Project: Hotel Booking

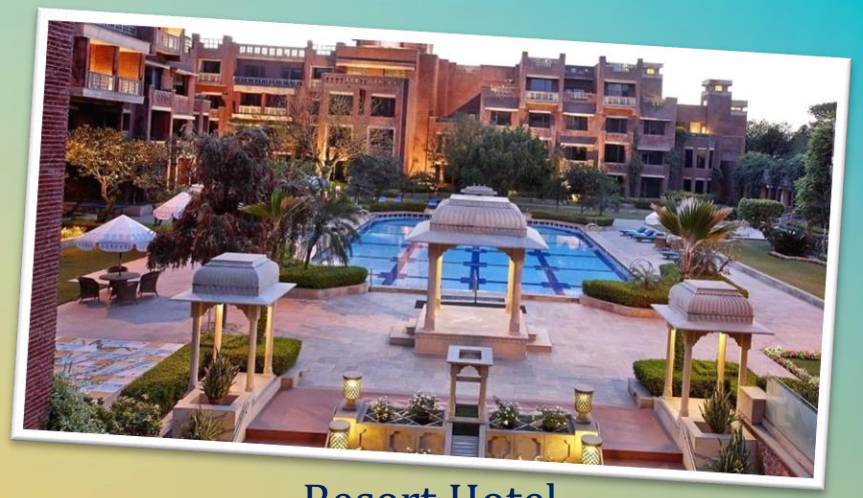
The complete process……

Problem Statement → Assumption → Research Questions → Hypothesis → Data Loading → Exploratory Data Analysis → Data Analysis → Final Conclusion

Business Problem

- In recent years, 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.
- This is the hotel booking dataset. The dataset contains two types of hotels: Resort(Expensive) hotel and city(affordable) hotel.
- Consequently, lowering cancellation rates is both hotels' primary goal to increase their revenue efficiency, and for us to offer thorough business advice to address this problem.
- The main topics of this report are the analysis of hotel booking cancellations as well as other factors that have no bearing on their business and yearly revenue generation.

City Hotel



Resort Hotel

Assumptions

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

Research Question

1. What are the variables that affect hotel reservation cancellations?
2. How can we make hotel reservation cancellations better?
3. How will hotels be assisted in making pricing and promotional decisions?
4. Which months face the highest number of cancellations?
5. Which country facing the most numbers of cancellation problems?
6. From which segment has the most number of guests coming?

Hypothesis

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

We'll be solving this problem with the help of python

Now its time to import Python libraries into jupyter notebook

Importing Libraries

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

After Importing the dataset, this is how our data look like. It contains 1,18,897 rows and 32 columns.

```
In [49]: # Overview of the dataset
         df
```

Out[49]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	st
0	Resort Hotel	0	342	2015	July	27	1	0	
1	Resort Hotel	0	737	2015	July	27	1	0	
2	Resort Hotel	0	7	2015	July	27	1	0	
3	Resort Hotel	0	13	2015	July	27	1	0	
4	Resort Hotel	0	14	2015	July	27	1	0	
...	...	...	...	...	...	...	...	...	...
119385	City Hotel	0	23	2017	August	35	30	2	
119386	City Hotel	0	102	2017	August	35	31	2	
119387	City Hotel	0	34	2017	August	35	31	2	
119388	City Hotel	0	109	2017	August	35	31	2	
119389	City Hotel	0	205	2017	August	35	29	2	

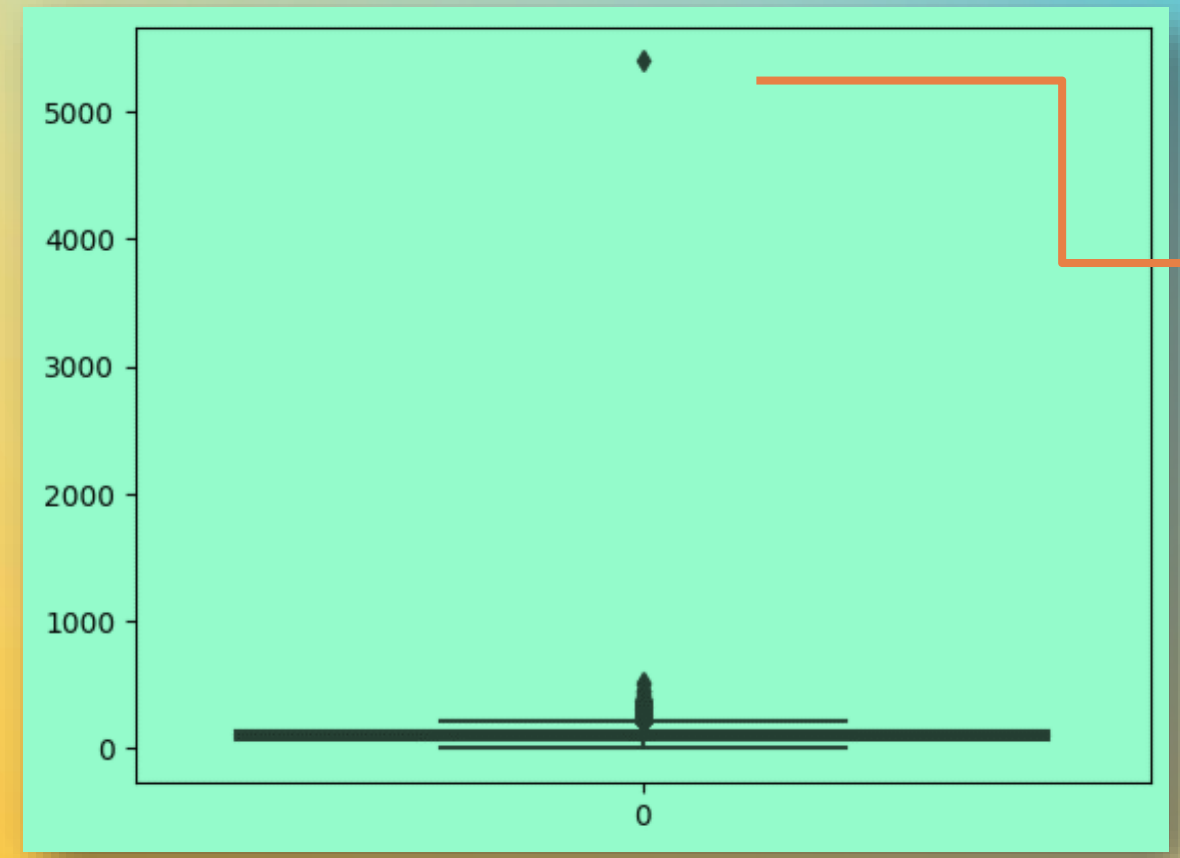
118897 rows × 32 columns

Now we'll perform some operations like Data cleaning and Exploratory data analysis.

After checking the information in the dataset, I came to know:

- **Columns:** *children, country, agent,* and *company*, all these columns contain null values
- **Column:** *reservation_status_date*: It is a date data type but it was showing an object.
- we can't remove such columns which contain fewer null values. we can remove the company and agent columns because they contain huge null values otherwise it will affect our analysis.
- ** If too much data are missing in a single column then remove columns(well, in this analysis we don't need company and agent columns so we can remove these)
- ** If less data are missing then remove rows only.

Checking Outliers of average daily rate (if any) using a box plot.



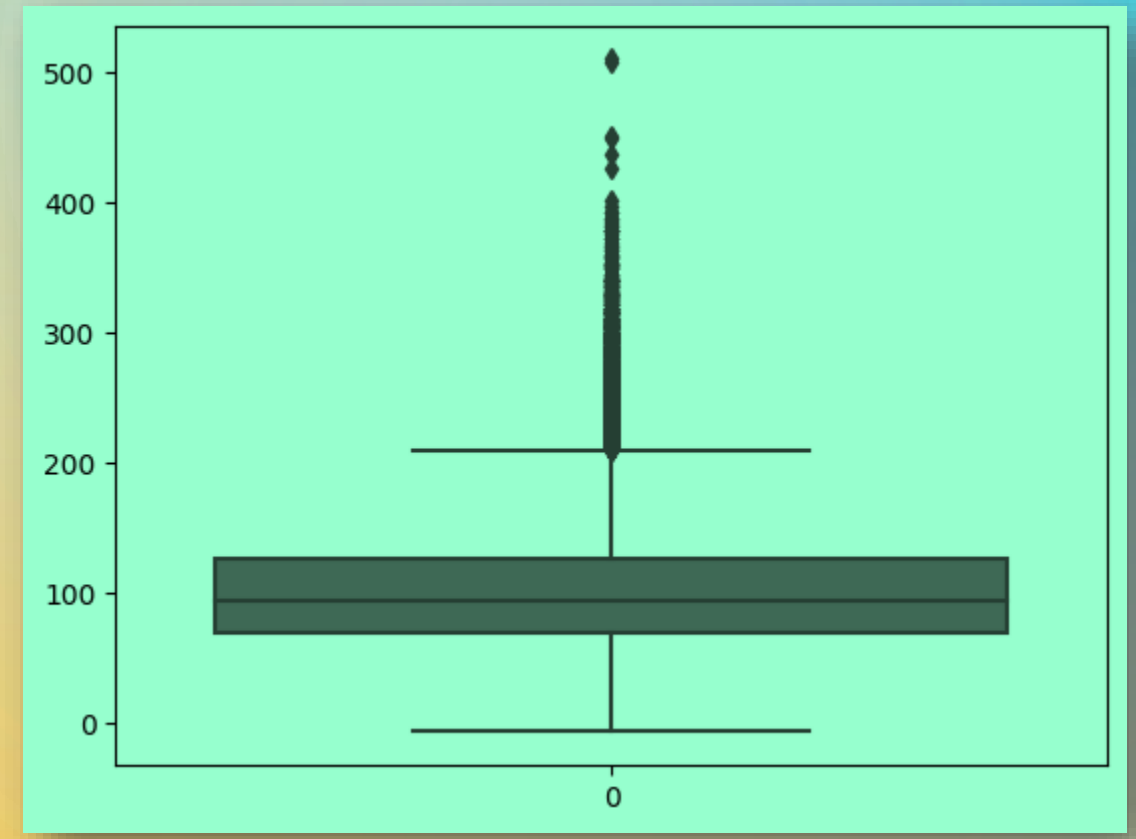
$$ADR = \frac{\text{Room revenue}}{\text{Number of rooms sold}}$$

Outliers

These are data points that significantly differ from the rest of the data in a dataset.
In statistics and data analysis, outliers are observations that lie far away from the central tendency of the data, often much higher or lower than the average or median values.

Here we can see the adr(Average Daily Rate) columns contains outlier above 5000.
Therefore we decided to exclude values above 5000.

Graph After clearing the outlier.

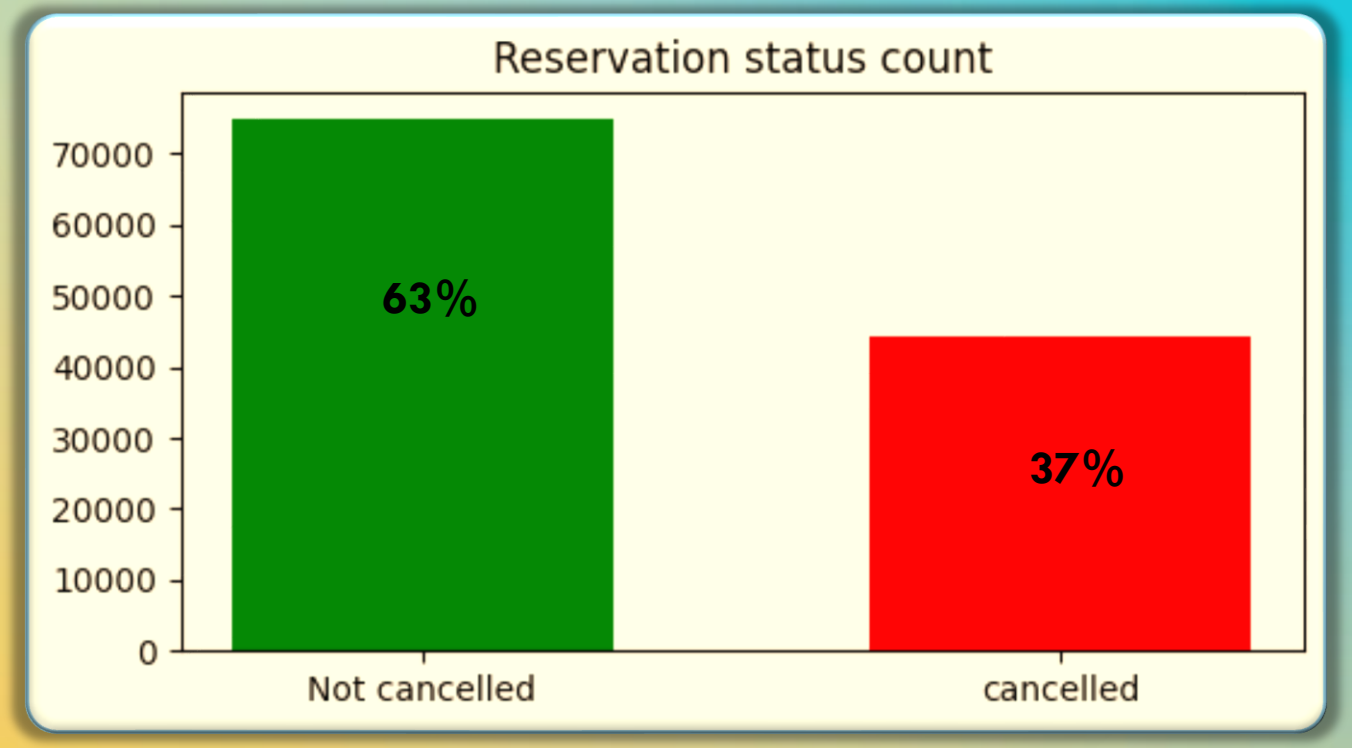


Now it's much better, still outliers but we can perform operations.

Almost EDA part is completed, now we can move to the **analysis** part. We'll tackle each problem thoroughly.

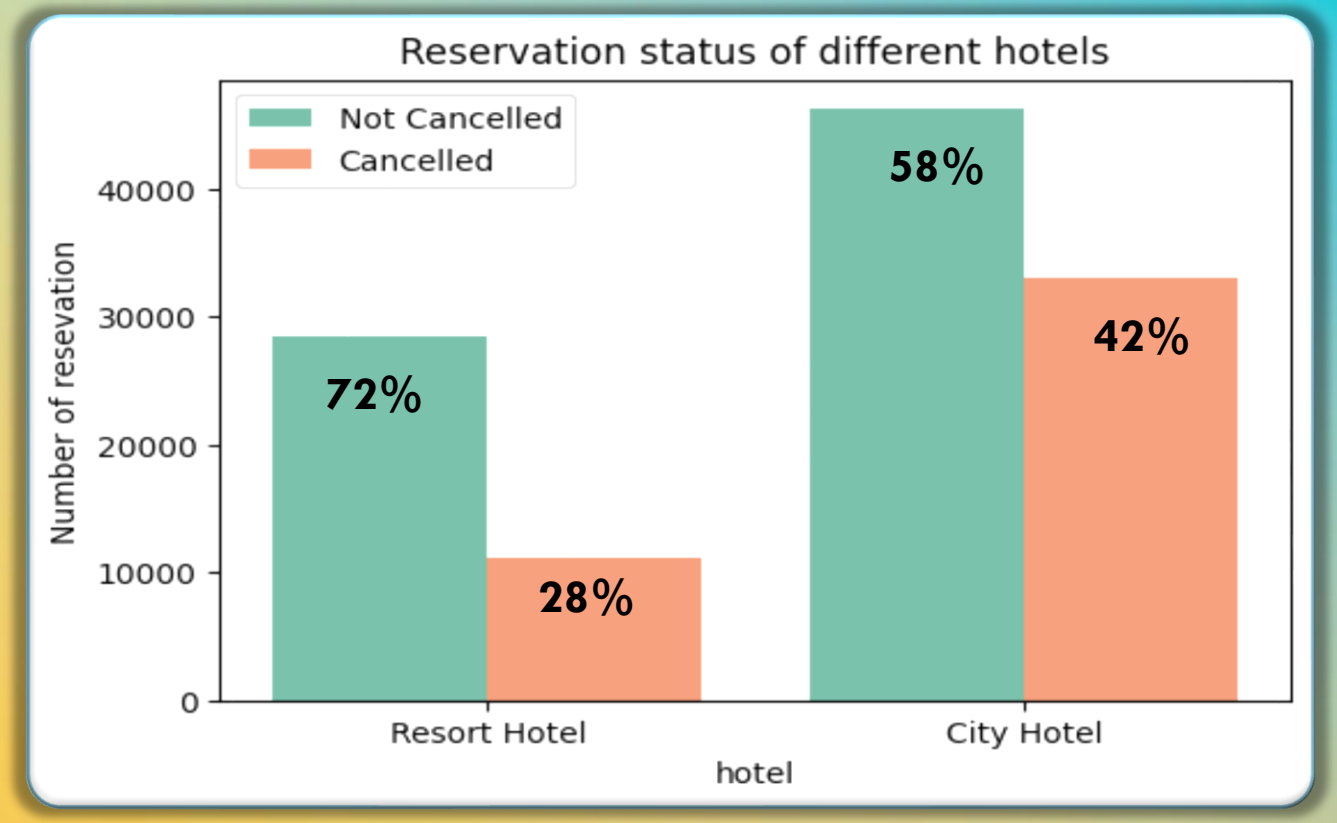
- Our primary goal was to examine the cancellation rates at both hotels. We sought to understand why many visitors were choosing to cancel their reservations.
- We also identified the top 10 countries where this challenge is most prominent. As data analysts, we need to address this issue effectively.
- To achieve this, we will investigate various potential factors contributing to cancellations, such as pricing, accommodation facilities, discounts, hygiene standards, and other relevant aspects.
- Through this analysis, we aim to uncover the key drivers of cancellations and will find out what's causing most of the cancellations. Then, we'll come up with plans to fix this issue.

What percent of bookings are getting canceled?



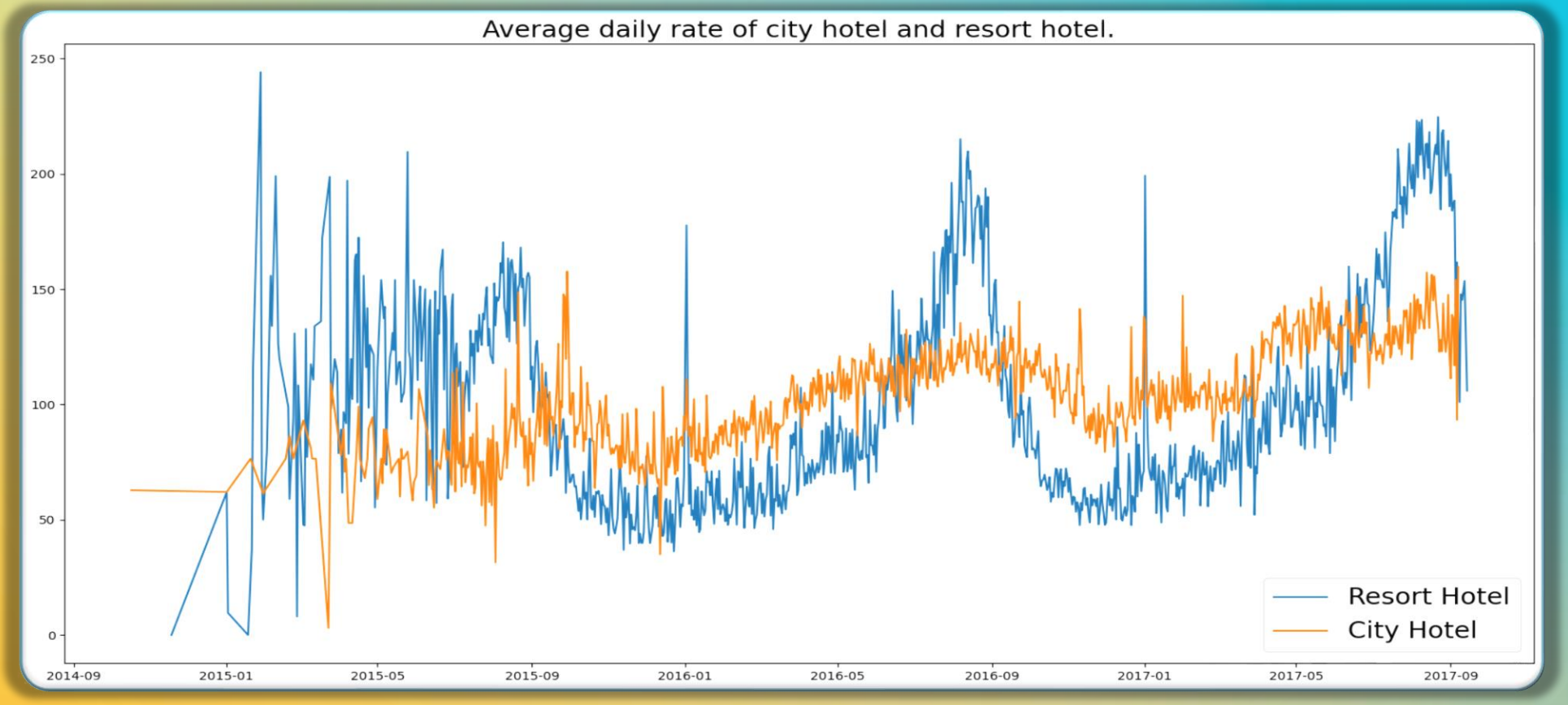
Almost **37%** of the bookings are being cancelled.
Let's now find out the reasons behind this.

Which type of hotels are facing this type of issues?



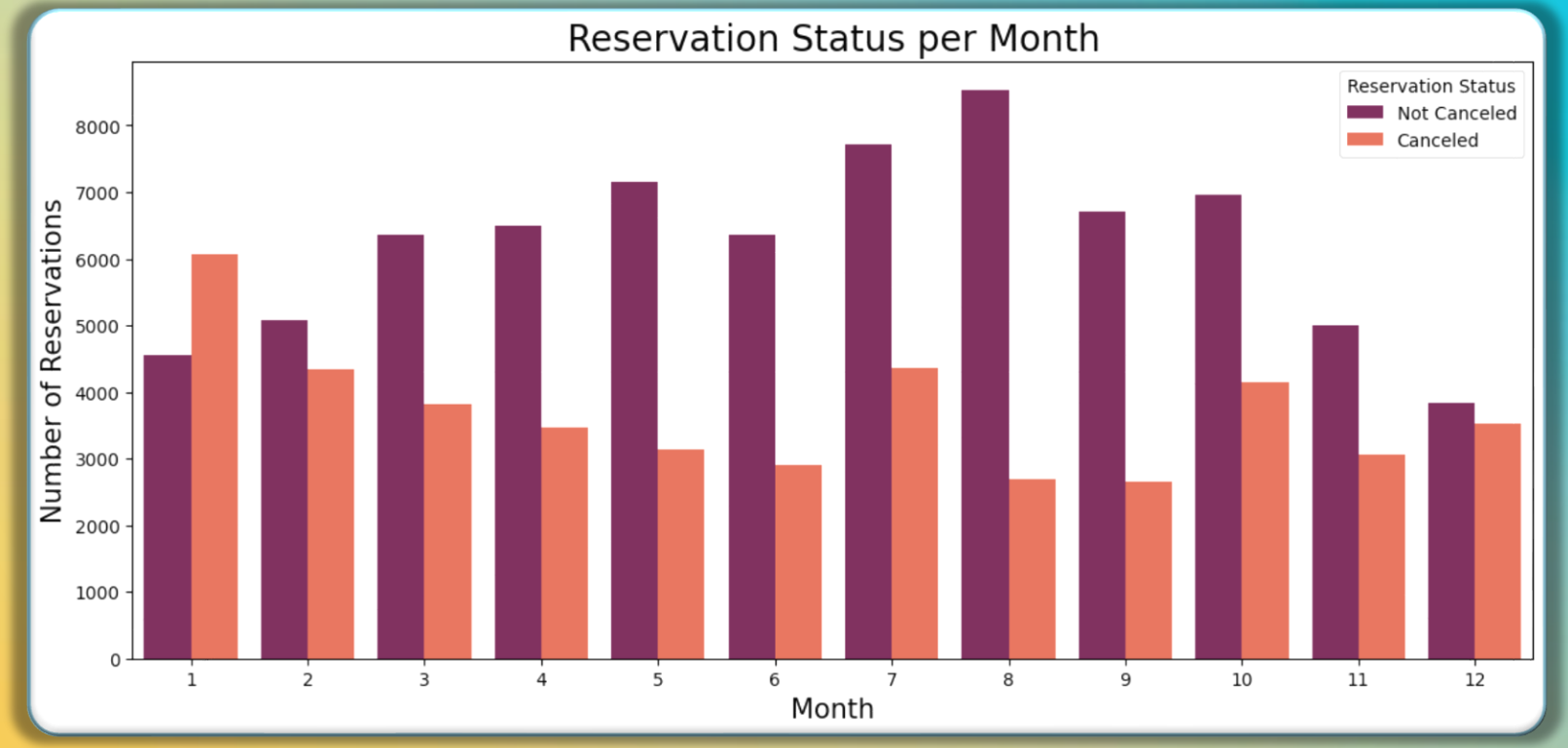
According to the below chart we can see most cancellations are from **City Hotels.** Around **42 %** more than half. It is a major issue.

Now check the *adr* of both hotels.



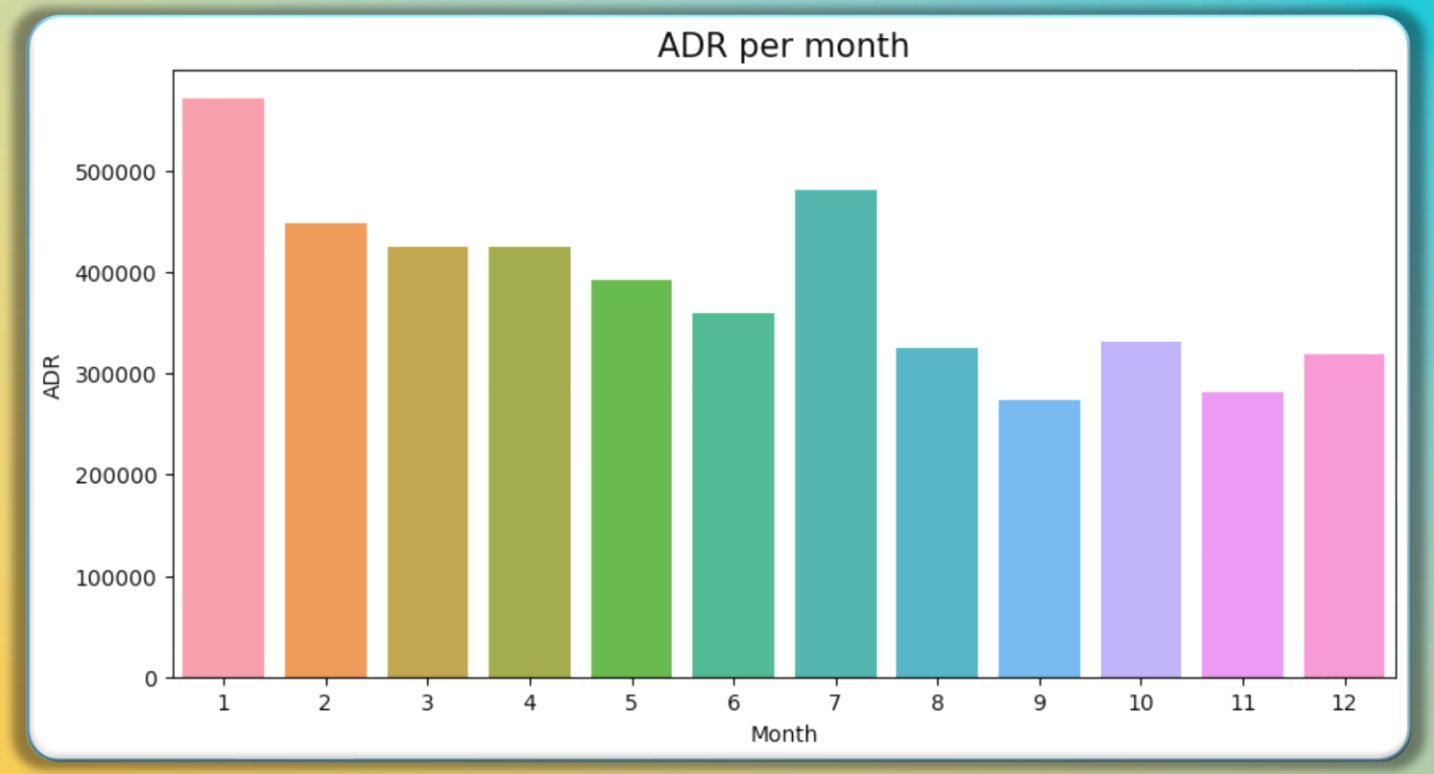
- We can see the average price of the City Hotel is less as compared to Resort Hotel throughout the years.
- Maybe due to low prices, people book and then cancel when they don't find up to the mark.
- The Resort Hotel may be on the weekends and Holidays to raise their price.

In which month most numbers of cancellations are being done?



- It is a ratio of confirmed bookings and the number of cancellations in any particular month.
- Here it's clearly showing **January** is the highest rate of cancellations.
- And **August(8)** and **September(9)** are the month with the highest number of confirmed bookings.

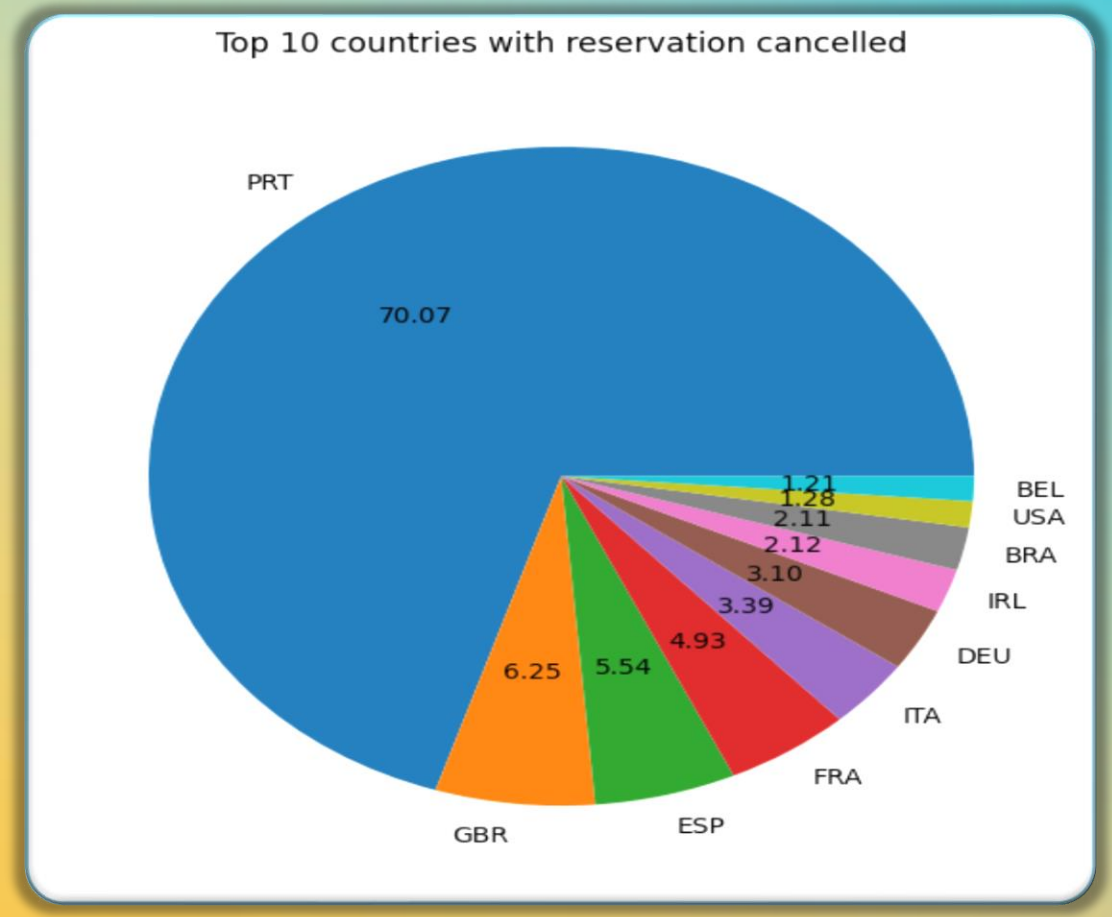
Now check the pricing for each month.



Now look here, on January *adr*(Average daily rate) is very high.
And in August and September, the *adr* is very low.
So, one thing we can clearly say:

- ✓ High Price leads to high cancellations chances.

Top 10 countries with the most number of cancellations.



Portugal is number one, 70 % of cancellations are from this country

We were also asked, From which segment most numbers of guests are visiting the hotel.

```
In [30]: (df['market_segment'].value_counts(normalize = True)*100).round(2)

Out[30]: Online TA        47.44
         Offline TA/TO    20.32
         Groups           16.66
         Direct           10.47
         Corporate         4.30
         Complementary     0.62
         Aviation          0.20
         Name: market_segment, dtype: float64

In [31]: (cancelled_data['market_segment'].value_counts(normalize = True)*100).round(2)

Out[31]: Online TA        46.97
         Groups           27.40
         Offline TA/TO    18.75
         Direct            4.35
         Corporate         2.22
         Complementary     0.20
         Aviation          0.12
         Name: market_segment, dtype: float64
```

From the above analysis, we came to know.

Online Travel Agency - 47 %

Offline Travel Agency - 20 %

and most numbers of cancellations are from.

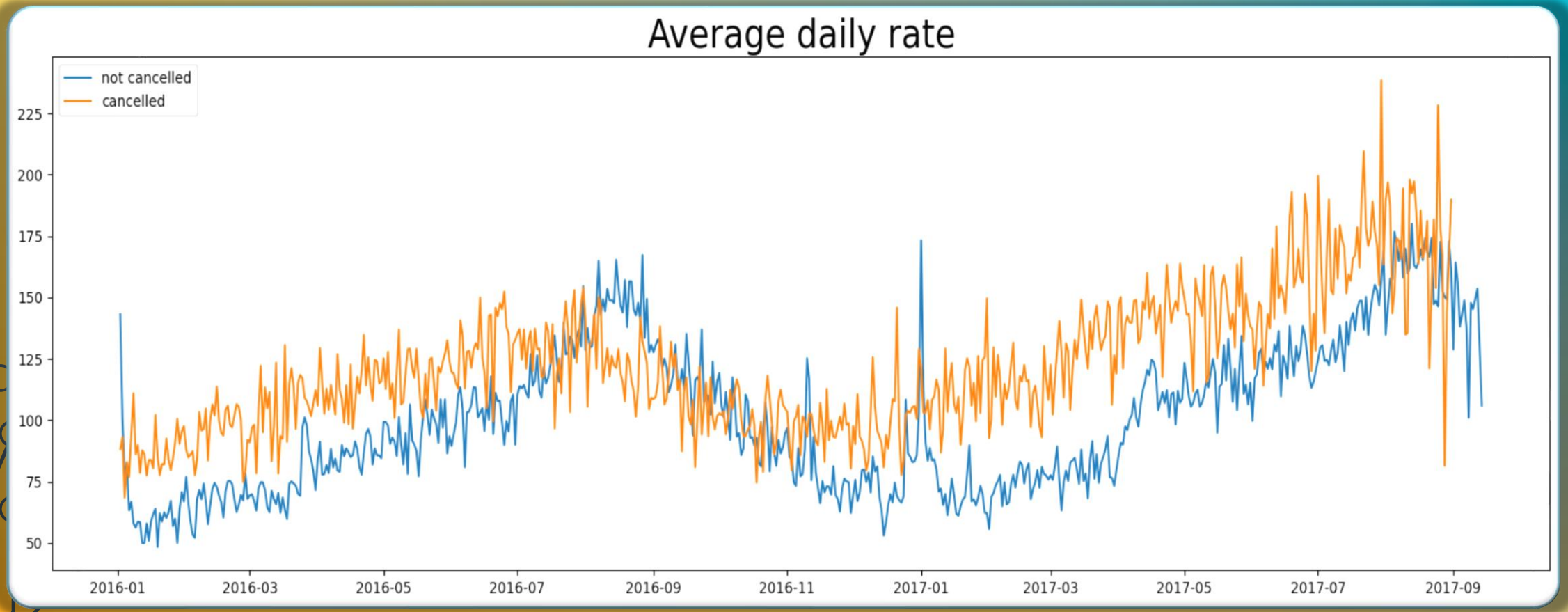
Online Travel Agency - 47%

Groups - 27 %

This graph displays cancellations based on the average daily rate.

Whenever the price is high, we can see more cancellations happening.

- ✓ It is now clear that most of the booking cancellations happen due to high prices.
- ✓ So, we conclude that a high price leads to high cancellations.
- ✓ And a low price leads to a low cancellation rate.



Final Conclusions and Suggestions

- ✓ In order to prevent cancellations hotels should work on their pricing strategy. As much as possible keep the pricing low.
- ✓ January experiences a higher number of cancellations. Providing special offers and discounts during this month could help generate more revenue.
- ✓ It's essential to enhance the overall quality and cleanliness of the hotel. Online bookings and online cancellations are high.
- ✓ Maintaining the hotel to match the expectations as promised on the websites.
- ✓ Most probably guests booked and when they visit the hotel they don't find it as shown on the websites. So prevent dissatisfaction.
- ✓ Graph spikes indicate that hotel prices are significantly increased on weekends.
- ✓ Consider avoiding drastic price hikes during weekends to encourage more bookings and minimize cancellations.
- ✓ Focus on Portugal Country because this country has a 70 % cancellation rate.

Last but not least.....

I would like to thank my mentor **Ayushi Ma'am**.
I followed her guidelines on YouTube to make this project successful.

I would also like to express my sincere gratitude to all the mentors I've had the privilege of learning from on YouTube. These remarkable individuals have been instrumental in expanding my knowledge and skills. While I may not have had the chance to know them personally, their influence on my career journey has been truly significant.

I extend my heartfelt appreciation to **Dhaval Patel**, **Hemanand Vadivel**, **Rishabh Mishra**, **Shakra Shamim**, **Krish Naik**, and **Shashank Mishra** for their invaluable guidance and unwavering support. I watch their video on YouTube.