obvious in the before graph, has been removed because the title of the graph itself shows it is about the month and we can see the label of each month in x-axis.

Iteration 3:

Before:

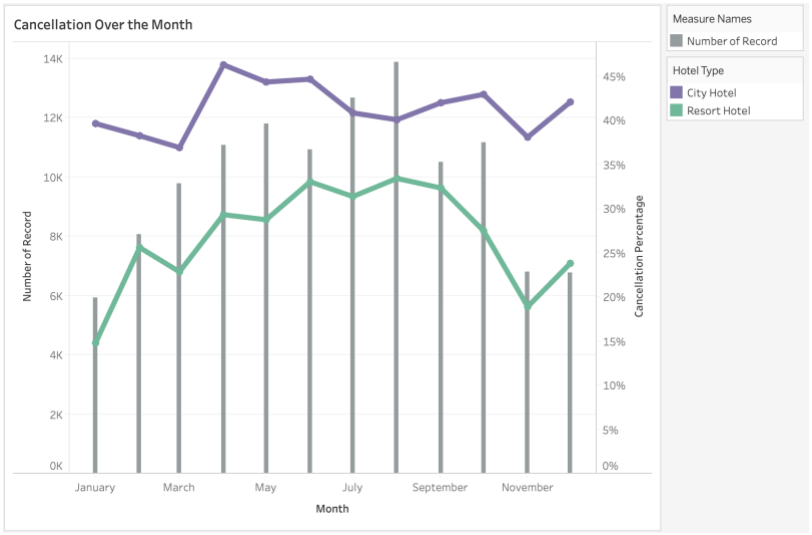


After:

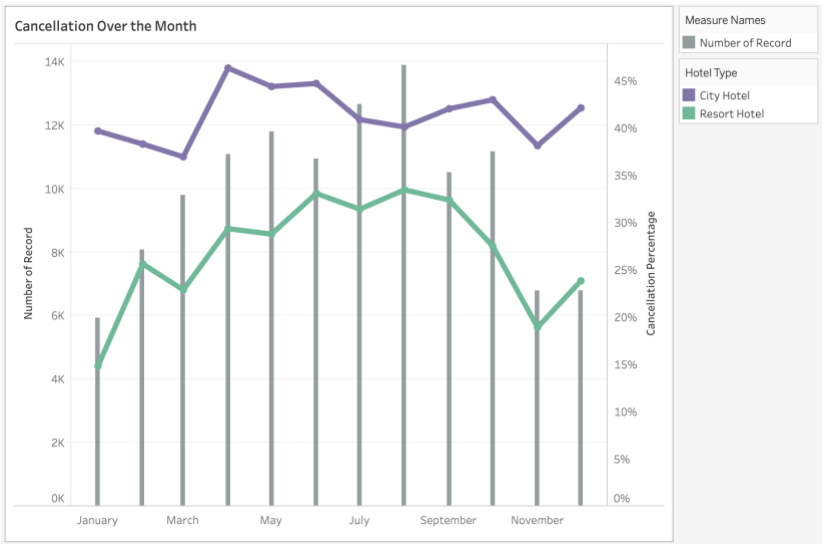


Title of x-axis (Hotel Type) that is obvious in the before graph, has been removed because the title of the graph itself shows it is about hotel type.
“Hotel type” filter box that shows the color of each hotel has been removed from the graph because x-axis labels clearly describe the sectors being represented.

Iteration 4:
Before:



After:

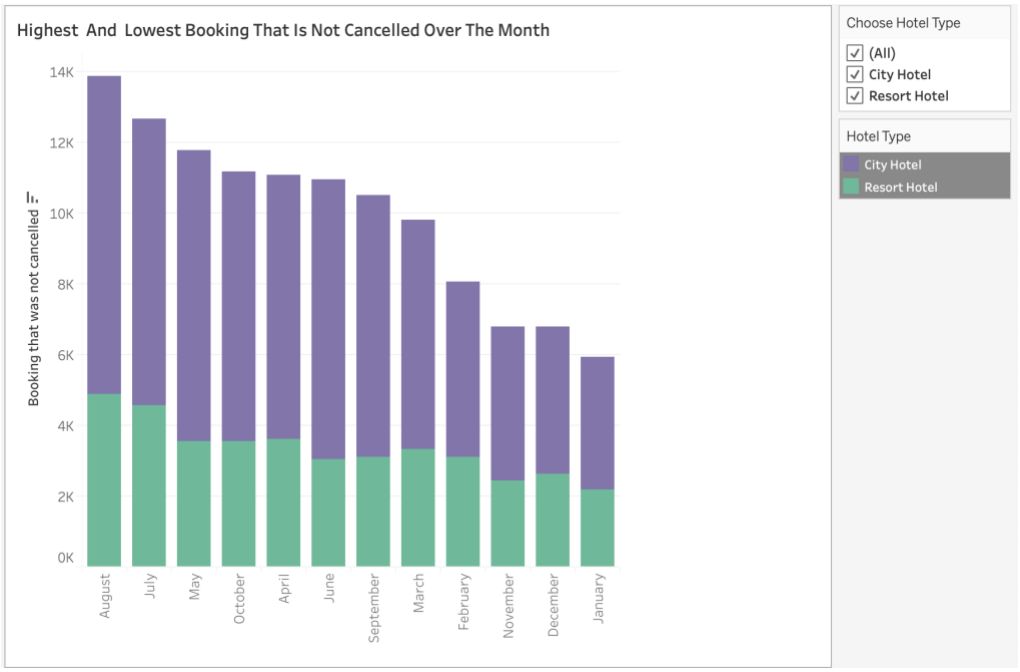


Title of x-axis (Month) that is obvious in the before graph, has been removed because the title of the graph itself shows it is about the month and we can see the label of each month in x-axis.

Iteration 5:

Before:

After:

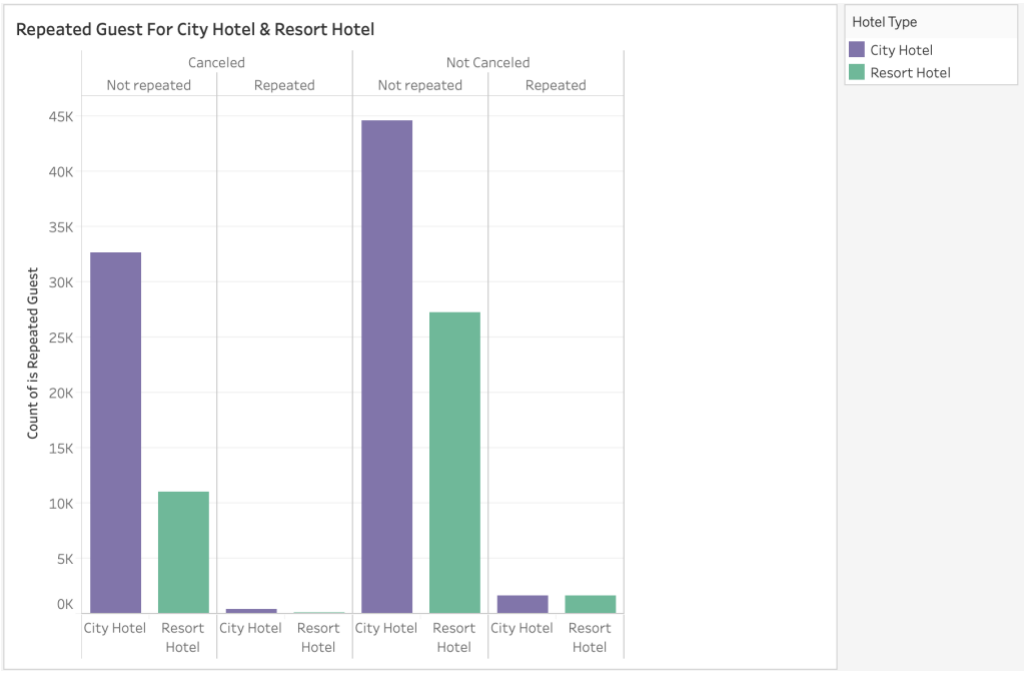


Title of x-axis (Arrival Date Month) that is obvious in the before graph, has been removed because the title of the graph itself shows it is about the month and we can see the label of each month in x-axis.

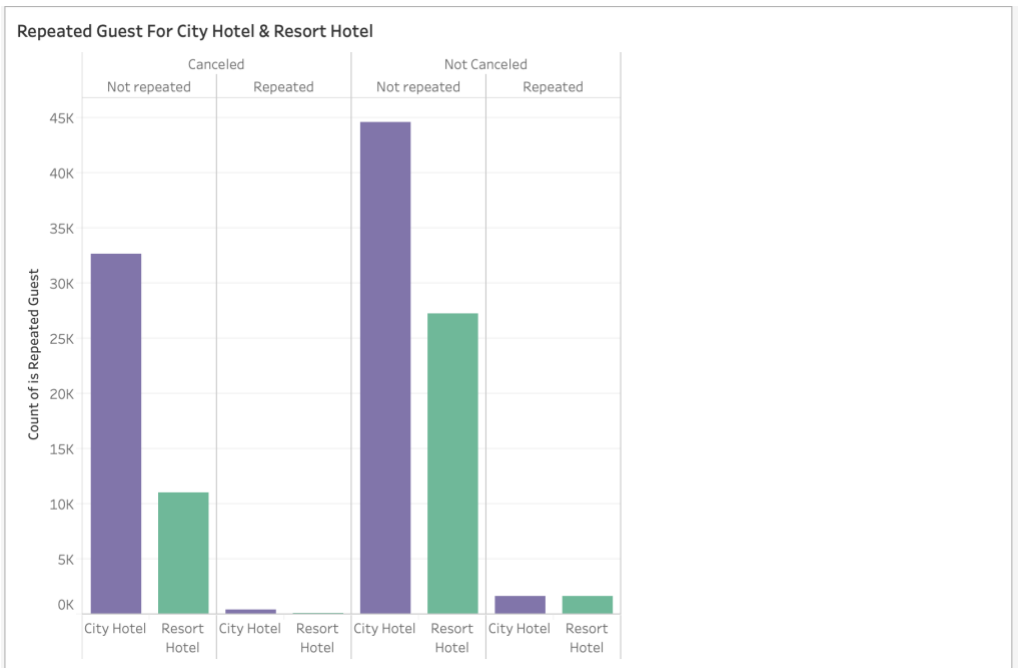
“Cancellation Status” filter box has been removed from the graph because the title of y-axis shows these bookings are about not cancelled reservations.

Iteration 6:

Before:



After:

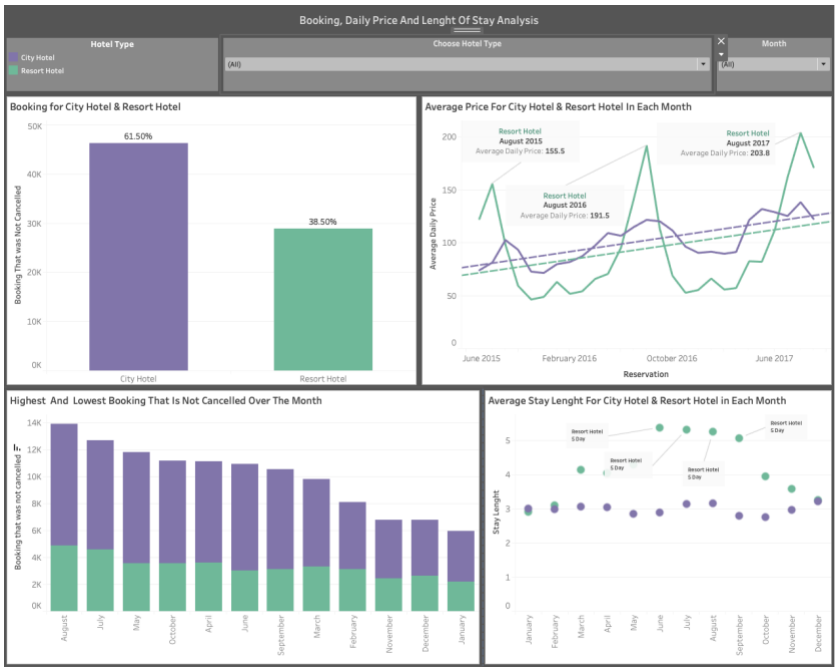


“Hotel type” filter box that shows the color of each hotel has been removed from the graph because x-axis labels clearly describe the sectors being represented.

6. Alternative

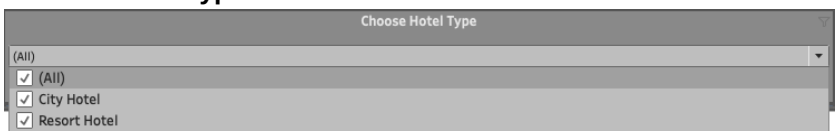
Alternatives for Dashboard 1

Dashboard 1:

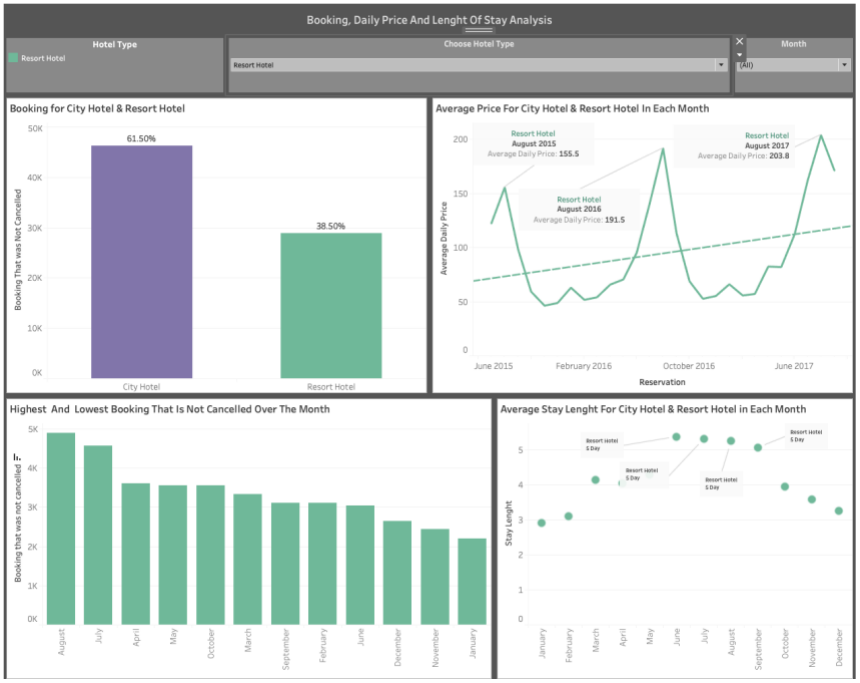


Alternatives 1:

Choose Hotel Type Filter Box:



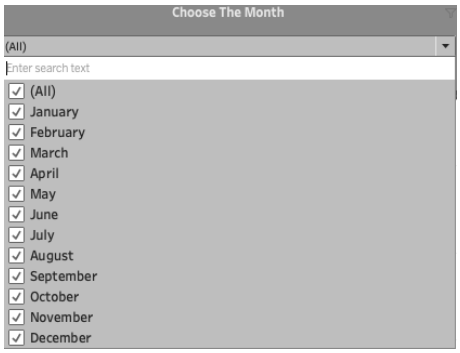
Example: Choose Resort Hotel



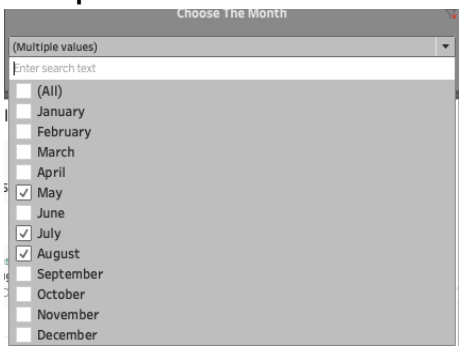
After Choosing Resort hotel, the 3 graphs in Dashboard 1 show the information just for the resort hotel. It also can be changed by the City hotel as well.

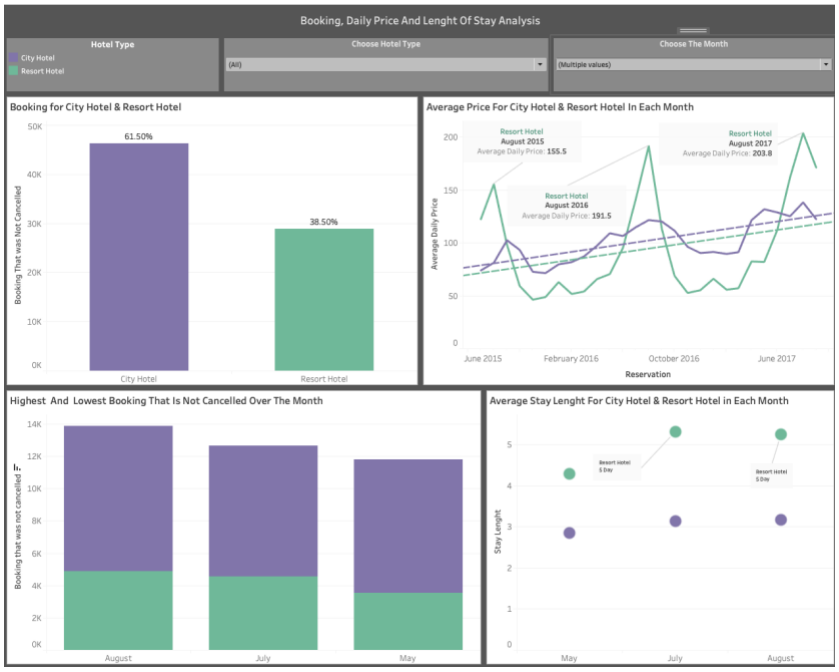
Alternatives 2:

ChooseThe Month Filter Box:



Example: choose 3 months



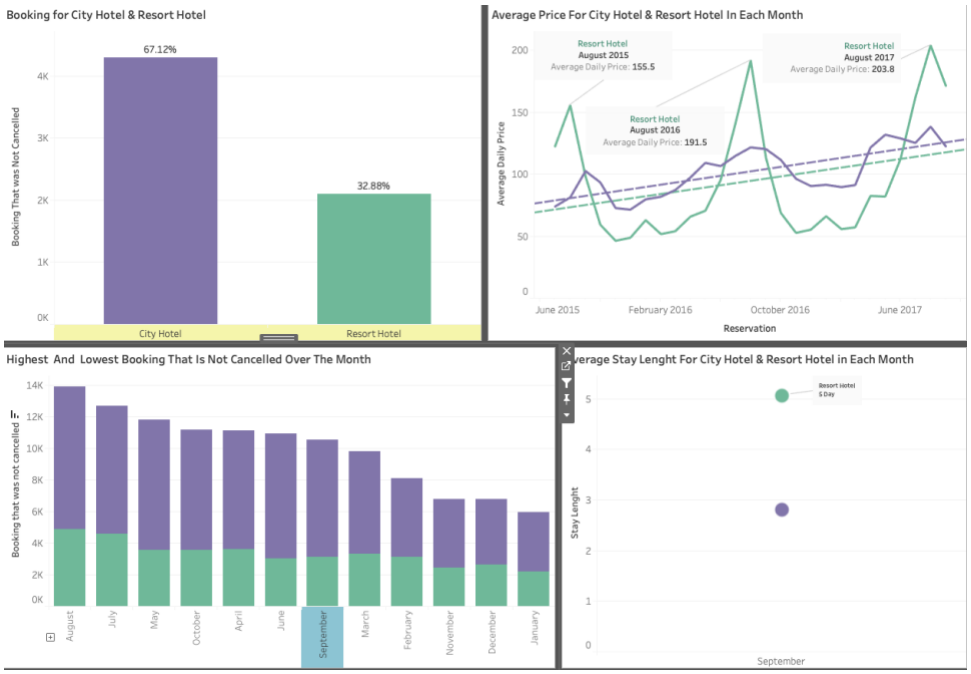


After Choosing 3 months, the first graph and the two bottom graphs show the information about those 3 months.

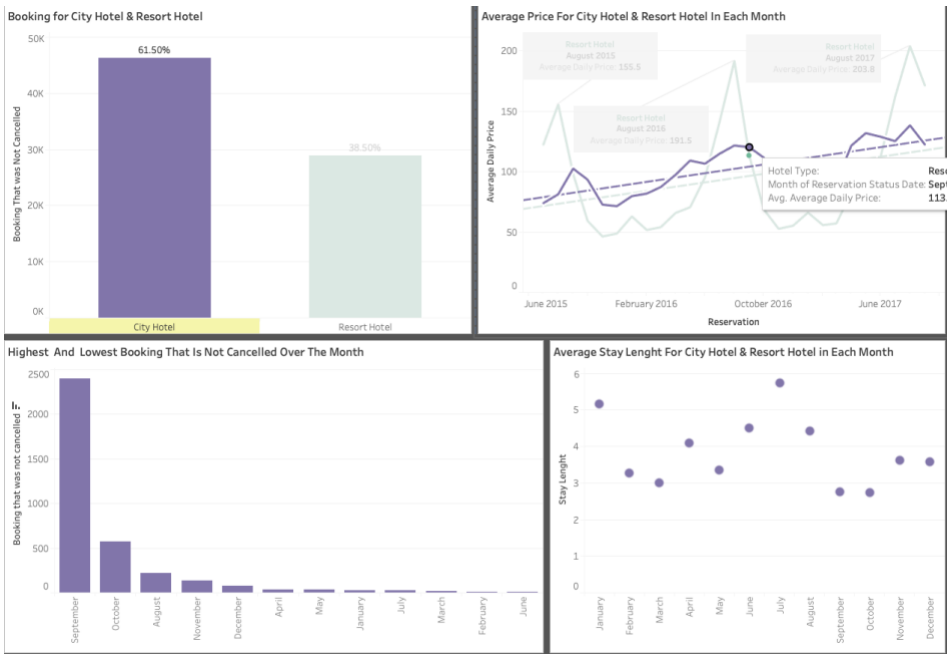
Alternatives 3:

Choose the Month and the Hotel type from the graphs and affected on all others graphs:

1. By choosing one month from the left bottom graph, the 3 graphs show that month information except the right top graph.



2. By choosing the city hotel from the top right graph, all the graphs show that hotel information.

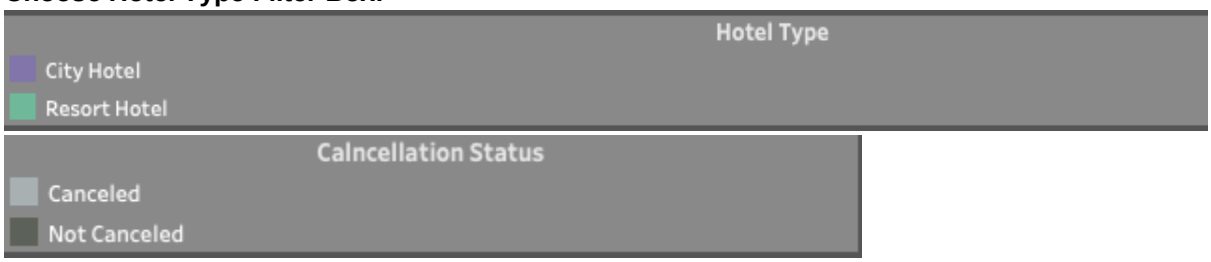


Alternatives for Dashboard 2

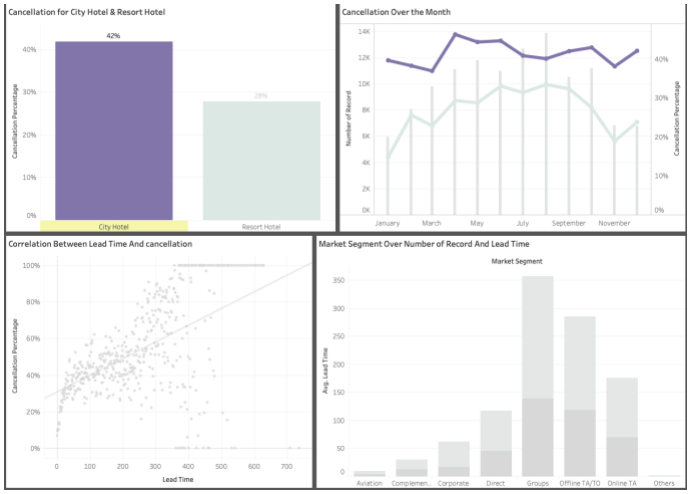
Dashboard 2:

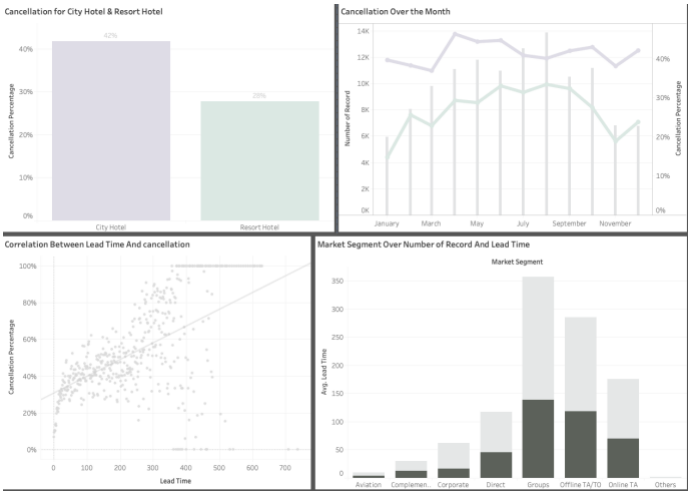
Alternatives 1:

Choose Hotel Type Filter Box:



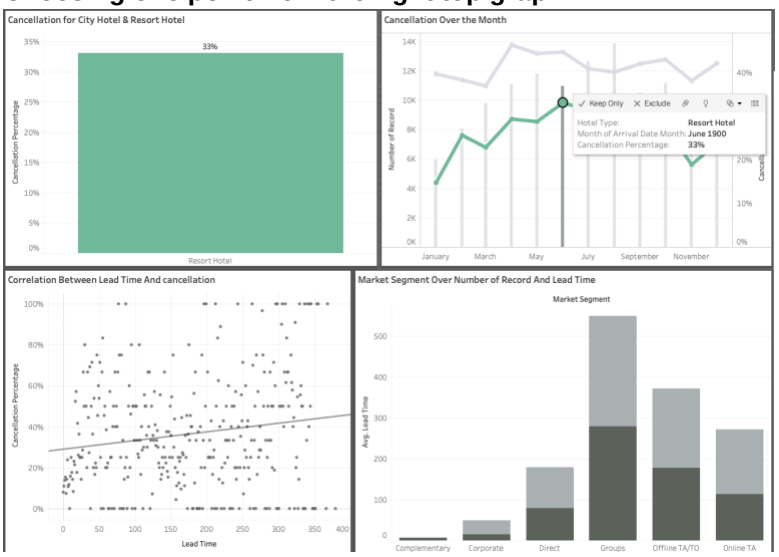
Example: by choosing City hotel and Not cancelled status you can see that color and information in the dashboard





Alternatives 2:

Choosing one point from the right top graph:



After Choosing one point, all the graphs change to show that point information on their graph.