

Capstone Project-1

Hotel Booking Analysis

Team Members

Amir Khan

Suhail Shaikh

Bachina Murali Krishna

Het Kothari

Aleem Khan

Contents

- **Defining problem statement**
- **Dataset observation**
- **Data preparation & cleaning**
- **Data visualization**

Problem Statement

- Hotel industry is a very volatile industry and the bookings depend on variety of factors such as type of hotels, seasonality, days of week and many more which makes analyzing the patterns available in the past data more important to help the hotels plan better for the future bookings.
- Using the historical data, hotels can perform various campaigns to boost the business. We can use data visualization to predict various insights to grow the hotel industry.
- We will be using the data available to analyze the factors affecting the hotel bookings. These factors can be used for reporting the trends and predict the future bookings.

Dataset Observation

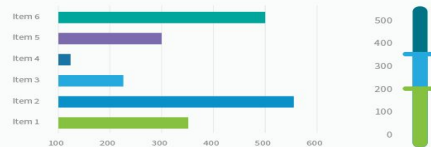
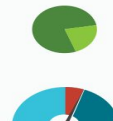
- Our Hotel Booking Analysis Dataset has total 1,19,390 rows and 32 columns where we observed 4 null values in children column, 488 in country column, 16,340 in agent column and 112,593 in company column, resp.
- We have the following columns for which the further analysis is done:

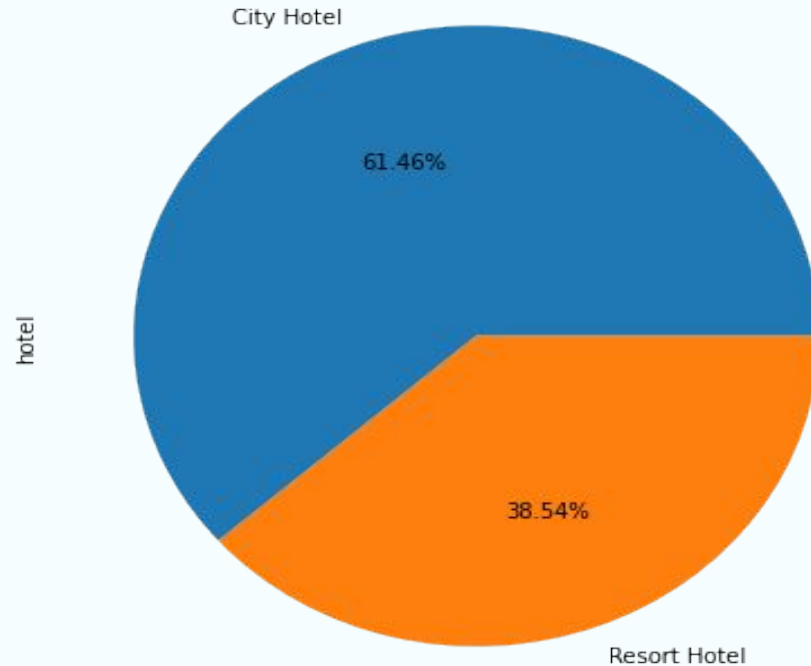
is_canceled, lead_time, arrival_date_year, arrival_date_month, arrival_date_week_number, arrival_date_day_of_month, stays_in_weekend_nights, stays_in_week_nights, adults, children, babies, meal, country, market_segment, distribution_channel, is_repeated_guest, previous_cancellations, previous_bookings_not_canceled, reserved_room_type, assigned_room_type, booking_changes, deposit_type, agent, company, days_in_waiting_list, customer_type, adr, required_car_parking_spaces, total_of_special_requests, reservation_status, reservation_status_date.

Data Preparation & Cleaning

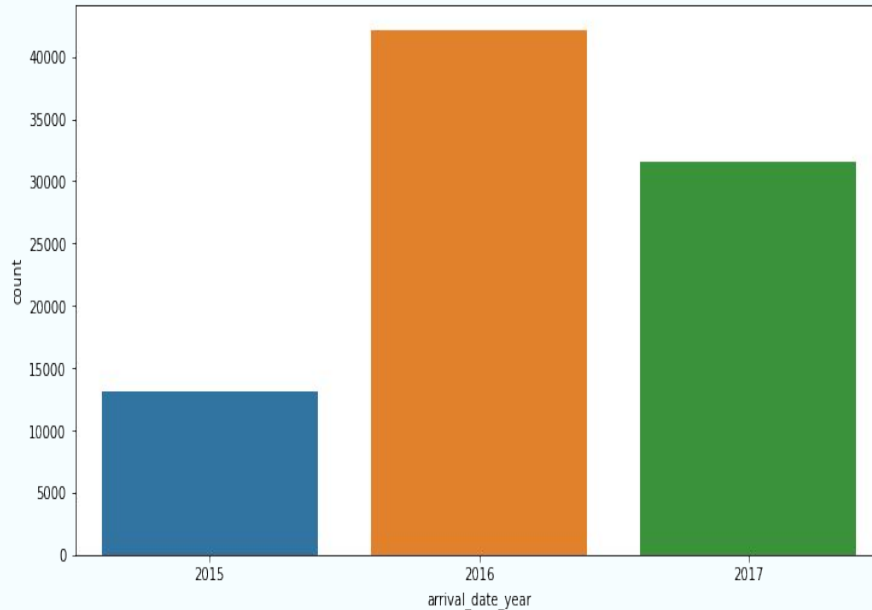
- We import various Python inbuilt libraries - NumPy, Pandas, Matplotlib and Seaborn to work on the dataset.
- Since most of the values in column `agent` and column `company` are null values, dropping these columns won't affect our dataset.
- We drop 31,994 duplicate values from our dataset for clear understanding.
- Next, we deal with outliers using Boxplot where we get a clear picture about the outliers in numeric columns.
- Replace the outliers with some specific values to prepare our dataset for further visualization.

Data Visualization

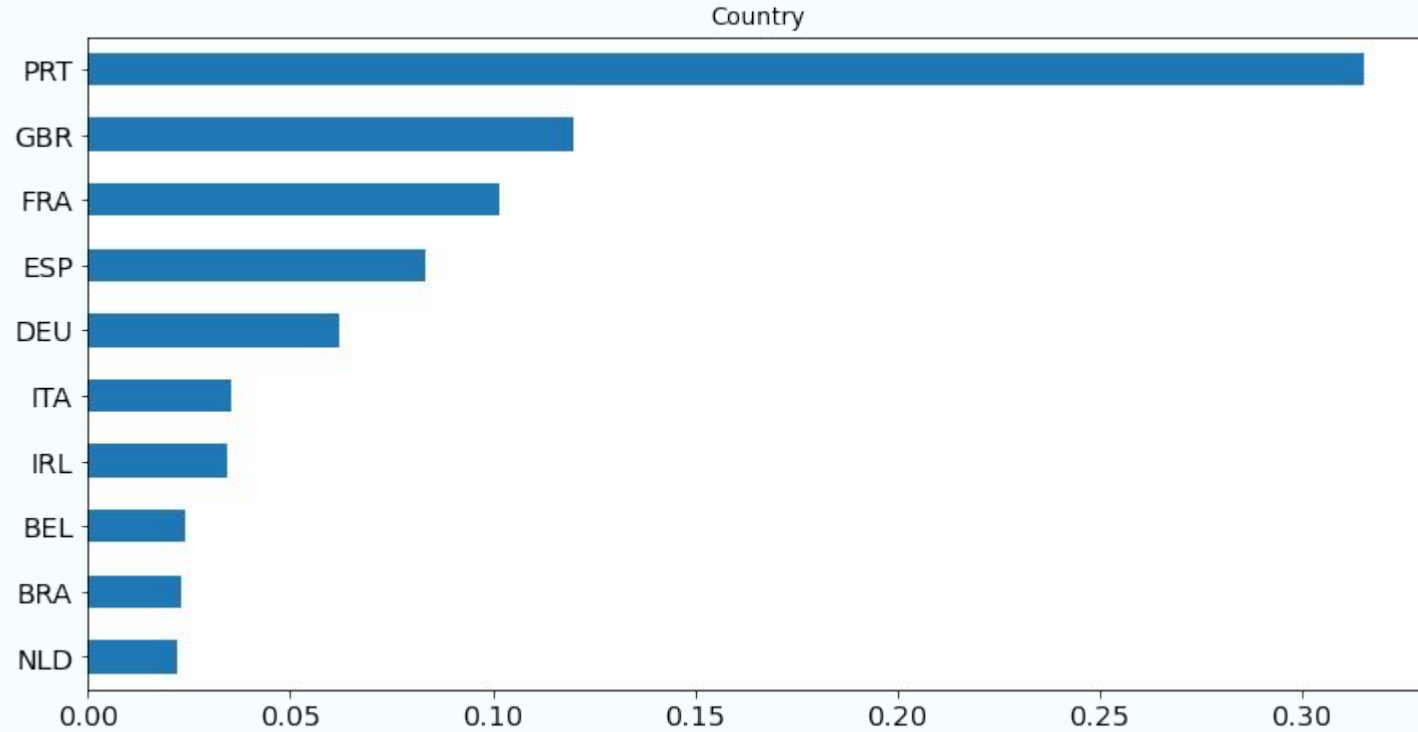




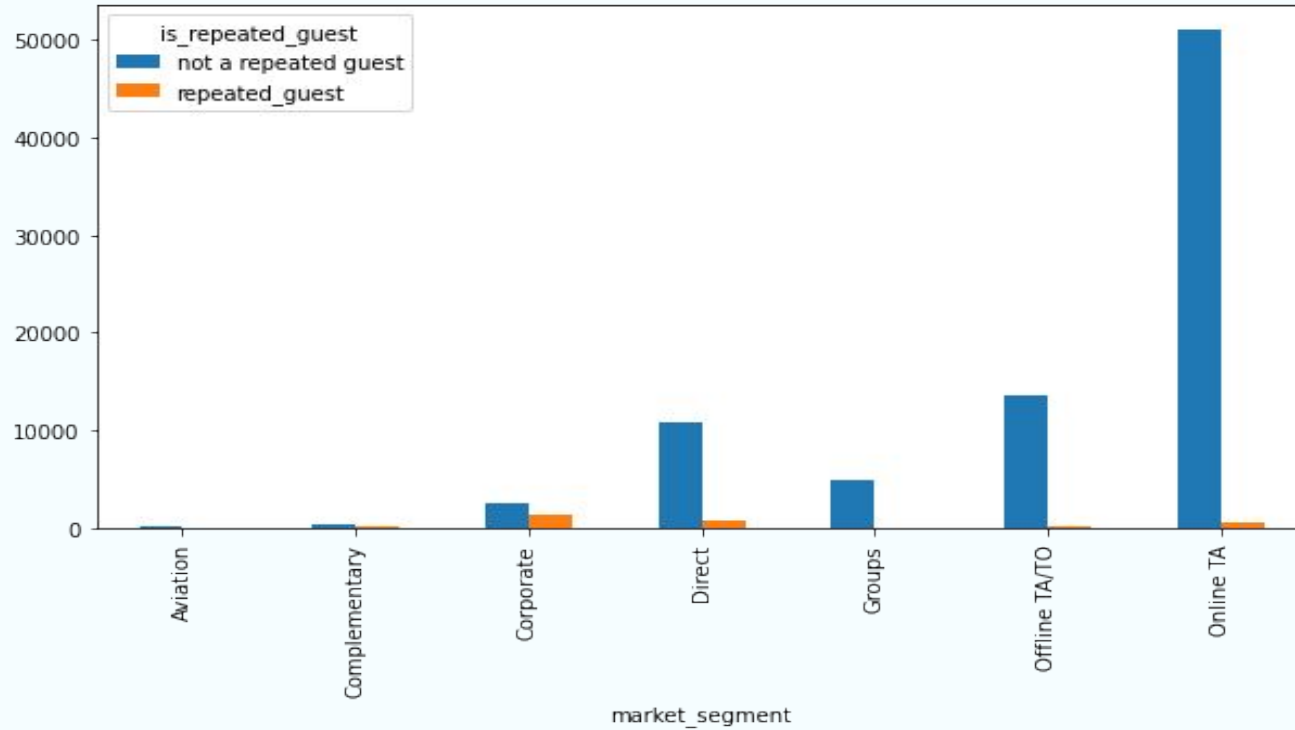
- ❑ 61.1% bookings is done for City Hotel and 38.9% for Resort Hotel. The reason for this difference could be the purpose of visit of the guests and/or the capacity of the hotels.



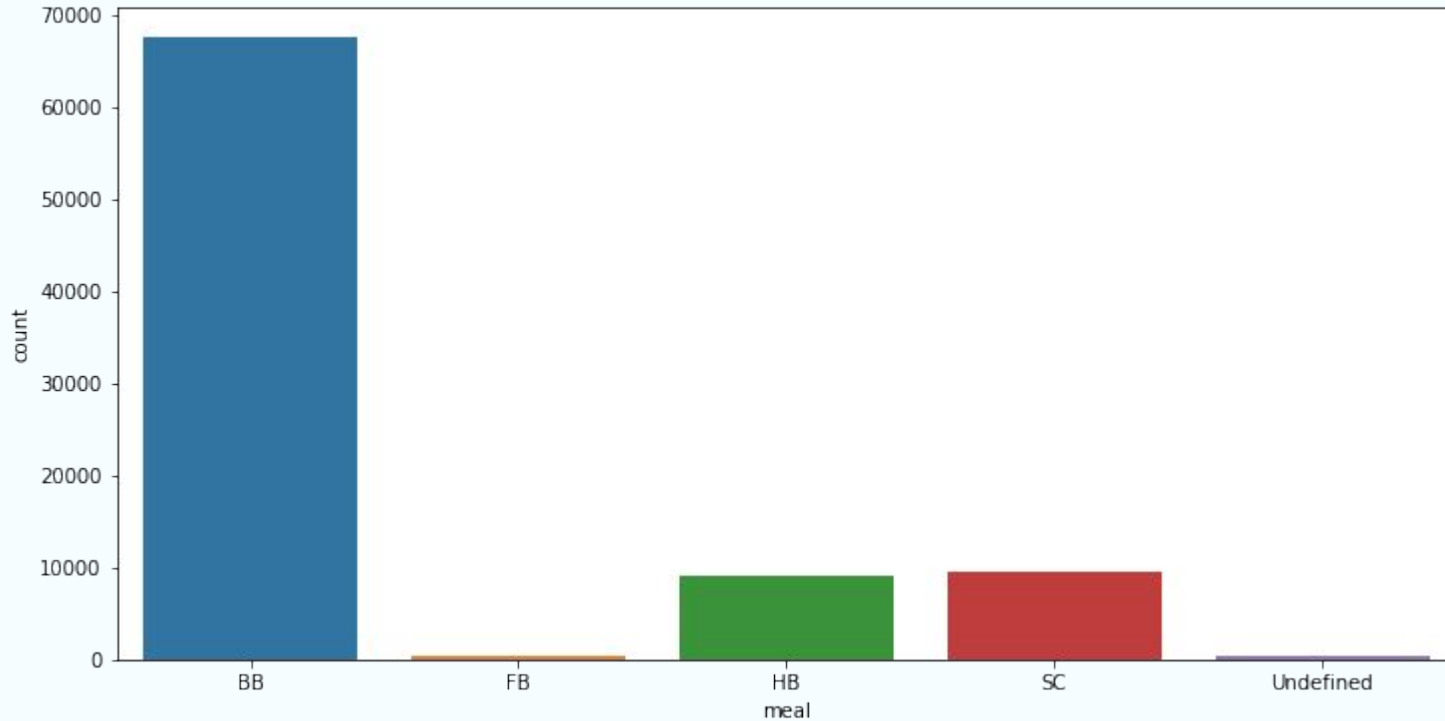
- ❑ We saw the rise and fall of bookings over the year and the reason for that was found to be incompleteness of data collected in 2015 and 2017



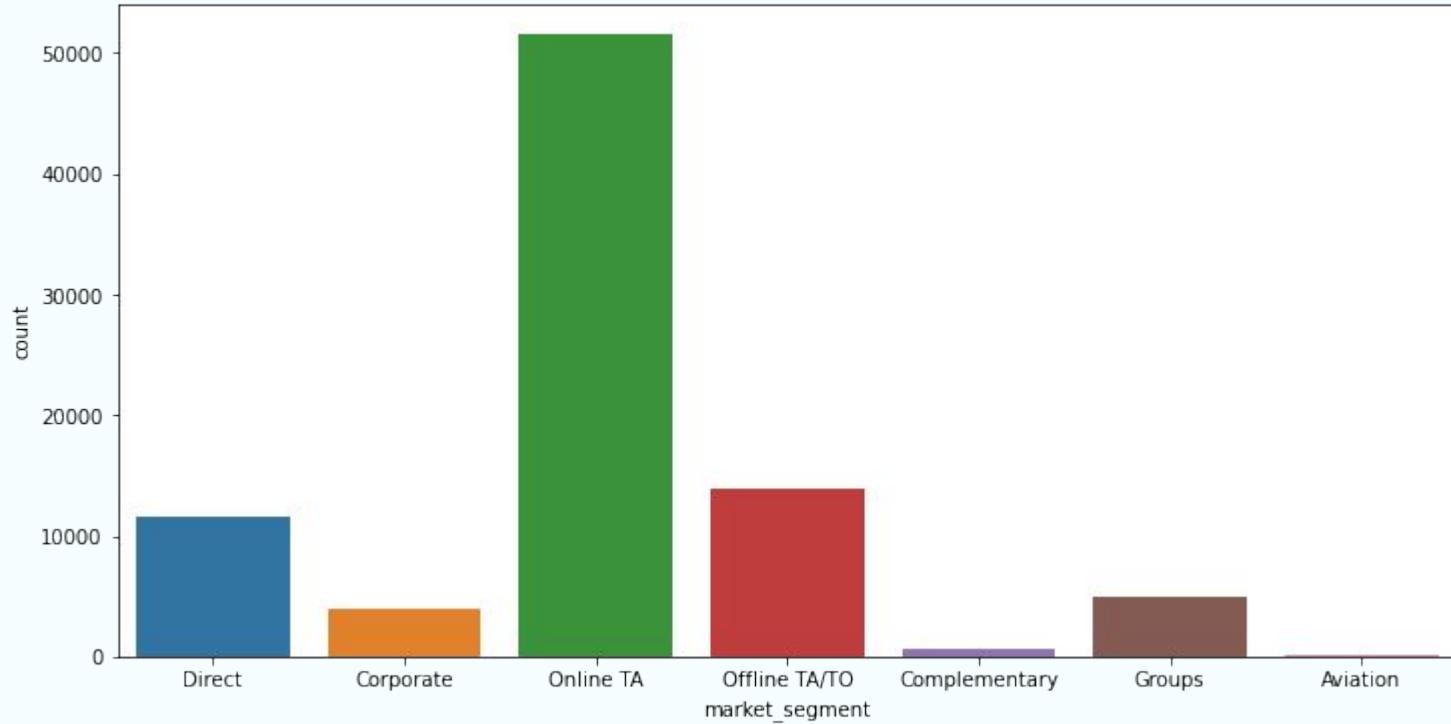
- Portugal is one of the top country followed by Great Britain and France from where most people booked the hotels.



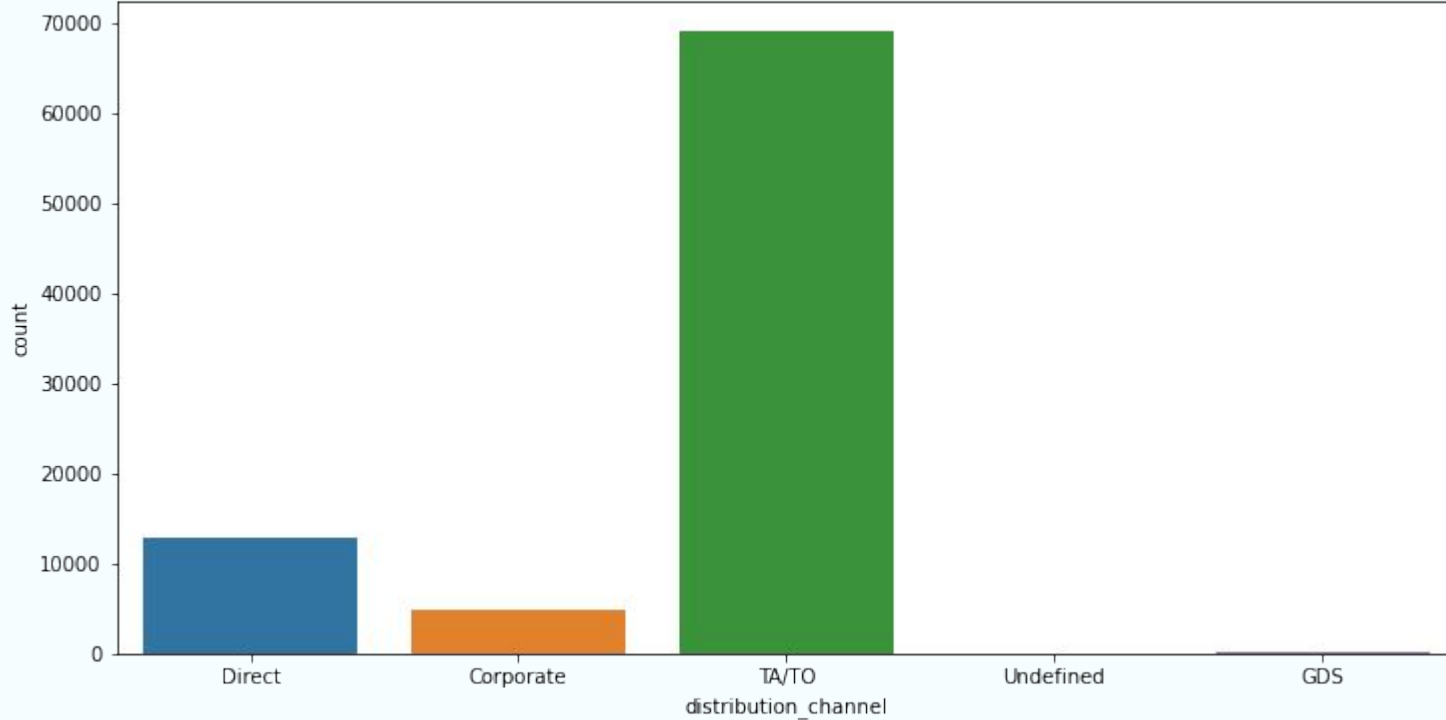
- ❑ Highest proportion of repeated guest was found from the complimentary segment, but the number is very few. Second highest proportion of repeated guest is from the corporate segment



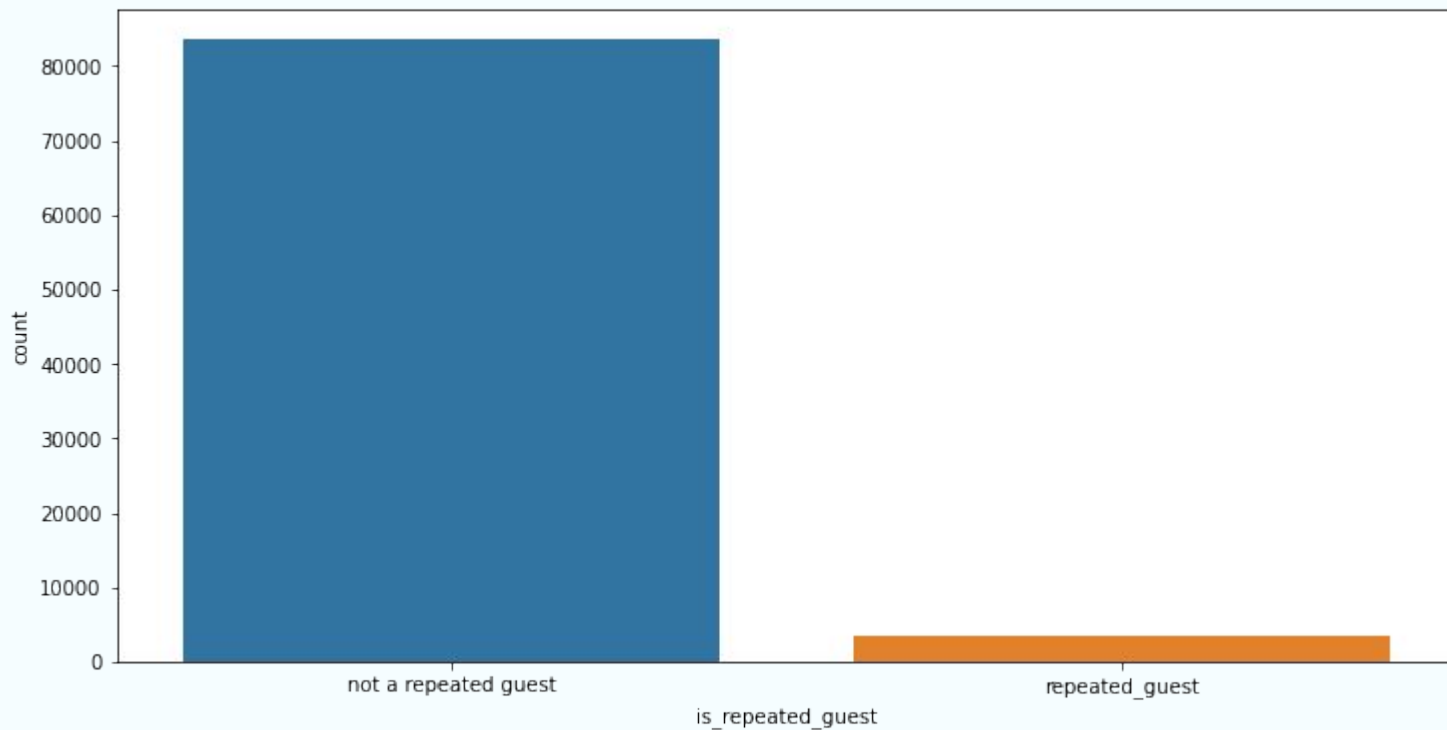
- ❑ Most of the people prefer bed breakfast BB as their meal type during the stay.



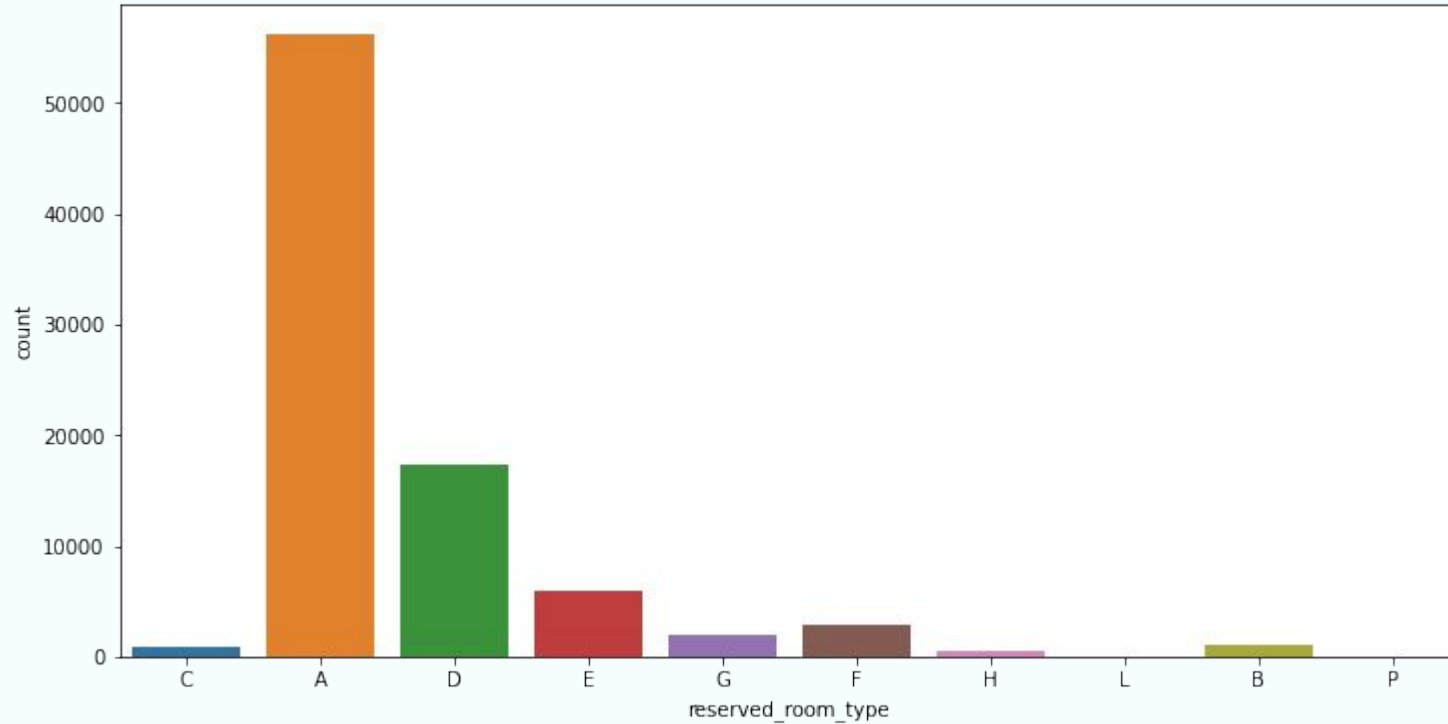
❑ Major market segment for hotel booking is Online Travel Agents.



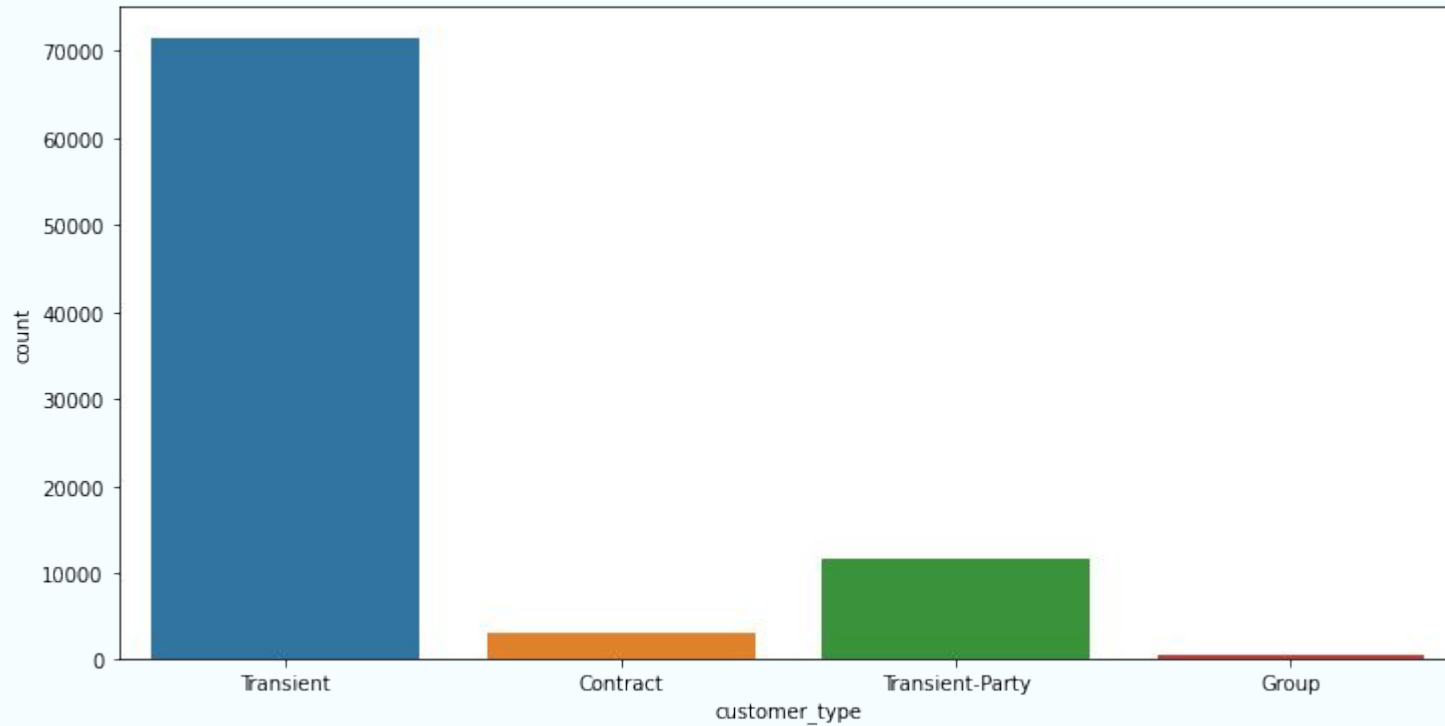
☐ Most distributed channel is Online TA/TO.



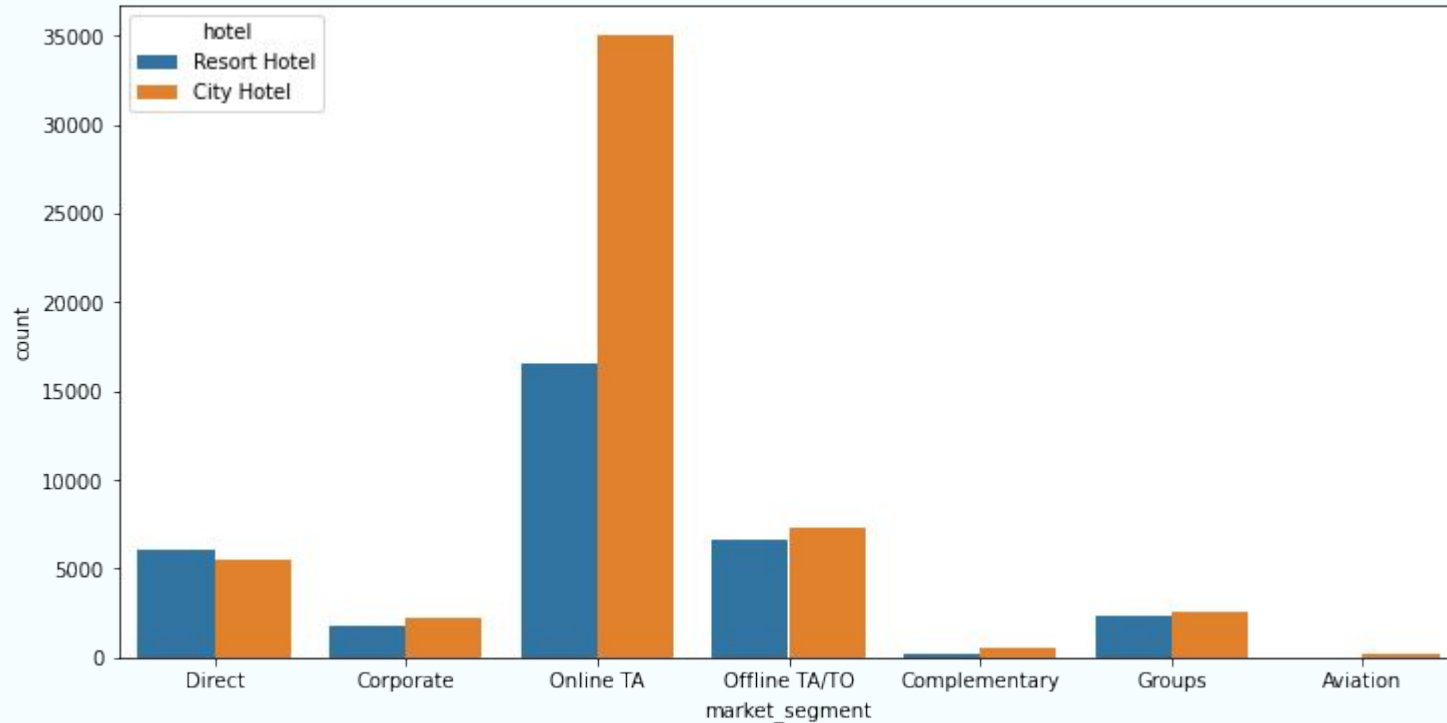
□ About 96% of guest are not repeated.



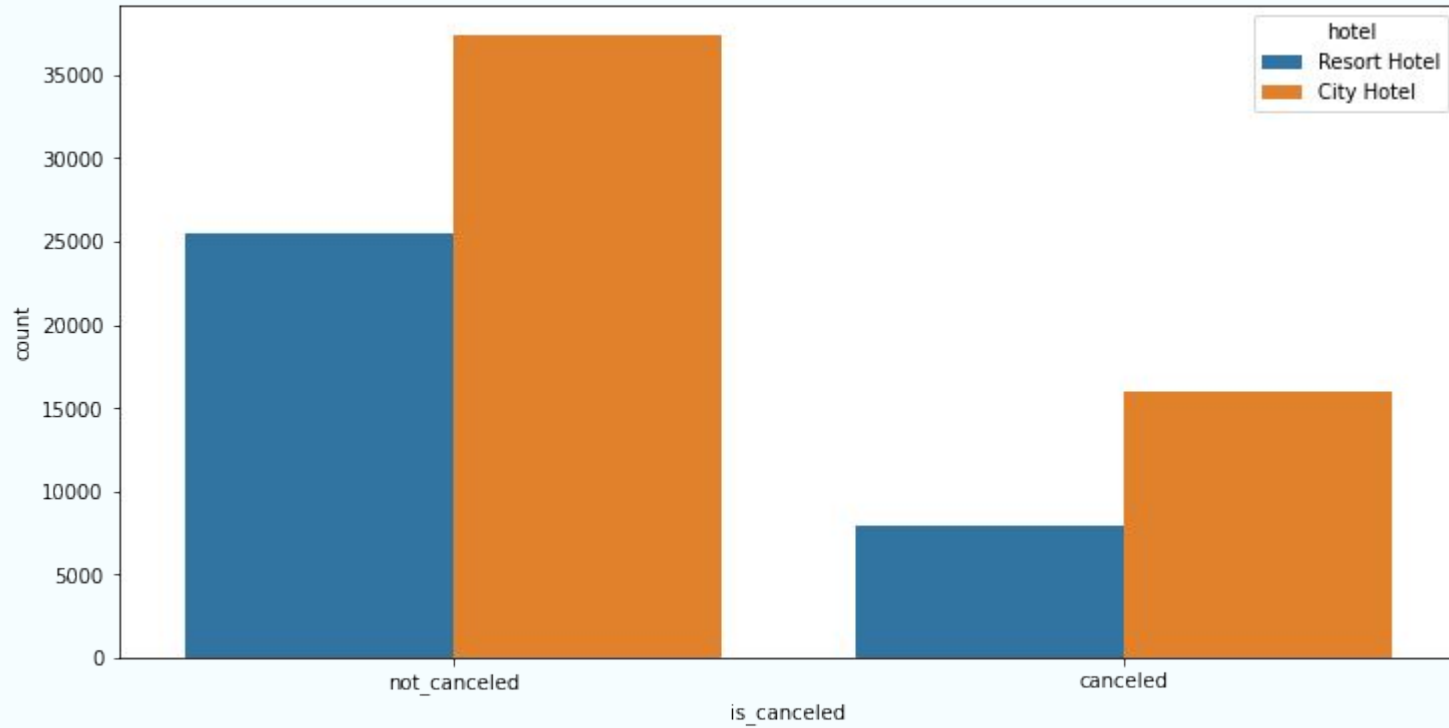
- ❑ The 'A' room type is the most popular among the clients, with 62.62% of the reservations. A very little percentage of booking was held for room type 'P'.



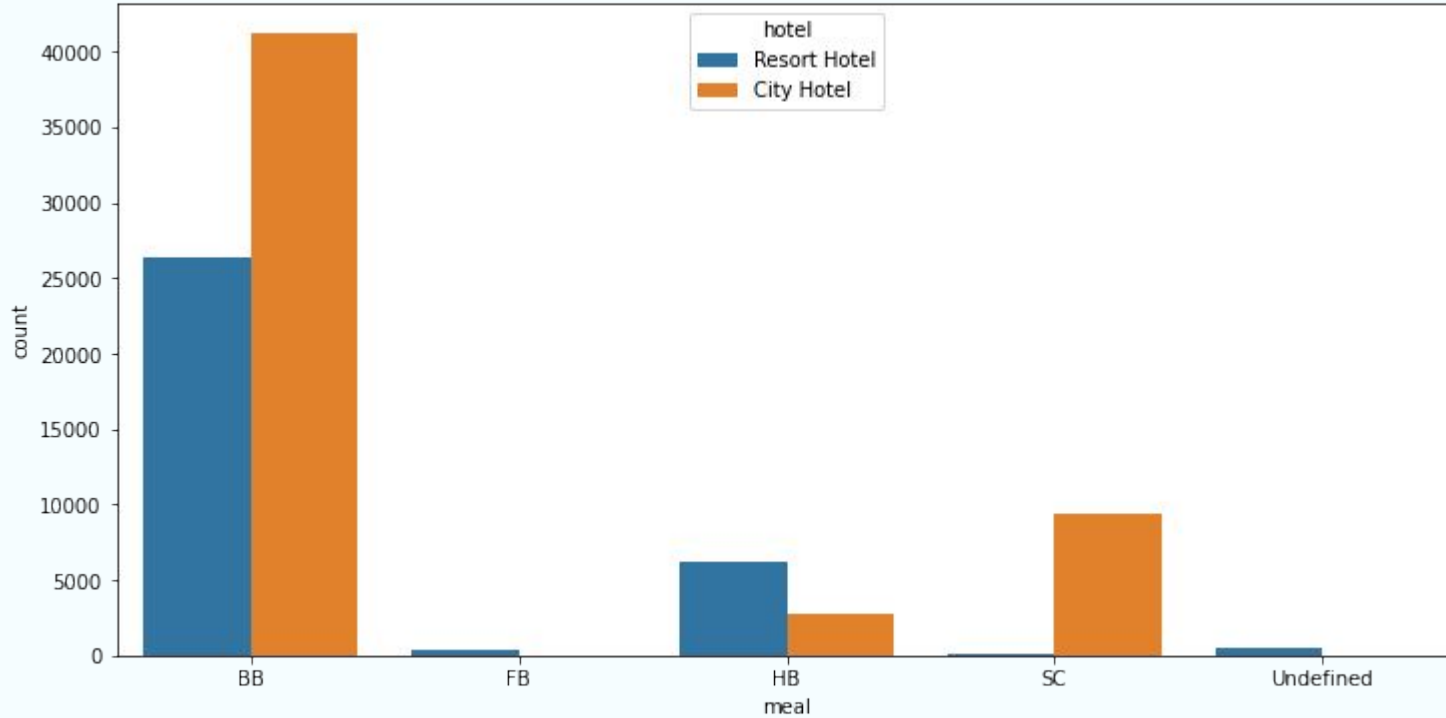
❑ Transients are the most common type of customer.



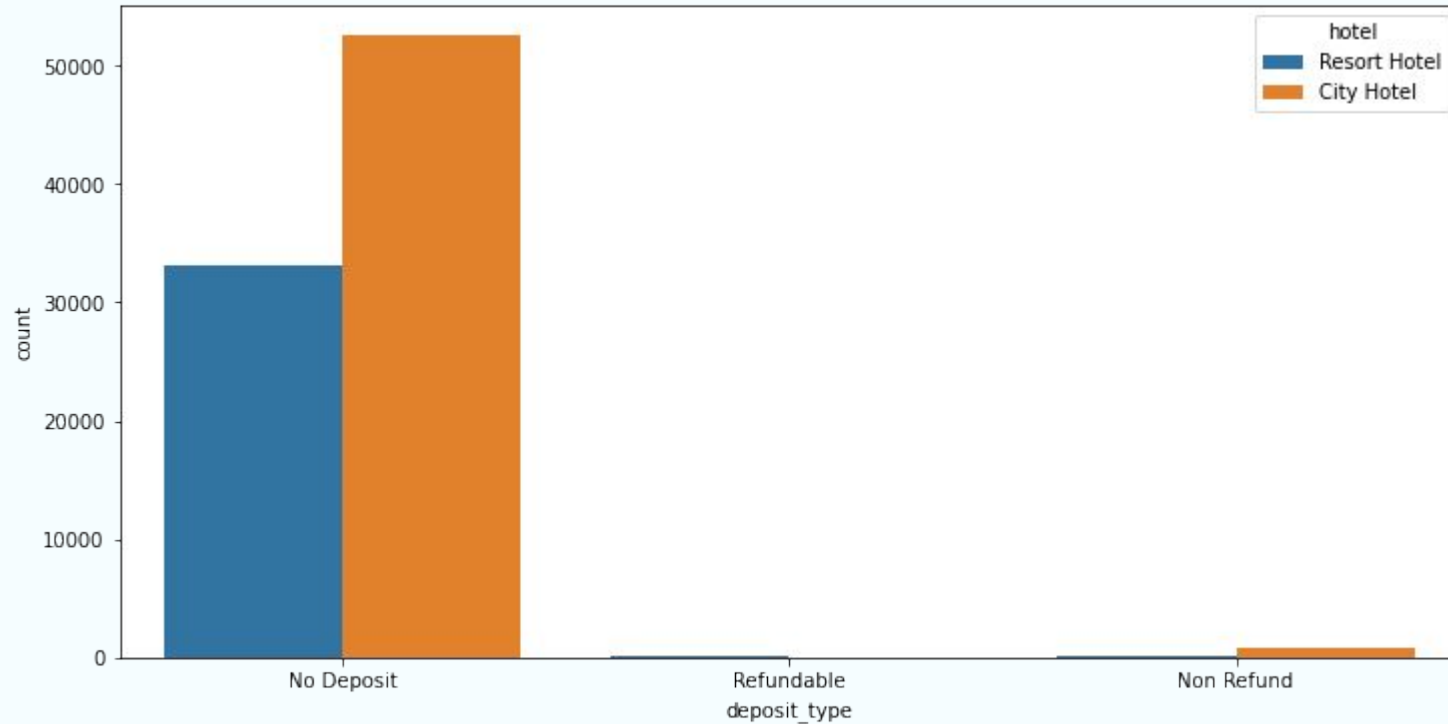
- ❑ Major segment in market is Online TA for both Resort and City Hotel.



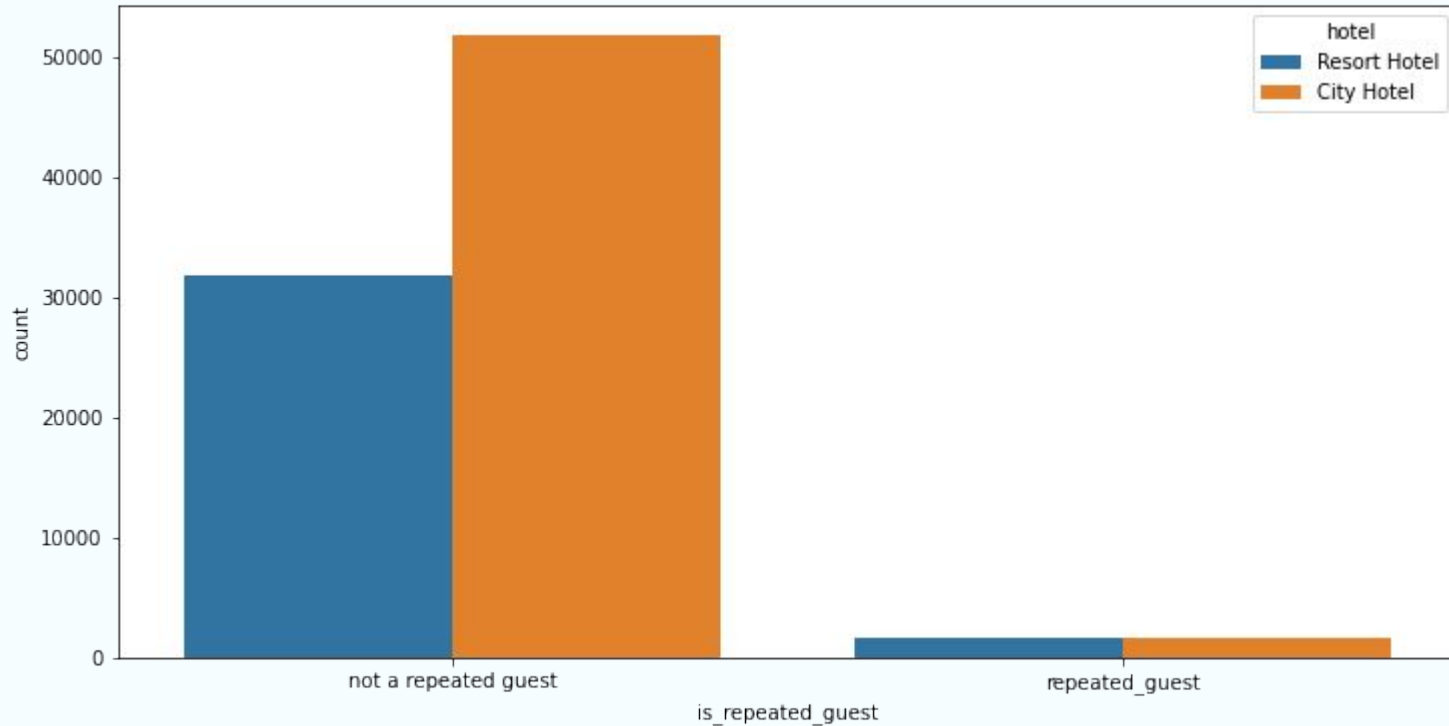
- ❑ City Hotel has more number of cancellations than the Resort Hotel.



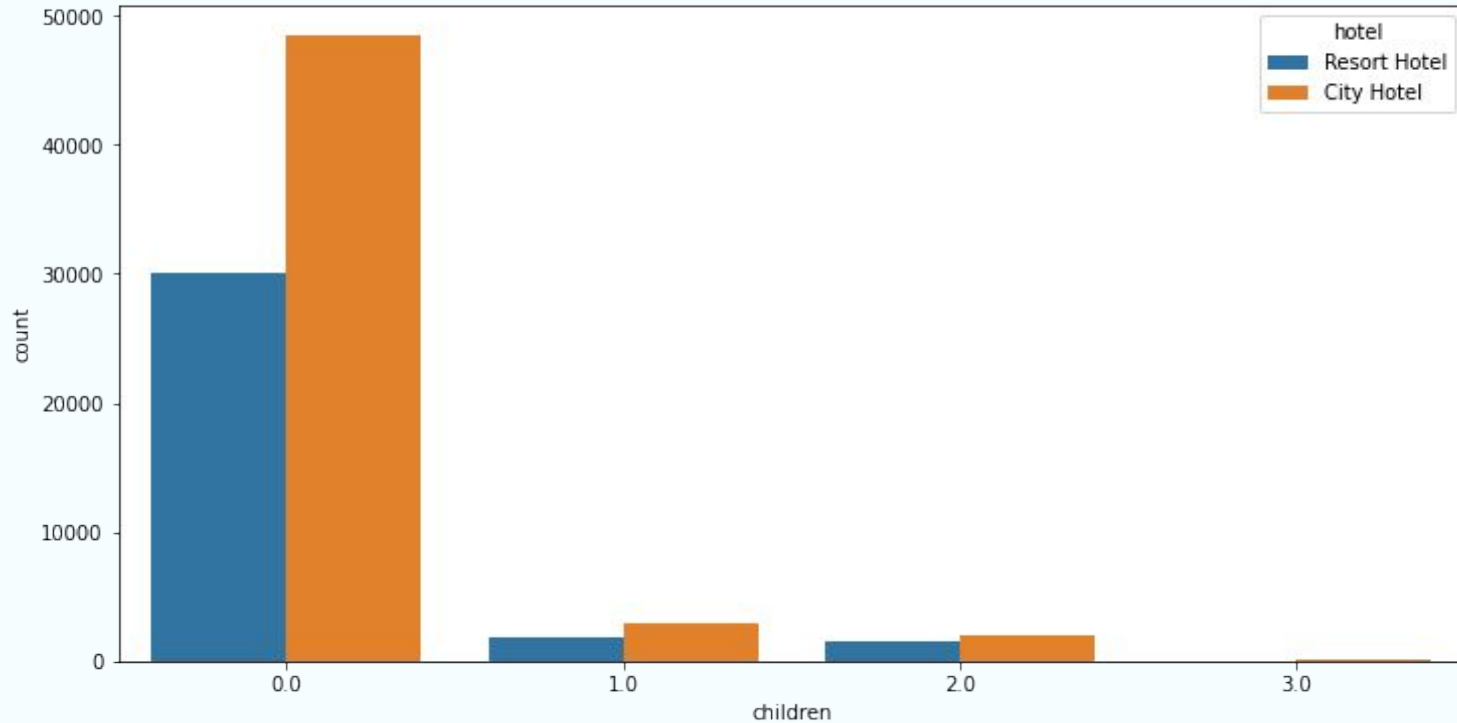
- ❑ 'BB' meal is the most booked meal by the guest in both City Hotel and Resort Hotel.



- ❑ City Hotel takes more non-refundable deposits as compared to Resort Hotel.



- ❑ In City Hotel there are more non-repeated guest as compare to Resort Hotel. Also, as a proportion, repeated guests are higher for the Resort hotel.



- ❑ As the number of children increases, the preference for the Resort hotel seems to increase

Some recommendations to increase the sales of Hotel industry

- ❑ Cancellation policies should be improved, by applying non refundable charges or some penalty.
- ❑ Percentage of online booking cancellation is more, encourage Direct Bookings by offering special discounts.
- ❑ Customer base from foreign countries like Portugal, Great Britain and France is higher. There could be a marketing team to handle that customer base.
- ❑ Hotels should consider the maximum number of special requests from guests to reduce the possibility of cancellations which will eventually help in better customer experience.
- ❑ Keep an eye on which market segment cancellations are coming more and why.
- ❑ City hotel could have more attractions for kids in order to increase its base.

Thank You