DECLARATION: I understand that this is an **individual** assessment and that collaboration is not permitted. I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar. I understand that by returning this declaration with my work, I am agreeing with the above statement.

1 Have you ever created a visualization of data, however small?

Yes, I have produced a data visualisation. Analysis of hotel reservations that are cancelled but have no impact on annual revenue creation is one of the main objectives of this visualisation.

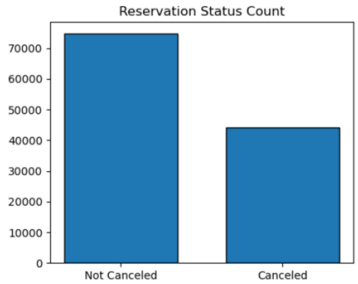
Description About Data :
Hotel cancellation rates in resorts and cities have recently been very high. Due to this, each hotel is currently faced with a variety of problems, such as decreased profits and less than optimum hotel room usage. As a result, both hotels want to boost their efficiency in generating money, therefore lowering cancellation rates is a top priority. This is why we provide extensive data visualisation.
I have downloaded the Kaggle hotel booking datasets, which include booking information from July 2015 to August 2017. This dataset initially has 36 columns and 119390 rows of observations for a city hotel and a resort hotel. However, I removed 5 columns that I didn't need for my study. For the sake of my data analysis, I use a variety of columns. For instance, the **hotel** columns only include the two values of city hotel and resorts hotel. and the columns **is_canceled**, which has two values of 0 and 1. The value 1 denotes a cancelled reservation, while a value of 0 implies no cancellation. Additionally, one of the columns with values used for data analysis is **reservation_status**. It has two values : check-out: the client has arrived but has already left; No-Show: The customer checked in but did not arrive. On top of that, I made a few assumptions about the datasets and a few particular hotels. Below is a list of them.

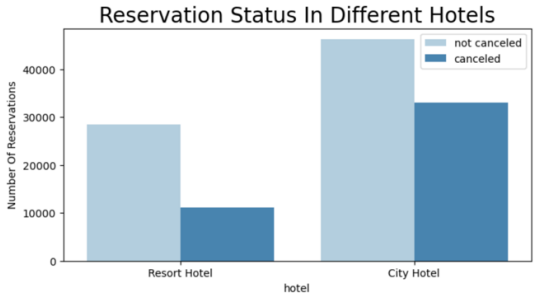
1. Between 2015 and 2017, there were no unusual occurrences that would significantly affect the data used.
2. Booking cancellations are the main element impacting how well one may generate income.
3. Customers book hotels the same year they cancel their bookings.
4. The hotels are not using any of the suggested solutions

Tools Used:

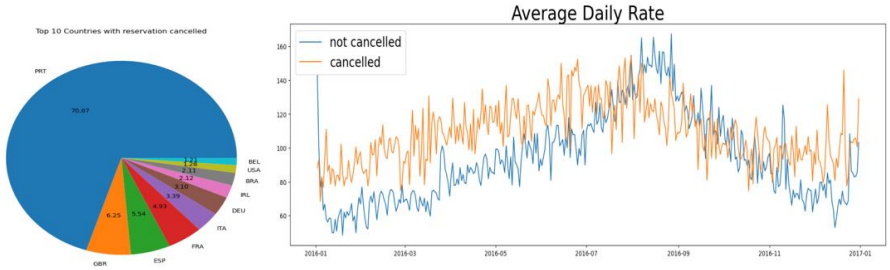
1. **Pandas:** Python's Pandas package is used to manipulate data sets. It provides resources for exploring, organising, analysing, and manipulating data.
2. **Matplotlib.pyplot:** Pyplot is a Matplotlib module that offers straightforward methods for including plot elements like lines, pictures, text, and other things.
3. **Seaborn:** Python's Seaborn package allows us to create statistical visuals. It incorporates tightly with Pandas data structures and is built upon Matplotlib. Seaborn helps in our data exploration and comprehension.

Data Analysis and Visualisations:

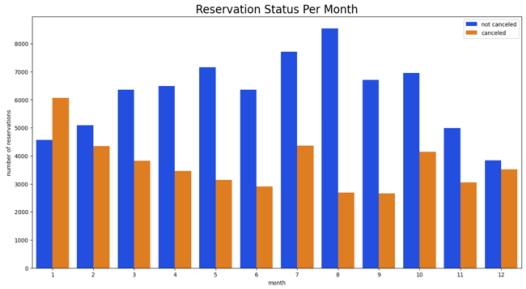




According to the first bar graph, 37% of clients cancel their reservations, which has a big influence on the hotel's revenue. In the second bar graph, we note that city hotels receive more reservations than resort hotels. We can assume that hotel rooms at resorts will cost more money than those in city hotels.



The pie chart shows that Portugal is the top country with highest number of cancellations. The line graph above shows that on some days, the average daily fee for a city hotel is less than that of a resort hotel, and on other days, it is even less.



The above bar graph analyses the months with the highest and lowest levels of reservations based on the status of the reservations. As we can see, August has the highest number of confirmed reservations as well as the lowest amount of cancellations, while January has the highest number of cancellations.

2 Is the visualization Explanatory or Exploratory? Discuss why you think so.

The above visualization is exploratory data visualizations. First we identified the problem statement and the data we want to study even though we had little knowledge of the datasets at first. Then we investigate how to clean the datasets by removing superfluous columns and null values from the datasets. Then, using bar graphs, pie charts, and line graphs, we study the data to gain meaningful insights.

3 In the example above, what are some of the reasons for why visualization might be required or beneficial? Could the objective/purpose be achieved without visualization.

Here are some of the reasons why visualization is beneficial. They are as follows:

1. Visualizations helps us in understanding more complex data. visualizations become a crucial tool for comprehending data when we have enormous datasets. For instance: Since there are initially about 119390 rows and 36 columns, understanding the data would be a very taxing effort.
2. Visualizations also helps in making informed decisions. From the above example we can suggest hotels owners to start campaigns or marketing with a reasonable amount to increase their revenues as the cancellation is highest in the month of January.

In summary, By simply looking at the data directly, it would be very difficult to give decisions. visualization is a flexible tool that is used for a variety of tasks in a wide range of fields, It enables users to view, interact with, and comprehend data more effectively.

References

1. https://www.kaggle.com/datasets/mojtaba142/hotel-booking/