Hotel Booking Demand Analysis

Team Members:
Shazmeen Shaikh
Tafseer Khan
Zainab Zaveri

1) Describe the scenario.

In a world where almost everything is digitized and anything can be done with a few clicks, including booking and cancellation of hotels, businesses are forced to adapt and accommodate consumers' rapidly changing plans. The service industry especially has the unique challenge of ensuring they have the appropriate amount of resources available at the right time to meet their customers' needs. To stay in competition, the hotels are providing free cancellation fees. This tactic encourages people who are uncertain of their plans to go ahead and book the room because there is no negative consequence if they aren't particularly serious about their trip and must cancel. While there might not be any risk for the consumer, there is a substantial amount of concern around cancellations from the hotel's perspective.

2) What is the problem you are trying to solve?

The rate of cancellations is rising because third party vendors often persuade individuals to use their services by advertising the lack of cancellation fees. Matter of fact, in 2018 the hotel cancellation rate seemed to reach its peak at around 40% [Hospitality Tech News, 2020]. To alleviate the cancellation, the hotels overbook the rooms in order to compensate for any cancelled room. The issue with overbooking is that the hotels don't know how many customers will cancel. Overbooking might be to the extent that the hotel would have to turn customers away. This can result in damaging their brand image and the hotel will end up losing those customers forever. In other cases, There could be fifty empty rooms each night if there is no overbooking. This problem could be solved by using predictive analytics which will identify cancellations and overbooking. That will allow us to overbook at an appropriate rate to ensure we have enough rooms available for the customers, while keeping empty rooms to a minimum.

3) Who is the intended audience?

The Hotel industry can make a lot of use from this analysis. Once the Executives and manager understand the problem, they can appoint a data analyst to find solutions and personalize it according to the needs of the hotel. Along with this, students and professionals who want to understand exploratory data analysis (EDA) or get started in building predictive models can use this case.

4) Where did your data come from?

The Hotel booking demand data was taken from Kaggle https://www.kaggle.com/jessemostipak/hotel-booking-demand but The data is originally from the article Hotel Booking Demand Datasets, written by Nuno Antonio, Ana Almeida, and Luis Nunes for Data in Brief, Volume 22, February 2019.

5) What is included in the data?

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 has 32 columns and 119390 rows.

7) Describe the steps taken in the project

Initially, we understood the data from both data and business perspective before doing any kind of analysis. There was a need for data preprocessing which included data cleaning. Analysis and visualization. After observing the data, we were able to develop recommendations.

The data cleaning was done in both R and excel but the final results were in excel. Data analysis and visualization was performed in tableau. The recommendations were given based on both observation of the dataset and the analysis.

8) Describe the analysis

The hotel industry comprises a fixed capacity, perishable goods, fixed number of hotel rooms. Revenue directly correlates with bookings and good consumption. Moreover, consumption is highly uncertain, driven in part by consumers' ability to cancel bookings. This specific analysis focuses on demand closure, the likelihood for a customer to book and later cancel the booking.

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.

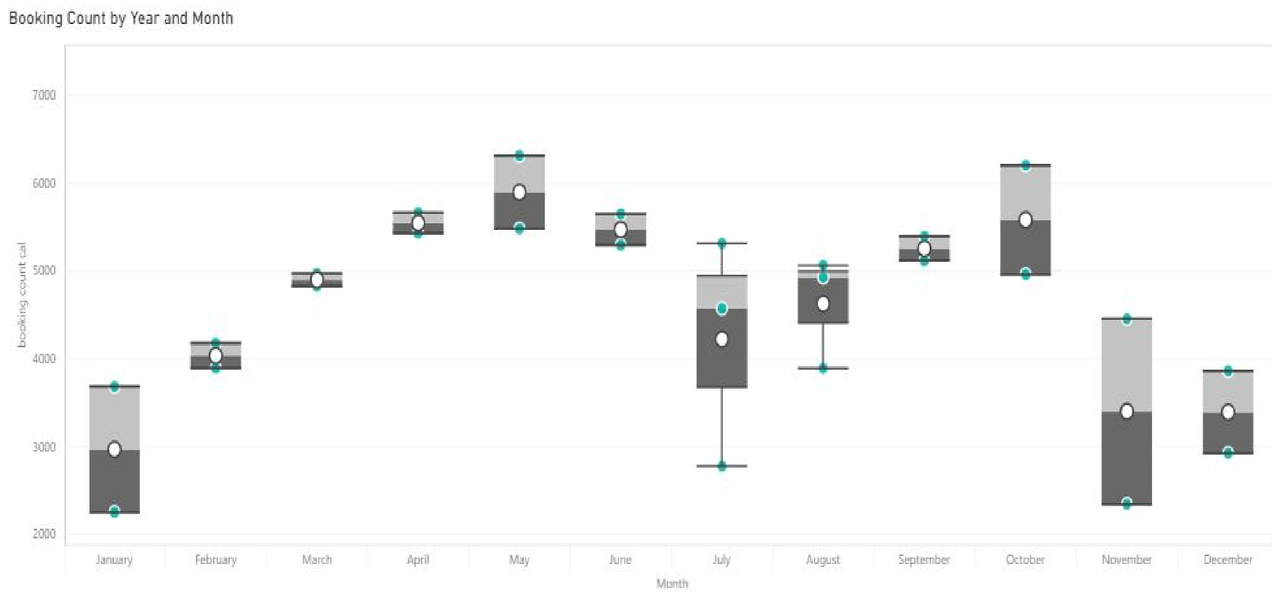
The analysis results in the following observations.

- There are a few values having ADR 0 or lesser which is not possible. There are a total of 1960 such values, these values were removed.
- There was minimal imbalance in the dataset between test and train
- More proportion of city hotels were cancelled than resorts.
- The average daily rent paid for resort booking that were NOT canceled is 90.8

- 14.5% of the revisiting guests cancelled their bookings.
- Portugal had the highest number of bookings compared to other countries.
- Most of the special requests were from assigned room ‘A’.
- The average daily rate (ADR) is 6.45% higher for guests on weekends.

9) Describe the discoveries (what did you learn from the data, etc.)

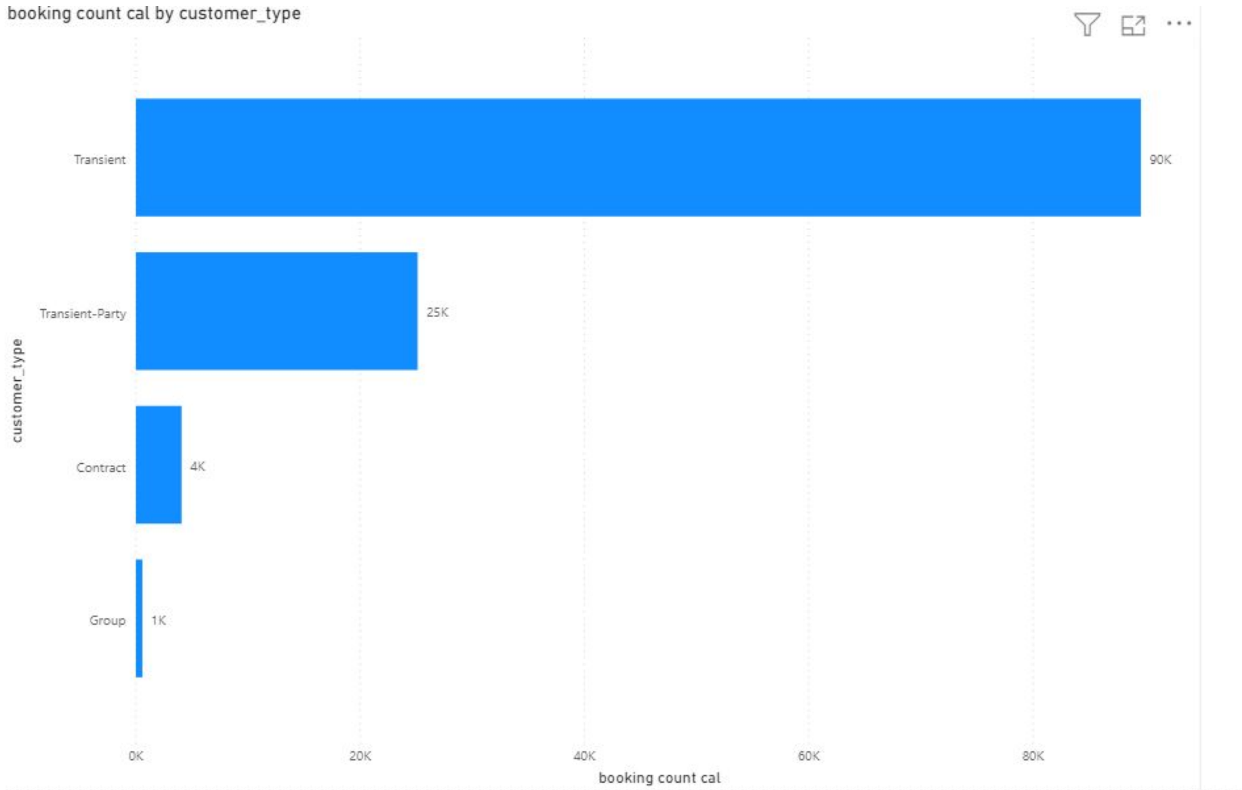
a. Distribution of number of bookings by arrival month



The graph above shows the bookings by arrival month and the box plots represent the variation in bookings of that month over the years. The bookings are lower in January, July and December and peaks in May and October. The box plots also show that the upper and lower months tend to have larger variations in bookings than the ascending and descending months in-between.

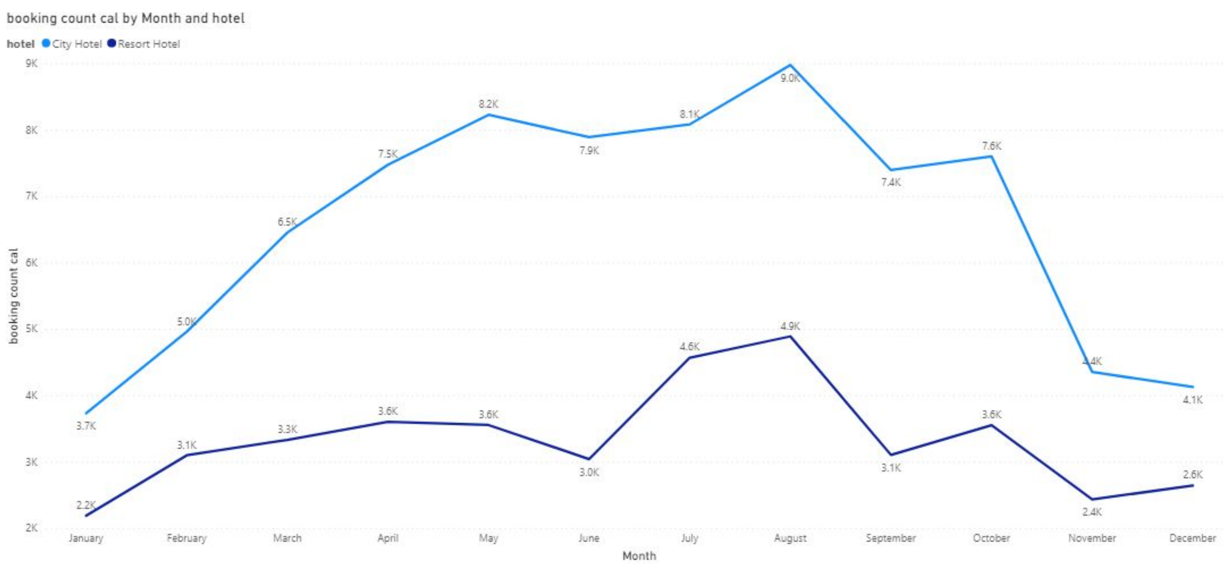
In this graph, bookings are plotted again by arrival month as a time series. We can see that there was an event between November 2015 and January 2016 that led to artificially depressed bookings at that time. May and October are clearly the best months for booking volume.

b. Distribution of number of bookings by customer segment



The bar chart shows the distribution of the four customer segments. It is clear from the graph that the majority of bookings come from the Transient segment or Transient Party. Group and Contract Bookings make up only about 4% of all bookings and thus variables that are heavily correlated to that group may skew the overall analysis or vice versa.

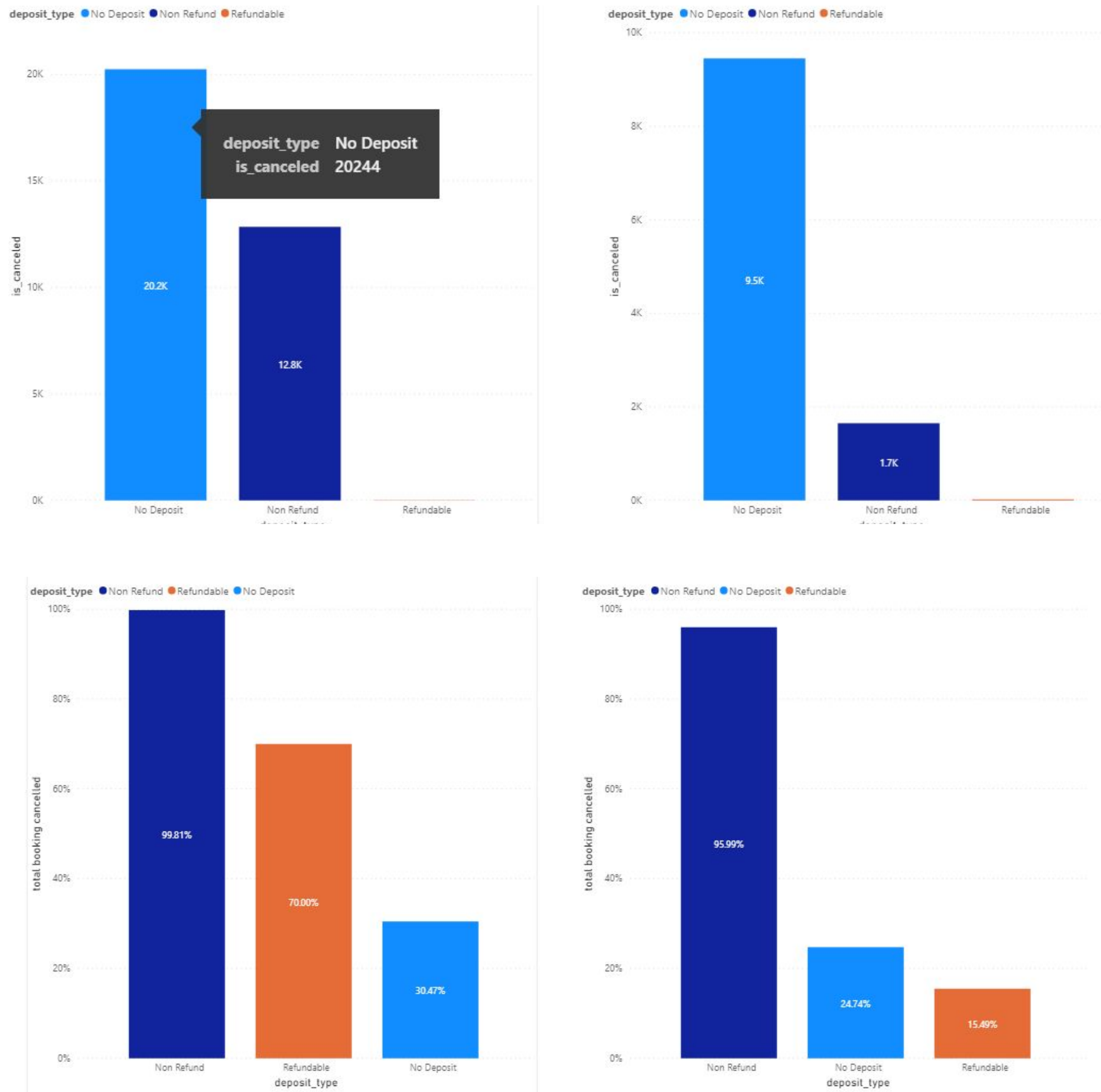
c. Distribution of cancellation by hotel and month



From the graphs above it is clear that the city hotel has a notably higher booking count and cancellation rate than the resort hotel. The city hotel is located closer to central Portugal so it's more easily accessible to a larger population. Large cities are often home to a plethora of businesses. It's likely that the customer base at the city hotel is composed of not only vacationers, but people travelling for business. For the resort hotel, the number of bookings and the number of cancellations is lowest during the holiday months - November, December, January. This makes sense because most vacation customers prefer to spend the holidays celebrating with their families rather than traveling. For the city hotel, since the customer base includes businessmen, the number of new bookings is low during these months, but the cancellation rate is still high. This could be because people still work during these months. For the resort hotel, the largest number of bookings occur in the months of July and August.

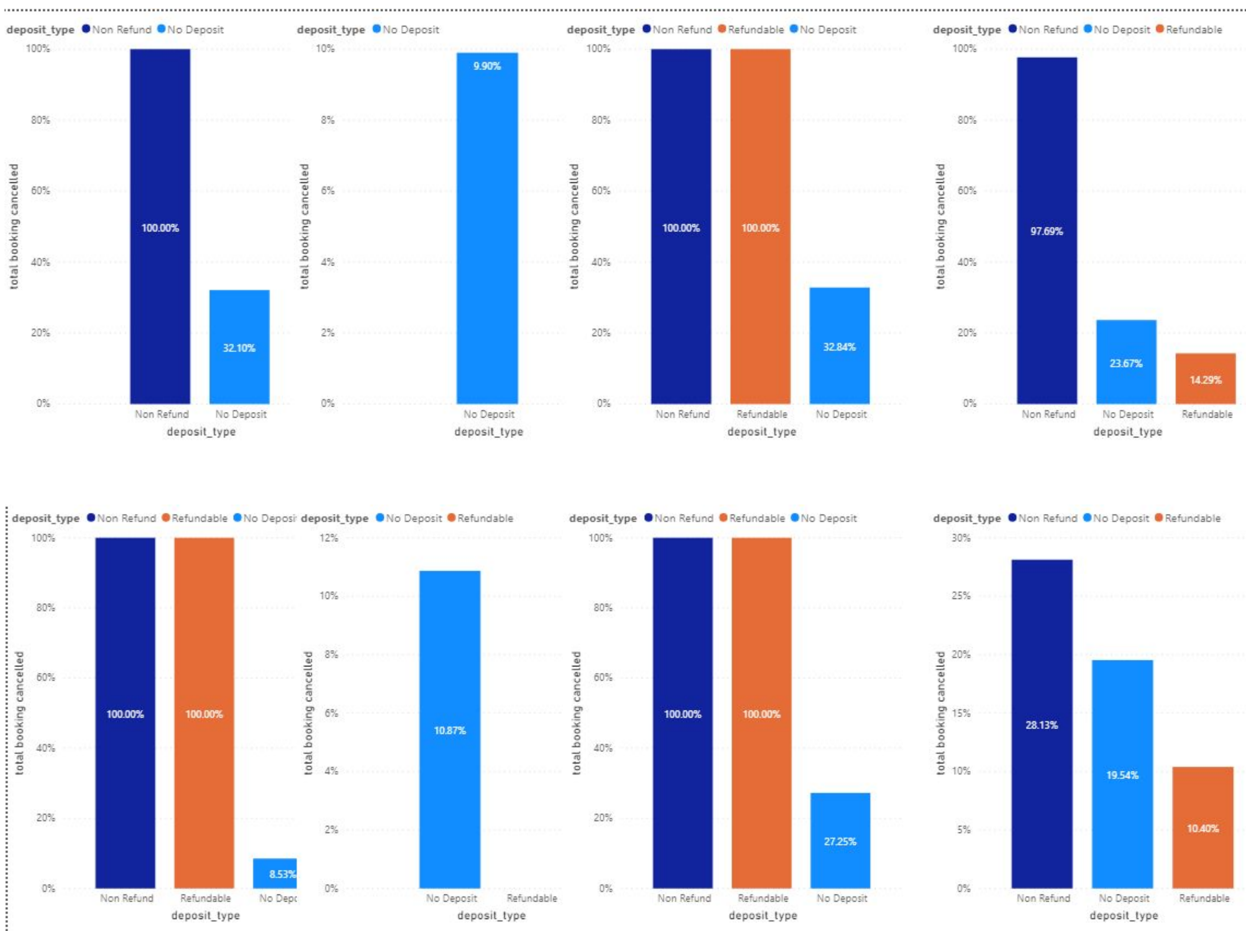
This is likely because students are on vacation during these months, so families are more likely to travel. Because those months are in high demand, it might make sense to increase cancellation rates during busy times such as student vacations. We see a similar spike in the city hotel, but their busy booking months are from May until August and their slow months are the first few and last few months of the year. While there is some fluctuation, the city hotel cancellation rate is constant throughout the year. This can be attributed to the idea that businesses are working all months out of the year so there will inevitably be bookings and cancellations to accommodate change of business plans. People traveling for leisure to the resort hotel, however, are more preoccupied with the holidays and less likely to book or cancel anything during these hectic holiday months. In this scenario, it might make sense to offer special sales during the holidays at the resort hotel to increase the number of bookings.

d. Effects of deposit on cancellations by hotel



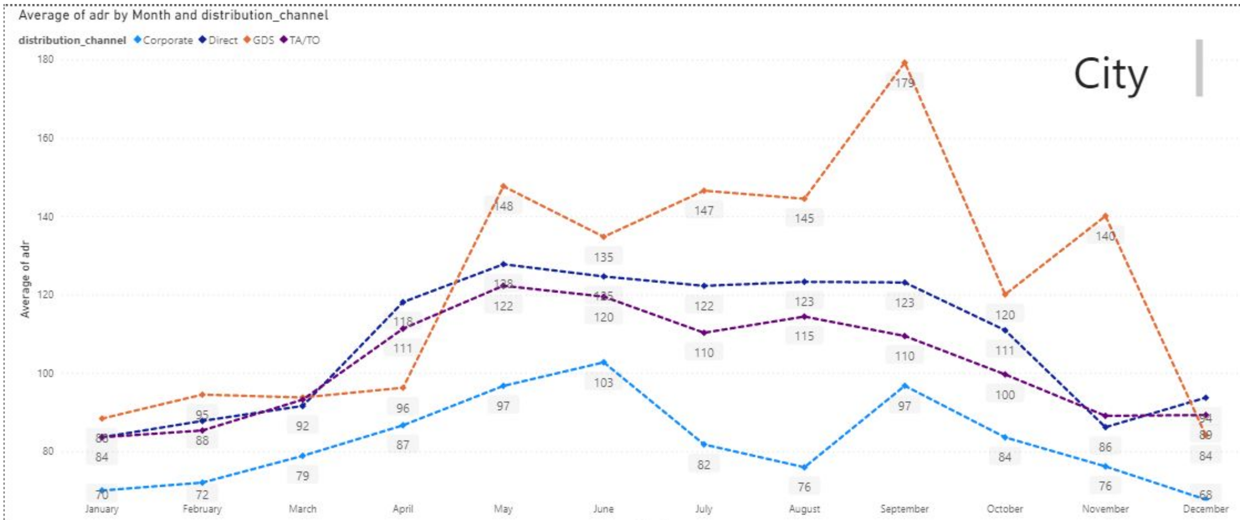
When it comes to cancellations, the effect of non-refundable deposit is the highest for both City and Resort hotels. When it comes to refundable deposits, they affect city hotels much more than they do for resort hotels (70.0% and 15.5% respectively). A cancellation from a no-deposit reservation has a 30.5% effect for city hotels and 24.7% effect for resort hotels. It is important to note that the most bookings for both hotel types come from no-deposit reservations, but that same type had the smallest effect on the cancellation rate. Non-refundable had the next highest bookings for both hotels (though very small compared to no-deposit), but they had the largest cancellation rate at 99.8% for the city and 96.0% for resort. Analyzing the type of deposit, a reservation has and the effect it has on cancellations is useful to help determine which reservations they can anticipate having cancellations and if overbooking should occur to offset the lost reservations.

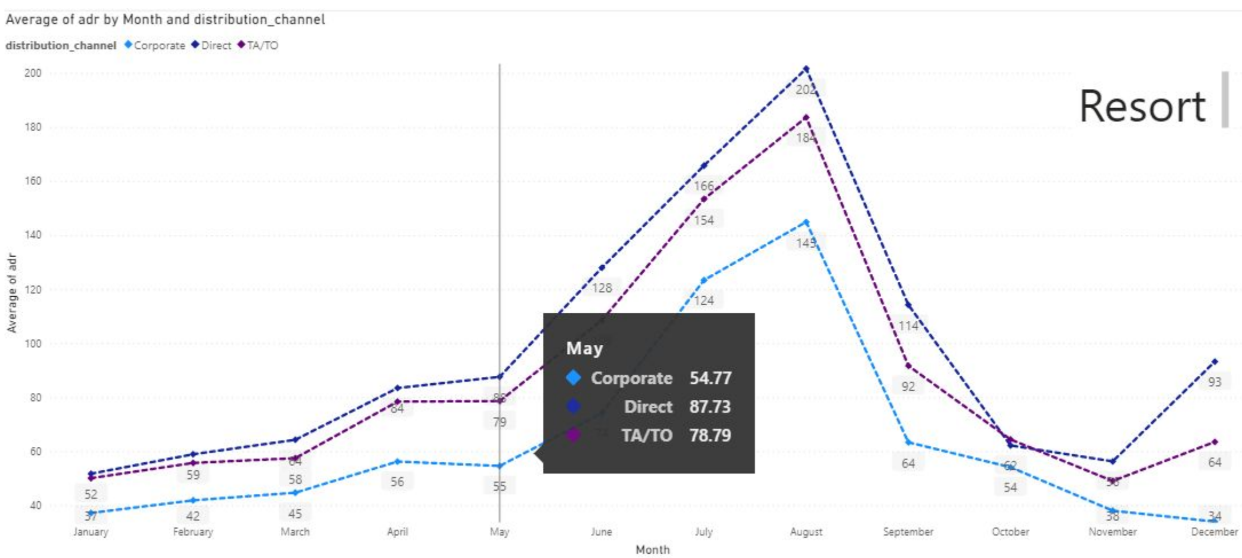
e. Effects of deposit on cancellations by customer segments



For City hotel cancellations, the effect of deposits was the highest for Transient . Contract had 100% effect and Transient-Party had 97.70% effect for non-refundable with the group being lower for both and refundable having only an effect on TransientParty. For Resort hotel cancellations, the effect of deposits was the highest for Transient and Contract (non-refundable and refundable at 100%). Transient-Party influenced all three deposit types, but they were in the lower range. with the group being lower for both and refundable having only an effect for Transient-Party. Group cancellations had the least effect, with only no deposit having a small effect. Overall, the Transient customer segments had the most similar effect profiles for both hotel types. Analyzing the type of deposit by each customer segment and the effect that has on cancellation rate is important as it could be helpful to see which customers have a higher chance of canceling and making sure cancellation and overbooking policies take the high canceling segments into account.

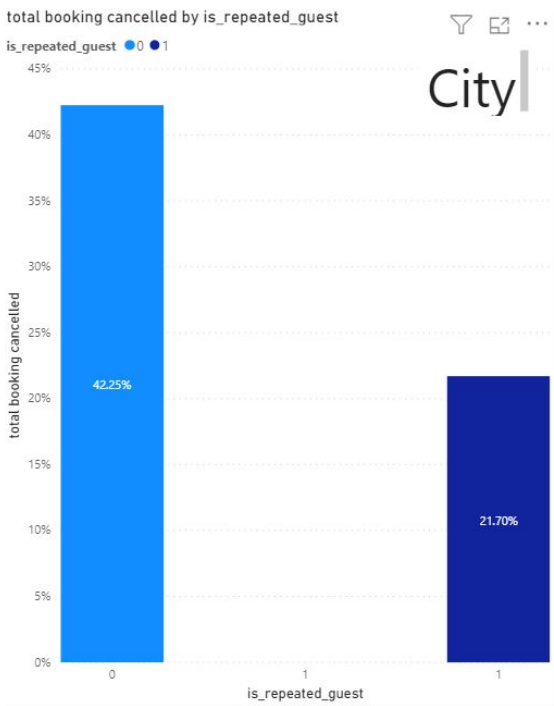
f. Distribution of ADR by hotel, month, and distribution channel

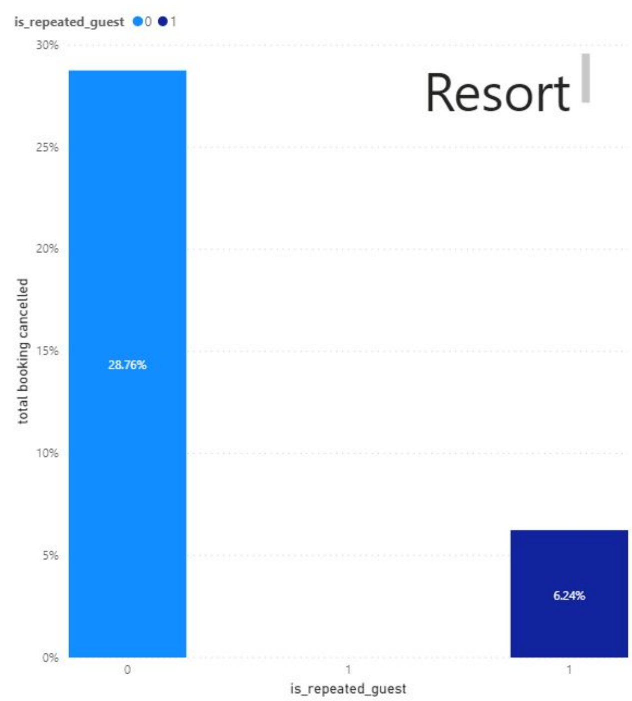




The monthly change in ADR across all distribution channels follows a similar pattern for the resort. That is, the rates gradually increase from January to May before significantly increasing during the summer months. After peaking in August, the ADR plummets in September and remains much lower than the summer rates through the end of the year. A large increase in ADR over the summer makes sense, as this is when consumer demand would be the highest. Not only are those the three warmest months of the year, but that time period also coincides with summer vacation for students, when families are most likely to take trips. For the city location, the three predominant distribution channels (Corporate, Direct, TA/TO) have more stable ADR throughout the year. Although the summer peak ADR for these channels is significantly higher at the resort location, the city hotel tends to have much higher ADR during the non-summer months. This indicates less fluctuation in seasonal demand for larger cities such as Lisbon, which has lots of travelers throughout the year. At both locations, the corporate channel has the lowest ADR each month. Corporations have the most bargaining power for room rates due to the scale of their business travel compared to that of individuals. Using the distribution channel as an independent variable for modeling purposes will require transformation. This can be done by creating 2 columns with binary responses (1-Yes, 0-No) for 2 of the 3 channels mentioned previously.

g. Relationship between repeated guest and cancellations





We can see from the plot above that a repeated guest is less likely to cancel the booking in both the city and resort hotels. In the city hotel, there is a 42% cancellation rate when compared to 22% for a repeated guest which is almost half of it. This difference seems to be a lot more in the resort as the likelihood of cancellation of a new guest is 29% which is 5 times more when compared to the numbers of a repeated guest at merely 6%. This can be explained by the fact that since the visitors at the resorts are more families and vacationers, repeated guests are already acquainted with the services provided and wouldn't cancel unless for a different reason when compared to new guests. SAs city hotels have more business travelers there is a high possibility of cancellation due to various factors even if they are repeated customers like canceling or rescheduling of meetings or conferences. Analyzing this information is helpful for the analysis as we can predict cancellations better based on whether a guest is new or a repeated customer and have flexible cancellation policies based on this information.

10) Describe any challenges encountered and how you resolved the challenge

It was hard to understand what kind of graphs would be best to display the information. While we tried a variety of different forms of charts and graphs, Bar and Line chart were most used. The major reason being it's one of the most popular ones and so it's easy to understand by the general audience.

The data seemed pretty clean when we first started making graphs but as we visualized in tableau, it turned out the dates weren't in date format but in text so all the graphs we received were in alphabetical order as opposed to date (month or year) order. We created a date hierarchy and removed null values to get more accurate results.

11) Describe any adjustments you had to make from the original plan

Originally, we planned on executing the whole project in using a programming language and a visualization tool. The plan was executed successfully but we used R studio instead of python and PowerBI instead of Tableau. The main reason for doing so was to learn rather than implement it in a language we already know. We also wanted to provide recommendations based on prediction results in .R.

12) Recommendations

- Dynamic customer policies: This is based on profiling customers and adjusting pricing or cancellation policies. The model can predict a specific customer's likelihood to cancel based on model prediction and adjust or offer options.
- Dynamic inventory management: As time progresses closer to a given date, the certainty of the booking would improve, dynamically adjusting inventory levels and pricing allows the hotel to maximize revenue.
- Strategic Planning and Operations – The number of ways a hotel management can strategize, build processes and campaigns around the model are limitless. From browser cache, IP addresses, loyalty programs tons of additional data can be gathered for targeting ads, pricing options, cancellation policies, or partnership programs.
- The overbooking can be changed based on season for example, November, December and January has the lowest bookings so here, more reservations can be taken and thus more bookings.
- City hotel has a higher rate of bookings and cancellations then the resort hotel. It is recommended that the overbooking in city hotels can be slightly higher.
- Many cancellations (29,718) came from reservations that did not require a deposit. Overbooking this type of reservation will allow the hotels to offset a loss of revenue. This model, when taking other variables into account, will predict the number of cancellations that are expected to allow the hotel to overbook at a rate that will likely not be noticed by the consumer.

- Repeated guests should not be overbooked as they are likely to cancel in both types of hotels. Plus, the brand would not want to risk their loyalty by booking them at another accommodation.

13) Deliverables
- Excel file (Kaggle) [Provided kaggle link and .csv file in drive]
- Tableau workbook[provide in drive link and submission]
- Report

Reference:

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