Capstone Project
Hotel Booking Analysis

By
Sayesh Ankaram

Contents

Dataset introduction

- **Hotel booking Analysis** :
  - This dataset contains booking information for City hotels & Resort hotels. It also includes information such as when the booking was made, length of stay, number of adults, children, babies and the types of rooms, meals and much more.

- **Dataset Use-Case**:
  - This hotel dataset can help to answer so many questions like best time of the year to book a hotel room, optimal length of stay in order to get the best daily rates, predicting whether or not a hotel is likely to receive huge amounts of booking requests etc.

Dataset Synopsis

- **Categorical Variables**:

- *agent* – ID of the travel agency that made the booking.
- *arrival_date_month* – Month of arrival date with 12 categories: January to December.
- *assigned_room_type* – Type of room assigned to the booking: A,B,C,D,E,F,G,L.
- *company* – ID of the company/entity that made the booking or responsible for paying the booking.
- *country* – country of origin.
- *customer_type* – Customer type based on mode of booking: Contract, Transient, Transient-Party, Group.
- *deposit_type* – Deposit paid by customer: No deposit, Refundable, Non-Refundable.
- *distribution_channel* – Booking distribution channel: TA/TO, Corporate, Direct, GDS, Undefined.
- *is_canceled* – cancellation status: if canceled (1) else (0).
- *is_repeated_guest* – If guest is repeated: (1) if yes or (0).
- *market_segment* – Market segment distribution: Direct, Corporate, Online TA, Offline TA/TO, Complementary, Groups, Undefined, Aviation.
- *meal* – Type of meal booked: BB, FB, HB, SC,Undefined.

Dataset Synopsis

- **Categorical Variables**:

- *reserved_room_type* – Type of room reserved for the booking: A,B,C,D,E,F,G,L.
- *arrival_date_month* – Month of arrival date with 12 categories: January to December.
- *assigned_room_type* – Type of room assigned to the booking: A,B,C,D,E,F,G,L.
- *reservation_status* – status of reservation: Check-Out, Canceled, No-Show.

Dataset Synopsis

- **Numerical Variables**:

- *adr* – Average daily rate.
- *adults* – Number of adults.
- *arrival_date_day_of_month* – Day of month of arrival date.
- *arrival_date_week_number* – Week number of the arrival date.
- *arrival_date_year* – Year of arrival date.
- *babies* – Number of babies.
- *booking_changes* – Number of changes made to the booking.
- *children* – Number of children.
- *days_in_waiting_list* – Number of days the booking was in waiting list.
- *lead_time* – Number of days booking got elapsed.
- *previous_bookings_not_canceled* – Numebr of previous bookings not canceled.
- *previous_ cancellations* – Number of previous bookings canceled.
- *required_car_parking_spaces* – Number of car parking spaces required.
- *stays_in_weekend_nights* – Number of stays in weekend nights.

Dataset Synopsis

- **Numerical Variables**:

- *stays_in_week_nights* – Number of stays in week nights.
- *total_of_special_requests* – Number of special request made by the customer.

Fundamental Approach

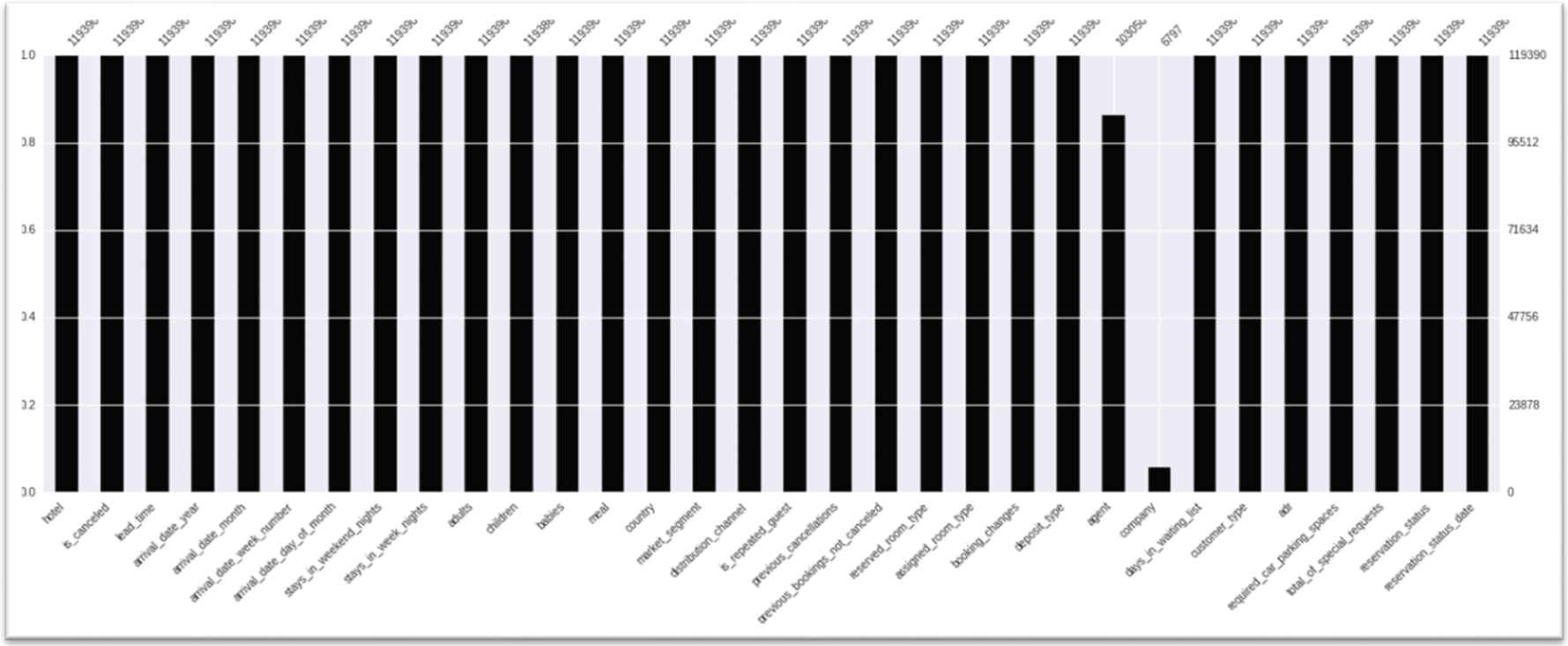
- **Instance Creation**: Creating an instance of the dataset so that the original dataset will not get disturbed.
- **Dataset Description**: Obtaining a described info about the dataset we're working with (by using .shape(), .info() & .describe() methods to gain better insights about the data).
- **Prime Variable Selection**: Selecting a prime variable/feature/column with which large scale insights can be obtained (the prime variable is *'is_canceled'* , for obtaining active data columns with bookings that are not canceled).
- **Dataset Processing**: Cleaning the data by obtaining Null values, Dropping unnecessary rows and columns with Null values, Altering dtypes of columns.
- **Categorical Dataframe Creation**: Creating a Categorical dataframe having all the categorical columns in it, so that we can perform various operations to get a clear understanding about each and every aspect related to the hotel data itself.
- **EDA & Data Visualizations**: Performing exploratory data analysis over the cleaned data in order to get final conclusions and plotting the analyzed data using different graphs.

Overview

- **After creating an Instance of the dataset, Describing the dataset, Selecting Prime Variable, Processing the dataset, Creating a Categorical dataframe out of that dataset and performing various operations on categorical data.. Here's some information that we've gathered:**
  - The original dataset consists of 119390 indexes and 32 columns.
  - Few columns supposed to have index with 'Int64' dtypes but instead they're in 'float64', such columns are: *'children'.*
  - There are a total of 125425 Null values.
  - After cleaning the data, the columns left with max number of Null values are: *'company'* with 94.36 % of rows being Null and *'agent'* with 13.68 % of rows being Null.

Overview

- Null Values from the dataset.



Overview

- The data present in this dataset ranges from year 2015 to 2017 with confirmed bookings: 2015 – (21996 bookings), 2016 – (56707 bookings) & 2017 – (40687 bookings).
- Over these 3 years:
  - Most booked hotel is: City hotel – (79163 bookings).
  - Most preferred booking with meal type is: Breakfast & Bed – (92236 bookings).
  - Max number of bookings for a country is: Portugal – (48483 bookings).
  - Most preferred mode for booking is: Online Travel Agents – (56408 bookings).
  - Most preferred distribution channel is: Travel Agents/Tour Operators – (97750 bookings).
  - Most reserved & assigned room type is: A – (reserved: 85873, assigned: 74020).

Overview

- Most bookings with deposit type are: No deposit – (104461 bookings).
- Customer type with most bookings is: Transient – (89476 bookings).
- Maximum number of booking status is: Check-Out – (75011 bookings).

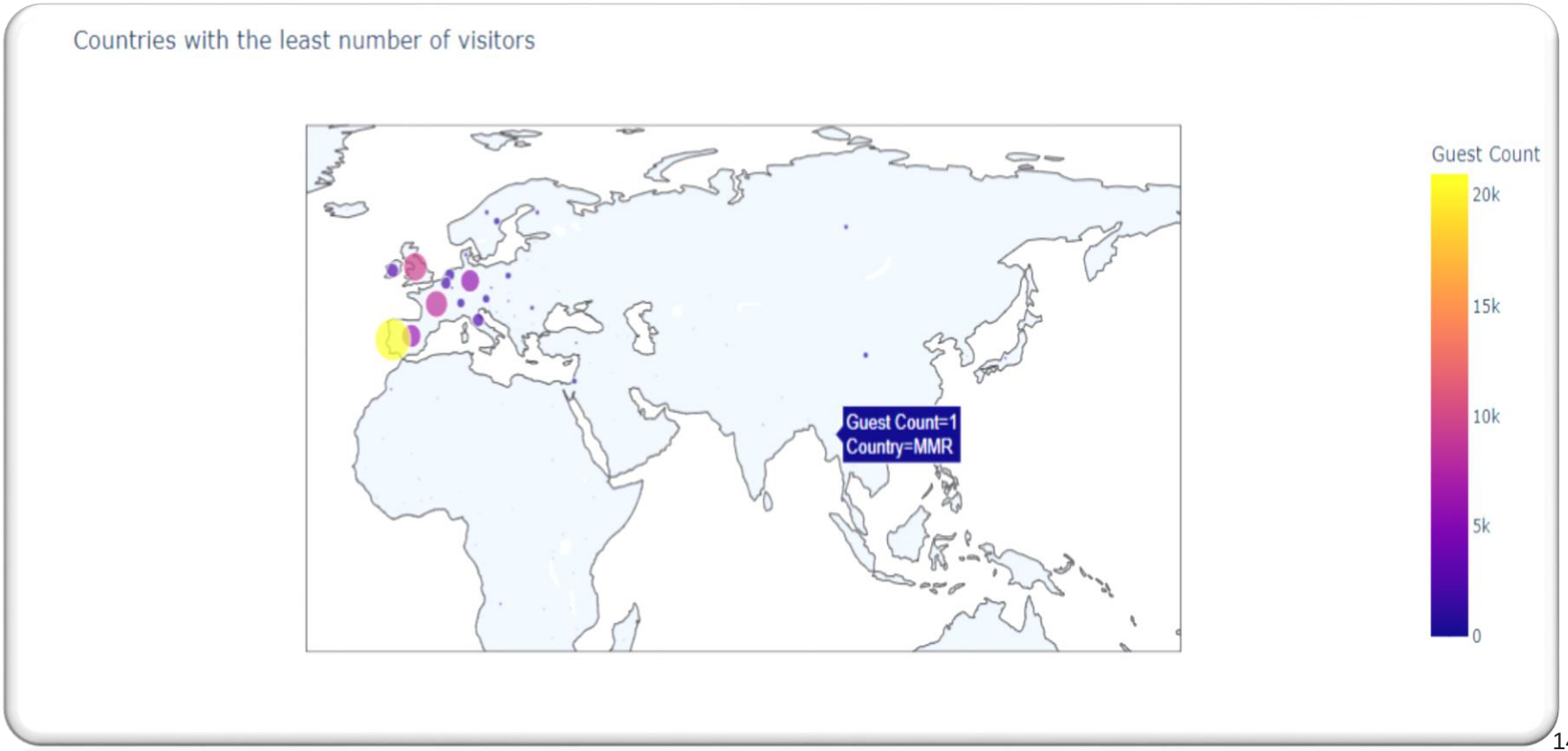
EDA – (Exploratory data Analysis)

- **After obtaining an overview of the dataset, It's time to dive deep inside and explore the dataset. Below are the 10 crucial questions that carry deeper insights about the data, We will explore them all by Visualizing them onto a graph.**

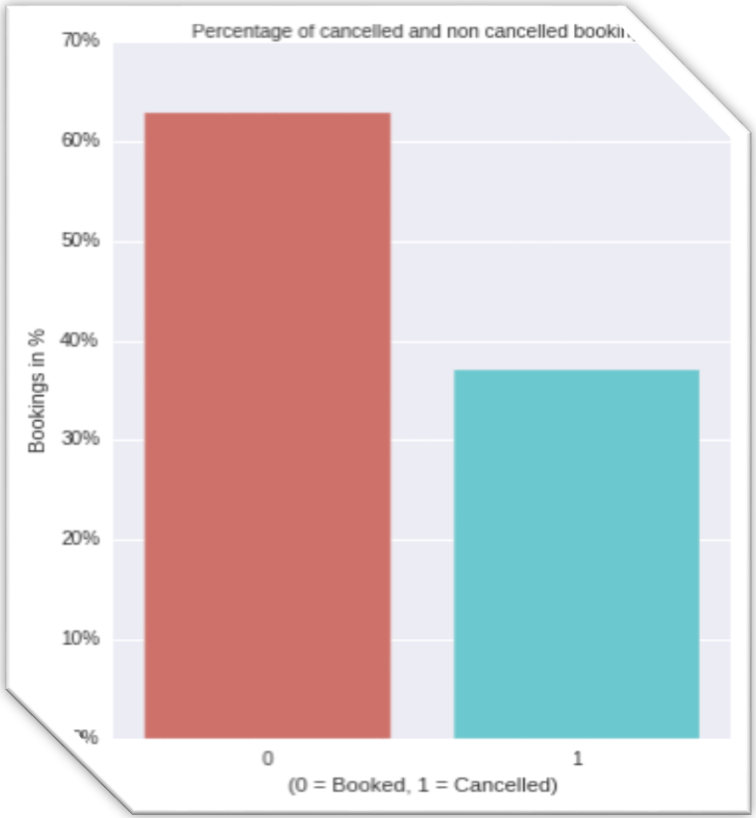
1. Which country has the most number of visitors?
2. Which country has the least number of visitors?
3. What is the percentage of bookings that were both confirmed and canceled?
4. Which mode of booking was most frequently used?
5. What is the hotel price per stay over a year?
6. What are the hotel rates paid by customers per night based on different room types?
7. How long do people stay in the hotel rooms?
8. Which meal type is most preferred and booked by customers?
9. Which is the most booked type of Accommodation?
10. What is the Reservation Status of hotels?

Data Visualization

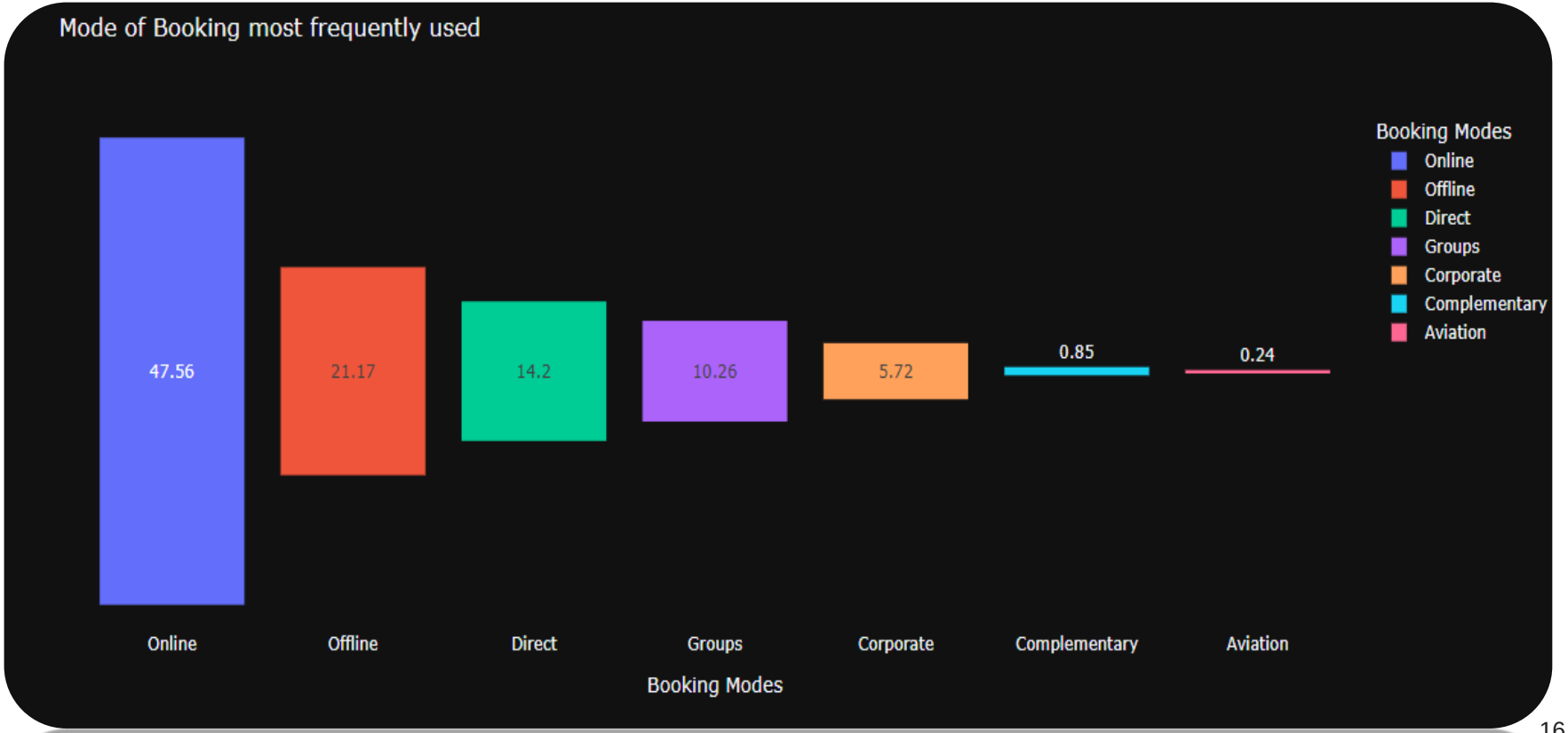
Data Visualization



Data Visualization

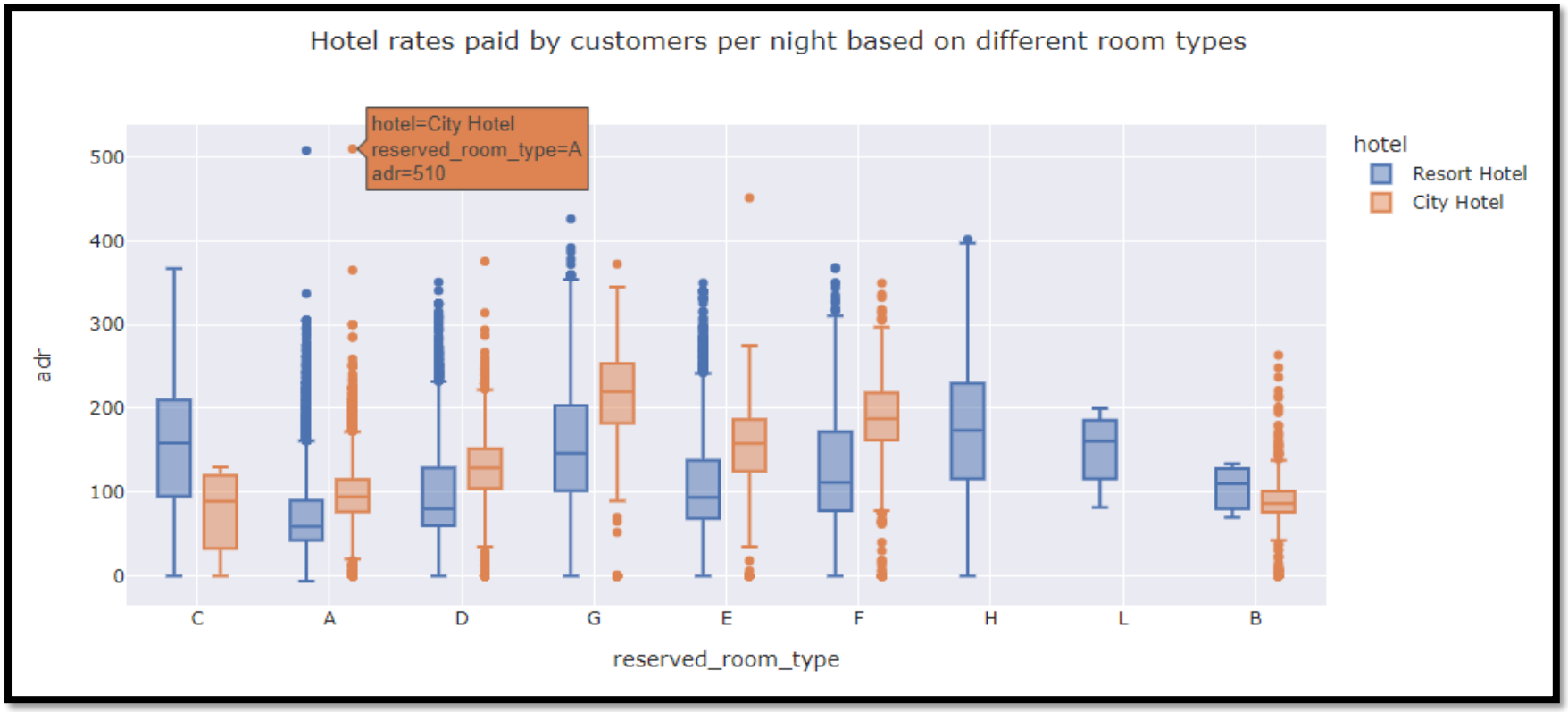


EDA – (Exploratory data Analysis)

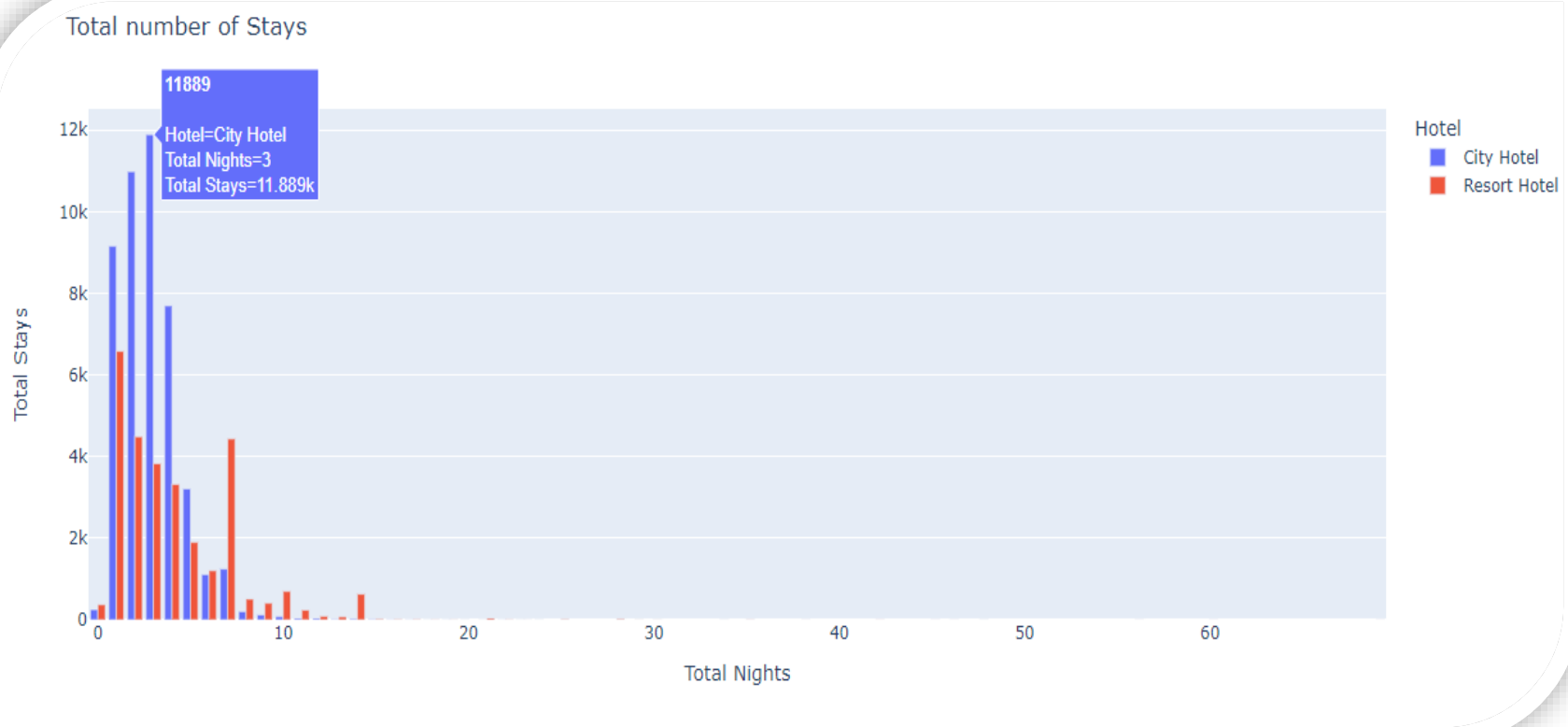


Data Visualization

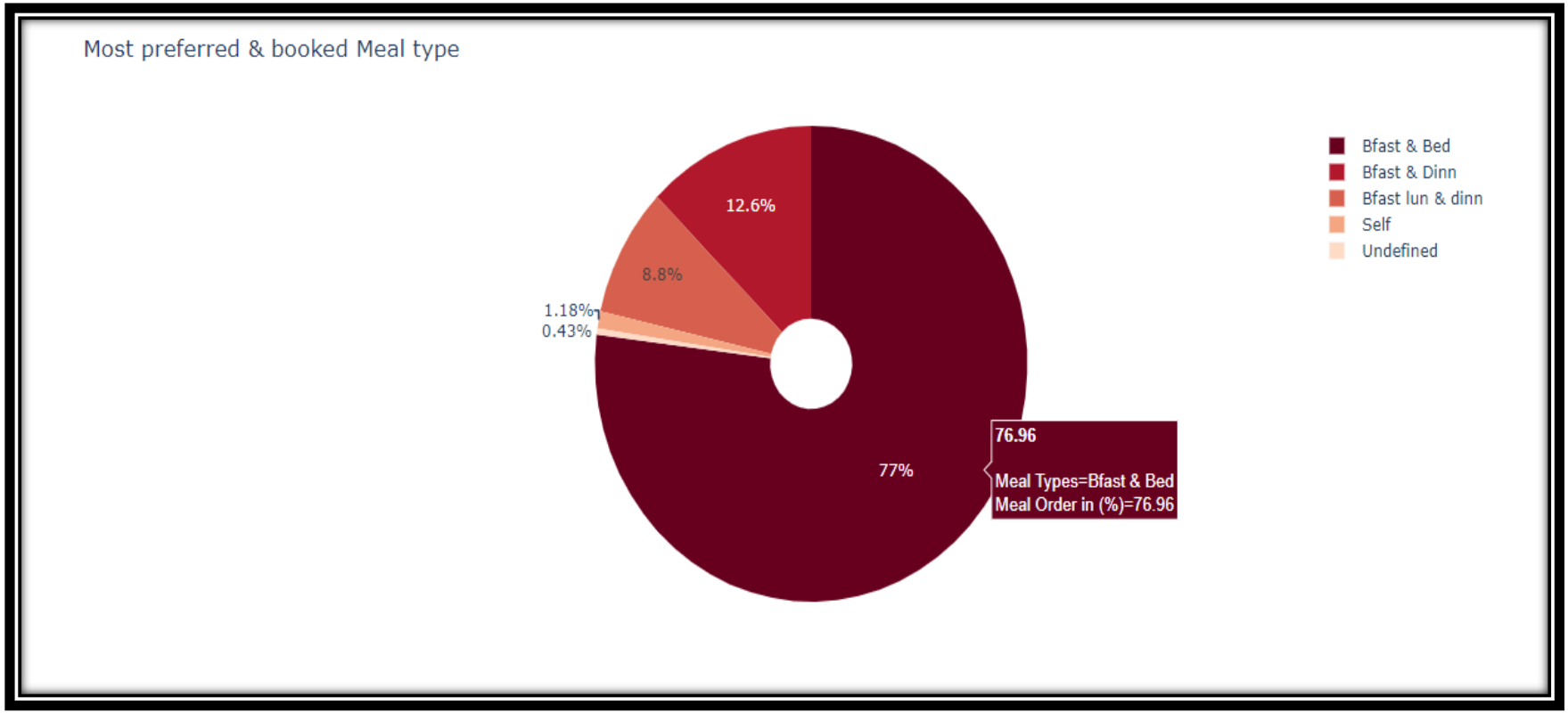
Data Visualization



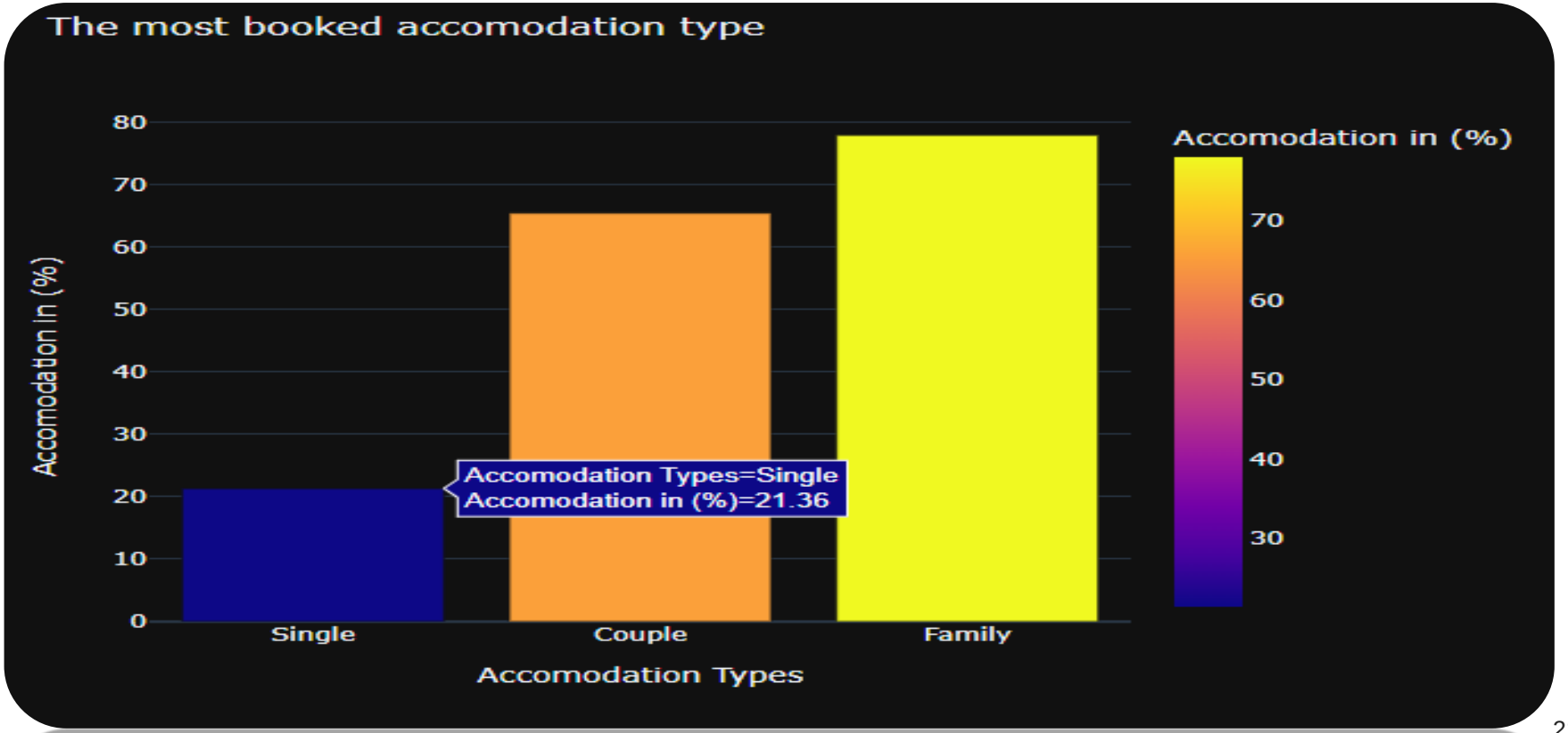
Data Visualization



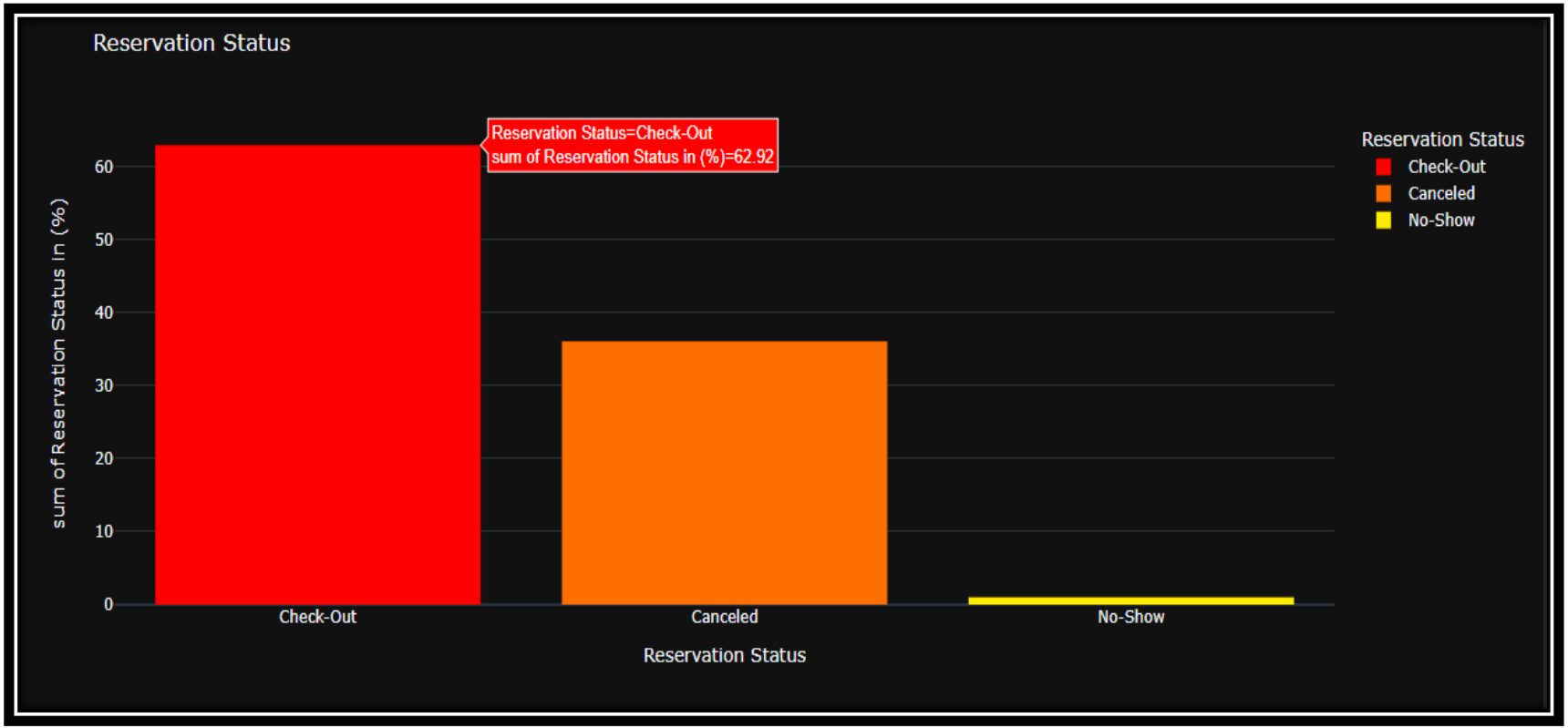
Data Visualization



Data Visualization



Data Visualization



Conclusion

- **Based on the data analysis done so far, we can conclude that:**
  - There are about 63% of confirmed bookings & 37% of canceled bookings.
  - For 1-3 days, majority of people prefer City Hotels and for 4-7 days, majority of people prefer Resort Hotels.
  - The Highest Rates for a hotel room in both City & Resort hotels is for Room Type A.
  - Hotel rates for Resort hotels are at peak in Summer season (May-August) but in Autumn season (September-November) the rates become cheaper.
  - Hotel rates for City hotels have a decent growth in Spring season (March-May) but in Autumn season (September-November) the hotel rates slightly fell down.
  - 47.56 % of customers prefer Online mode for hotel booking & 21.17 % of customers prefer Offline mode.

Conclusion

- 76.96 % of customers prefer Breakfast & Bed & 12.6 % of customers prefer Breakfast & Dinner as the Meal Type while Booking hotels.
- 77.85 % of customers prefer Family Type of Accommodation & 65.51 % of customers prefer Couple Type of Accommodation.
- 62.92 % of customers Checked-Out, 36.06 % of customers Canceled the Booking & 1.10 % of customers didn’t show up.
- Countries with the most number of Guests are Portugal (Guest Count: 20977) followed by Great Britain (Guest Count: 9668) and France (Guest Count: 8468).
- 87.6 % of customers prefer No – deposit while booking hotel rooms.
- 66.4 % of customers prefer City hotels & 33.5 % of customers prefer Resort hotels.

Thank you