

Investigate Business Hotel using Data Visualization

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



Created by:

Muhammad Oky Hariawan

okyhariawan@gmail.com

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

“A bachelor's degree in business management with 2 year's expertise in Customer Service and Workforce Management. A data-driven & tech-savvy person who has huge interest in data analytics and currently learning data science. Skilled at analyse data, keen attention to detail, ability to effectively prioritize and execute tasks in a high-pressure environment. Proficient using python, SQL, Microsoft Power BI, Excel VBA, & other statistical tools.”

“Sangat penting bagi suatu perusahaan untuk selalu menganalisa performa bisnisnya. Pada kesempatan kali ini, kita akan lebih mendalami bisnis dalam bidang perhotelan. Fokus yang kita tuju adalah untuk mengetahui bagaimana perilaku pelanggan kita dalam melakukan pemesanan hotel, dan hubungannya terhadap tingkat pembatalan pemesanan hotel. Hasil dari insight yang kita temukan akan kita sajikan dalam bentuk data visualisasi agar lebih mudah dipahami dan bersifat lebih persuasif.”

From the data descriptive information:

- Untuk kolom `children` karena merupakan data jumlah anak, maka tipe data yang sesuai adalah int64 bukan float64
- Untuk kolom `agent` dan `company` karena merupakan data ID, maka tipe data yang sesuai adalah int64 bukan float64

Then we can see that:

- Dalam dataframe ini memiliki 119390 baris dan 29 kolom.
- Dalam dataframe ini memiliki nilai **null** di kolom `children`, `country`, `agent`, `company` dan memiliki persentase yang besar sehingga perlu treatment lebih lanjut.
- Dalam kolom `adults` terdapat nilai 0, maka kita perlu melakukan analisis lebih lanjut dikarenakan tidak mungkin pemesanan hotel dilakukan tanpa orang dewasa
- Kolom `adr` (average daily rate) memiliki nilai negatif, kita perlu melakukan analisis lebih lanjut dikarenakan `adr` tidak mungkin memiliki nilai negative
- Kolom `adr` juga memiliki nilai yang sangat tinggi dengan 5400

▼ Handle missing values

```
[ ] data_missing_value = df_clean.isnull().sum().reset_index()
data_missing_value.columns = ['feature', 'missing_value']
data_missing_value['percentage'] = round((data_missing_value['missing_value']/len(df_clean))*100,3)
data_missing_value = data_missing_value.sort_values('percentage', ascending=False).reset_index(drop=True)
data_missing_value = data_missing_value[data_missing_value['percentage']>0]
data_missing_value
```

	feature	missing_value	percentage
0	company	112593	94.307
1	agent	16340	13.686
2	city	488	0.409
3	children	4	0.003

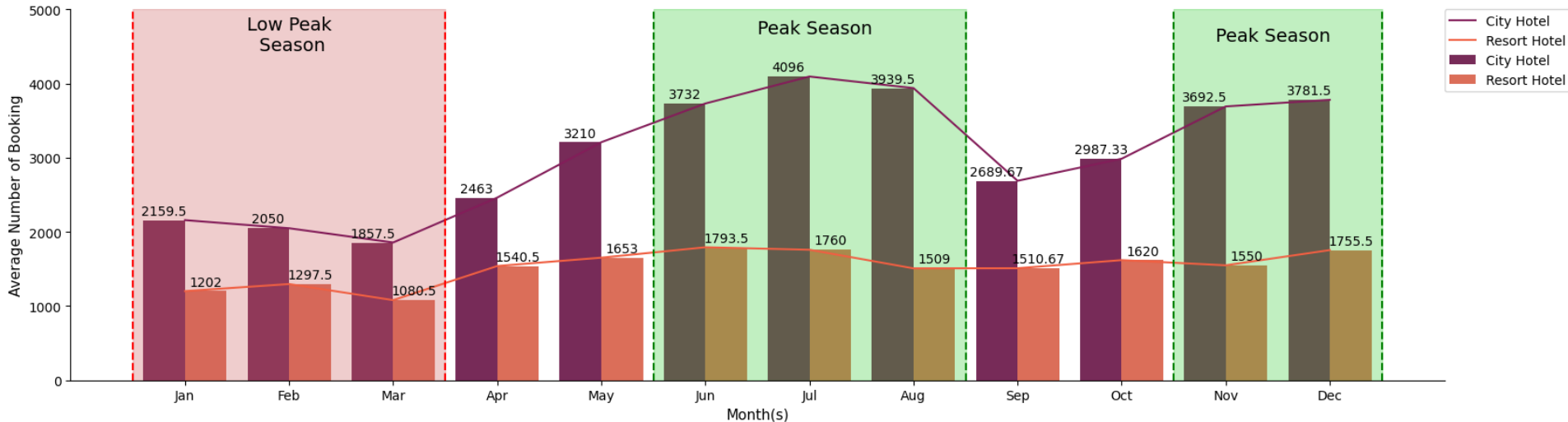
Treatment:

- Kolom `children` karena hanya memiliki 4 baris data yang kosong maka missing value tersebut akan di hapus
- Dikarenakan ada 333 nilai unique dalam Kolom `agent` dan imputasi akan menjadi langkah yang ceroboh maka feature ini akan di drop
- Kolom `company` karena memiliki persentase missing value (94%) maka feature `company` dapat di drop
- Kolom `city` karena sebagian besar nilainya adalah Kota Denpasar dan missing value rasionya kecil, maka missing value tersebut akan diisi dengan nilai Kota Denpasar

Monthly Hotel Booking Analysis Based on Hotel Type

Average Number of Hotel Bookings per Month Based on Hotel Type

June and July are peak season for hotel bookings. City hotel reached highest average number of hotel bookings on July and resort hotel on June. There're another growth on average number of hotel bookings in December for city hotel and resort hotel. It could be cause by christmas and new year's eve holiday.



Key Takeaways:

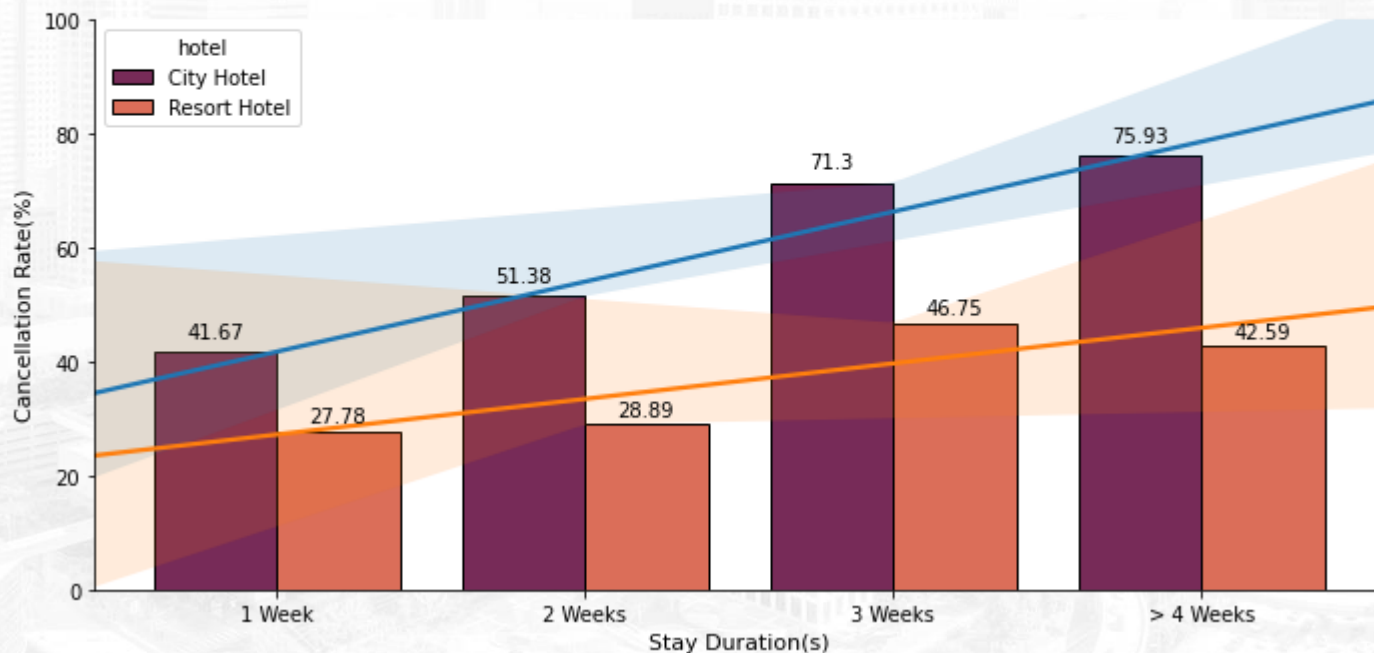
- **High peak season** terjadi pada bulan Juni - Agustus untuk city & resort hotel. Ini dapat terjadi dikarenakan rentang bulan ini adalah libur semester ganjil untuk siswa di Indonesia, rentang yang panjang ini dapat digunakan untuk liburan.
- Terlihat pula **High peak season** terjadi pada bulan November – Desember yang dapat terjadi karena memasuki musim libur hari natal dan tahun baru.
- Kemudian **Low peak season** terjadi pada bulan Januari - Maret. Ini dapat terjadi karena rentang ini adalah libur semester ganjil dengan waktu libur yang singkat.
- Untuk dapat mengoptimalkan pemesanan kamar hotel, maka pihak hotel dapat menerapkan promo tahun baru sehingga pada low peak season dapat lebih optimal.

Cancellation Rate of Hotel Bookings per Stay Duration Based on Hotel Type

The longer customer stayed, the higher the percentage of booking being canceled

The most canceled hotel bookings on city hotel was more than four weeks stay duration (75.93%).

The most canceled hotel bookings on resort hotel was on three weeks stay duration (46.75%).



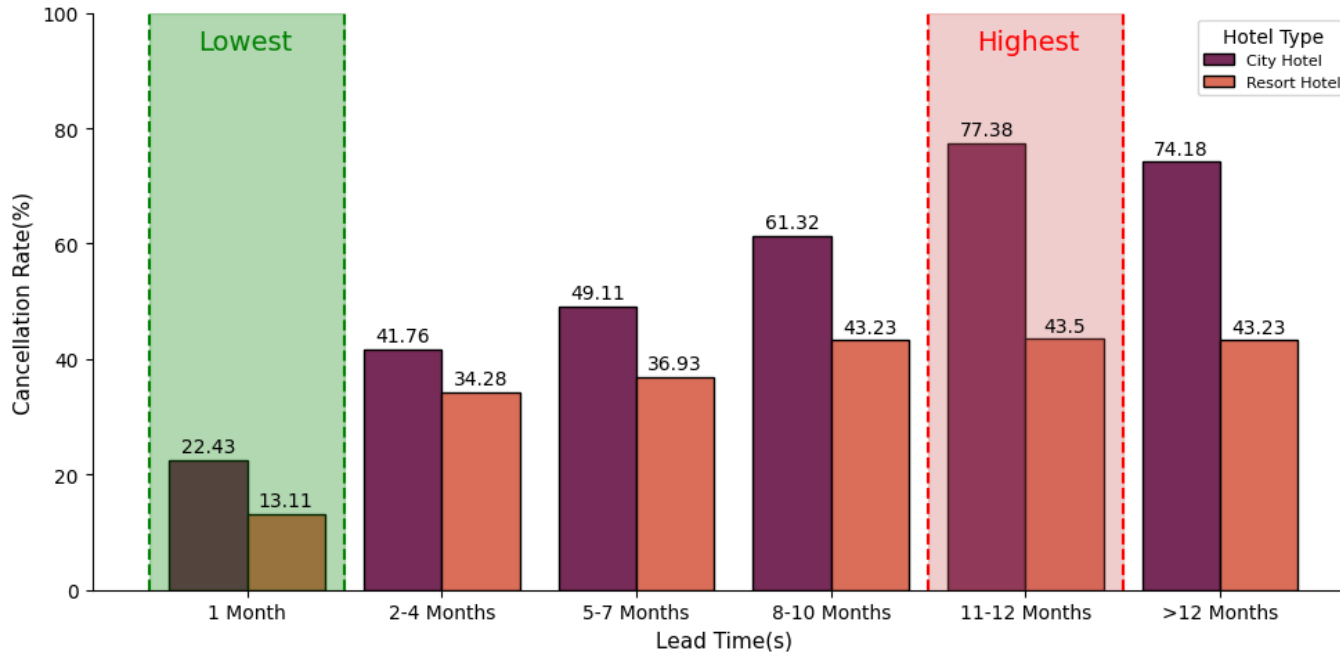
Key Takeaways:

- Terlihat bila angka cancellation di dominasi oleh city hotel.
- Secara umum semakin lama durasi menginap maka angka cancellation akan semakin meningkat untuk kedua tipe hotel. Kecuali untuk minggu ke 5 pada resort hotel (dapat dilakukan analisa lebih lanjut)
- Untuk dapat mengurangi angka cancellation rate pada kedua tipe hotel tampaknya perlu ada kebijakan pembayaran non-refund untuk customer dengan waktu menginap lebih dari 4 minggu
- Selain itu dapat pula dilakukan kebijakan pembayaran deposit atau pembayran dimuka untuk tamu yang menginap lebih dari 4 minggu guna mengurangi angka cancellation.

Impact Analysis of Lead Time on Hotel Bookings Cancellation Rate

Cancellation Rate of Hotel Bookings per Lead Time Based on Hotel Type

*Both hotel types has lowest cancellation rate on 1 month lead time
and highest cancellation rate on 11-12 months lead time.*



Key Takeaways:

- Secara umum semakin lama lead time maka kemungkinan pembatalan pesanan akan semakin tinggi
- Guna mengurangi angka pembatalan pihak hotel dapat menerapkan maksimum lead time & deposit non refund. Menggabungkan kebijakan jenis deposit non refund dan maksimum lead time dapat membantu hotel untuk mendapatkan customer tanpa potensi pembatalan.
- Hotel dapat menerapkan deposit non refund untuk pemesanan lebih dari 210 hari atau diatas 8 bulan untuk city hotel dikarenakan pada visualisasi terlihat tingkat pembatalan sudah lebih dari 50% dan lebih riskan akan terjadinya pembatalan.