Investigate Business Hotel using Data Visualization

Created by:
Destya Febiolla

destya.febiolla@gmail.com
https://www.linkedin.com/in/destya-febiolla
https://github.com/Destyf
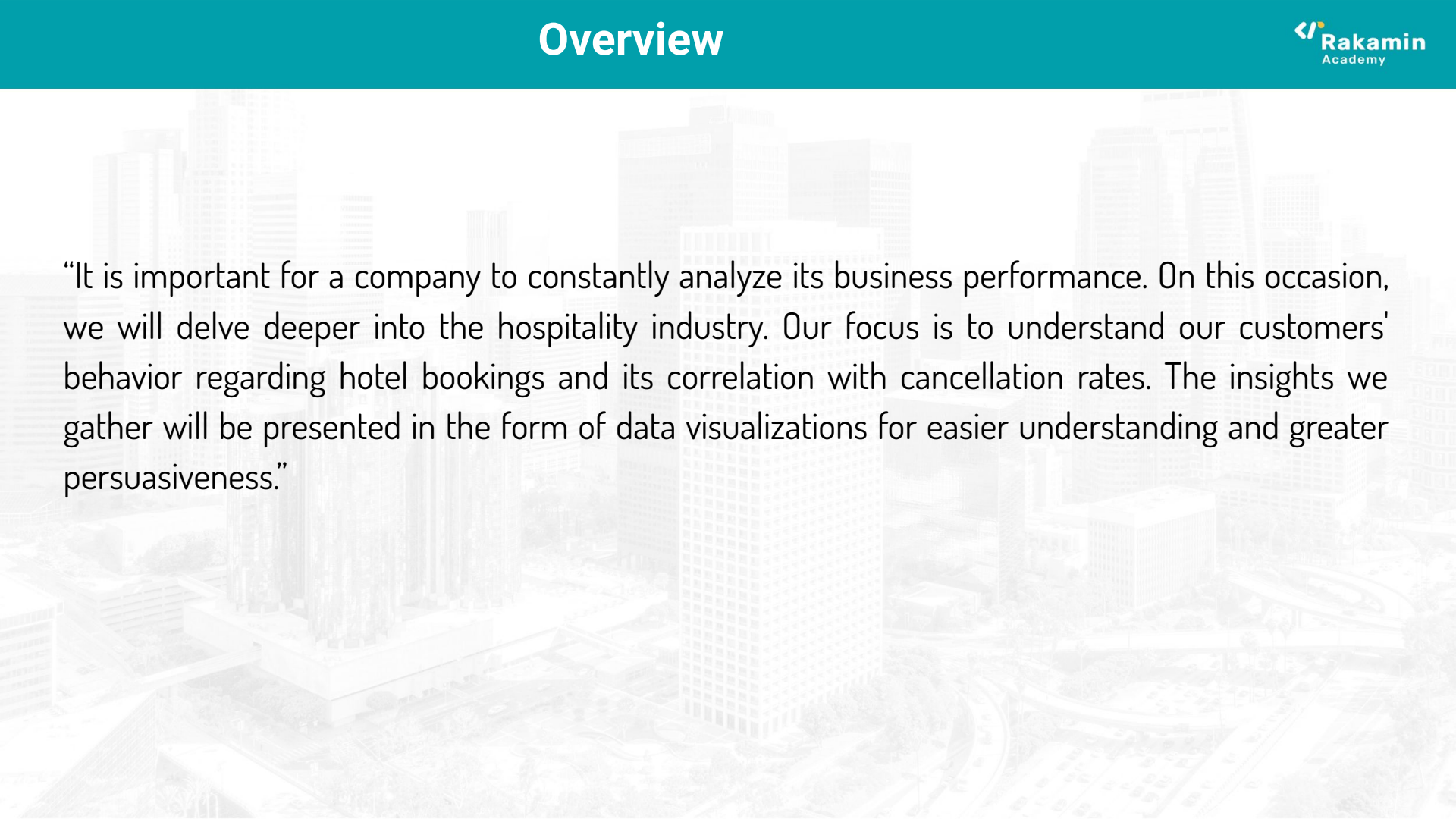
Hello! I'm Destya, a career shifting enthusiast with a strong interest in Data Science. Equipped with a solid foundation from a Data Science Bootcamp, I specialize in data preprocessing and machine learning. With it I am determined to make a meaningful impact as a Data Scientist, and eager to collaborate with fellow data enthusiasts and organizations seeking to harness the power of data for informed decision-making.

Rakamin Academy

"It is important for a company to constantly analyze its business performance. On this occasion, we will delve deeper into the hospitality industry. Our focus is to understand our customers' behavior regarding hotel bookings and its correlation with cancellation rates. The insights we gather will be presented in the form of data visualizations for easier understanding and greater persuasiveness."

Dataset

```
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 29 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   hotel                           119390 non-null  object
 1   is_canceled                     119390 non-null  int64
 2   lead_time                       119390 non-null  int64
 3   arrival_date_year               119390 non-null  int64
 4   arrival_date_month              119390 non-null  object
 5   arrival_date_week_number        119390 non-null  int64
 6   arrival_date_day_of_month       119390 non-null  int64
 7   stays_in_weekend_nights         119390 non-null  int64
 8   stays_in_weekdays_nights        119390 non-null  int64
 9   adults                          119390 non-null  int64
 10  children                        119386 non-null  float64
 11  babies                          119390 non-null  int64
 12  meal                            119390 non-null  object
 13  city                            118902 non-null  object
 14  market_segment                  119390 non-null  object
 15  distribution_channel            119390 non-null  object
 16  is_repeated_guest               119390 non-null  int64
 17  previous_cancellations          119390 non-null  int64
 18  previous_bookings_not_canceled  119390 non-null  int64
 19  booking_changes                 119390 non-null  int64
 20  deposit_type                    119390 non-null  object
 21  agent                           103050 non-null  float64
 22  company                         6797 non-null    float64
 23  days_in_waiting_list            119390 non-null  int64
 24  customer_type                   119390 non-null  object
 25  adr                             119390 non-null  float64
 26  required_car_parking_spaces     119390 non-null  int64
 27  total_of_special_requests       119390 non-null  int64
 28  reservation_status              119390 non-null  object
dtypes: float64(4), int64(16), object(9)
```

- Rows : 119.390
- Features : 29
- Missing Value : children, city, agent, company

children	4
city	488
agent	16.430
company	112.593

For the detail of my codes, you can see here here

Data Preprocessing

01 Handling Missing Value

There are 4 features with missing values:
- Company : The 'company' feature has values in IDs from Hotel. The missing values assumed to indicate the customers did not orders through Hotel, so they are filled with 0.
- Agent: The 'agent' feature has values in IDs from Hotel. The missing values assumed to indicate that the customers did not orders through the Hotel, so they are filled with 0.
- City: The missing values is 0.409%. It is decided to drop them because they do not significantly affect the results.
- Children: The missing values is 0.003%. It is decided to drop them because they do not significantly affect the results.

02 Correcting Dtypes

Change Type Data 'float64' to 'int64' that have no decimal value
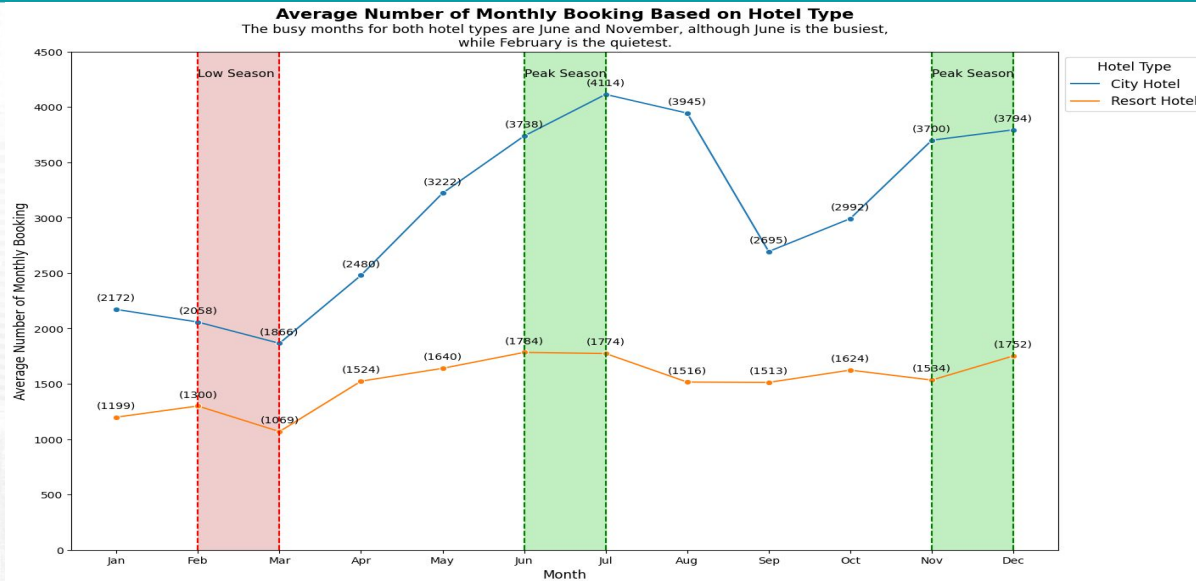
03 Replace Inappropriate Value

There is a value 'Undefined' in the 'meal','market_segment', & 'distribution_channel' feature that has no any clear category. It is assumed that customers did not order Breakfast, Dinner, or Full Board (Breakfast+Dinner) in 'meal', so it will be changed to 'No Meal'. The 'Undefined' in feature 'market_segment' & 'distribution_channel' will be replaced with mode value in that feature.

04 Remove Unnecessary Data

Remove value of 'adr' that below 0 and unnecessary data with 0 Guest and 0 Stay Days

For the detail of my codes, you can see here here

Average Number of Monthly Booking Based on Hotel Type
The busy months for both hotel types are June and November, although June is the busiest,
while February is the quietest.

- Peak Season: June and November are identified as the peak months for high demand in hotel accommodations for both City and Resort Hotels. This means that these months witness a significant influx of guests, resulting in higher occupancy rates and potentially higher room prices. Factors such as holidays, special events, or favorable weather conditions contribute to peak season.
- Low Season: February is considered the quietest month with the lowest demand for hotel accommodations for both City and Resort Hotels. During the low season, hotels experience a decrease in bookings, often due to factors like unfavorable weather, fewer holidays, or less popular travel periods. To attract guests and fill their rooms during this period, hotels might offer promotions, discounts, or special packages.

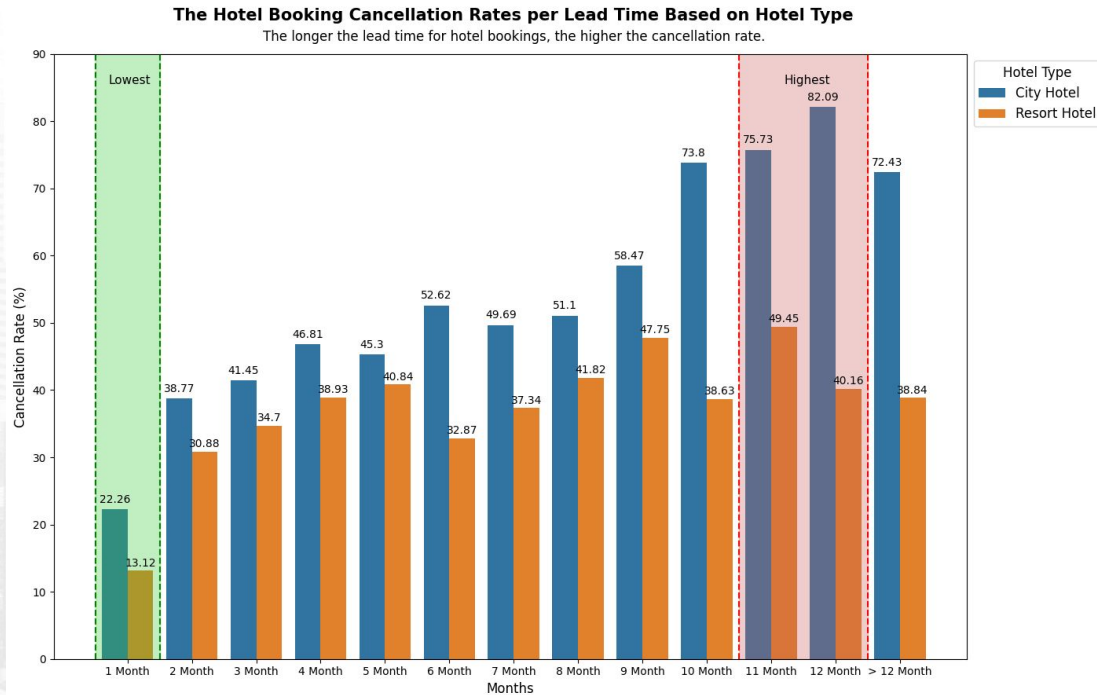# Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates

The Trend of Hotel Booking Cancellation Rates per Stay Durations Based on Hotel Type

The longer the stay duration booked, the higher the cancellation rate.
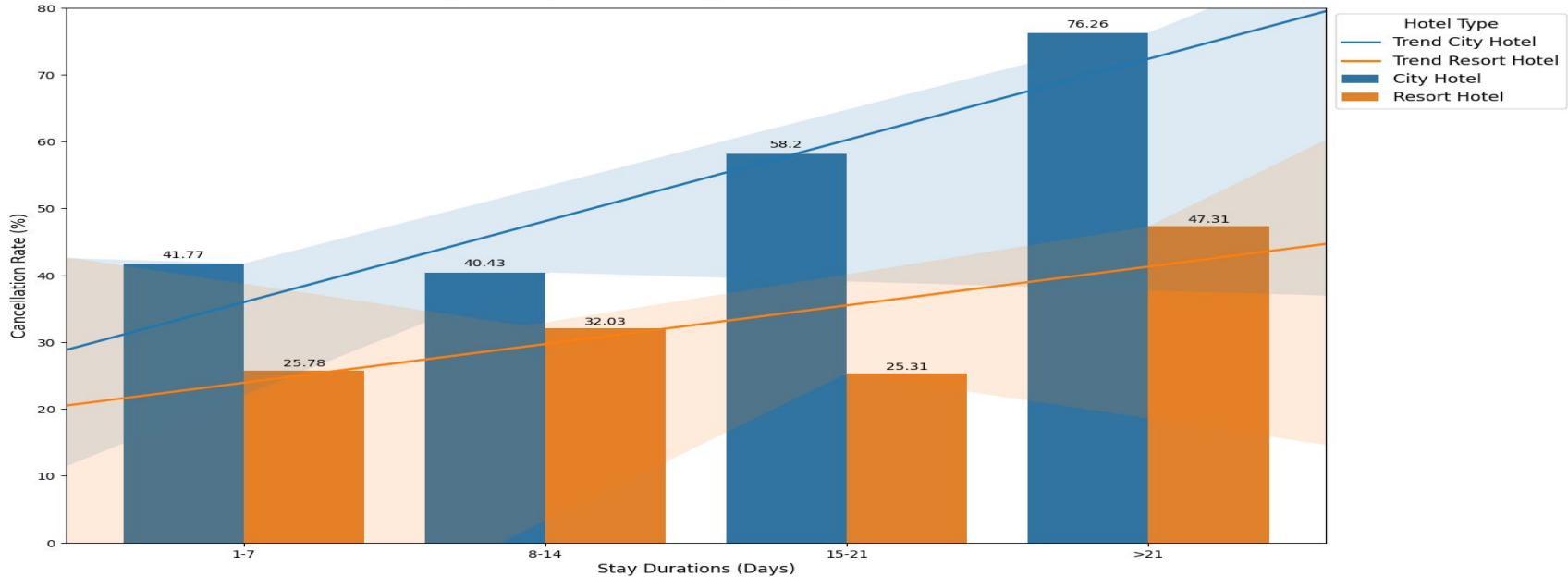
- City Hotel dan Resort : Both types of hotels, City Hotel as well as Resort, show an increase in cancellation rates as the guests' duration of stay extends. This indicates that the longer guests stay, the higher the likelihood of them canceling their reservations.
- City Hotel vs. Resort : The cancellation rate in City Hotel is higher than in Resort. Cancellations in City Hotel reach 76.26% when the duration of stay exceeds 21 days, but it is at its lowest at 40.43% during a stay duration of 8-14 days. On the other hand, Resort has its highest cancellation rate at around 47.31% when the duration of stay also exceeds 21 days, and it is lowest at 25.78% during a stay duration of 15-21 days.

The Hotel Booking Cancellation Rates per Lead Time Based on Hotel Type
The longer the lead time for hotel bookings, the higher the cancellation rate.

- There is a correlation between the lead time for booking and the cancellation rate, especially for city and resort hotels. As the lead time increases, the cancellation rate also tends to rise.
- The lowest cancellation rates are observed at a lead time of 1 month, with city hotels having a cancellation rate of 22.26% and resort hotels at 13.12%.
- The highest cancellation rates are between the lead times of 11 and 12 months. In month 12, city hotels experience the highest cancellation rate of 82.09%, while resort hotels have a peak cancellation rate of 49.45% in month 11.