INVESTIGATE HOTEL BUSINESS USING DATA VISUALIZATION

Created by:
Putrini Nur Amalina H.
putrininur@gmail.com
linkedin.com/in/putrininur
github.com/putrini

A Diponegoro University graduate who experienced working in the Finance and Accounting Department at a Food Distribution company. A data-driven and tech-savvy person who has huge interest in data analytics who is skilled in SQL, Python, and data visualization using Google Data Studio. Highly skilled in Microsoft Excel and able to actively communicate in English fluently. Currently looking at opportunities in data fields.

Supported by :
Rakamin Academy
Career Acceleration School
www.rakamin.com

OVERVIEW

Data visualization is a powerful tools for data analytics because of the ability to turn large and unorganized data into visual context through plots or graphs that gives clear idea of the information and makes it easier to identify trend, pattern and outlier. Data visualization can help delivering data in most efficient way possible.

In recent years, data analytics have been adopted to many industries, including hospitality industry. To be able to adapt in competitive era, hospitality industry like hotel should be able to manage and analyze data to generate insightful business recommendation that can redefine the way hotel conduct business. Using data visualization can help company to identify which area need to be improved, which factors affect customers satisfaction, etc. Therefore in this paper we will analyze Hotel Business Performance using Data Visualization.

GOAL

The goal is to analyze hotel business performance using business metrics related to monthly booking amount, cancellation rate relation to stay duration and lead time. By these information, we can find patterns that will lead to business recommendation based on data.

TOOLS

Python as
Programming language

Python library used are
- Pandas
- Numpy
- Matplotlib
- Seaborn

Google colab as
notebook

EXPLORATORY DATA ANALYSIS

```
<class 'pandas.core.frame.DataFrame'>
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
```

Insight :

- Data consist of 119390 rows and 29 columns
- There are missing value in `children`, `city`, `agent`, `company` columns
- Data have int64 (16), float64 (4), object(9) data types.

click to show the complete query

DATA PREPROCESSING

HANDLE NULL VALUES

How to handle :

- City = imputation with ‘unknown’ value because the origin city is unknown.
- Children = imputation with 0 value to state the guest have no children
- Agent = imputation with 0 value as undefined value
- Company = imputation with 0 value as undefined value

```
#handle missing value berdasarkan strategi
df['city']=df['city'].fillna('unknown')
df['children']=df['children'].fillna(0)
df['agent']=df['agent'].fillna(0)
df['company']=df['company'].fillna(0)
```

HANDLE UNDEFINED VALUE

There are undefined value in meal column. We have to handle this value to avoid misinterpretation.

How to handle :

- Meal = imputation with ‘No Meal’ value that implies the guest have no meal order.

```
df['meal']=df['meal'].replace(to_replace ='Undefined', value ='No Meal')
```

FEATURE ENGINEERING

- Create new column to define total guest and stay duration value.

```
# make new column : guest and stay_duration
df['guest'] = df['adults'] +
              df['children']+
              df['babies']
df['stay_duration'] = df['stays_in_weekend_nights'] +
                      df['stays_in_weekdays_nights']
```

MONTHLY HOTEL BOOKING ANALYSIS BASED ON HOTEL TYPE

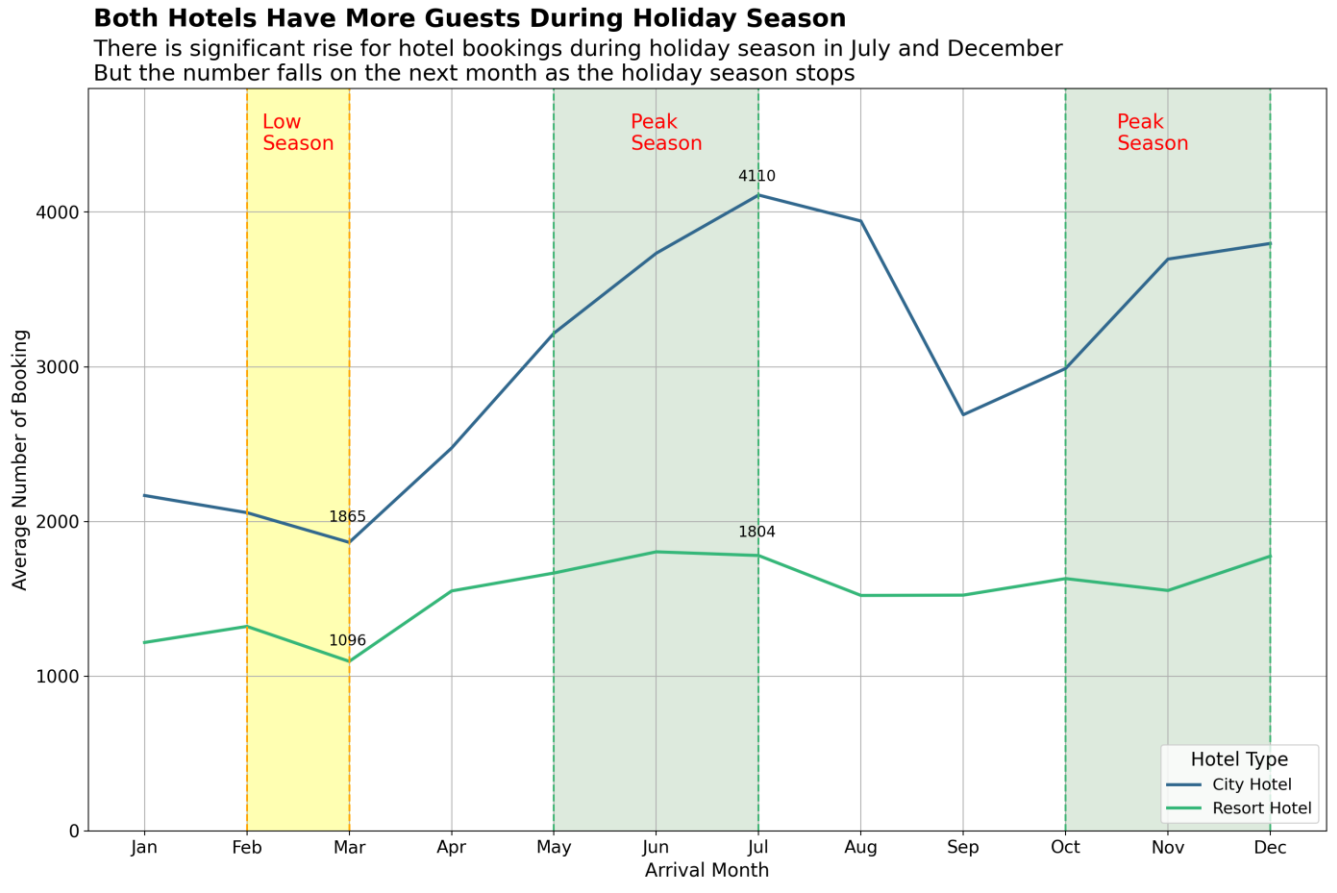
The number of bookings per month represent the traffic of hotel activity in a year. These numbers reflects on the amount of revenue. That's why monthly booking is one of the key goals in hotel industry.

By analyzing monthly hotel booking we can also recognize the peak season and low season as well as the customer behavior during key times of the year.

Objectives :

- Make predictions for forecasting the expectation in terms of demand for accommodation in the hotels and the best price-value ratio for their guests company.
- Develop sales forecast for strategic planning to estimate future demand for products and services
- Help business to operate more efficient by managing money.

MONTHLY HOTEL BOOKING ANALYSIS BASED ON HOTEL TYPE



- Peak season occurs in the middle of the year which is June and July where there is significant increase in average hotel booking for both hotel type. This would probably impacted by the holiday season.
- There is also an increasing amount of hotel bookings in November – December for both hotel type. This also probably impacted by the holiday season which is christmas and New year holiday.
- But the number falls on the next month which is on March and September.
- The lowest rate of the year occurs in March for both hotel type.

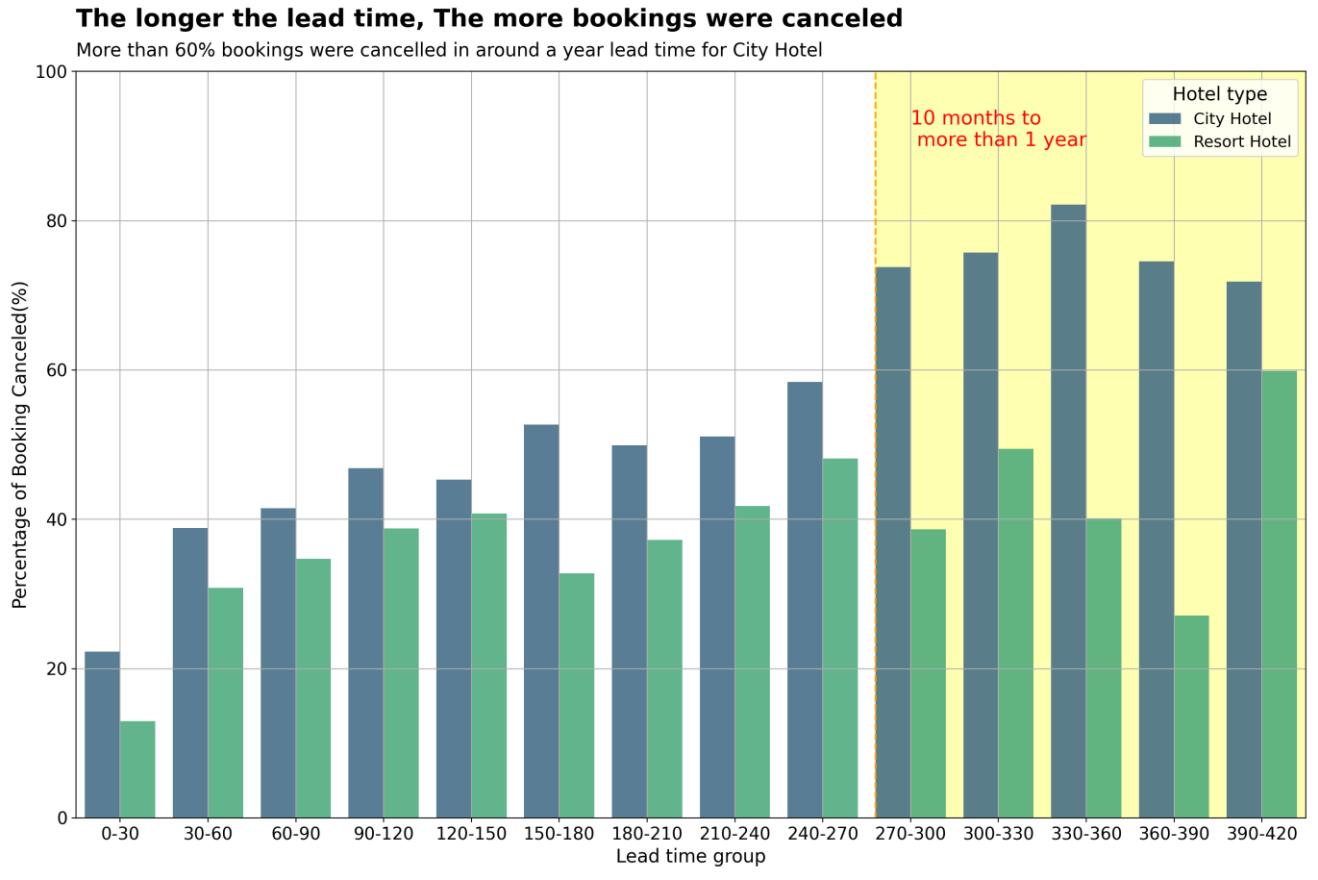
ANALYSIS ON HOTEL BOOKINGS CANCELLATION RATES

According to D-Edge Hospitality Solutions, part of the Accor-owned hotel technology group, global hotel cancellation rates on bookings have reached 40% on average. Cancellations can have a bad effect on the hotels involved. A loss of income occurs as a result of unsold rooms and no-shows. A no-show is a cancellation without notice. It is thus imperative from a cost-saving perspective to find out what causes hotel booking cancellations to rise, and how to mitigate this rise.

Factor of cancellations is the lead days, i.e. the number of days between booking and check-in. Bookings with more lead days have a higher likelihood of being cancelled because the customer having more time to change their mind about the booking or that something else would have happened within that period of lead days that would disrupt travel plans. Or maybe they'll just forget about the booking.

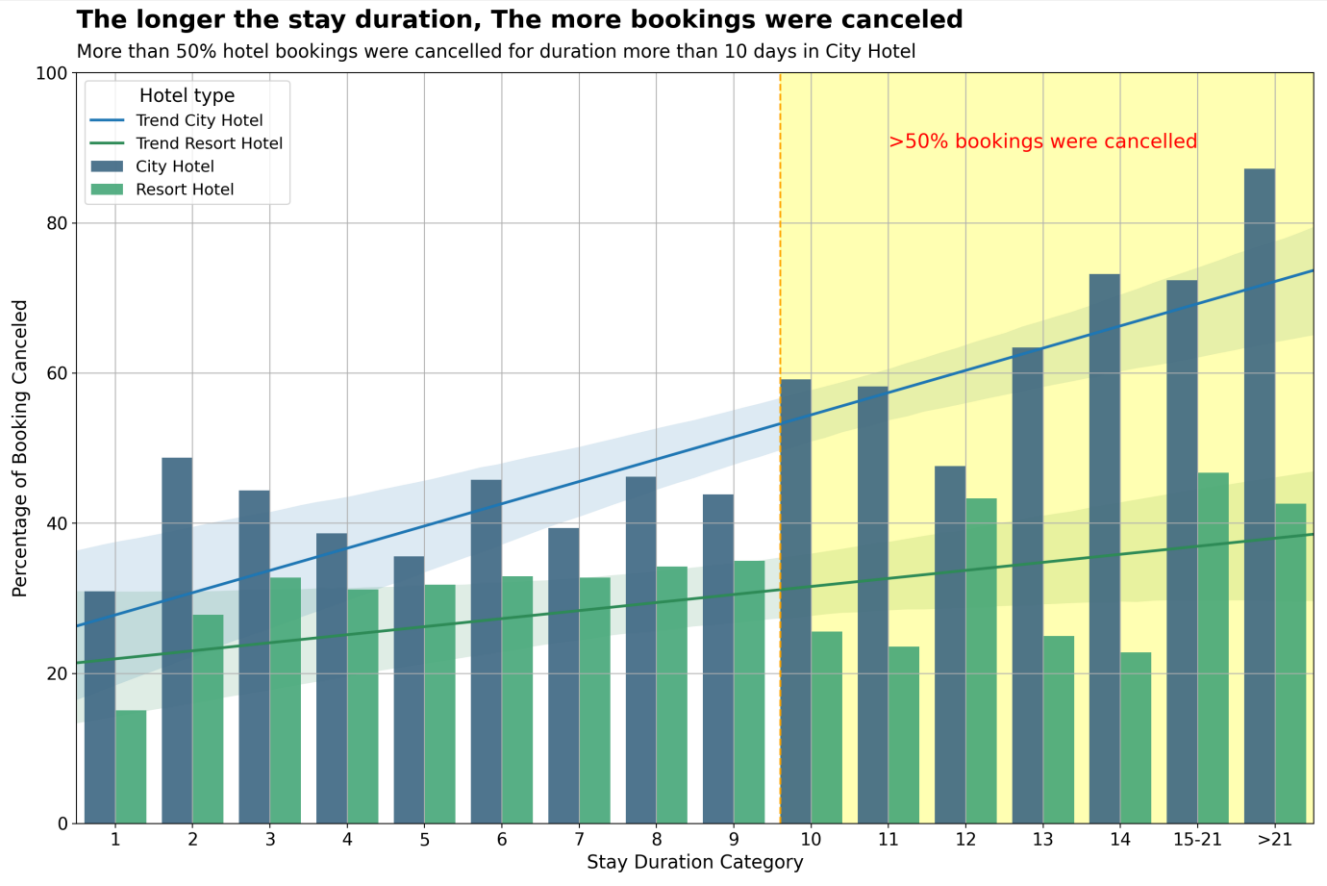
Another factor of cancellation is stay duration. The longer the stay duration, the pricier the hotel bills. Customers probably checking out another alternative as they knew the price they have to pay and this leads to hotel booking cancellation.

ANALYSIS OF LEAD TIME ON HOTEL BOOKINGS CANCELLATION RATE



- There is positive linear relation between cancelled booking and lead time. The more longer the lead time, the more bookings were cancelled.
- City type hotel with lead time more than 270 days (9 months) have 60% probability to be cancelled.

ANALYSIS OF STAY DURATION ON HOTEL BOOKINGS CANCELLATION RATES



- Both hotel types have positive linear relation for stay duration and cancelled booking. But the relation is more significant for city hotel type.
- City hotel type have average more than 50% probability to be cancelled when the stay duration is more than 10 days