

# Investigate Business Hotel using Data Visualization

By Joni Syofian





# About Me

Joni is a fresh graduate student from Bandung Institute of Technology. He is interested in data science, data analytics, and ocean issues. To improve his skills in the field of data, he took several courses and just completed a data science bootcamp with a good grade.

# Background

It is critical for a company to constantly assess its business performance. The company can take appropriate action to deal with a problem by analyzing business performance.

On this occasion, we will examine the hotel industry's performance. The primary goal of this project is to identify the customer actors involved in hotel bookings and their relationship to hotel cancellation rates. So three things will be analyzed in this project:

- 1. Monthly Hotel Booking Analysis Based on Hotel Type**
- 2. Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates**
- 3. Impact Analysis of Lead Time on Hotel Bookings Cancellation Rate**

Based on the results of the insights obtained, visualization will be carried out so that they can be easily understood.



# Data and Steps

The data used in this project consists of 29 columns and 119390 rows. This project's creation makes the assumption that the data is either from an order placed in 119390 or that there are no duplicates.

The project will go through several stages as follows:

1. Load data: load data from source
2. Data Preprocessing: handling missing values and incorrect values from numerical and categorical data
3. Insight and Visualization: Visualization and analysis of the visualization
4. Business Recommendation: Provide business recommendations from the results of the analysis



# Data Preprocessing

Before analyzing data, data preprocessing is first carried out so that the data can be processed and insights can be obtained. In addition, in the raw data, there may also be missing values and values that are not appropriate. Therefore, preprocessing data is required. The stages of data processing this time will be carried out in several ways, as follows:

1. Handling missing values
2. Change data type
3. Handling incorrect values (categorical and numerical data)



# = Data Preprocessing: Handling Missing Values & Change Data Type

```
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)
memory usage: 26.4+ MB
```

From the information on the side, it can be seen that there are 4 features or columns that have missing values: company, agent, city, and children.

- the company and agent feature, input the value 0 on the missing data. This indicates that the customer placed an order without an agent or company.
- the city feature, input "Unknown." because there is no clearer information.
- children feature, input a value of 0 assuming that the customer does not bring their child.

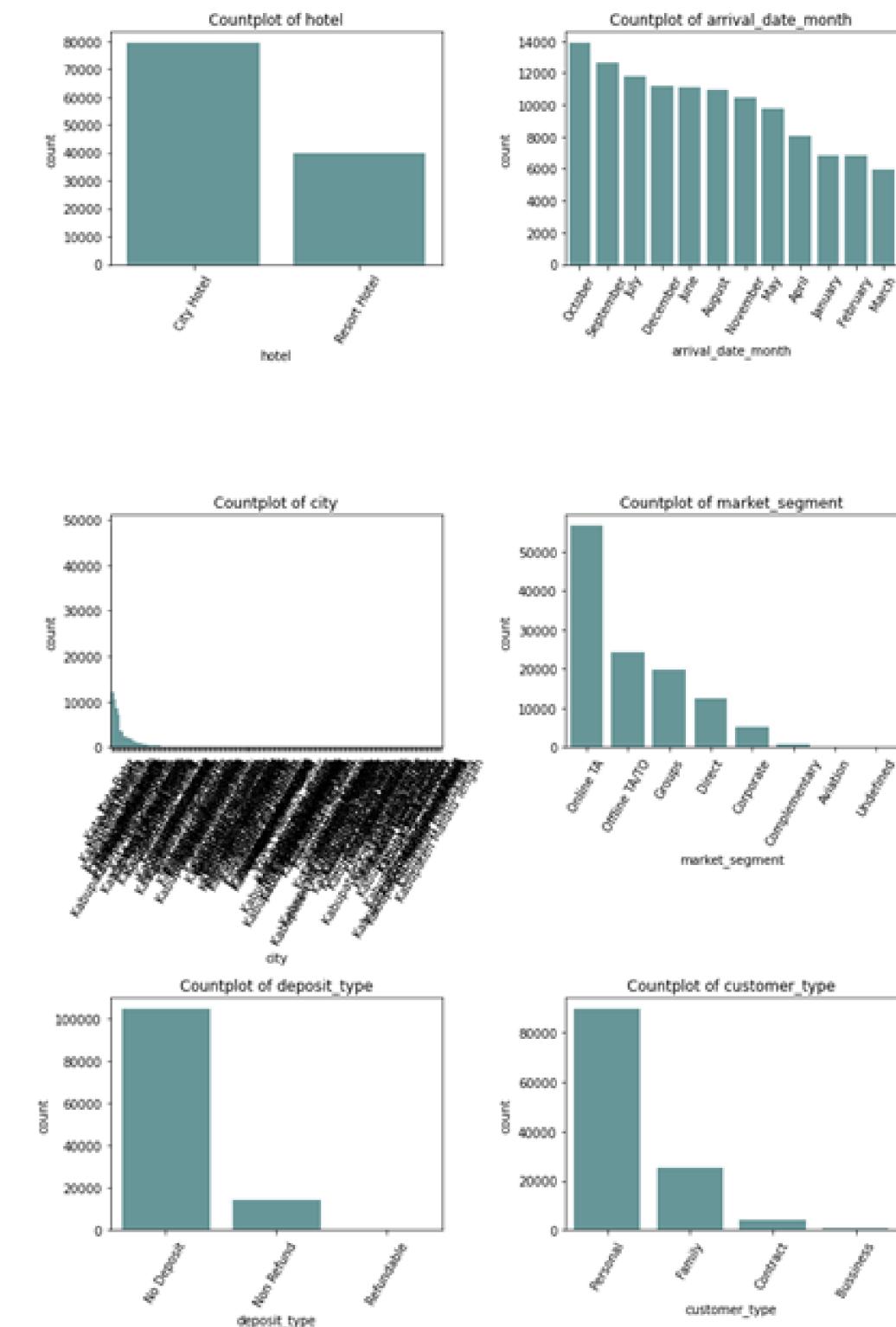
```
# handling missing values (null values)
df['children'] = df['children'].fillna(0)
df['city'] = df['city'].fillna('Unknown')
df['agent'] = df['agent'].fillna(0)
df['company'] = df['company'].fillna(0)
```

In addition to missing values, the information also contains data type changes in some features, such as the children, agents, and company features, from float to integer.

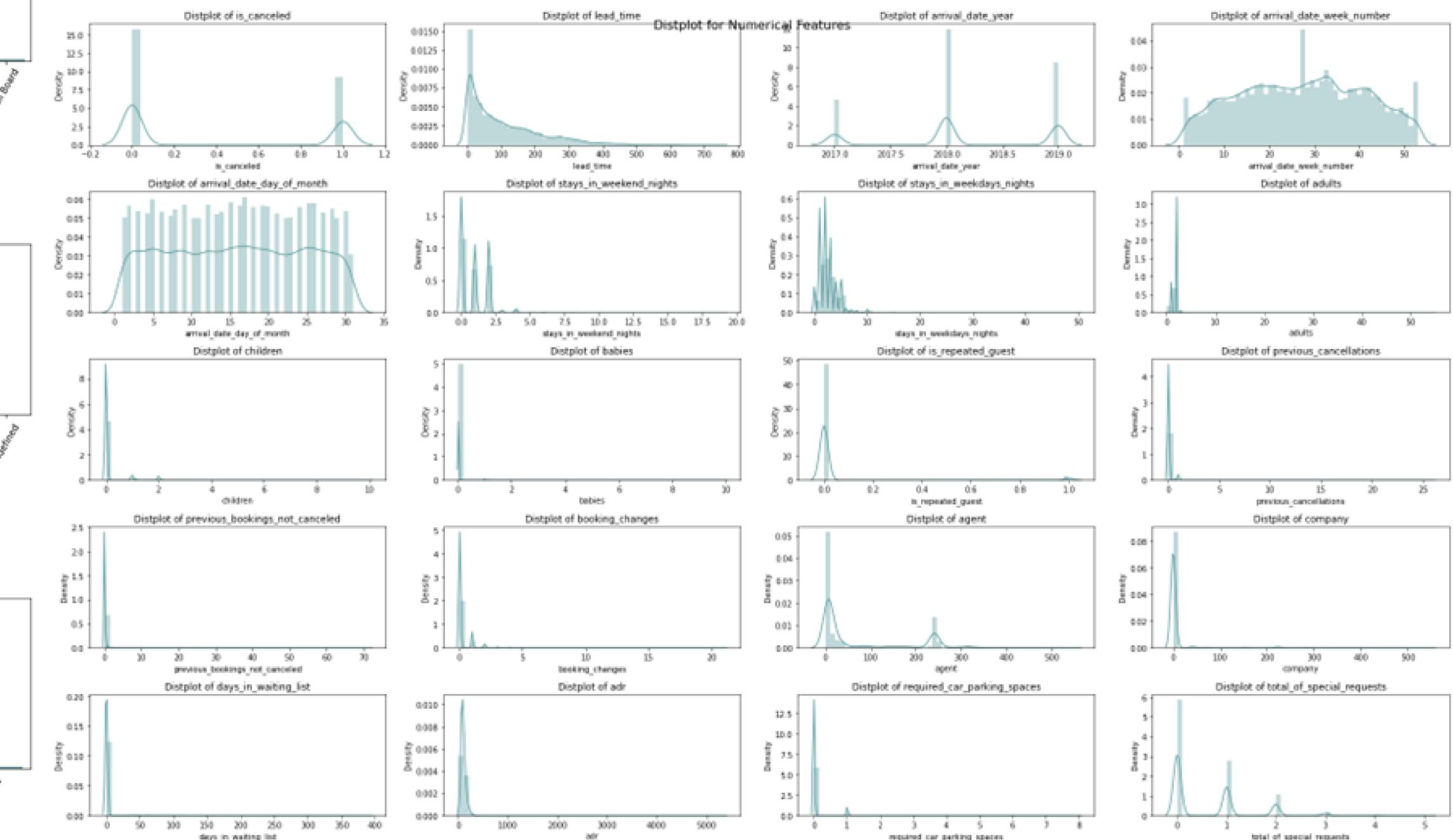
```
# change the data type of children, agent and company features to integer instead of float
df['children'] = df['children'].astype('int64')
df['agent'] = df['agent'].astype('int64')
df['company'] = df['company'].astype('int64')
```

# = Data Preprocessing: Handling Incorrect value (categorical & numerical data)

## □ Categorical data



## □ Numerical data



# = Data Preprocessing: Handling Incorrect value (categorical & numerical data)

From the observed category and numeric data, there are several features that have unsuitable values, so they need to be replaced or dropped.

## ❑ Categorical data

In the category data, there are several features that have incorrect values, such as the "market\_segment," "distribution\_channel," and "meal" features, which have an "undefined" value, so it is necessary to handle these values.

- In the features "market\_segment" and 'distribution\_channel' the value of 'Undefined' is replaced with the most values, namely 'Online TA' in the 'market\_segment' feature and 'TA/TO' in the 'distribution\_channel' feature.
- In the 'meal' feature, the value of 'Undefined' is replaced with 'No Meal' because it has the same meaning.

```
# Replace 'Undefined' in 'meal' feature with 'No Meal'  
df['meal'] = df['meal'].replace(['Undefined'], 'No Meal')  
  
# Replace 'Undefined' in 'market_segment' and 'distribution_channel' features with the most value in that features.  
df['market_segment'] = df['market_segment'].replace(['Undefined'], 'Online TA')  
df['distribution_channel'] = df['distribution_channel'].replace(['Undefined'], 'TA/TO')
```

## ❑ Numerical data

On numerical data is carried out unnecessary data deletion with 0 guests and 0 nights with the following steps:

- First, we create a new feature, 'total\_guests' from the combination of the 'adults', 'children' and 'babies' features.
- Then a new feature 'stay\_duration' is created from the combination of 'stays\_in\_weekend\_nights' and 'stays\_in\_weekdays\_nights'.
- Then we deleted the data with 'stay\_duration' worth 0 and 'total\_guests' worth 0.

```
df['total_guests'] = df['adults'] + df['children'] + df['babies']  
print('Data with 0 guest: {0} out of {1} all data'.format(df[df['total_guests'] == 0].shape[0], df.shape[0]))  
df['stay_duration'] = df['stays_in_weekend_nights'] + df['stays_in_weekdays_nights']  
print('Data with 0 night: {0} out of {1} all data'.format(df[df['stay_duration'] == 0].shape[0], df.shape[0]))  
# Remove unnecessary data  
df_clean = df[(df['total_guests'] > 0) & (df['stay_duration'] > 0)].copy()  
print('before pre-processing:', df.shape[0])  
print('after pre-processing:', df_clean.shape[0])
```



# Monthly Hotel Booking Analysis Based on Hotel Type

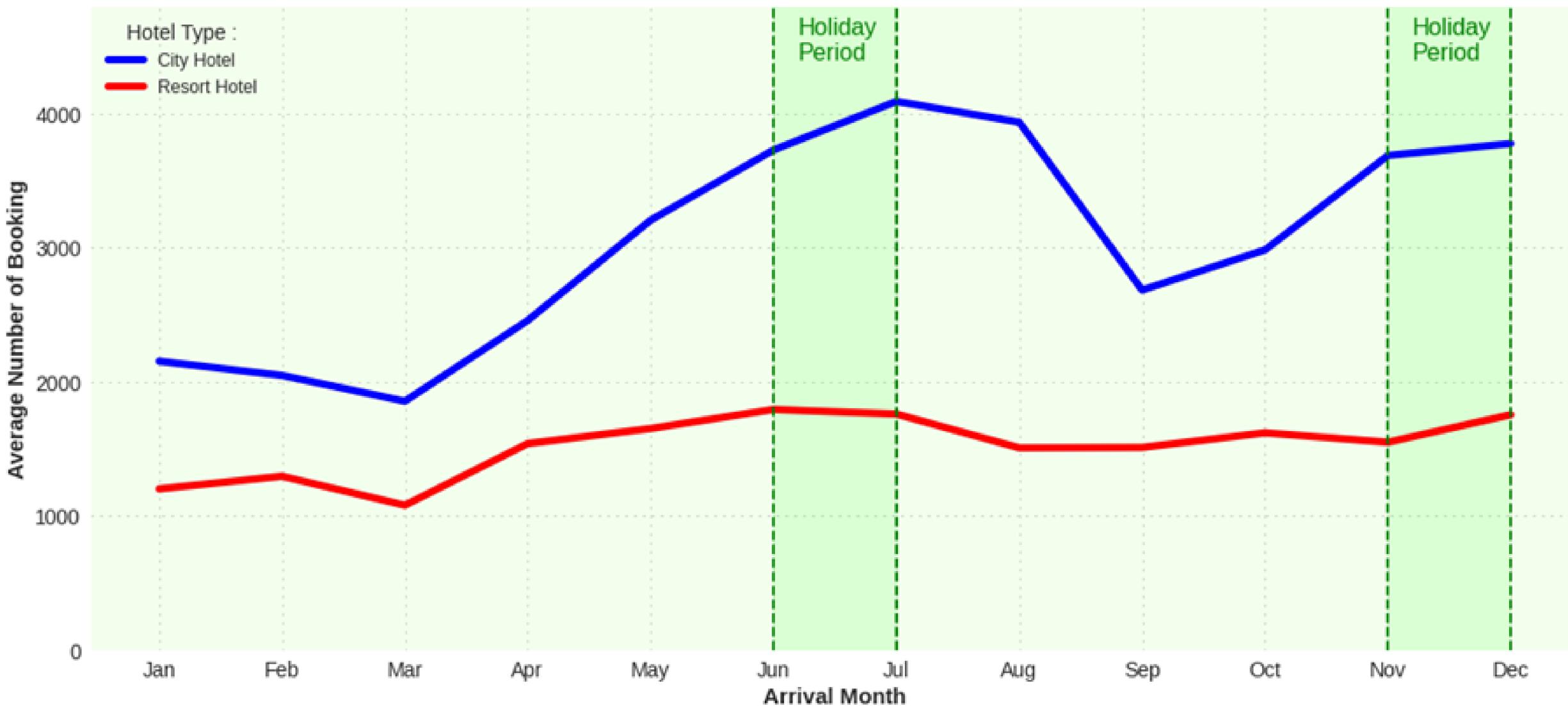
In the hospitality business, of course, it is closely related to the customer. The more customers who order, the more income the company has. Therefore, analyzing the behavior of customers when booking hotels is very important. For example, we can identify what types of hotels are most in demand by customers, and we can also associate this with the season conditions when these hotels are booked. So at this stage, we will look for a comparison of the number of hotel bookings every month based on the type of hotel.



# = Monthly Hotel Booking Analysis Based on Hotel Type

## The number of guests from both types of hotels increased during the holidays

In August and September, the number of guests at the "City Hotel" decreased, while both hotels have less guests during not holiday (Jan-Mar)



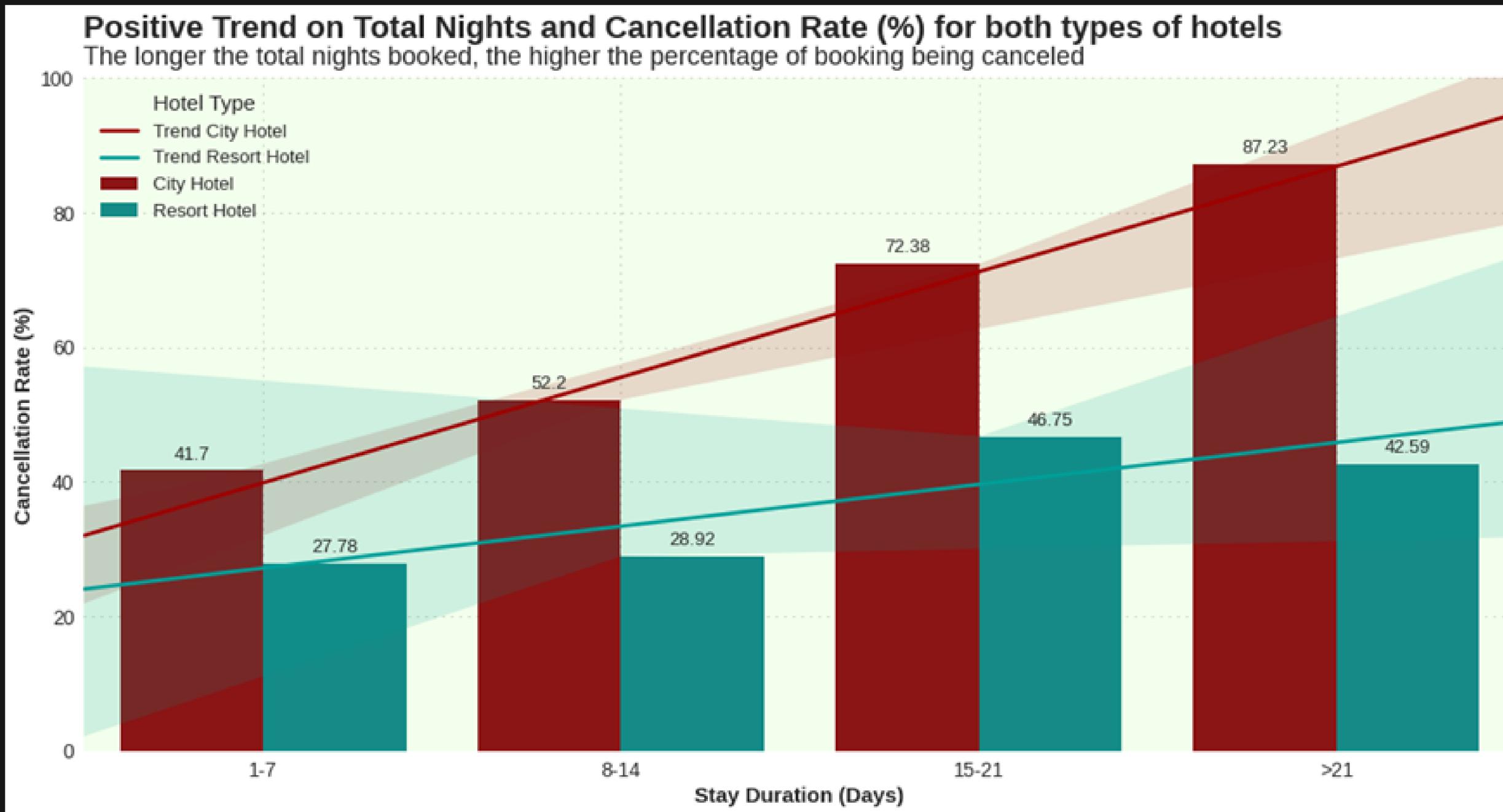
- During the holiday period, both hotels have more guests, with the City Hotel having more than the Resort Hotel.
- In Indonesia, most schools have holidays from June through July. Typically, a lot of families would travel together at that time. This explains why June and July saw an increase in visitors at both hotels. This can be used as one of the reasons for the timing of marketing implementation.
- From August to September, City Hotel's guest number decreases significantly.
- It's because of the school period. so that the number of guests has decreased from the school holiday period.
- The lowest value of average number of booking at March (for both hotel types).

# Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates

In addition to analyzing customer behavior in booking hotels, to measure the success of a hotel business, we can see the level of order cancellations. If many customers cancel their orders, this will adversely affect the hotel's business performance. Therefore, we need to find out what factors affect the cancellation of orders. At this stage we will investigate how the length of stay can affect the cancellation rate of hotel bookings.



# = Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates



- The longer the total number of nights booked, the higher the cancellation rate (positive trend); City hotels have a steeper trend than Resort hotels.
- The higher cancellation rate for both hotels; city hotel on >21 days stay duration (87.23%) and resort hotel on 15-21 days stay duration (46.75%)
- To prevent this from happening, the hotels should implement a cancellation policy. The longer the total number of nights booked, the higher the cancellation fee.



# Impact Analysis of Lead Time on Hotel Bookings Cancellation Rate

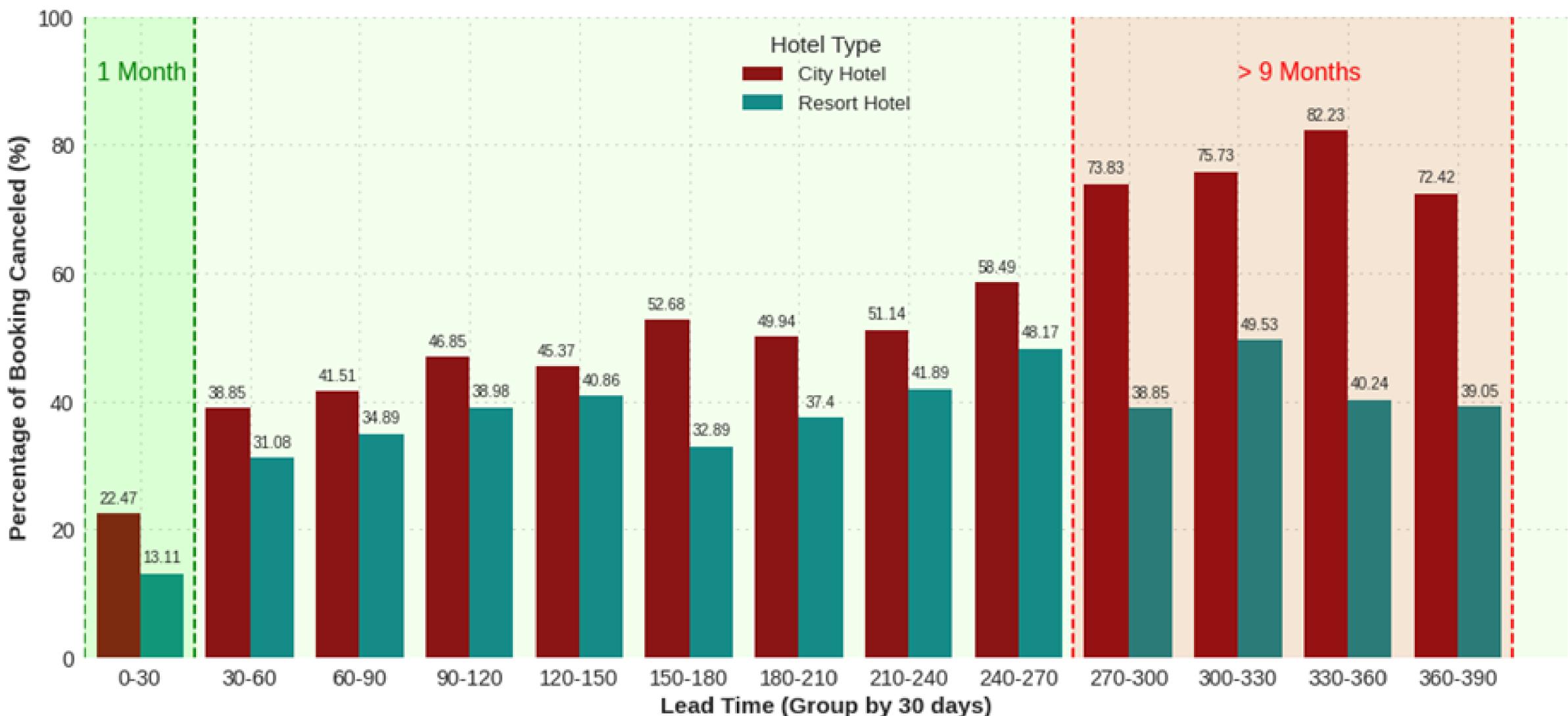
The hotel business usually allows customers to book hotels before the day of their arrival. The time intervals that occur also vary, some are only a few days, and there are also customers who book hotels up to several months before the day of arrival. So, we will check whether the time interval between hotel reservations and the customer's arrival day has an effect on the rate of cancellation of hotel bookings.

# = Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates

- Both hotel types has lowest cancellation rate of bookings on 1 month lead time; city hotel (22.47%) and resort hotel (13.11%).
- When booking more than 9 months, city hotels have a high percentage of cancellations reaching more than 70%.
- Resort Hotel has quite stagnant (in around 40%).
- Significant growth of cancellation rate for city hotels each month from around 20% to around 70%, also it happens for resort hotels each month from around 10% to around 40%.
- This growth of cancellation rate could be happened because customer vacation plan canceled or the customer forgets to have booked a hotel. In order to prevent them from canceling their reservations, the hotel could send them reminders. Additionally, the hotel could prevent this by making the cancellation policy applicable to all bookings.

The time interval between hotel reservations and the customer's arrival day has an influence on the cancellation rate of hotel bookings

Both hotel types has lowest cancellation rate of bookings on 1 month lead time; city hotel (22.47%) and resort hotel (13.11%)  
When booking more than 9 months, city hotels have a high percentage of cancellations reaching more than 70%.



# Business Recommendation

Based on the visualization and insight gained, the following businesses can be recommended to the hospitality industry:



1. Hospitality can implement a penalty system to the cancellation of hotel bookings carried out by the order to be able to reduce the cancellation rate of ordering.



2. Hospitality can apply the maximum term or order distance of no more than 3 months in order to reduce the cancellation rate.



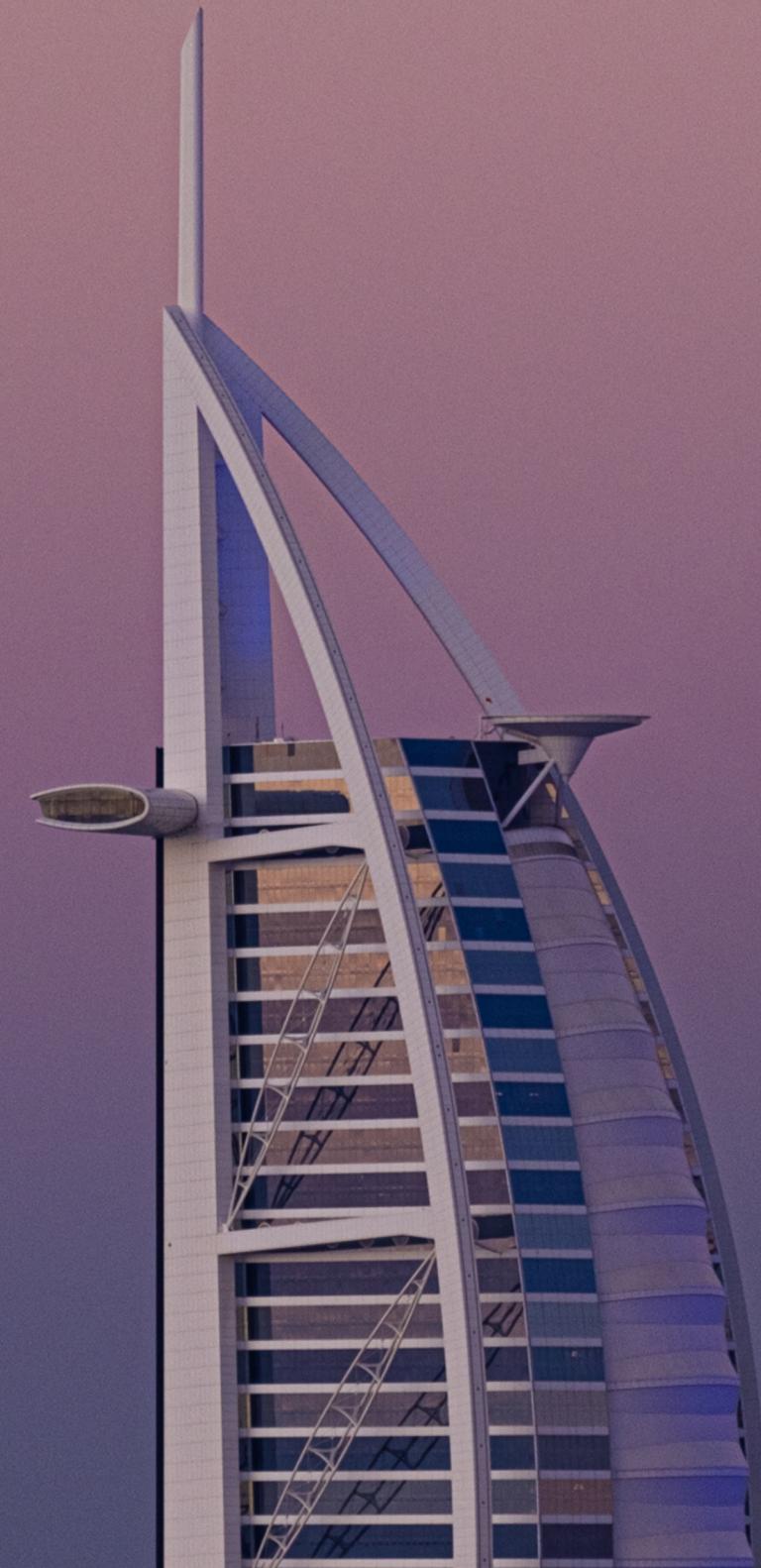
3. Hospitality can implement a reminder system, especially on the order that makes an order for a long time from the time of order.



4. The hotel can carry out marketing or offer products/or services in June-July and November-December where there are many customers.



# Thank You



**Email**

jonisyofian14@gmail.com

**LinkedIn**

<https://www.linkedin.com/in/jonisyofian/>