

EDA Capstone Project

Hotel Booking Analysis

By – Ranajay Biswas

Points of Discussion :

In this exploratory data analysis project, we going to focus on these following topics -

- Loading the data in our work environment.
- Summarizing dataset.
- Cleaning & Imputation.
- Data Wrangling.
- Visualization.
- Discover key factors and relationships among the features and attributes.
- Solutions to Business objectives.
- Conclusion.

Reading the dataset

We have the data as a csv file format. We will use Python Pandas `read_csv()` function to read the dataset and convert it to tabular format. This tabular data is also called Pandas dataframe.

Now we can perform EDA on our data.



Cleaning and Filling :

This step is crucial for removing faulty data and filling in missing values.

- Performing a duplicate check, we found almost 32k duplicate observations.
- We found missing values for many columns like 'Company' , 'Agent', 'Country', 'Children' etc. We checked the data type and distributions for these columns.
- We imputed the missing values with median and mode values when the number of missing observations were within reasonable range.
- 'Company' column had more than 90% missing values. So, we dropped that feature.

After that we performed other checks to see if we had all the valid data or not.

Visualization :

This is the part where we are plotted different diagrams and charts to visualize our data.

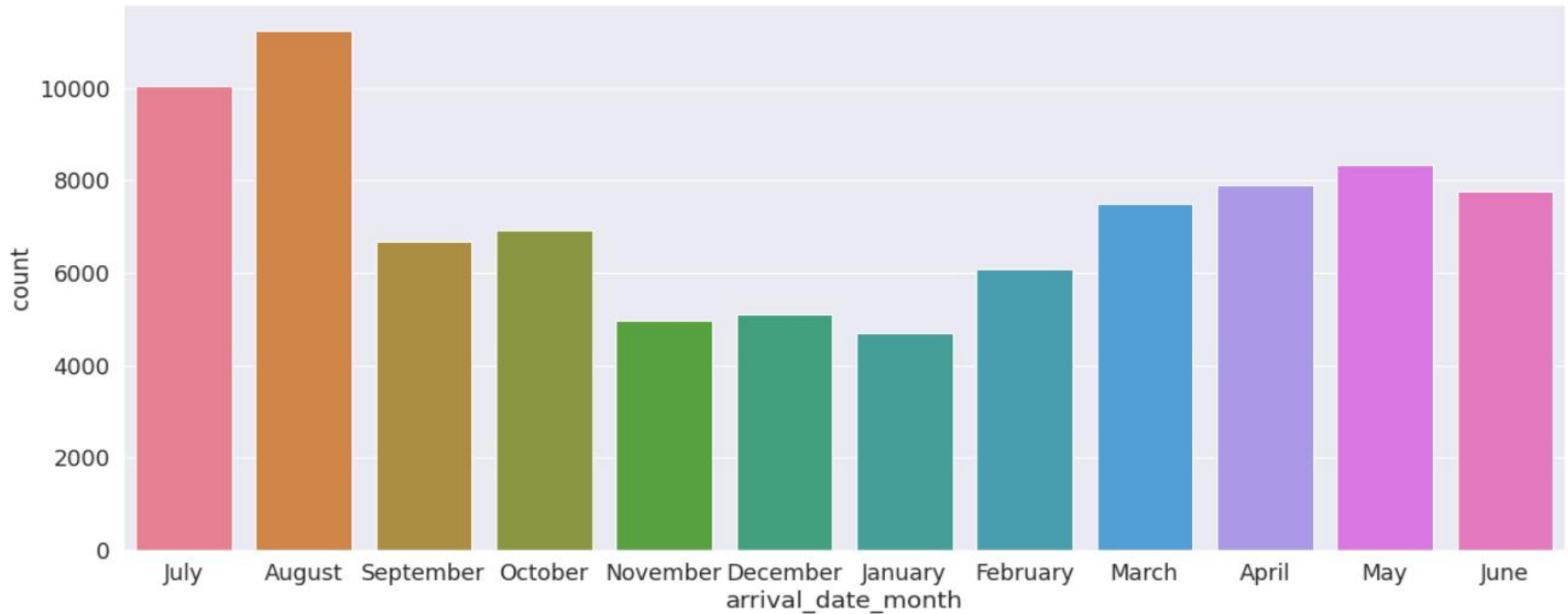
Let's check for some trends and patterns.

For this we used python libraries like –
Matplotlib, Seaborn and Plotly.



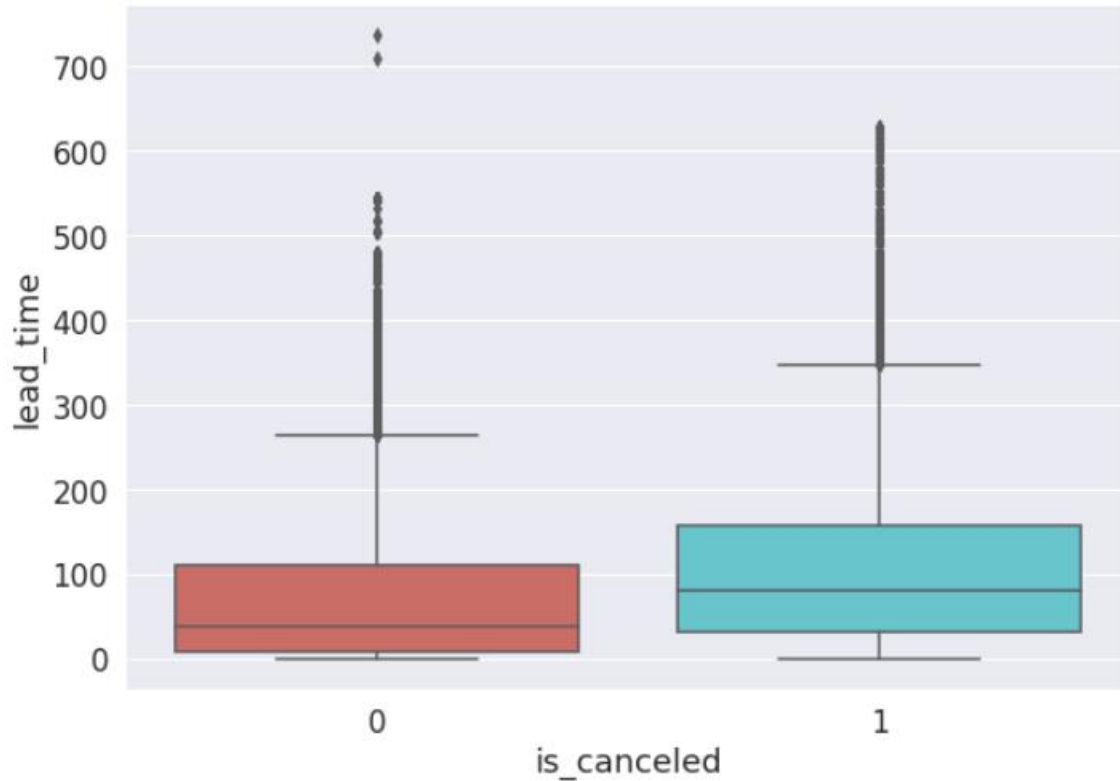


- About 72.5% of the bookings were not canceled, while 27.5% were canceled.
- Most of the bookings are for City hotels.

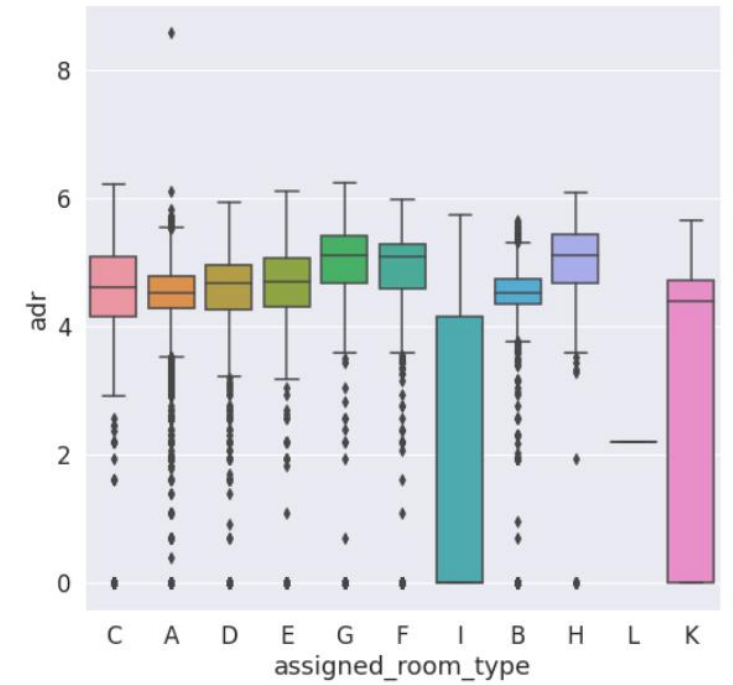
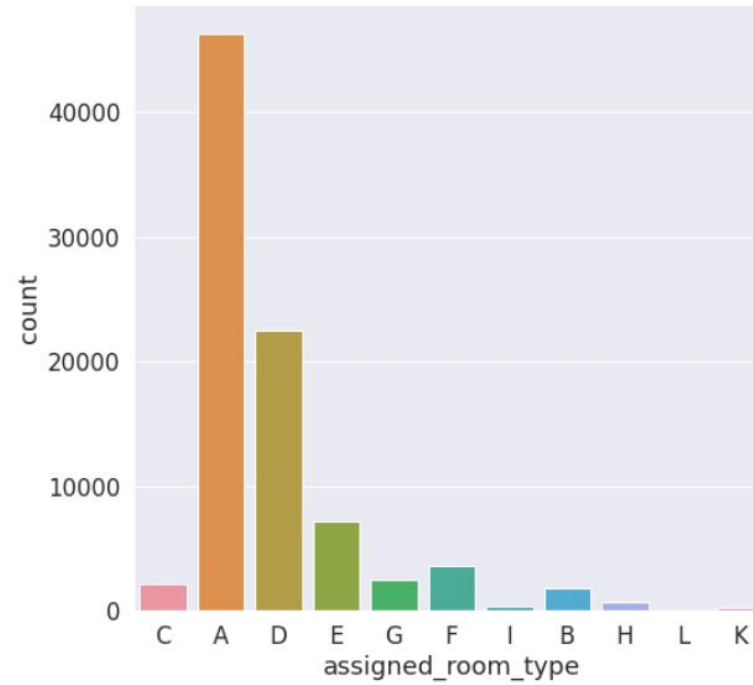
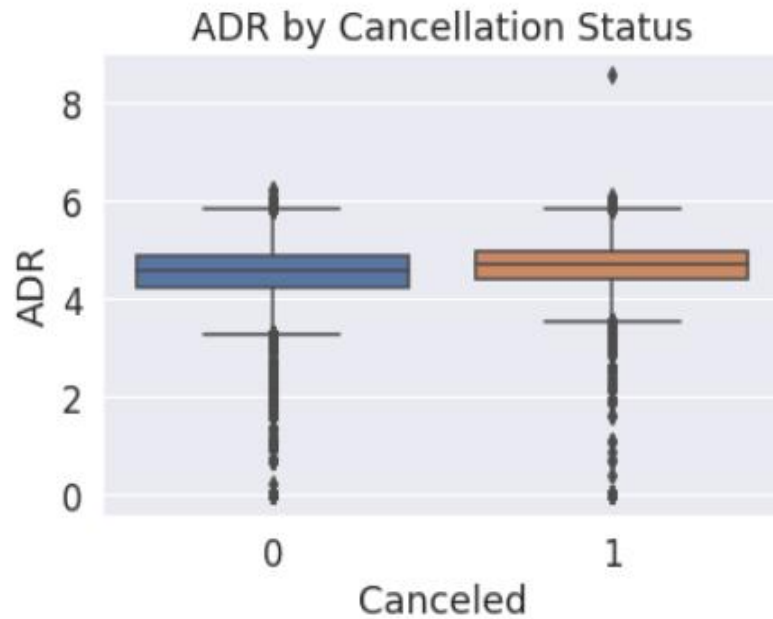


- Most of the bookings are for the summer months (June, July, and August).
- November, December and January have the lowest number of bookings.

- We can see that most of the bookings are for the Online Travel Agents market segment.
- for Aviation, we have the least booking.

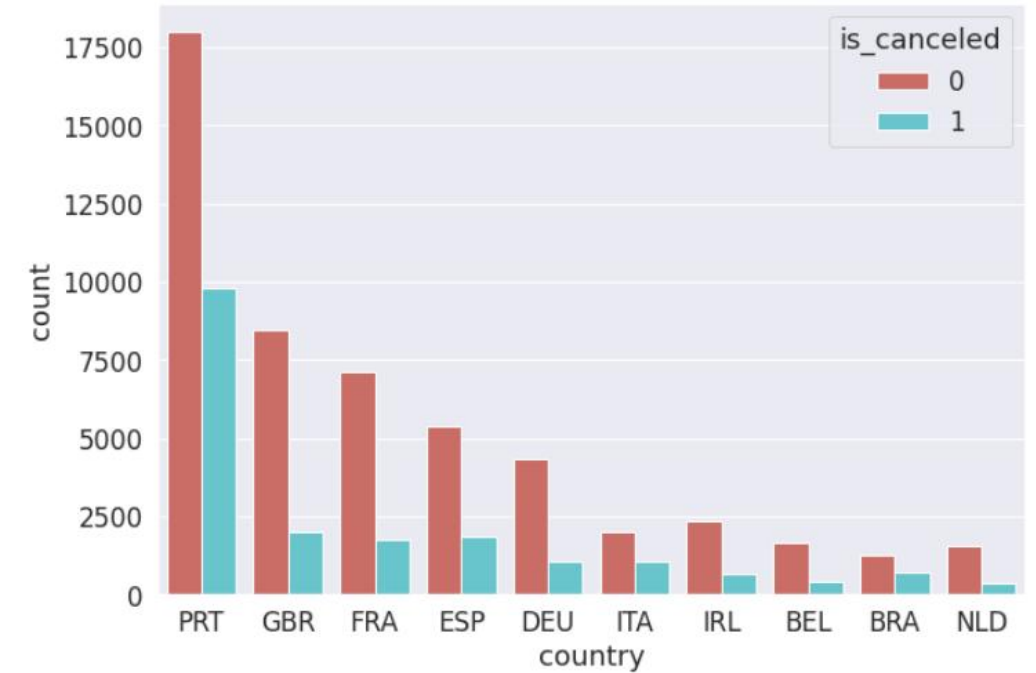
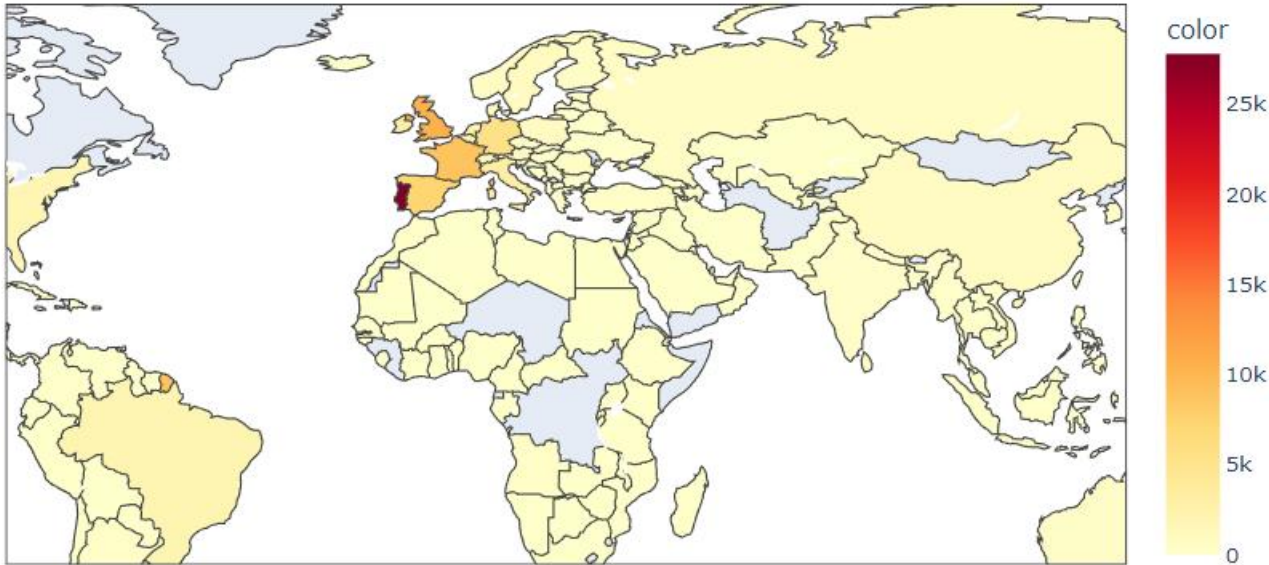


- The lead time is higher for canceled bookings compared to non-canceled bookings.



- No obvious connection is visible between Daily Rate and Cancellation.
- Room type A and D were assigned the more than 60% of the times to guests.
- Except for Room type I and K, all other rooms have similar kind of average daily rate.

Home Country of Guests



- Most of the bookings are from Portugal, followed by the United Kingdom and France.
- We can also see that the cancellation rate varies widely across different countries.

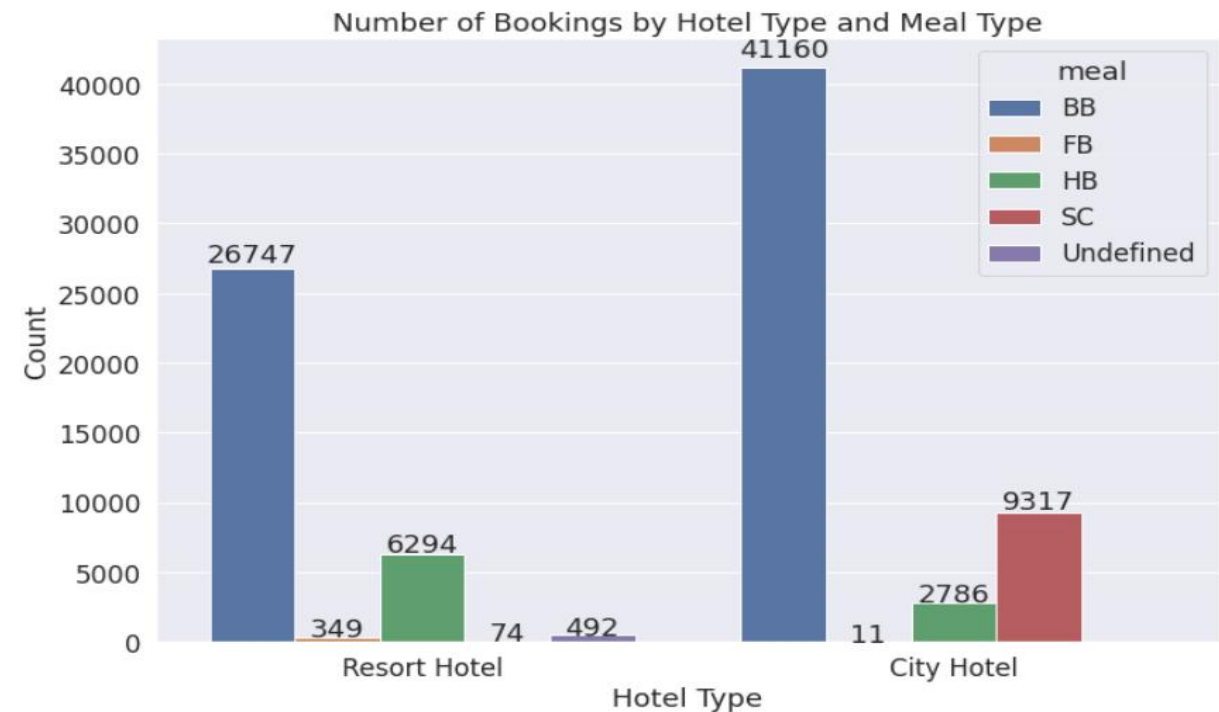
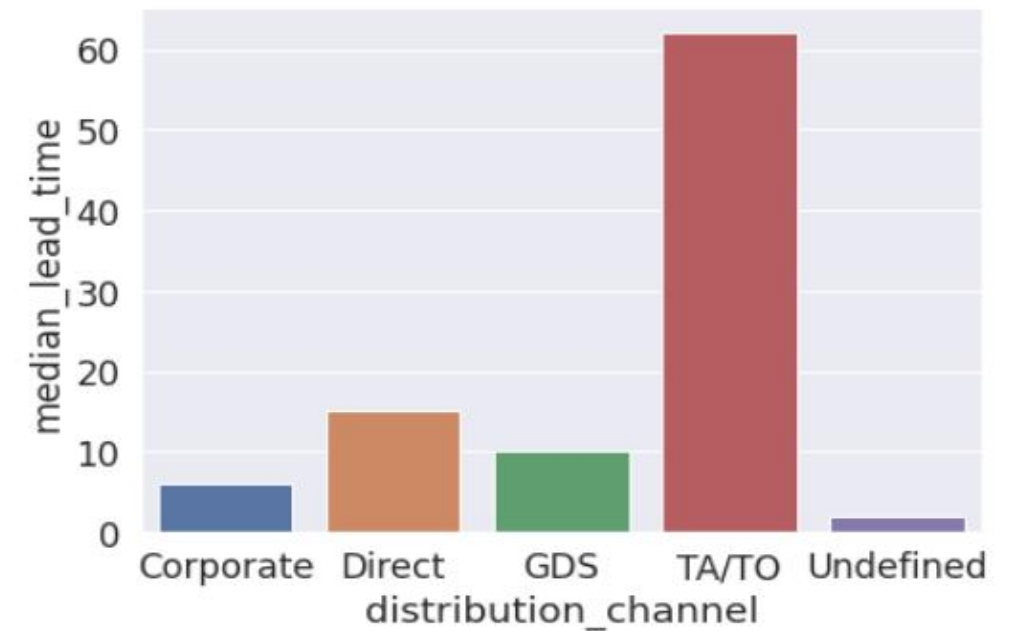
Room Prices per night Over the Months

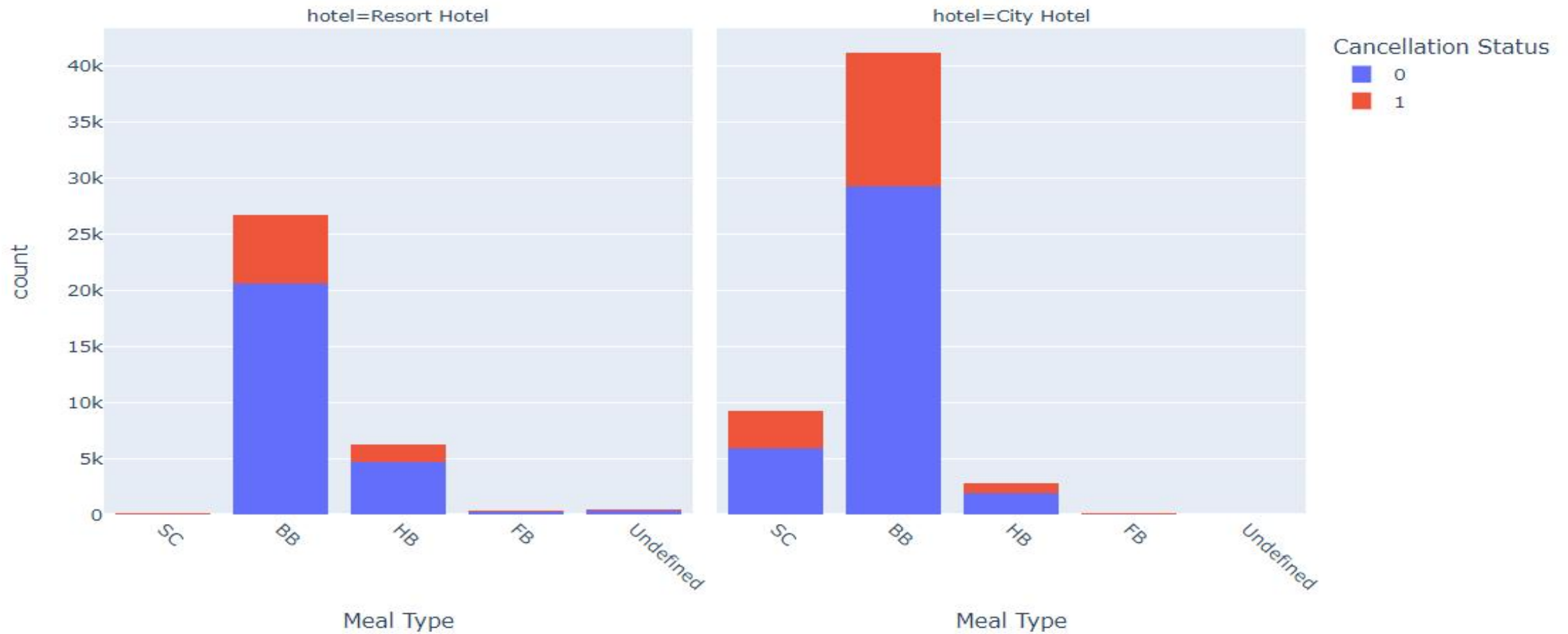


The line chart tells us about the up & down trend in average daily price for City hotels and Resorts.

- Prices of resorts' are highest in the month of August, followed by July.
- City hotels charge the most on May-July.
- From November-January, the resort prices are lowest.

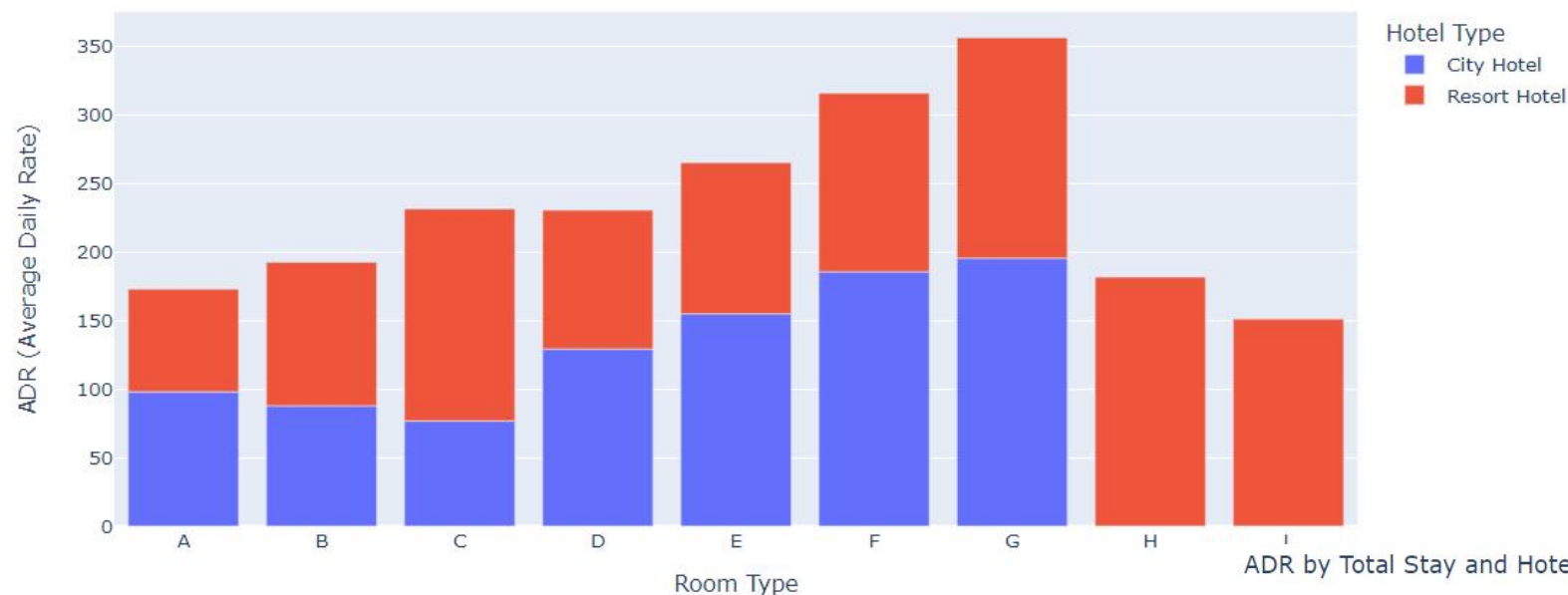
- We have more than 2x new guests compared to the repeating guests.
- The booking cancellation number is also higher for the new guests as well.
- TA/TO is mostly used for planning Hotel visits ahead of time. But for sudden visits other mediums are most preferred.
- We have most number of BB: Bed & Breakfast in the meal type for City hotels and Resorts.
- For Resorts' HB is the 2nd most popular. HB means Half Board. It includes bed, breakfast and evening meal.





- Most bookings and cancelations, both are for hotels with meal type - 'BB'

ADR by Room Type and Hotel Type



This stacked bar chart gives us the idea of how prices vary for different room types.

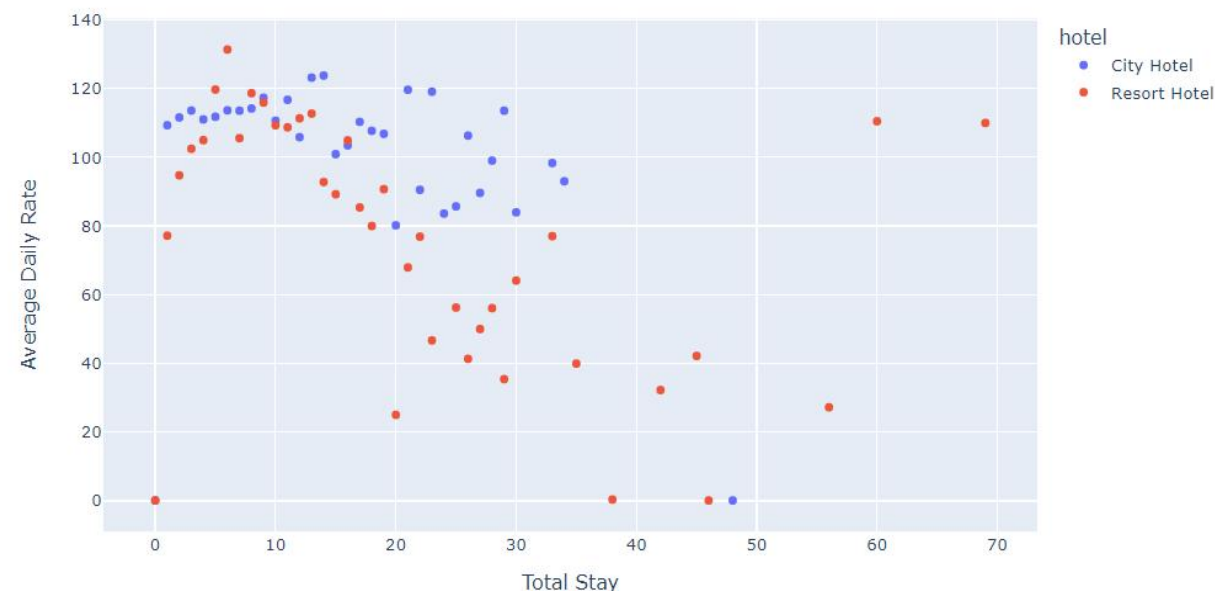
- Resorts charge the most daily for H type rooms.
- City hotels charge the most for G type rooms.

We get the total stay by adding the number of nights people stayed in weekends and weekdays.

From the scatter plot,

- We see most people stayed between 0-30 days for both resorts and city hotels, where the ADR ranges from 80-120 euros.

ADR by Total Stay and Hotel Type



Solutions to Business Objective :

- From December - January, the number of bookings are low. In order to attract more guests, better prices should be offered.
- Half Board and Full Board meal types are underperforming. So, we need to look out for opportunities and ask for feedback from the guests to improve their experiences.
- We can target those guests more who are staying with the hotels for longer period of times.
- We get more cancelations when the lead time is higher. So, we need to find strategies to reduce this metric.
- Returning guests have a lower chance of canceling a booking, so we should reward them with discounts and offers to secure more future bookings in our hotels.
- Stay competitive.
- Ask for feedback often.
- Periodically throw Offers to retain customers.

Conclusion :

- More guests are booking hotels in summer months (June, July, and August). Whereas, November, December and January have the lowest bookings.
- Most people book through Online Travel Agents market segment. for Aviation, there is least booking.
- TA/TO distribution channel is mostly used for planning Hotel visits ahead of time.
- Higher lead_time leads to more cancelations.
- Returning guests tend to cancel less often.
- Room type A and D were assigned the more than 60% of the times to guests.
- There is no obvious connection between booking cancellation and average daily rate.

Thank You...