

Investigate Business Hotel using Data Visualization

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



Created by:

Rheza Paleva Uyanto

uyantorheza@gmail.com

<https://www.linkedin.com/in/rheza-uyanto/>

Lulus dari Program Pendidikan Profesi Apoteker Universitas Surabaya pada tahun 2017, memiliki pengalaman praktik kefarmasian di Rumah Sakit selama 4 tahun. Kemampuan komunikasi dan managerial yang baik, dan mampu bekerja sama dalam tim ataupun secara mandiri. Sangat termotivasi dalam bidang mentoring dan pengembangan diri. Mampu dalam penggunaan Microsoft Office. Memiliki ketertarikan dalam bidang Data Analitik, sehingga mengikuti Rakamin Bootcamp Data Science Batch 24. Melalui portofolio ini, saya dapat meningkatkan skill data visualisasi menggunakan Python.

Overview

“Sangat penting bagi suatu perusahaan untuk selalu menganalisa performa bisnisnya. Pada kesempatan kali ini, kita sebagai data analis 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. ”

Exploration Data Analysis (EDA)

Beberapa langkah EDA yang dilakukan :

- Dengan menggunakan google colab, melakukan import library dan dataset.
 - Dataset : hotel_bookings_data.csv
- Melakukan analisis deskriptif
 - Mengetahui struktur dataset : 29 kolom dan 119.390 baris.
 - Tipe data dari tiap kolom : tipe string ada 9 kolom, tipe integer/float : 20 kolom.

#	Column	Non-Null Count	Dtype
0	hotel	119390	object
1	is_canceled	119390	non-null
2	lead_time	119390	int64
3	arrival_date_year	119390	non-null
4	arrival_date_month	119390	non-null
5	arrival_date_week_number	119390	non-null
6	arrival_date_day_of_month	119390	non-null
7	stays_in_weekend_nights	119390	non-null
8	stays_in_weekdays_nights	119390	non-null
9	adults	119390	non-null
10	children	119386	non-null
11	babies	119390	non-null
12	meal	119390	non-null
13	city	118902	non-null
14	market_segment	119390	non-null
15	distribution_channel	119390	non-null
16	is_repeated_guest	119390	non-null
17	previous_cancellations	119390	non-null
18	previous_bookings_not_canceled	119390	non-null
19	booking_changes	119390	non-null
20	deposit_type	119390	non-null
21	agent	103050	non-null
22	company	6797	non-null
23	days_in_waiting_list	119390	non-null
24	customer_type	119390	non-null
25	adr	119390	non-null
26	required_car_parking_spaces	119390	non-null
27	total_of_special_requests	119390	non-null
28	reservation_status	119390	non-null

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Exploration Data Analysis (EDA)

- Melihat Statistic Summary

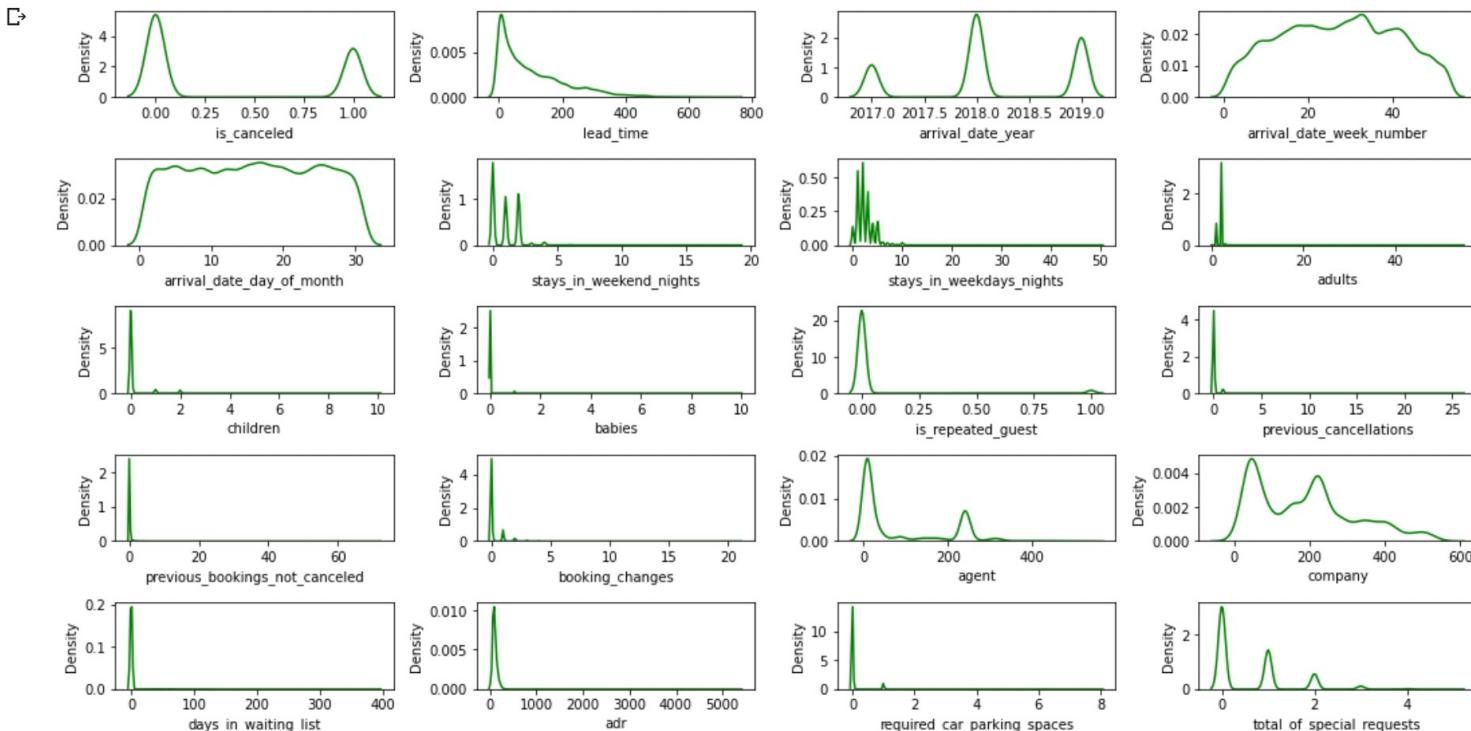
- Untuk Data Numerik menggunakan histogram, untuk memudahkan mengetahui sebaran nilai

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_weekdays_nights	adults	children	babies	is_repeated_guest
count	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119386.000000	119390.000000	119390.000000
mean	0.370416	104.011416	2018.156554	27.165173	15.798241	0.927599	2.500302	1.856403	0.103890	0.007949	0.031912
std	0.482918	106.863097	0.707476	13.605138	8.780829	0.998613	1.908286	0.579261	0.398561	0.097436	0.175767
min	0.000000	0.000000	2017.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	18.000000	2018.000000	16.000000	8.000000	0.000000	1.000000	2.000000	0.000000	0.000000	0.000000
50%	0.000000	69.000000	2018.000000	28.000000	16.000000	1.000000	2.000000	2.000000	0.000000	0.000000	0.000000
75%	1.000000	160.000000	2019.000000	38.000000	23.000000	2.000000	3.000000	2.000000	0.000000	0.000000	0.000000
max	1.000000	737.000000	2019.000000	53.000000	31.000000	19.000000	50.000000	55.000000	10.000000	10.000000	1.000000
	previous_cancellations	previous_bookings_not_canceled	booking_changes	agent	company	days_in_waiting_list	adr	required_car_parking_spaces	total_of_special_requests		
count	119390.000000	119390.000000	119390.000000	103050.000000	6797.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000
mean	0.087118	0.137097	0.221124	86.693382	189.266735	2.321149	101.831122	0.062518	0.571363		
std	0.844336	1.497437	0.652306	110.774548	131.655015	17.594721	50.535790	0.245291	0.792798		
min	0.000000	0.000000	0.000000	1.000000	6.000000	0.000000	-6.380000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	9.000000	62.000000	0.000000	69.290000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000	14.000000	179.000000	0.000000	94.575000	0.000000	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000	229.000000	270.000000	0.000000	126.000000	0.000000	0.000000	1.000000	
max	26.000000	72.000000	21.000000	535.000000	543.000000	391.000000	5400.000000	8.000000	5.000000		

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Exploration Data Analysis (EDA)

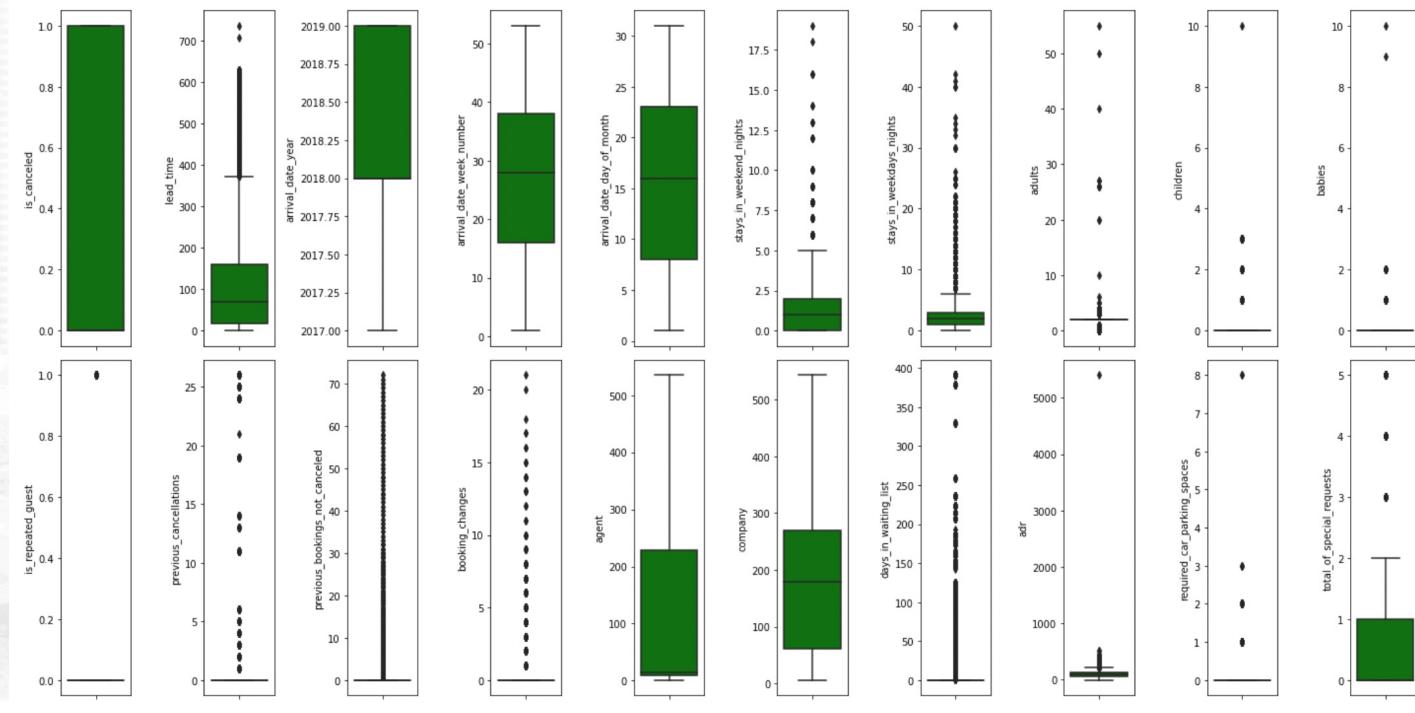
- Melihat Statistic Summary
 - Untuk Data Numerik menggunakan histogram, untuk memudahkan mengetahui sebaran nilai



Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Exploration Data Analysis (EDA)

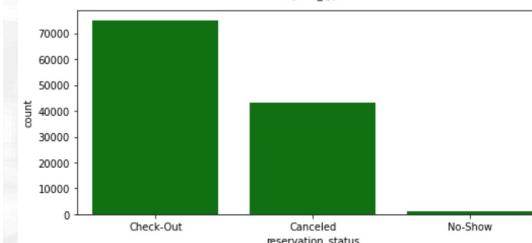
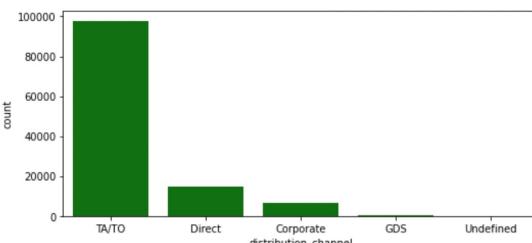
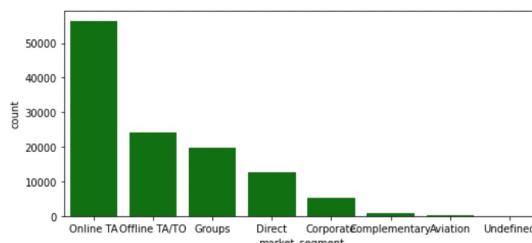
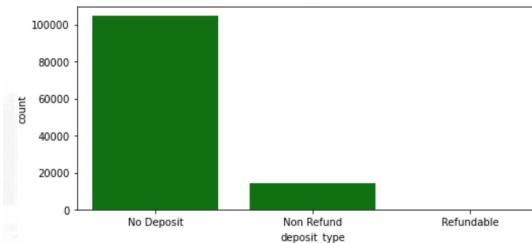
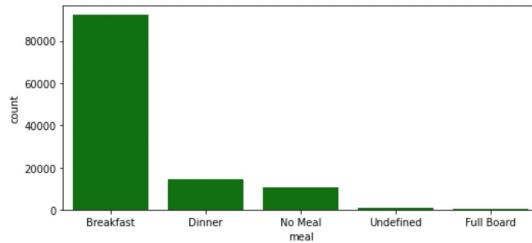
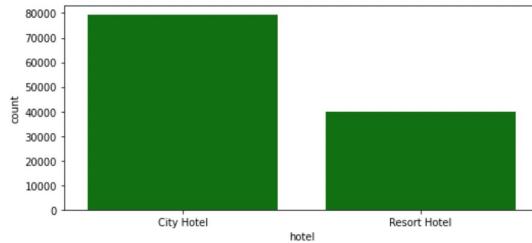
- Melihat Statistic Summary
 - Untuk Data Numerik menggunakan boxplot, untuk memudahkan mengetahui sebaran nilai



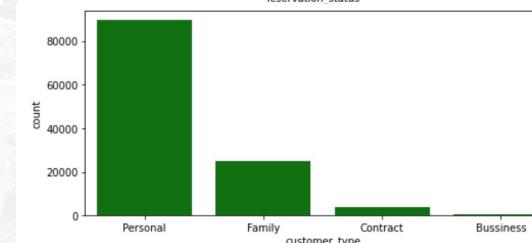
Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Exploration Data Analysis (EDA)

- Melihat Statistic Summary
 - Untuk Data Kategorik menggunakan barplot, untuk memudahkan mengetahui sebaran nilai



	hotel	arrival_date_month	meal	city	market_segment	distribution_channel	deposit_type	customer_type	reservation_status
count	119390	119390	119390	118902	119390	119390	119390	119390	119390
unique	2	12	5	177	8	5	3	4	3
top	City Hotel	October	Breakfast	Kota Denpasar	Online TA	TA/TO	No Deposit	Personal	Check-Out
freq	79330	13877	92310	48590	56477	97870	104641	89613	75166



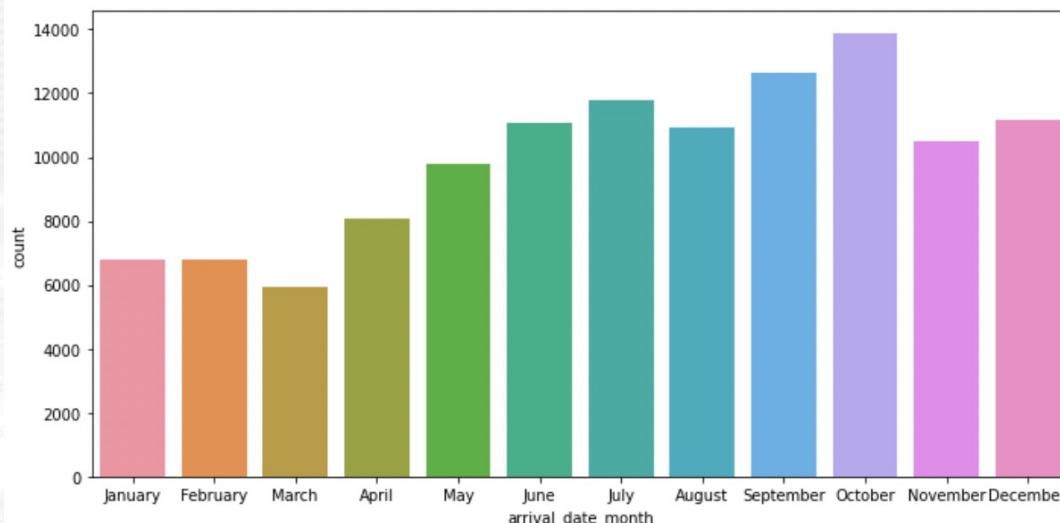
Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Exploration Data Analysis (EDA)

- Melihat Statistic Summary

- Untuk Data Kategorik menggunakan barplot, untuk memudahkan mengetahui sebaran nilai

	hotel	arrival_date_month	meal	city	market_segment	distribution_channel	deposit_type	customer_type	reservation_status
count	119390	119390	119390	118902	119390	119390	119390	119390	119390
unique	2	12	5	177	8		5	3	4
top	City Hotel	October	Breakfast	Kota Denpasar	Online TA		TA/TO	No Deposit	Personal
freq	79330	13877	92310	48590	56477		97870	104641	89613
									75166



Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Data Preprocessing

Beberapa langkah data preprocessing yang dilakukan :

- Null Value
 - Children = 4, diisi modus
 - City = 488, karena kardinalitas tinggi, yakni 177 item, kolom yang Null Value diisi Unknown
- Kolom yang didrop
 - Agent = 16340, kolom tidak bermakna, di drop
 - Company = 112593, kolom tidak bermakna, di drop
- Kolom yang dinormalisasi :
 - Isi kolom meal : No Meal dan Undefined dijadikan satu menjadi No Meal.

df.isna().sum()	
hotel	0
is_canceled	0
lead_time	0
arrival_date_year	0
arrival_date_month	0
arrival_date_week_number	0
arrival_date_day_of_month	0
stays_in_weekend_nights	0
stays_in_weekdays_nights	0
adults	0
children	4
babies	0
meal	0
city	488
market_segment	0
distribution_channel	0
is_repeated_guest	0
previous_cancellations	0
previous_bookings_not_canceled	0
booking_changes	0
deposit_type	0
agent	16340
company	112593
days_in_waiting_list	0
customer_type	0
adr	0
required_car_parking_spaces	0
total_of_special_requests	0
reservation_status	0
dtype: int64	

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Monthly Hotel Booking Analysis Based on Hotel Type

Untuk melakukan analisis booking hotel jenis per bulannya, dilakukan pengelompokan tiap bulannya

	hotel	arrival_date_month	nunique	count		hotel	arrival_date_month	nunique	count
0	City Hotel	April	2	4965	12	Resort Hotel	April	2	3103
1	City Hotel	August	2	7894	13	Resort Hotel	August	2	3045
2	City Hotel	December	2	7605	14	Resort Hotel	December	2	3555
3	City Hotel	February	2	4132	15	Resort Hotel	February	2	2648
4	City Hotel	January	2	4357	16	Resort Hotel	January	2	2437
5	City Hotel	July	2	8232	17	Resort Hotel	July	2	3559
6	City Hotel	June	2	7480	18	Resort Hotel	June	2	3609
7	City Hotel	March	2	3736	19	Resort Hotel	March	2	2193
8	City Hotel	May	2	6458	20	Resort Hotel	May	2	3336
9	City Hotel	November	2	7400	21	Resort Hotel	November	2	3108
10	City Hotel	October	3	8983	22	Resort Hotel	October	3	4894
11	City Hotel	September	3	8088	23	Resort Hotel	September	3	4573

Data bulan Oktober dan September berasal dari tiga tahun, data lainnya hanya dari 2 tahun

Sehingga perlu dilakukan Normalisasi

Monthly Hotel Booking Analysis Based on Hotel Type

Normalisasinya : membuat kolom berisi nilai rata-rata booking perbulan.

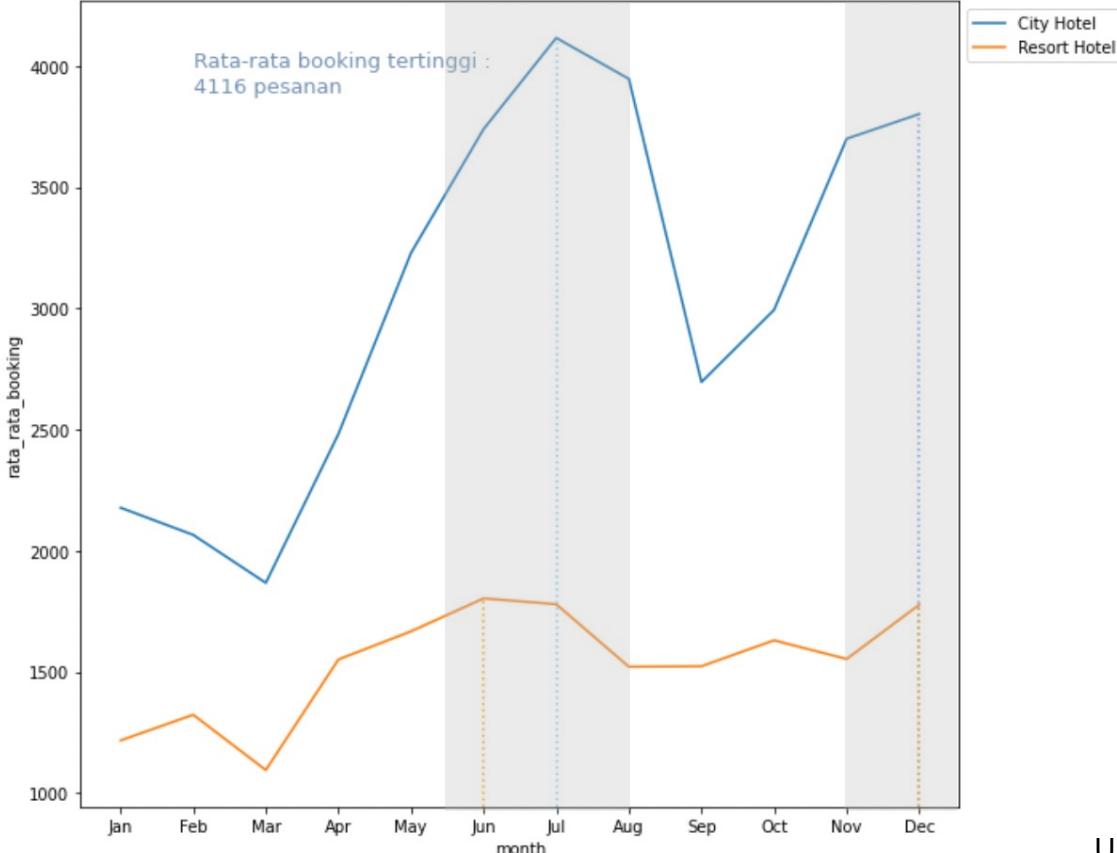
Tabel rata-rata booking hotel perbulannya berdasarkan jenis hotel

	hotel	arrival_date_month	nunique	count	rata_rata_booking		hotel	arrival_date_month	nunique	count	rata_rata_booking
0	City Hotel	April	2	4965	2482.0	12	Resort Hotel	April	2	3103	1552.0
1	City Hotel	August	2	7894	3947.0	13	Resort Hotel	August	2	3045	1522.0
2	City Hotel	December	2	7605	3802.0	14	Resort Hotel	December	2	3555	1778.0
3	City Hotel	February	2	4132	2066.0	15	Resort Hotel	February	2	2648	1324.0
4	City Hotel	January	2	4357	2178.0	16	Resort Hotel	January	2	2437	1218.0
5	City Hotel	July	2	8232	4116.0	17	Resort Hotel	July	2	3559	1780.0
6	City Hotel	June	2	7480	3740.0	18	Resort Hotel	June	2	3609	1804.0
7	City Hotel	March	2	3736	1868.0	19	Resort Hotel	March	2	2193	1096.0
8	City Hotel	May	2	6458	3229.0	20	Resort Hotel	May	2	3336	1668.0
9	City Hotel	November	2	7400	3700.0	21	Resort Hotel	November	2	3108	1554.0
10	City Hotel	October	3	8983	2994.0	22	Resort Hotel	October	3	4894	1631.0
11	City Hotel	September	3	8088	2696.0	23	Resort Hotel	September	3	4573	1524.0

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Monthly Hotel Booking Analysis Based on Hotel Type

Grafik Rata-rata Booking Hotel perbulan Berdasarkan Jenis Hotel



- Kedua jenis hotel mengalami peningkatan jumlah booking pada periode Juni-Juli dan Periode Desember.
- Peningkatan jumlah booking pada periode Juni-Juli berkaitan dengan liburan semesteran anak sekolah, sedangkan Periode Desember berkaitan dengan liburan natal dan tahun baru.

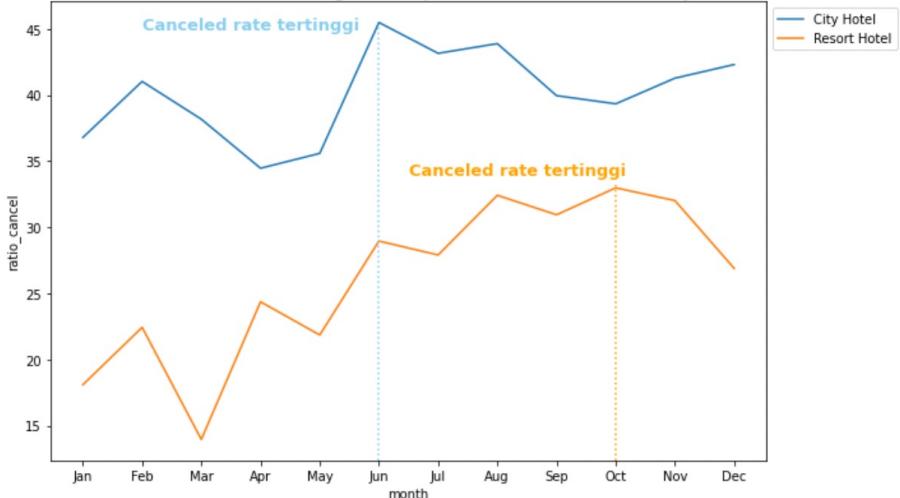
Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates

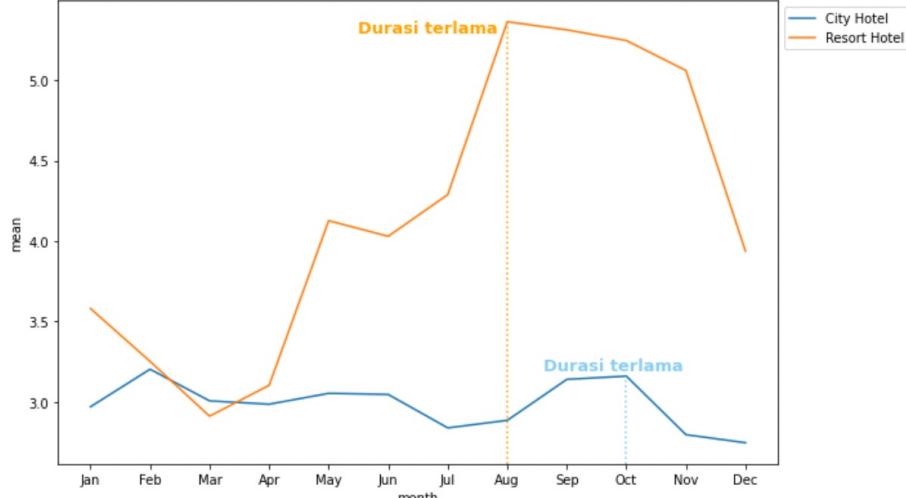
Hal yang dilakukan :

- Melakukan filtering jumlah reservasi yang cancel, digrup berdasarkan bulan dan jenis hotel
- Melakukan agregasi jumlah reservasi per bulan dari masing-masing hotel untuk melihat jumlah total reservasi.
- Melakukan merge, dan agregasi kembali untuk mendapatkan ratio reservasi yang cancel dibandingkan total reservasi.
- Melakukan visualisasi cancelled rate ratio
- Melakukan agregasi, dengan menambahkan jumlah hari stay weekeend + weekday
- Melakukan grouping berdasarkan bulan, dari masing-masing jenis hotel, didapatkan rata-rata jumlah hari stay

Grafik Canceled rate Booking Hotel perbulan Berdasarkan Jenis Hotel



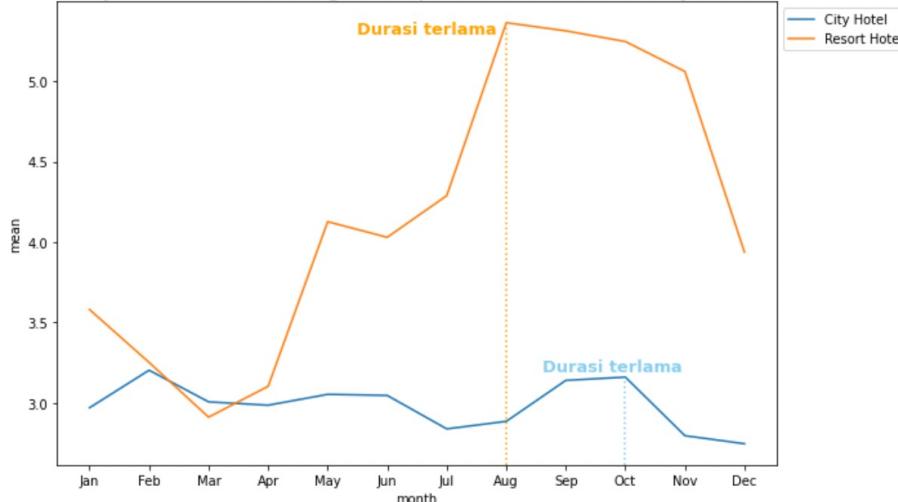
Grafik Jumlah Hari Booking Hotel perbulan Berdasarkan Jenis Hotel



Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates

Grafik Jumlah Hari Booking Hotel perbulan Berdasarkan Jenis Hotel



Grafik Canceled rate Booking Hotel perbulan Berdasarkan Jenis Hotel



Interpretasi :

- Berdasarkan durasi lama menginap : Pada jenis Resort Hotel, pengunjung menginap paling lama 5-6 hari pada bulan Agustus. Durasi tersebut paling lama di antara bulan lainnya. Hal ini berkaitan dengan libur sekolah,
- Sedangkan untuk jenis City Hotel, pengunjung rata-rata menginap selama 3 hari sepanjang tahun. Tidak terjadi peningkatan yang signifikan terkait lama durasi menginap.

- Tetapi perlu diingat, tingkat cancel rate pada hotel Resort juga meningkat pada musim liburan, tercatat tingkat cancel rate pada bulan Agustus (32%) meningkat dibanding bulan-bulan sebelumnya sampai mencapai puncaknya pada bulan Oktober (33%).
- Tingkat cancel rate pada Resort Hotel cenderung berfluktuasi, berbeda dengan City Hotel yang cenderung bertahun di angka 35-45% sepanjang tahunnya. Tetapi perlu dicatat, tingkat cancelled rate City Hotel cukup tinggi.

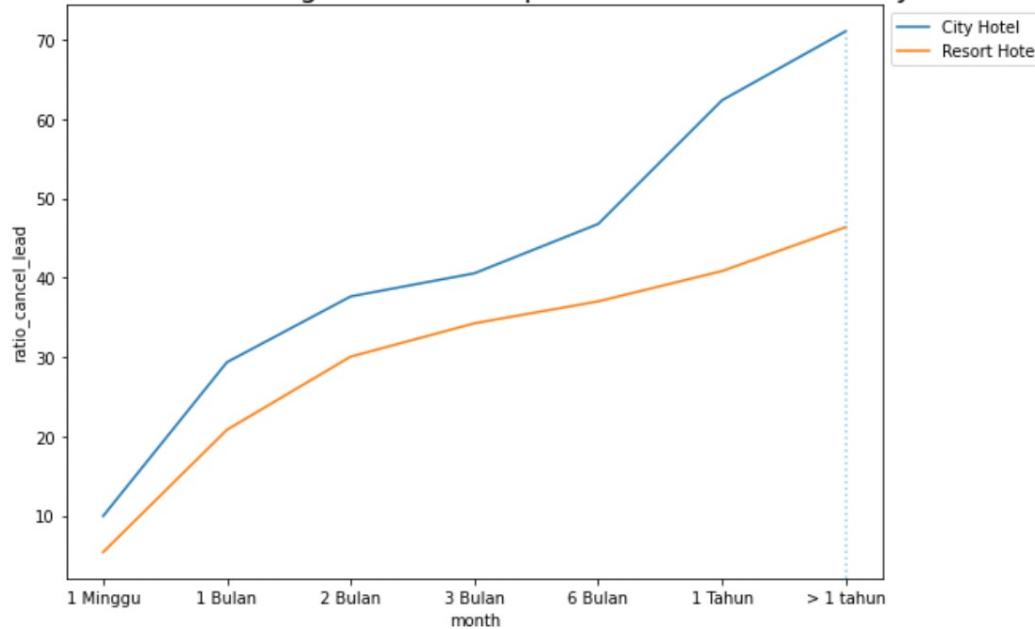
Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Impact Analysis of Lead Time on Hotel Bookings Cancellation Rate

Hal yang sudah dilakukan:

- membuat kategori baru untuk lead time dengan menggunakan np.where
- menghitung jumlah reservasi cancel dan seluruhnya per kategori, dan dari masing-masing jenis hotel
- membuat visualisasi

Grafik Canceled rate Booking Hotel terhadap Lead Time Berdasarkan Jenis Hotel



Interpretasi:

- Dari masing-masing kategori lead time, semakin lama lead time nya semakin tinggi pula ratio cancelnya. Hal ini berlaku untuk kedua jenis hotel.
- Cancel ratio paling tinggi terjadi pada pelanggan dengan lead time > 1 tahun, yakni > dari 70% pada city hotel, dan 45% pada resort hotel.