HOTEL BOOKING ANALYSIS



**Data Analysis Steps**

**A blue and white graphic with a magnifying glass and a computer

Description automatically generated**

1. Create a Problem statement
2. Identify the data you want to analyse.
3. Explore and clean the data.
4. Analyse the data to get useful insights.
5. Present the data in terms of reports or dashboards using visualisation tools.

**Tool and Technology Used**

* Python

**Business Problem**

* City hotel and Resort hotel have seen high cancellation rates. Each hotel is now dealing with a number of issues as a result, including fewer revenues and less than ideal hotel room use. Consequently, lowering cancellation rates is both hotels’ primary goal in order to increase their efficiency in generating revenue, and for this analysis, to offer thorough business advice to address this problem.
* The analysis of hotel booking cancellations as well as other factors that it has bearing on (their business and yearly revenue generation) are the main topics of this report.

**Assumptions**

* No unusual occurrences between 2015 and 2017 will have a substantial impact on the data for the analysis.
* The information is still current and can be used to analyse the hotel’s possible plans in an efficient manner.
* There are no unanticipated negatives to the hotel employing any advised technique.
* The hotels are not currently using any of the suggested solutions.
* The biggest factor affecting the effectiveness of the earning income is booking cancellations.
* Cancellations result in vacant rooms for the booked length of time.
* Clients make hotel reservations the same year they make cancellations.

**Focus of the Analysis**

* What are the variables that affect hotel reservation cancellations?
* How can we try to improve hotel reservations cancellations?
* How will hotels be assisted in making pricing and promotional decisions?

**Hypothesis**

* More cancellations occur when prices are higher.
* When there is a longer waiting list, customers tend to cancel more frequently.
* The majority of clients are coming from offline travel agents to make their reservations.

**Python libraries and functions used**

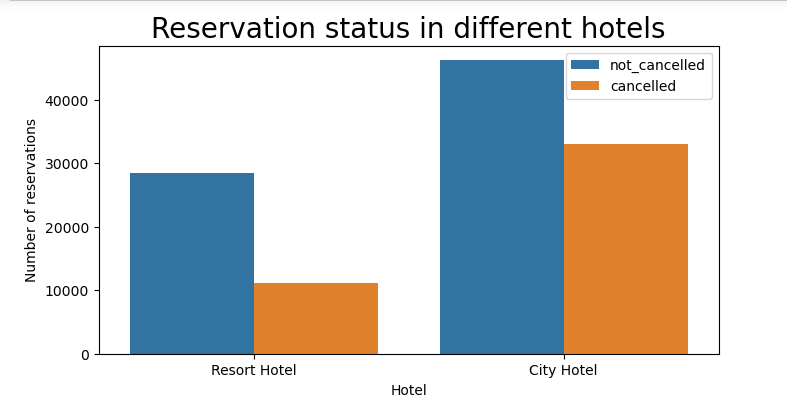
* Libraries used:

1. Pandas
2. Seaborn
3. Matplotlib
4. Warnings

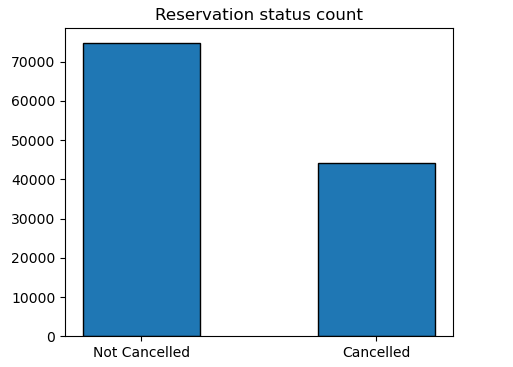
* Different functions and properties used for EDA:
  1. read\_csv()
  2. head()
  3. shape
  4. columns
  5. info()
  6. to\_datetime()
  7. describe()
  8. isnull()
  9. sum()
  10. drop()
  11. dropna()
  12. plot()
  13. value\_counts()

**Analysis and Findings**

The following bar graph shows the number of reservations that are cancelled and those that are not. It is obvious that there are still a significant number of reservations that have not been cancelled.

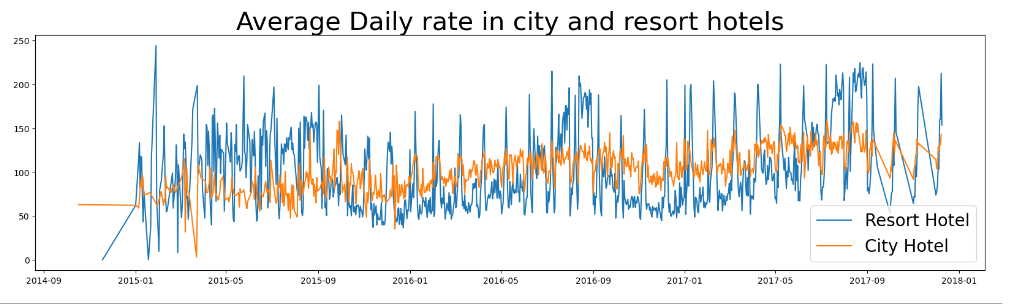


There are still 37% of clients who cancelled their reservation, which has a significant impact on the hotels’ earnings.

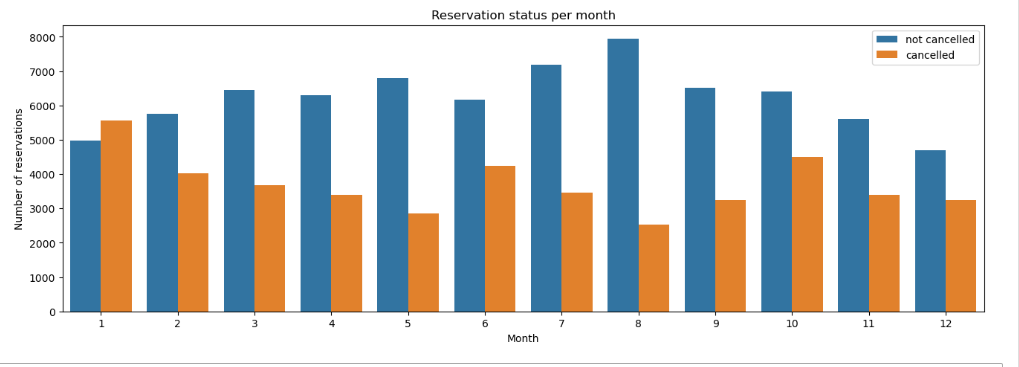


Also, it can be deduced that in comparison to resort hotels, city hotels have more bookings. It is possilbe that resort hotels are more expensive than those in cities.

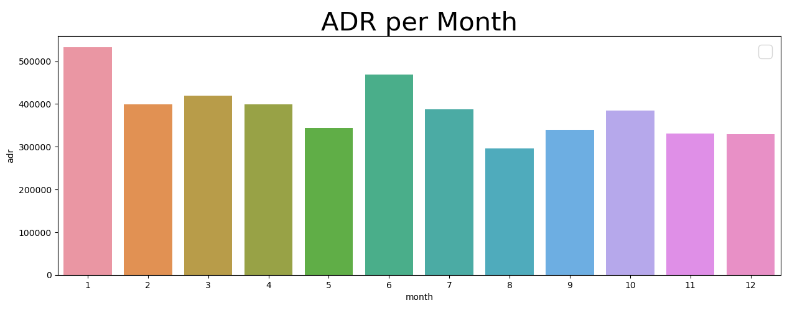
On analysing the following line graph, which shows the average daily rate(ADR), it can be stated that on ceratin days, the ADR for a city hotel is less than that of a resort hotel. It goes without saying that weekends and holidays may see a rise in resort hotel rates.



There is also a grouped bar graph to analyse the months with the highest reservation levels according to reservation status.

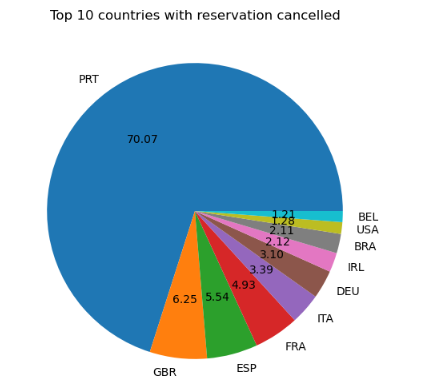


As can be seen, both the number of confirmed reservations and the number of cancelled reservations are the largest in the month of August, whereas January is the month with the most reservation cancellations.



The above bar graph demonstrates that cancellations are most common when prices are the greatest and are least common when they are the lowest. Therefore, the cost of the accomodation is solely responsible for the cancellations.

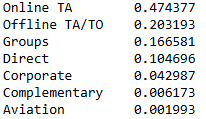
The following is the distribution of reservation cancellations in different countries.



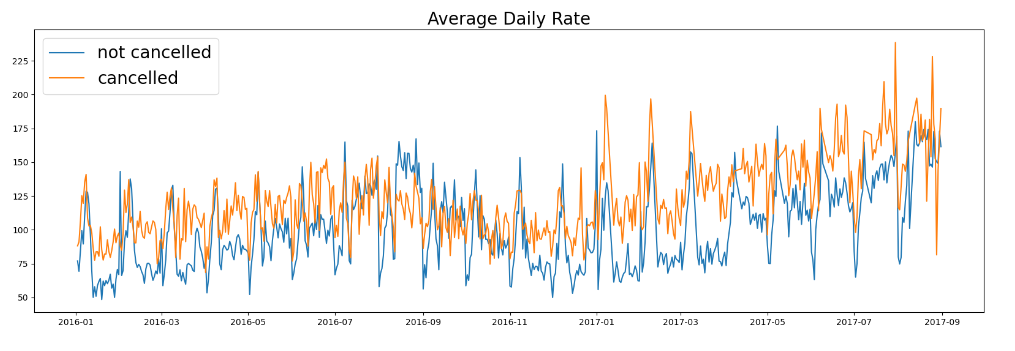
The country with the code of ‘PRT’ has the highest number of reservation cancellations with is approximately 70%.

The following is the distribution capturing the segments from where the guests are visiting the hotels and making their reservations. Are they coming from ‘Direct or Groups’, ‘Online’ or ‘Offline Travel Agents’?

**Market segment distribution**



From the above results, we find that around 46% of the clients come from online travel agencies, whereas 27% come from groups. Only a small number of them, around 4% of the clients book hotels directly by visiting them (Offline mode) and making reservations, this disproves one of our assumption that majority of them are from the ‘Offline mode’.



Further, from studying the above graph, it can be stated that reservations are cancelled when the ADR is higher than when it is not cancelled. It clearly proves all the above analysis all the more that the higher price leads to higher cancellation.

**Suggestions**

1. Cancellation rates follow the trend of the prices. In order to prevent cancellations of reservations, hotels could work on their pricing strategies and try to lower the rates or keep them in a lower margin for specific hotels based on locations. They can also use discounts to attract the potential consumers.
2. As the ratio of the cancellation and not cancellation of the resort hotels is higher than that of the city hotels, the hotels should provide a reasonable discount on the room prices on weekends or on holidays.
3. In the month of January, hotels can start campaigns or marketing with a reasonable amount to increase their revenue as the cancellation is the highest in this month.
4. Improving the quality of their hotels can also help, with a special focus on the country with the code ‘PRT’. This can significantly attract more customers, give a postive word of mouth and in turn reduce the reveservation cancellations in the future.