

Investigate Hotel Business Using Data Visualization

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



Created by:
Limatan Luviar
Limatan.junior@gmail.com
[Limatan Luviar](#)

Graduated from Universitas Sriwijaya majoring in Informatics Management in 2018 and interned at a government institution. Enrolling in and completing the Data Science Bootcamp at Rakamin Academy in 2024 was driven by my interest in data analysis. Currently, I am actively seeking job opportunities and have completed various data-driven projects at Rakamin. My focus is on Data Analyst roles, where I can utilize my analytical skills and business understanding to make a significant impact

“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.”

1. Null Values:

- Kolom 'children' memiliki 4 nilai null (0.003%) dan kolom 'city' memiliki 488 nilai null (0.409%). Untuk kolom 'children', nilai null bisa didrop saja. Sedangkan untuk kolom 'city', mengisi nilai null dengan nilai yang paling sering muncul (mode).
- Kolom 'agent' memiliki 16,340 nilai null (13.686%) dan kolom 'company' memiliki 112,593 nilai null (94.307%). Kedua kolom ini memiliki persentase nilai null yang cukup tinggi. Untuk kolom 'agent', jumlah nilai unique sebanyak 333. Hal ini membuat menggantikan missing values dengan nilai rata-rata atau median tidaklah cocok, karena hal itu dapat mengaburkan informasi yang ada dalam kolom 'agent'. Sebagai gantinya, Anda dapat melakukan drop pada baris-baris yang memiliki nilai hilang tersebut agar tidak mengganggu analisis yang akan dilakukan pada kolom 'agent'. Sedangkan untuk kolom 'company', karena persentase nilai null yang sangat tinggi, bisa dipertimbangkan untuk menghapus kolom tersebut dari analisis.

2. Kolom dengan Nilai Tidak Normal:

- Kolom 'children' memiliki nilai 10.0 yang tidak masuk akal, karena jumlah anak biasanya tidak mencapai 10. Nilai ini bisa dianggap sebagai outlier dan perlu ditinjau kembali.
- Dalam kolom adults terdapat nilai 0, maka perlu melakukan analisis lebih lanjut dikarenakan tidak mungkin pemesanan hotel dilakukan tanpa orang dewasa, Nilai ini bisa dianggap sebagai outlier dan perlu ditinjau kembali.
- Untuk value "Undefined" dalam kolom "meal", dapat mengganti nilai tersebut dengan nilai yang lebih bermakna atau representatif, seperti "No Meal"
- adr' (Average Daily Rate), terdapat beberapa nilai yang mungkin tidak masuk akal, seperti nilai negatif (-6.38) dan nilai yang terlalu tinggi (misalnya 5400.0).

3. Tipe Data yang tidak sesuai:

- Untuk kolom children karena merupakan data jumlah anak, jumlah anak biasanya dianggap sebagai bilangan bulat tanpa pecahan, sehingga tipe data int mungkin lebih sesuai
- Untuk kolom agent dan company karena merupakan data ID, maka tipe data yang sesuai adalah int

Check Missing Value

```
missing_count = df.isnull().sum()

# Menghitung persentase missing value pada setiap kolom
missing_percentage = (df.isnull().sum() / len(df)) * 100

# Menggabungkan kedua Series ke dalam DataFrame
missing_df = pd.concat([missing_count, missing_percentage], axis=1)
missing_df.columns = ['Jumlah Missing Value', 'Persentase Missing (%)']

# Menampilkan kolom-kolom yang memiliki nilai null beserta persentase missing valuenya
missing_df[missing_df['Jumlah Missing Value'] > 0]
```

✓ 0.5s

	Jumlah Missing Value	Persentase Missing (%)
children	4	0.003350
city	488	0.408744
agent	16340	13.686238
company	112593	94.306893

```
white_space = []
for col in df.columns:
    for val in df[col]:
        if isinstance(val, str) and ' ' in val:
            white_space.append(val)

# Output
print(white_space)
```

✓ 1.7s

[]

Kesimpulan dari check missing value:

1. Ditemukan 4 kolom dengan missing value: 'children' (4 missing values), 'city' (488 missing values), 'agent' (16,340 missing values), dan 'company' (112,593 missing values).
2. Persentase missing value tertinggi dimiliki oleh kolom 'company' dengan 94.31%, diikuti oleh 'agent' dengan 13.69%.
3. Tidak ditemukan nilai yang mengandung whitespace ganda dalam dataset.

Berdasarkan hasil check missing value, berikut adalah saran treatment yang tepat untuk null values:

1. **children:** Karena hanya terdapat 4 missing values pada kolom ini, kita bisa melakukan drop pada missing value.
2. **city:** Karena jumlah missing values cukup signifikan (488), dan mengingat kolom ini berisi informasi tentang kota, kita bisa mempertimbangkan untuk menggantinya dengan nilai yang paling sering muncul (mode).
3. **agent:** Dengan 16,340 missing values dan jumlah nilai yang berbeda sebanyak 333. Hal ini membuat menggantikan missing values dengan nilai rata-rata atau median tidaklah cocok, karena hal itu dapat mengaburkan informasi yang ada dalam kolom 'agent'. Sebagai gantinya, Anda dapat melakukan drop pada baris-baris yang memiliki nilai hilang tersebut agar tidak mengganggu analisis yang akan dilakukan pada kolom 'agent'.
4. **company :** Dengan 112,593 missing values, kemungkinan besar kolom ini tidak akan memberikan informasi yang signifikan. Sebaiknya, menghapus kolom ini dari dataset.

```
df.dropna(subset=['children'], inplace=True)
df['city'].fillna(df['city'].mode()[0], inplace=True)
df.dropna(subset=['agent'], inplace=True)
df.drop('company', axis=1, inplace=True)
```

✓ 0.1s

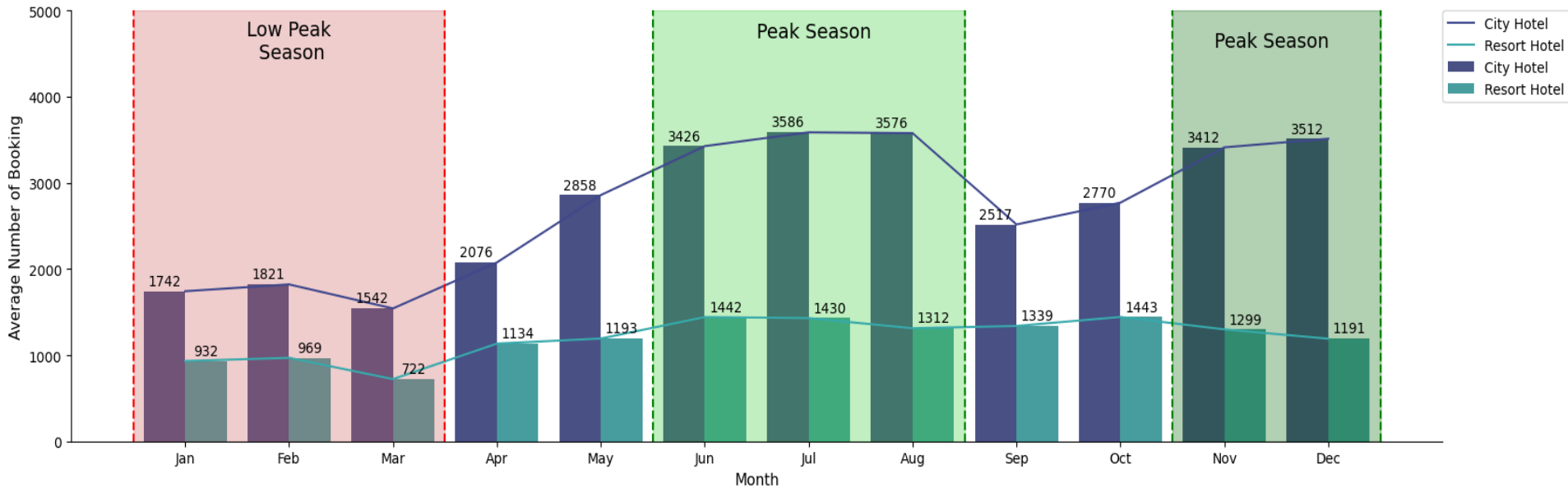
```
print('Apakah Masih ada nilai null?', df.isna().sum().a
```

✓ 0.3s

```
Apakah Masih ada nilai null? False
```


Rata-rata Jumlah Pemesanan Hotel per Bulan Berdasarkan Jenis Hotel

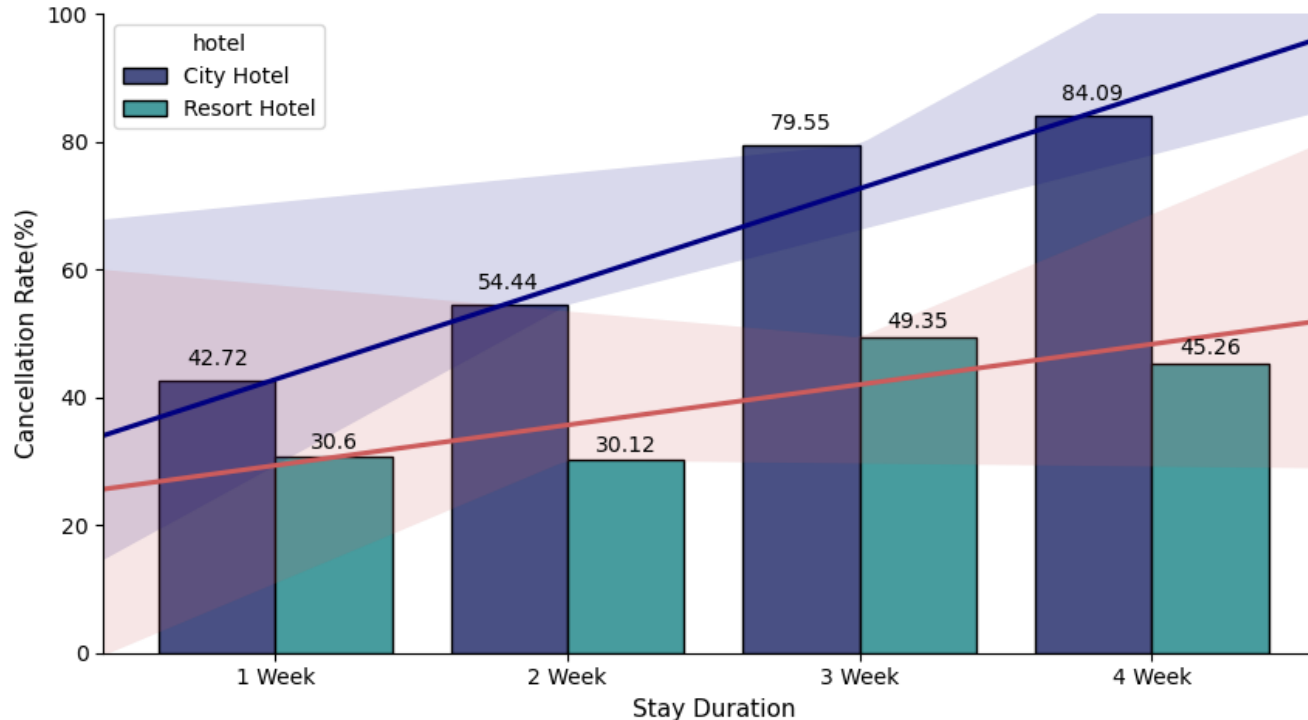
Bulan Juni dan Juli adalah periode tersibuk dalam pemesanan hotel, dengan puncaknya pada bulan Juli untuk hotel kota dan bulan Juni untuk hotel resort. Peningkatan signifikan juga terjadi pada bulan Desember, mungkin disebabkan oleh libur Natal dan Tahun Baru.



1. High peak season untuk hotel kota dan resort terjadi pada bulan Juni hingga Agustus. Hal ini mungkin disebabkan oleh masa liburan sekolah yang panjang di Indonesia, yang memungkinkan untuk liburan yang lebih lama
2. High peak season lainnya terjadi pada bulan November hingga Desember, kemungkinan karena musim liburan Natal dan Tahun Baru.
3. Di sisi lain, Low Peak season terjadi pada bulan Januari hingga Maret, kemungkinan disebabkan oleh pendeknya durasi liburan semester ganjil. Untuk mengoptimalkan pemesanan kamar hotel, hotel dapat menerapkan promosi Tahun Baru untuk meningkatkan pemesanan selama musim sepi.

Tingkat Pembatalan Pemesanan Hotel per Durasi Menginap Berdasarkan Jenis Hotel

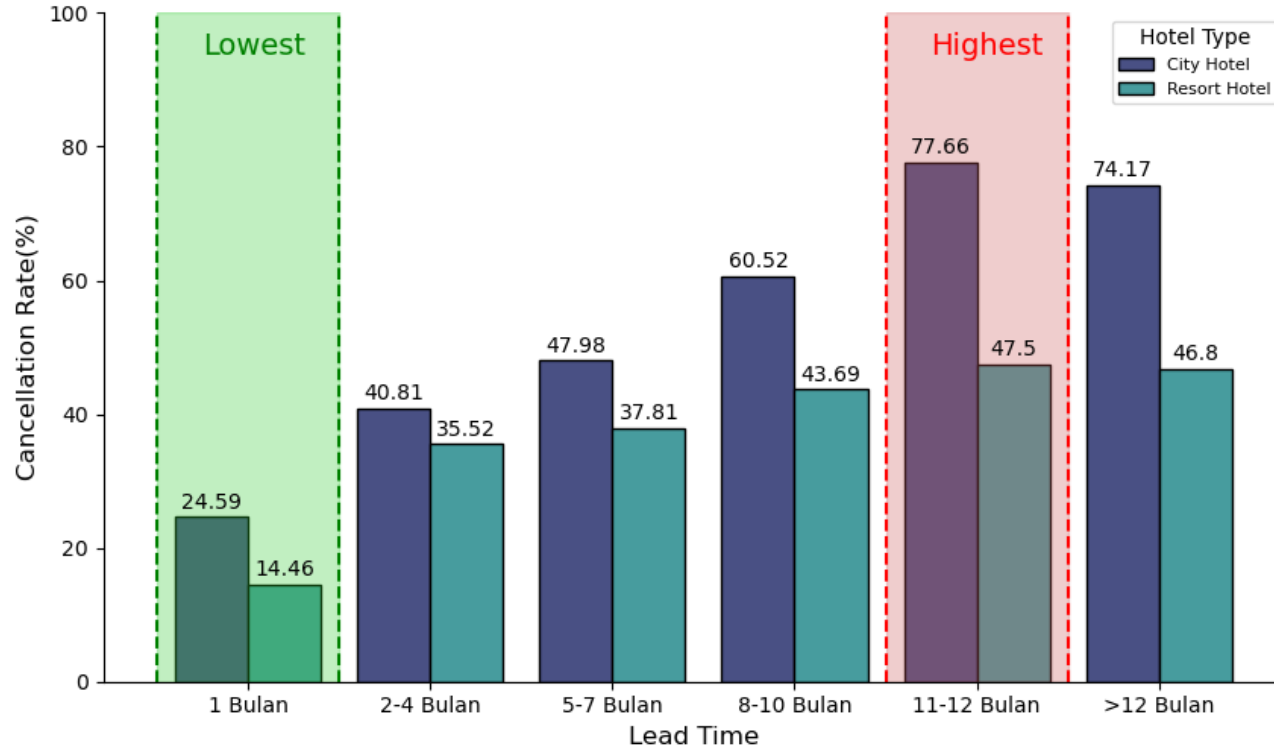
tingkat pembatalan (cancellation rate) tertinggi terjadi pada City Hotel untuk masa menginap 4 minggu (84.09%). diikuti oleh City Hotel untuk masa menginap 3 minggu (79.55%). Sedangkan tingkat pembatalan terendah terjadi pada Resort Hotel untuk masa menginap 4 minggu (45.26%). diikuti oleh Resort Hotel untuk masa menginap 1 minggu (30.6%).



1. Tingkat Pembatalan Tertinggi: City Hotel memiliki tingkat pembatalan yang tinggi, terutama untuk masa menginap 3 dan 4 minggu, dengan persentase pembatalan mencapai 79.55% dan 84.09% secara berturut-turut.
2. Tingkat Pembatalan Terendah: Resort Hotel memiliki tingkat pembatalan yang lebih rendah dibandingkan City Hotel, terutama untuk masa menginap 1 minggu, dengan persentase pembatalan mencapai 30.60%.
3. Rekomendasi: Untuk mengurangi tingkat pembatalan, City Hotel dapat memperbaiki pengalaman pelanggan dengan meningkatkan kualitas layanan dan menawarkan promosi khusus untuk masa menginap yang lebih lama. Sementara itu, Resort Hotel dapat mempertahankan strategi yang telah berhasil untuk menarik pelanggan untuk masa menginap singkat. Selain itu, penerapan kebijakan non-refundable payment untuk tamu yang menginap lebih dari 4 minggu dan kebijakan pembayaran deposit atau pembayaran dimuka dapat membantu mengurangi angka pembatalan pada kedua jenis hotel.

Tingkat Pembatalan Pemesanan Hotel per Lead Time Berdasarkan Jenis Hotel

Tingkat pembatalan paling rendah terjadi pada 2 jenis Hotel untuk tamu dengan lead time kurang dari 1 bulan. tingkat pembatalan tertinggi terjadi pada City Hotel untuk tamu dengan lead time 11-12 bulan



1. Semakin lama lead time (waktu pemesanan), kemungkinan pembatalan pesanan cenderung semakin tinggi.
2. Untuk mengurangi angka pembatalan, hotel dapat menerapkan kebijakan maksimum lead time dan deposit non-refundable.
3. Kombinasi kebijakan deposit non-refundable dan maksimum lead time dapat membantu hotel mendapatkan pelanggan tanpa potensi pembatalan.
4. Hotel dapat mempertimbangkan menerapkan deposit non-refundable untuk pemesanan lebih dari 210 hari atau di atas 8 bulan, terutama untuk City Hotel yang memiliki tingkat pembatalan tinggi untuk lead time 8-12 bulan.