Investigate Business Hotel using Data Visualization
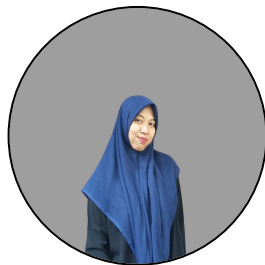
Created by:
Aisyah Khairunnisa Al Athifa
aisyah.k.athifa@gmail.com
LinkedIn Github Medium

"Aisyah is a machine learning and data enthusiast with a knowledge in data analysis, data cleansing, data visualization, and machine learning modelling. Seeking an entry level position of data analyst, data scientist or related role. "

"It is very important for a company to always analyze its business performance. Business performance analysis could help in finding problems and then put in place strategies to solve the issues, so that the company can reach their optimal potential.

This time, I will explore more about hotel business to find some insights. My focus is to analyze customer behaviour in making hotel reservations, and which factor has the impact of cancellation rate of the reservations. The results of the insights I found later will be presented through visualizations for better understanding.

The dataset that will be used today is a dummy data contains a single file which compares various booking information between two hotels in Indonesia: a city hotel and a resort hotel."

Data Pre-processing

```python
# Import dataset
df = pd.read_csv('hotel_bookings_data.csv')

# Handling 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)

# Fix data types
df['children'] = df['children'].astype('int64')
df['agent'] = df['agent'].astype('int64')
df['company'] = df['company'].astype('int64')

# Fix incorrect values
df['meal'] = df['meal'].replace(['Undefined'], 'No Meal')

# Remove unnecessary data with 0 guest and 0 night
df['total_guests'] = df['adults'] + df['children'] + df['babies']
df['stay_duration'] = df['stays_in_weekend_nights'] + df['stays_in_weekdays_nights']
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])
```

- **Handling missing values**

  There are 4 columns with missing values: children, city, agent, and company. Missing values in city column will be imputed with 'Unknown'. As for the rest will be imputed with 0.

- **Fix data types**

  Change data types from float to integer for children, agent, and company columns.
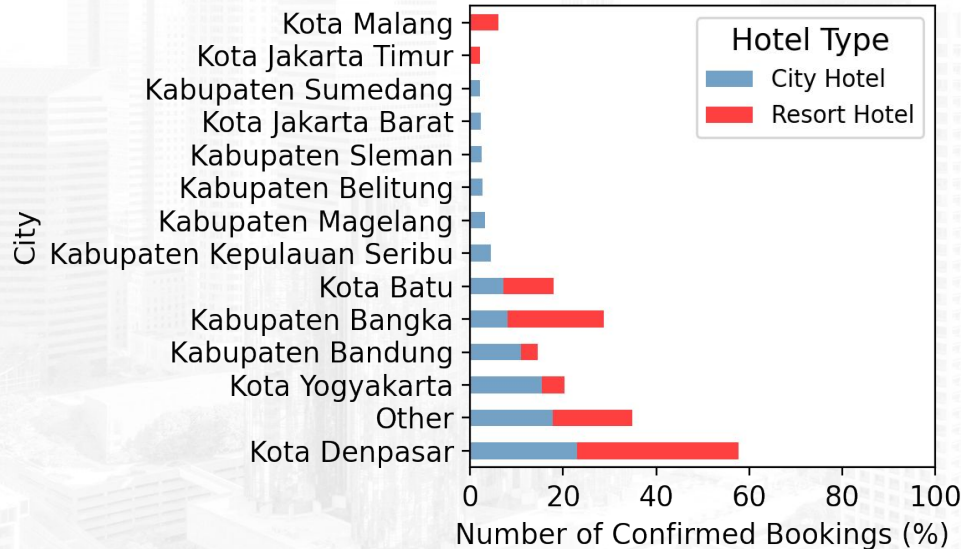
- **Fix incorrect values**

  Replace 'Undefined' value with 'No Meal' as they both have the same meaning.

- **Remove unnecessary data**

  Remove all data with 0 guest and 0 night.

Click here for full documentation

Number of Confirmed Bookings (%) by City

We can see that the majority of the guests were coming from Kota Denpasar. Also, city hotel has more varied guests in terms of their city of origin. The reasons could be that:

- There are more city hotels than resort hotels.
- There are more demand for city hotels than resort hotels.
- City Hotels are normally cheaper than resort hotels.

Click here for full documentation

Market segment analysis

**Number of Confirmed Bookings by Market Segment**
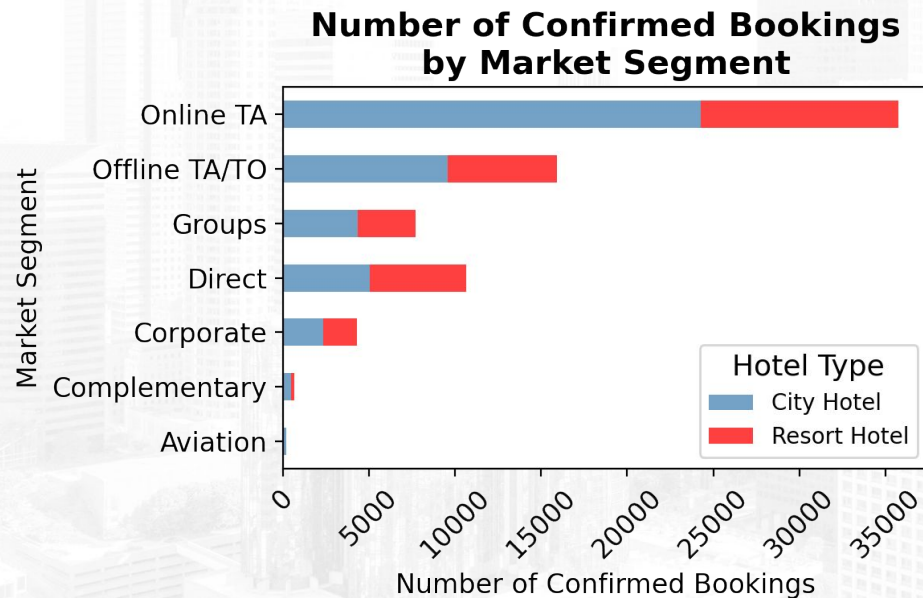
From this output, we can see that online TA has the highest number of confirmed bookings for both hotels. On the other hand, only city hotel that has received booking from aviation. That could be because nowadays book a hotel can easily be done through an app and or a website.

Rakamin
Academy

Room Price Per Night And Person Over The Year
Resort Hotel has the highest room price on October with IDR 84.20, while City Hotel has the
highest price IDR 70.54 on July
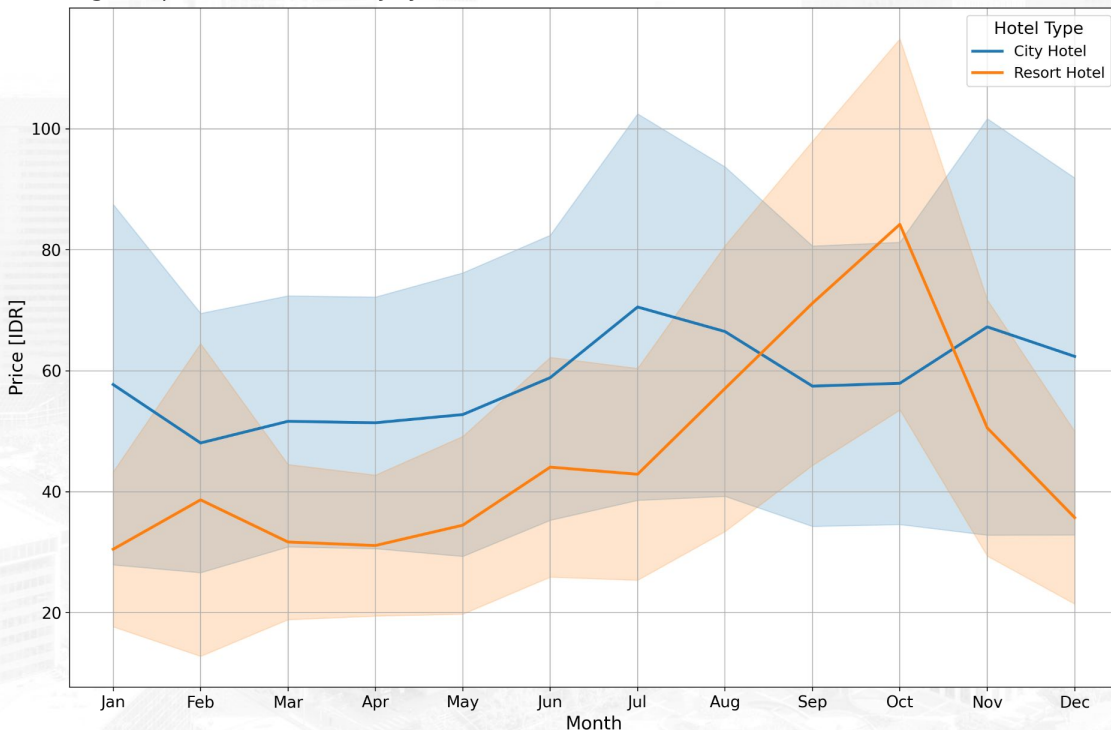
Normally, resort hotel has cheaper price than city hotel except for October. To know what cause this phenomenon we need to take a look at the number of booking which will be done next.

Click here for full documentation

What is the busiest month for the hotel?

What is the Busiest Month for the Hotel?

June is the busiest month for Resort Hotel, while July is the busiest month for City Hotel when the price is also highest
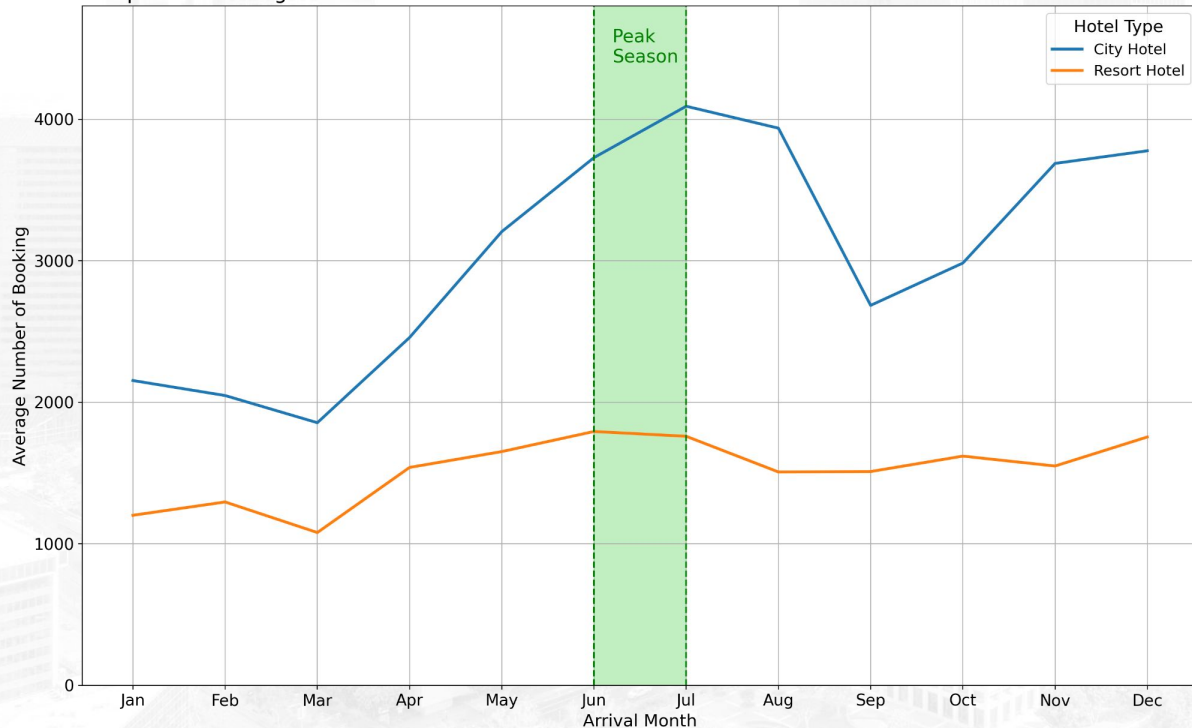

June-July is the holiday season for most schools in Indonesia. Usually there are many families who would go on a family trip at that time. This explains why June-July are the busiest months for both hotels.

As for the price, although the busiest month for resort hotel is June, apparently their highest price is in October.

Click here for full documentation

Rakamin Academy

Positive Trend on Total Nights and Cancellation Rate (%)
The longer the total nights booked, the higher the percentage of booking being canceled
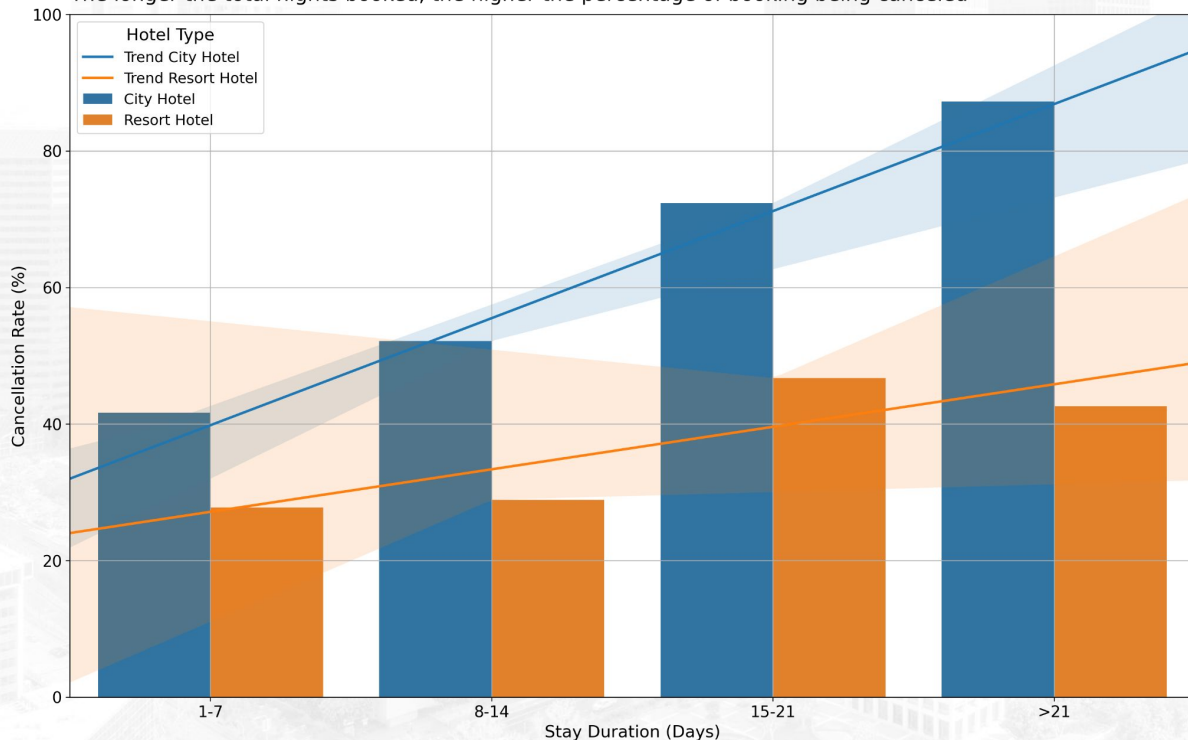

We can see that there is a positive trend on total nights and cancellation rate for both hotels. To prevent this from happening, the hotels should apply a cancellation policy. The longer the total nights booked, the higher the cancellation fee.

Click here for full documentation

Rakamin Academy

Both Hotels Have the Lowest Cancellation Rate within 1 Month of Lead Time
Resort Hotel has quite stagnant (in around 40%) canceled ratio, while City Hotel has quite high canceled ratio (above 70%) with lead time around 11-12 months
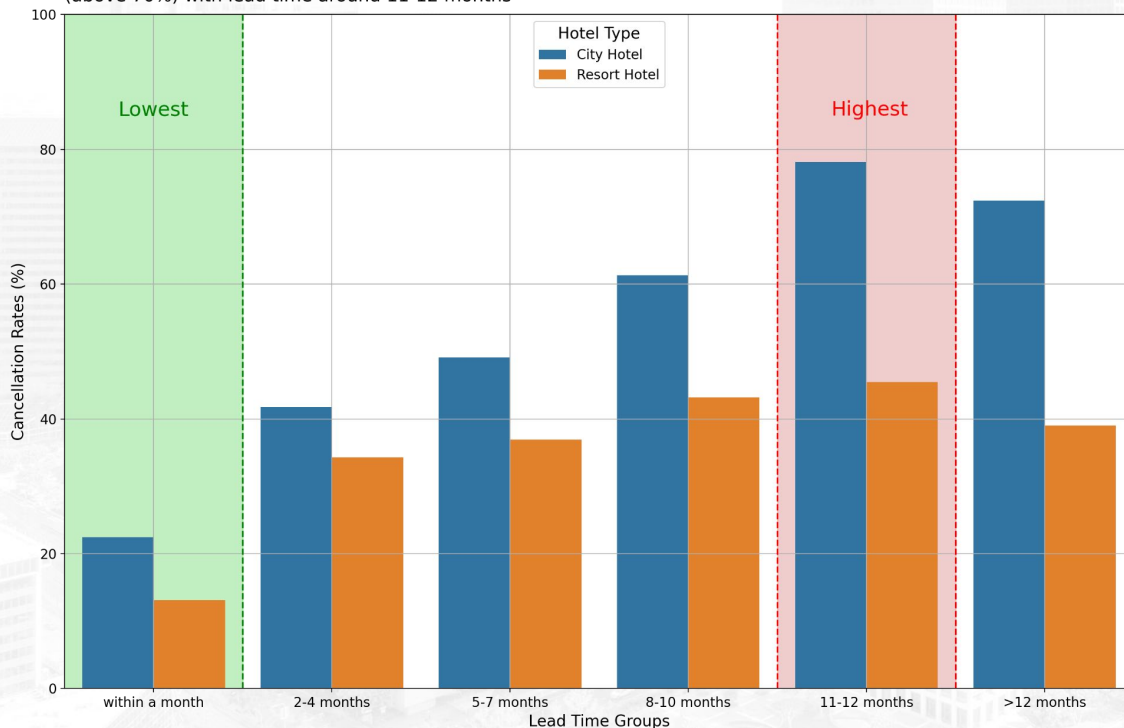
Both resort and city hotel have the highest cancellation rate around 11–12 months lead time. The customers probably forgot they have booked a hotel if the lead time is too long. The hotel could give reminders for them so that they would not cancel the reservation. Also, by applying cancellation policy on every reservation could helps the hotel to prevent this from happening.

Click here for full documentation