Investigate Hotel Business using Data Visualization

Created by:

Muhammad Miftah Thaha

muhammadmiftaht@gmail.com
https://www.linkedin.com/in/miftahthaha/
https://github.com/miftahthaha

Miftah is a recent graduate with a Bachelor's degree Electrical Engineer who have interest in Data Analytics and Science and have a strong foundation in statistical modeling, data analysis, and programming. As a Junior Data Scientist, he has experience through his final project in building and implementing machine learning models, analyzing complex data sets, and creating visualizations to communicate insights. He is a fast learner with excellent problem-solving skills and a passion for using data to drive business decisions. In addition, he possess strong communication and collaboration skills, having worked on multiple team projects during his studies. With a drive to excel in his field, Miftah is seeking an opportunity to contribute his skills and knowledge to a dynamic and innovative organization as a Junior Data Scientist.

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

Overview

"Very important for a company to always analyze its business performance. On this occasion, we will delve deeper into the hospitality business. The focus we aim for is to find out how our customers behave when making hotel reservations, and their relationship to the level of hotel reservation cancellations. The results of the insights we find will be presented in the form of visualized data to make them easier to understand and more persuasive."

Dataset

```
<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
```

- **Description**

  Dataset that contains information related to reservations made by guests at a hotel.

- **Shape**

  119390 Row and 29 Columns (Feature)

- **Datatypes**

  Float64 (4 Feature), Int64 (16 Feature), object (9 Feature)

- **Missing Values**

  Company, Agent, City, Children

For more details, you can see all file here and code here

Data Pre-Processing

Rakamin Academy

Data Pre-Processing Step:

1. **Import library** and **load** the dastaset.
2. Do **Descriptive Statistics** to look overview of the dataset.
3. **Handling Missing Values**.

   The hotel-bookings-data.csv file has missing values for 4 features:
   - company (112593 missing values) changed NaN to 0.
   - agent (16340 missing values) changed NaN to 0
   - city (488 missing values) changed NaN to ‘Unknown’
   - children (4 missing values) changed NaN to 0

1. **Handling Odd Values**.

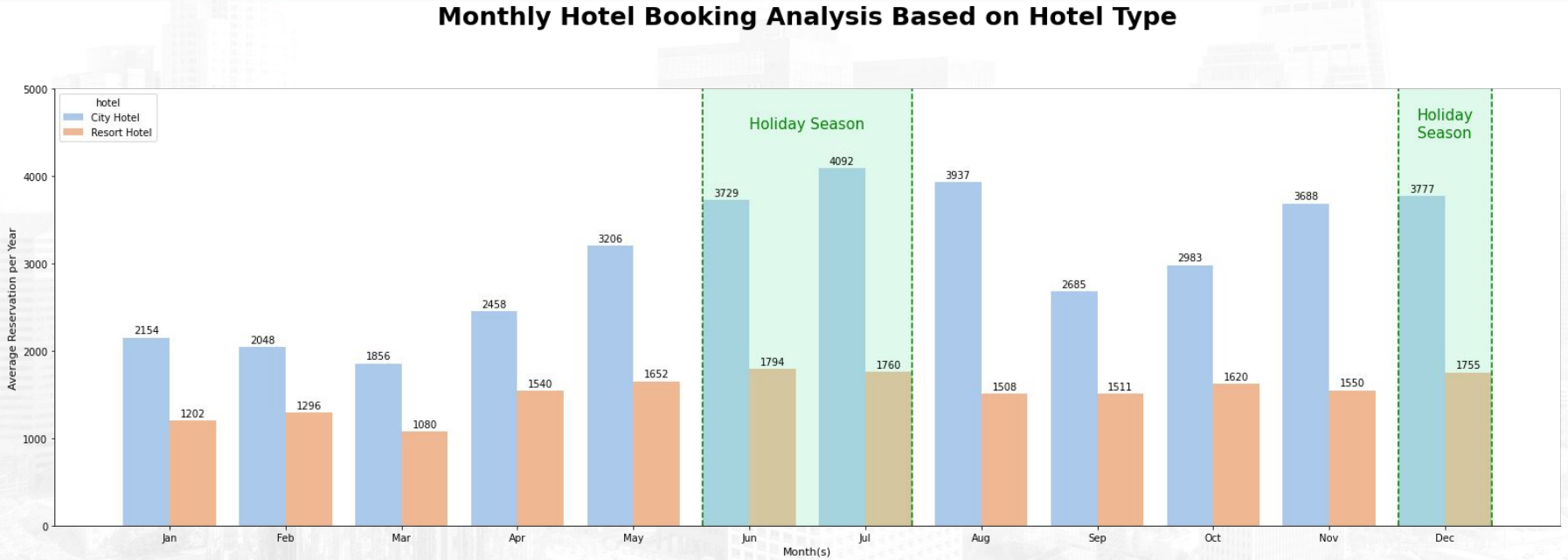
   The hotel-bookings-data.csv file has odd values for 3 features:
   - meal ‘Undefined’ changed to ‘No Meal’
   - market_segment ‘Undefined’ changed to ‘Online TA’
   - distribution_channel ‘Undefined’ changed to ‘TA/TO’

1. **Handling Unnecesary Values**.

   To get actual visitor, we must drop the data who have 0 Client Guest and 0 Occupation Night

For more details, you can see all file here and code here

Monthly Hotel Booking Analysis Based on Hotel Type



Based on the provided data from Monthly Hotel Booking Analysis Based on Hotel Type, here are some possible interpretations:

For more details, you can see all file here and code here

Monthly Hotel Booking Analysis Based on Hotel Type

1. The highest total number of reservations for both hotel types occurred in July (4029 City Hotel + 1760 Resort), followed closely by June (3729 City Hotel + 1974 Resort), and December(3777 City Hotel + 1755 Resort). This is because June, July, and December are the months of long holidays or school breaks (Holiday Season) in Indonesia, and customers tend to book hotels for recreational or tourism purposes during this period.
2. During the months of August to September and January to March, there is a considerable decrease in hotel bookings as customers have typically finished their long holiday or school break.

These are just a few possible interpretations that could be made from the given data. Further analysis could be done to explore these trends in more detail and identify potential reasons for these patterns.

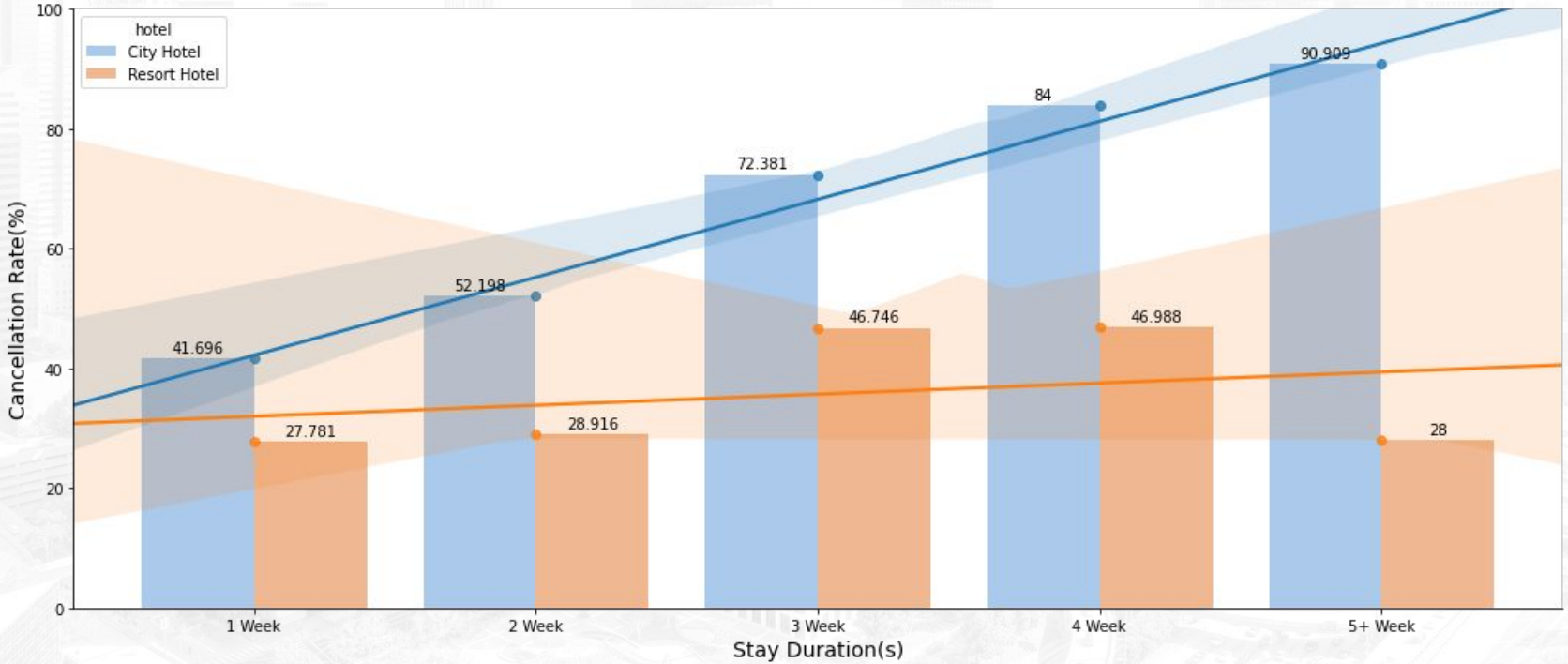
For more details, you can see all file here and code here

Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates

Rakamin Academy

The Cancelation Ratio of Bookings Trends to The Duration of Stay For Each Hotel Type: A Comparative Analysis

The rate of hotel booking cancellations tends to increase as the duration of stay increases.
Additionally, the trendline for City Hotels has a steeper slope compared to Resort Hotels.
The highest rate of canceled hotel bookings for City Hotels was for stays of 5 or more weeks (90.909%),
while for Resort Hotels it was for stays of 4 weeks (46.988%).



For more details, you can see all file here and code here

Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates

Rakamin Academy

Based on the provided data from Compare Cancelation Rate with Stay Duration on Hotel Type, here are some possible interpretations:

1. City Hotel have higher cancelation rates across all duration groups as compared to Resort Hotel.
2. In City hotel, cancelation rate increases with the duration of stay, with the highest percentage of 90.909% in the 5+ week duration group.
3. For Resort Hotel, cancelation rate is relatively consistent across all duration groups, with the highest percentage of 46.988% in the 4 week duration group.
4. The highest number of cancellations occurred in City Hotel in the 1 week duration group with 32,387 cancellations, while in Resort Hotel it was also in the 1 week duration group with 9,907 cancellations.
5. Resort Hotel has a lower overall cancel percentage as compared to City hotel.

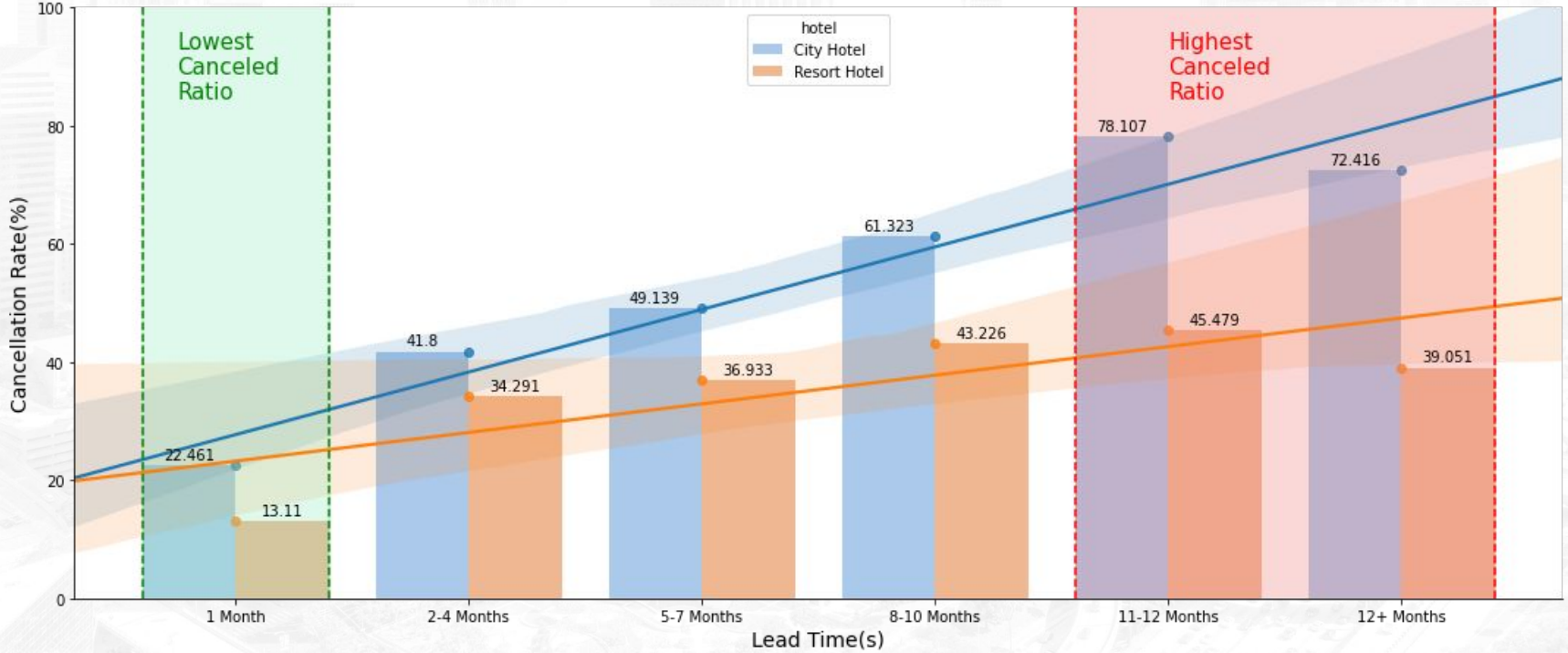
For more details, you can see all file here and code here

Impact Analysis of Lead Time on Hotel Bookings Cancellation Rate

Rakamin Academy

The Cancelation Ratio of Bookings Trends to The Lead Time For Each Hotel Type: A Comparative Analysis

The rate of hotel booking cancellations tends to increase as the lead time increases.
Additionally, the cancelation rate trendline for City Hotels has a slightly steeper slope compared to Resort Hotels.
The highest rate of canceled hotel bookings for City Hotels was for lead time of 11-12 Months (78.107%),
while for Resort Hotels it was for lead time of 11-12 Months (45.479%).



For more details, you can see all file here and code here

Impact Analysis of Lead Time on Hotel Bookings Cancellation Rate

Rakamin Academy

Based on the provided data from Compare Cancelation Rate with Lead Time on Hotel Type, here are some possible interpretations:

1. The longer the lead time for hotel bookings, the higher the cancellation rate tends to be for both City and Resort Hotels.
2. City Hotels show a more significant increase in cancellation rate as the lead time gets longer, with the highest cancellation rate observed for bookings made 11-12 months in advance (78.107%).
3. Resort Hotels have a lower cancellation rate in general, with the highest rate of 45.479% in the 11-12 months lead time.
4. The highest number of cancellations occurred in City Hotel in the 2-4 Months lead time with 11,783 cancellations, while in Resort Hotel it was also in the 2-4 Months lead time with 3,999 cancellations.
5. The lowest cancelation rate could be that bookings made for a shorter lead time tend to have a lower cancelation rate compared to those made further in advance. For example, the cancelation rate for bookings made for a 1-month lead time was the lowest for both City Hotel (22.461%) and Resort Hotel (13.110%). This could suggest that customers who book closer to their intended travel dates have a higher likelihood of following through with their plans.
6. Overall, these findings suggest that hotels should pay attention to the lead time when managing their booking policies and aim to provide more flexible options for customers who book further in advance.

For more details, you can see all file here and code here

Thank You!

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