

Investigate Business Hotel using Data Visualization



Created by:

Devia Febyanti

devia.febyanti99@gmail.com

<https://www.linkedin.com/in/devia-febyanti/>

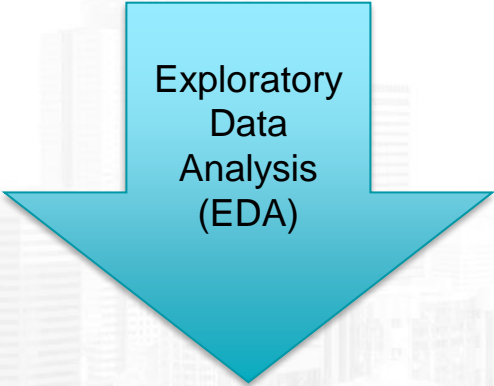
Experienced data operator with proven work history in the field of education. Currently motivated to become a Data Scientist to help driving financial inclusion through the use of big data and machine learning.



Bachelor of Geophysics
2018-2022

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

“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.”



Exploratory Data Analysis (EDA)

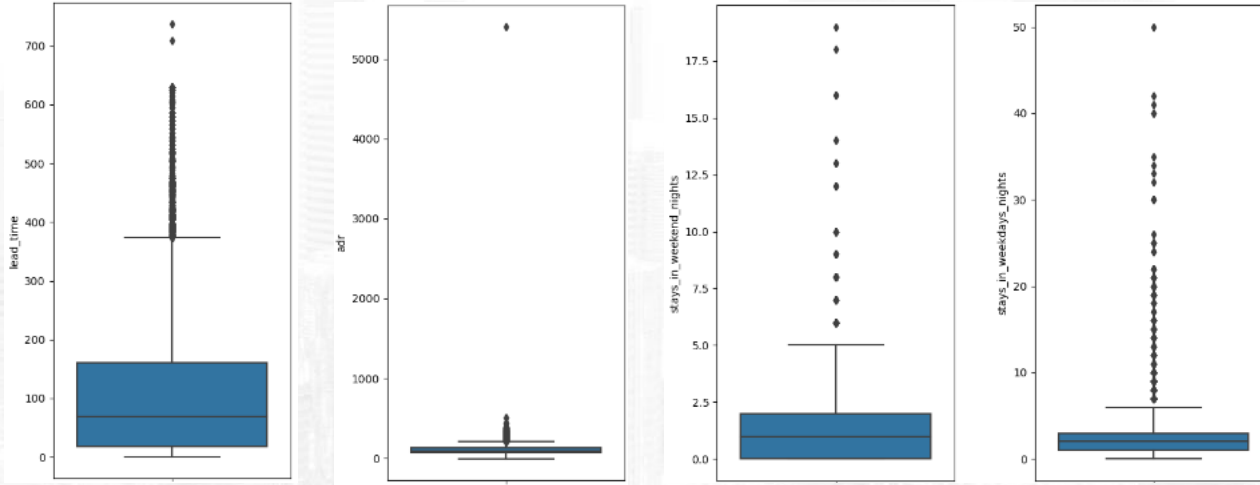
- Memisahkan antara fitur kategorikal dan numerical
- Melakukan analisis statistika deskriptif
- Univariate Analysis
- Multivariate Analysis



Data Preprocessing

- Handling missing value
- Handle duplicated data
- Mengganti value yang tidak sesuai
- Membuang data yang tidak diperlukan

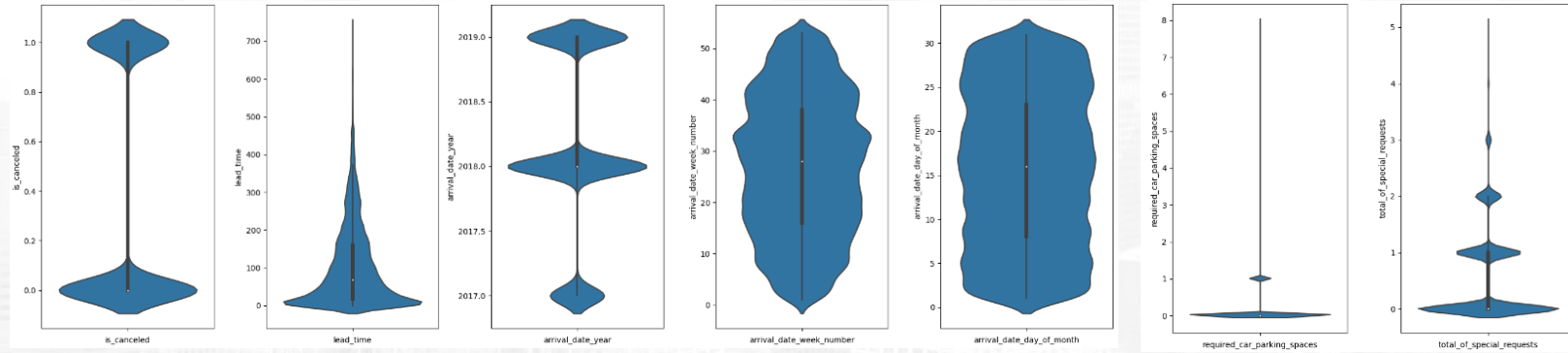
Univariate Analysis



Boxplot

- Diperoleh fitur-fitur yang memiliki outlier

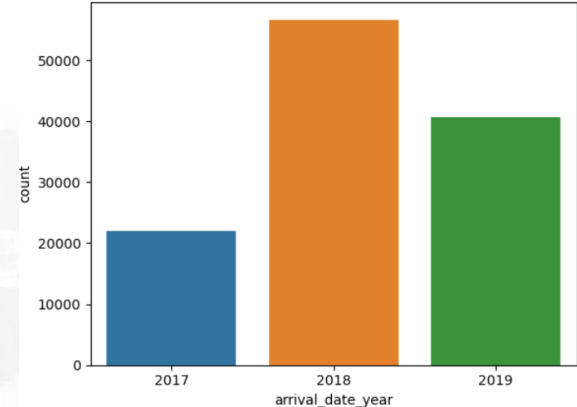
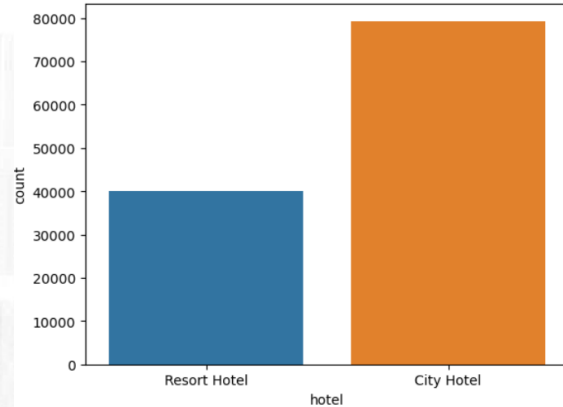
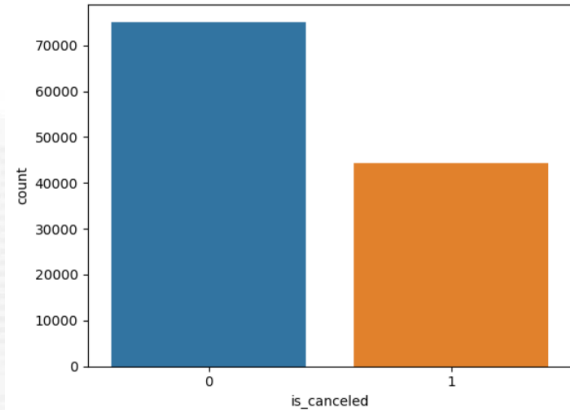
Univariate Analysis



Violin Plot

- `lead_time` (jumlah hari antara saat reservasi dilakukan dan saat kedatangan) di banyak 0 sampai 70 hari.
- `arrival_date_year` (tahun kedatangan) paling banyak di tahun 2018.
- `required_car_parking_spaces` (jumlah tempat parkir yang dibutuhkan oleh tamu) mayoritas 0 atau 1 space parkir.
- `total_of_special_requests` (jumlah permintaan khusus oleh tamu) paling banyak adalah tidak ada permintaan dan 1 permintaan.
- `arrival_date_week_number` (minggu ke berapa dalam tahun kedatangan) dan `arrival_date_day_of_month` (hari kedatangan dalam bulan) tersebar merata.
- Lebihnya banyak memiliki outlier yang harus dihandle.

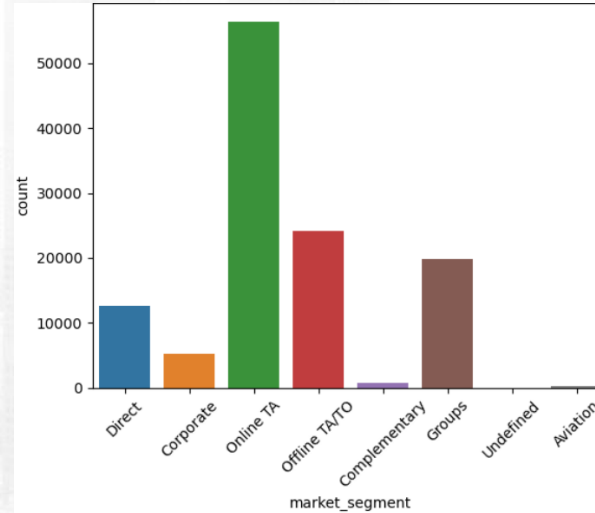
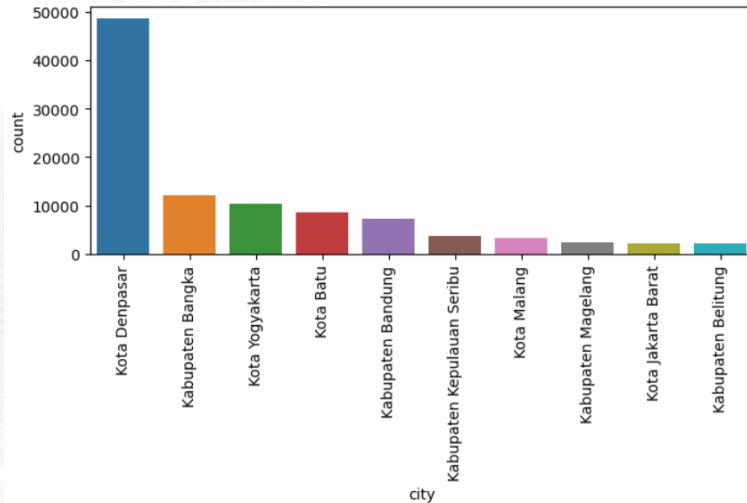
Univariate Analysis



Count Plot

- City Hotel lebih diminati pengunjung daripada Resort Hotel, hampir 2 kali lipat.
- Masih sangat banyak pengunjung yang melakukan pembatalan pemesanan hotel.
- Tahun kedatangan tiga tahun terakhir dari 2017 sampai 2019, paling tinggi pada 2018, diikuti tahun 2019 kemudian 2017.

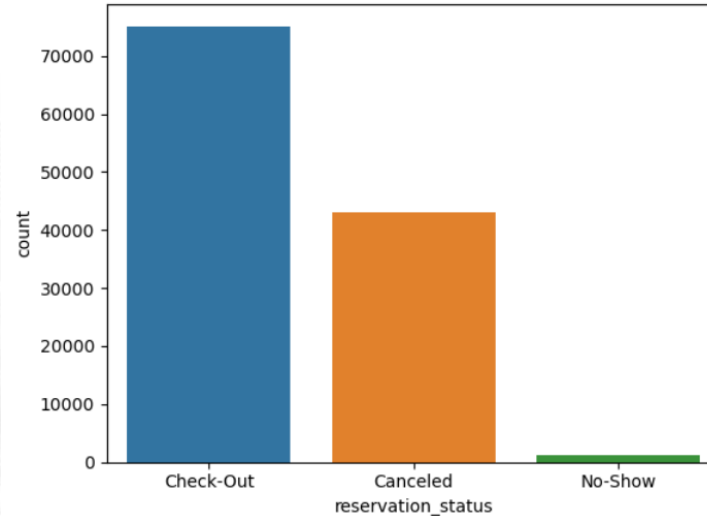
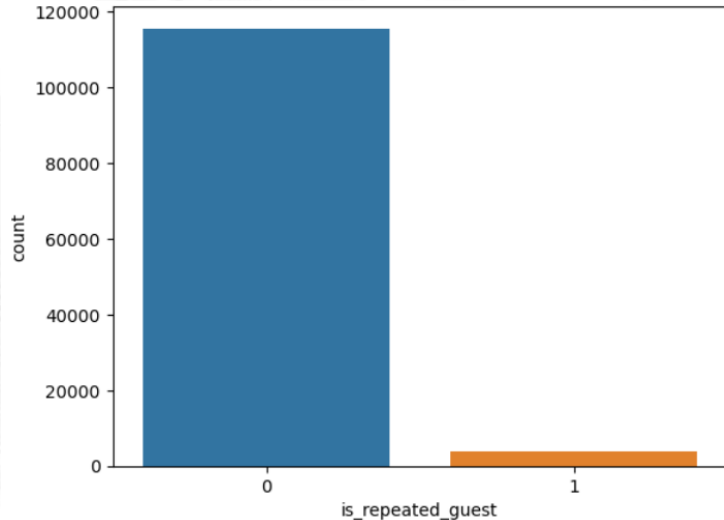
Univariate Analysis



Count Plot

- Hotel di kota Denpasar paling banyak dikunjungi kemungkinan karena lokasinya yang strategis sebagai pusat administrasi dan bisnis di Bali, akses yang baik ke Bandara Internasional Ngurah Rai, dan fasilitas yang lengkap seperti pusat perbelanjaan dan rumah sakit. Hotel paling banyak dikunjungi kedua adalah di Kabupaten Bangka, mungkin karena keindahan pantainya serta kekayaan budaya dan sejarahnya. Sementara itu, di Yogyakarta, kunjungan ke hotel didorong oleh daya tarik budaya, wisata sejarah, dan keberadaan destinasi terkenal seperti Candi Borobudur dan Keraton Yogyakarta, serta keramah tamahan masyarakatnya.
- Dari sisi market_segment, mayoritas pemesanan berasal dari Online TA (Travel Agents), diikuti oleh agen perjalanan offline (TA/TO), dan grup wisata, menunjukkan preferensi wisatawan terhadap kemudahan dan fleksibilitas yang ditawarkan oleh layanan pemesanan online.

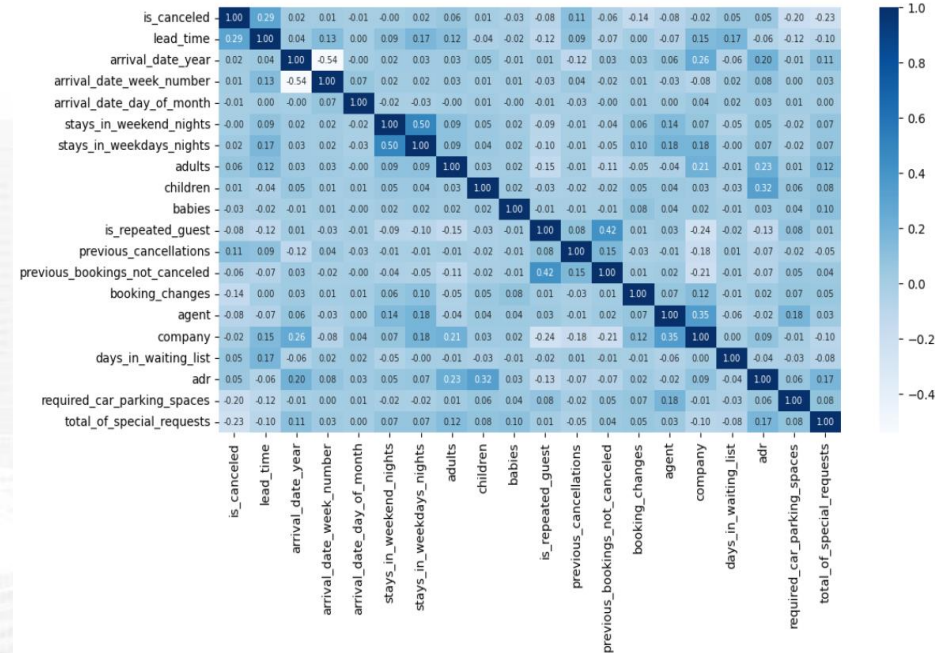
Univariate Analysis



Count Plot

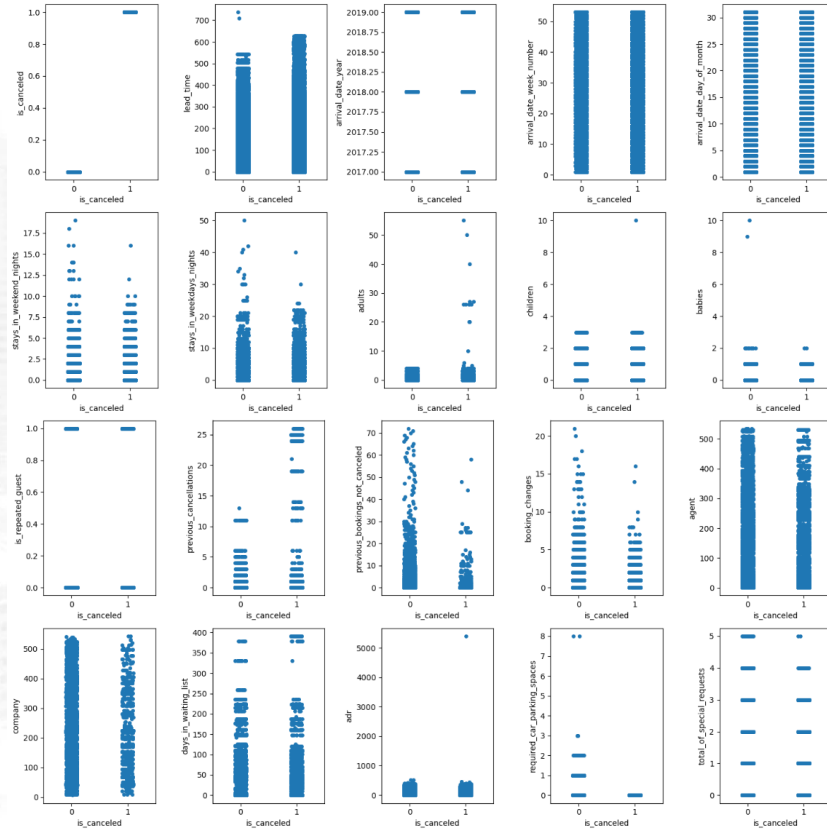
- Mayoritas tamu merupakan tamu baru (`is_repeated_guest` mayoritas 0 dibandingkan 1), menunjukkan bahwa hotel-hotel ini banyak menarik pengunjung baru yang mungkin tertarik dengan keunikan dan daya tarik daerah ini.
- `Reservation_status` yang paling banyak adalah check-out, yang menunjukkan bahwa sebagian besar reservasi tamu telah selesai dengan baik dan tamu telah meninggalkan hotel setelah periode menginap mereka. Diikuti oleh canceled, yang menunjukkan reservasi yang dibatalkan sebelum kedatangan tamu, dan no-show, yang mengindikasikan tamu yang tidak tiba tanpa memberi tahu hotel sebelumnya.

Multivariate Analysis



- Fitur target: is_canceled
- **is_canceled** dan **lead_time** berkorelasi positif lemah. Semakin lama rentang waktu pemesanan dan kedatangan maka semakin berpotensi untuk melakukan pembatalan pemesanan.
- **is_canceled** dan **previous_cancellations** berkorelasi positif lemah. Semakin banyak melakukan pembatalan di pemesanan sebelum-sebelumnya maka di pemesanan berikutnya berpotensi akan dibatalkan kembali.
- **is_repeated_guest** dan **previous_bookings_not_canceled** berkorelasi positif cukup kuat. Semakin banyak jumlah pemesanan sebelumnya yang tidak dibatalkan oleh tamu maka tamu tersebut semakin berpotensi akan melakukan pemesanan kembali.
- **adults** dan **adr** berkorelasi positif lemah. Ada kecenderungan bahwa semakin banyak jumlah orang dewasa dalam pemesanan, maka akan ada sedikit peningkatan dalam nilai adr (harga harian rata-rata). Namun, jumlah orang dewasa dalam pemesanan hanya memberikan kontribusi kecil terhadap variasi dalam nilai adr.
- **children** dan **adr** berkorelasi positif sedang. Ada kecenderungan bahwa semakin banyak jumlah anak-anak dalam pemesanan, ada sedikit peningkatan dalam nilai adr (harga harian rata-rata).
- **adr** dan **total_of_special_requests** berkorelasi positif lemah. Ada sedikit kecenderungan bahwa semakin banyak jumlah permintaan khusus oleh tamu, ada sedikit peningkatan dalam nilai adr. Namun, hubungan ini tidak terlalu kuat, yang berarti bahwa pengaruh total permintaan khusus terhadap nilai adr tidak signifikan.
- Nantinya ketika kolom kategorikal sudah di ubah ke numerik mungkin ada insight/informasi baru.

Multivariate Analysis



- Pembatalan lebih banyak terjadi ketika lead_time semakin tinggi.
- Pembatalan terjadi ketika pemesanan untuk jumlah orang dewasa cukup tinggi (rombongan).
- Pembatalan lebih sering terjadi ketika tamu pernah melakukan pembatalan sebelumnya.
- Pembatalan lebih sering tidak terjadi ketika tamu tidak melakukan pembatalan di pemesanan sebelumnya.
- Pembatalan tidak terjadi karena tamu banyak melakukan perubahan pemesanan.

Handle Missing Value

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

Mengisi nilai null fitur children menggunakan median
Mengisi nilai null kategorikal menggunakan modus
Mengisi fitur agent menggunakan 0
Drop fitur company karena data di dominasi dengan nilai null

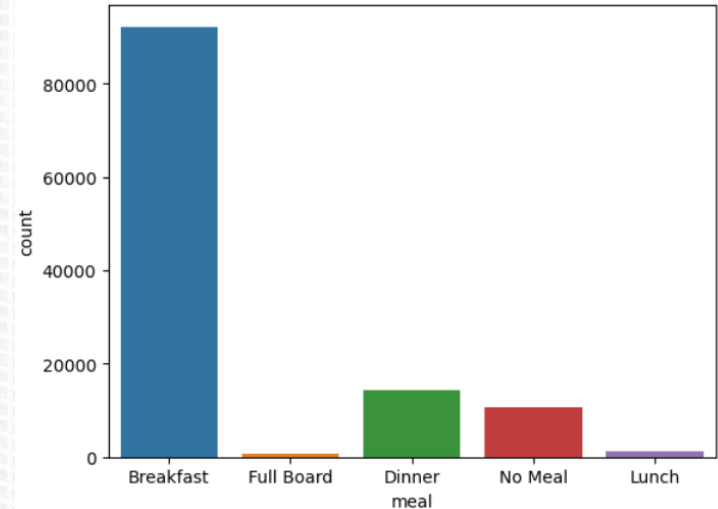
hotel	0
is_canceled	0
lead_time	0
arrival_date_year	0
arrival_date_month	0
arrival_date_week_number	0
arrival_date_day_of_month	0
stays_in_weekend_nights	0
stays_in_weekdays_nights	0
adults	0
children	0
babies	0
meal	0
city	0
market_segment	0
distribution_channel	0
is_repeated_guest	0
previous_cancellations	0
previous_bookings_not_canceled	0
booking_changes	0
deposit_type	0
agent	0
days_in_waiting_list	0
customer_type	0
adr	0
required_car_parking_spaces	0
total_of_special_requests	0
reservation_status	0

Handle Duplicated Data

Tidak melakukan handling duplikat karena tidak memiliki fitur ID yang bernilai unik yang dapat membantu identifikasi adanya duplikat data.

Mengganti value yang tidak sesuai

Mengganti Undefined menjadi Lunch karena belum ada di jenis meal yang tersedia.



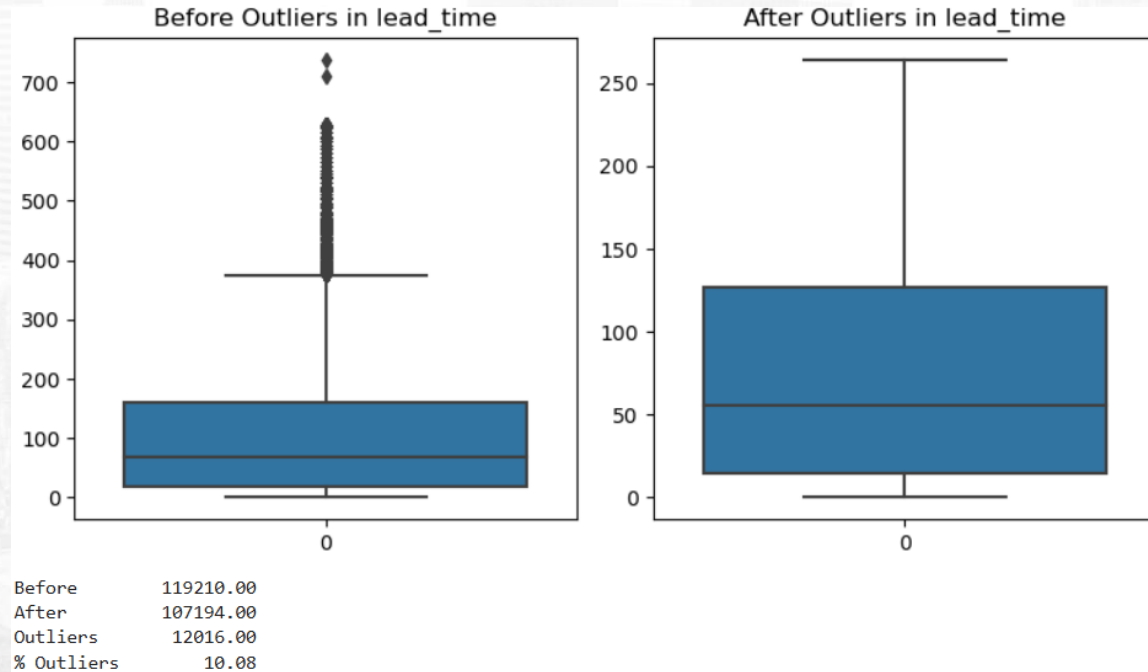
Membuang data yang tidak diperlukan

- Membuat fitur baru bernama `total_tamu` untuk mengetahui apakah terdapat baris data yang invalid, yaitu total tamu berjumlah 0.
- Total tamu merupakan jumlah dari `adults`, `children`, dan `babies`.
- Melakukan filter, hanya mengambil baris data yang total tamu lebih dari 0.
- Drop kembali fitur `total_tamu` agar kembali ke bentuk dataframe semula.

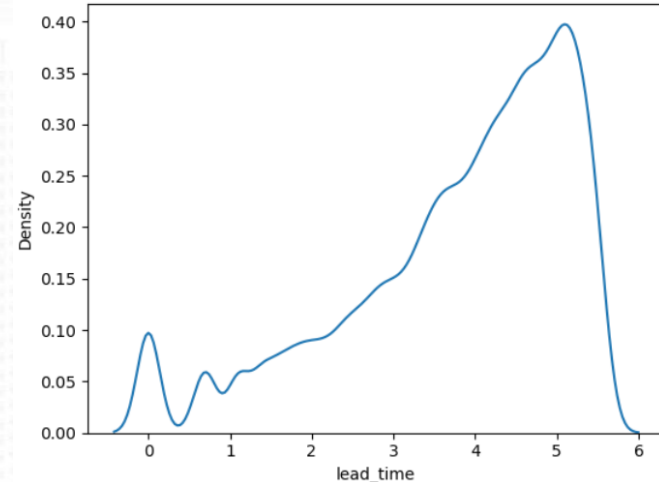
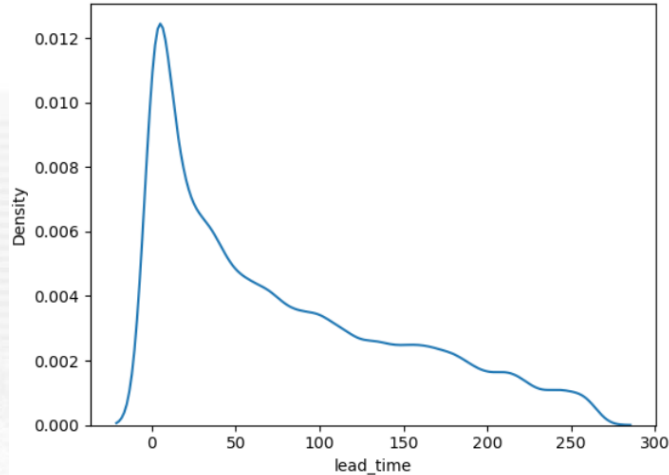
total_tamu	
2.0	82051
1.0	22581
3.0	10495
4.0	3929
0.0	180
5.0	137
26.0	5
12.0	2
27.0	2
20.0	2
10.0	2
40.0	1
50.0	1
55.0	1
6.0	1

Terdapat 180 data invalid atau pemesanan dengan jumlah tamu adalah nol.

Handling Outlier



Feature Transformation



Distribusi masih long right skew.

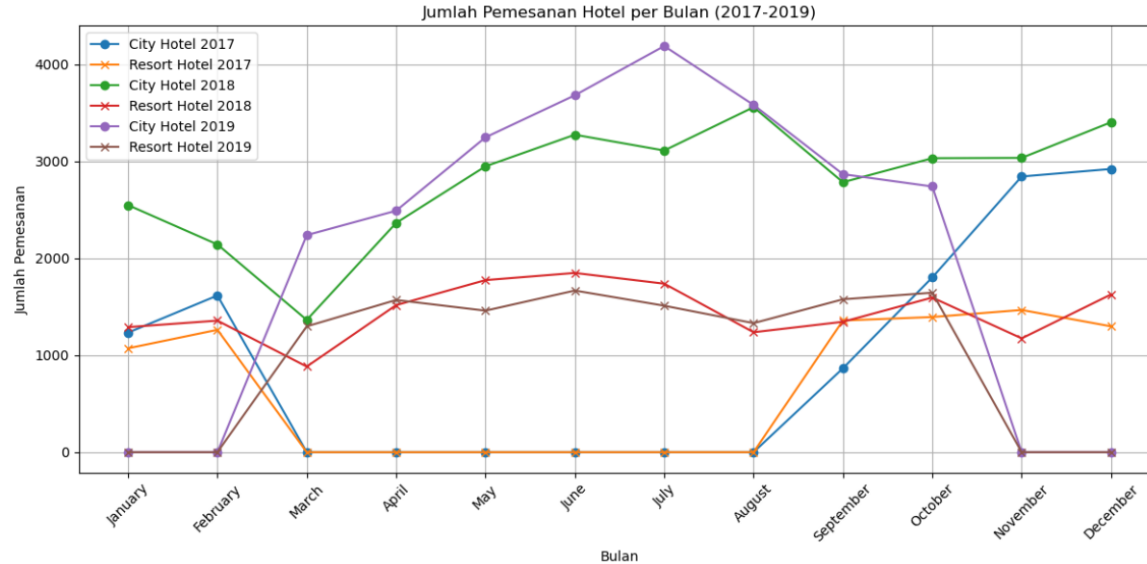
Lebih mendekati distribusi normal setelah dilakukan log transformation.

Monthly Hotel Booking Analysis Based on Hotel Type

arrival_date_month	2017		2018		2019	
	City Hotel	Resort Hotel	City Hotel	Resort Hotel	City Hotel	Resort Hotel
January	1233	1071	2549	1290	0	0
February	1616	1262	2143	1357	0	0
March	0	0	1364	883	2239	1298
April	0	0	2365	1517	2491	1570
May	0	0	2949	1775	3248	1460
June	0	0	3275	1849	3682	1667
July	0	0	3112	1738	4190	1511
August	0	0	3559	1237	3582	1331
September	865	1358	2785	1345	2868	1577
October	1803	1394	3033	1595	2741	1645
November	2844	1467	3036	1174	0	0
December	2922	1298	3404	1627	0	0

Nilai 0 menunjukkan tidak ada pemesanan pada bulan dan tahun tersebut.

Monthly Hotel Booking Analysis Based on Hotel Type



Tahun 2019

- Pemesanan hanya ada pada bulan Maret hingga Oktober.
- Bulan dengan pemesanan tertinggi pada City Hotel adalah Juli dan pada Resort Hotel adalah Juni.
- Bulan dengan pemesanan terendah pada City Hotel maupun Resort Hotel adalah Maret.

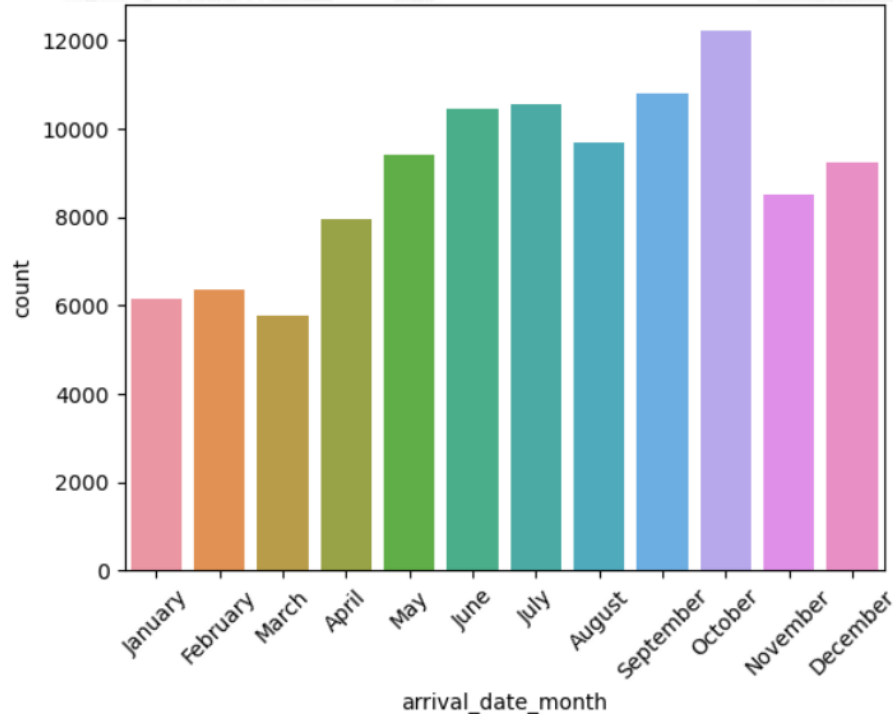
Tahun 2017

- Pemesanan hanya ada pada bulan Januari, Februari, September, Oktober, November, dan Desember.
- Bulan dengan pemesanan tertinggi pada City Hotel adalah Desember dan pada Resort Hotel adalah November.
- Bulan dengan pemesanan terendah pada City Hotel adalah September dan pada Resort Hotel adalah Januari.

Tahun 2018

- Pemesanan ada setiap bulannya dari Januari hingga Desember.
- Bulan dengan pemesanan tertinggi pada City Hotel adalah Agustus dan pada Resort Hotel adalah Juni.
- Bulan dengan pemesanan terendah pada City Hotel maupun Resort Hotel adalah Maret.

Mengurutkan Data Berdasarkan Bulan



Pemesanan Tertinggi

- Oktober: Mungkin dipengaruhi oleh event atau liburan akhir tahun.
- September dan Juli : Dipengaruhi oleh liburan sekolah dan event tertentu.

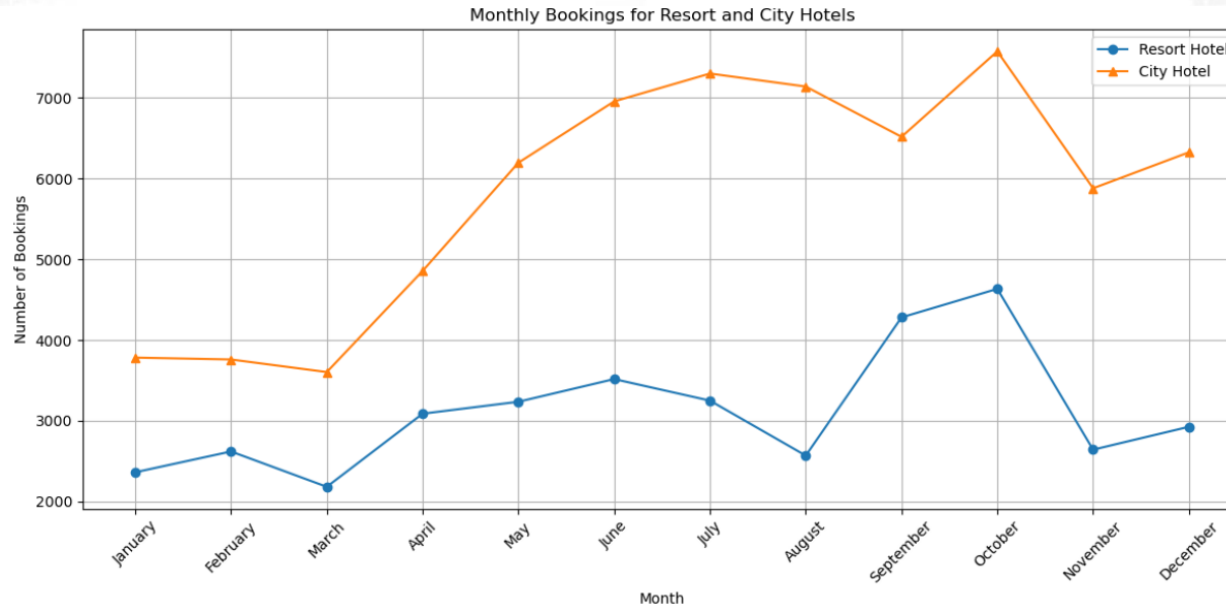
Pemesanan Menengah

- Juni dan Agustus: Liburan sekolah.
- Mei: Mungkin terkait dengan Hari Raya Idul Fitri.
- Desember: Liburan Natal dan akhir tahun.
- November: Persiapan liburan akhir tahun.

Pemesanan Rendah

- April: Bukan musim liburan.
- Februari dan Januari: Setelah liburan Tahun Baru.
- Maret: Tidak ada event besar.

Monthly Hotel Booking Analysis Based on Hotel Type

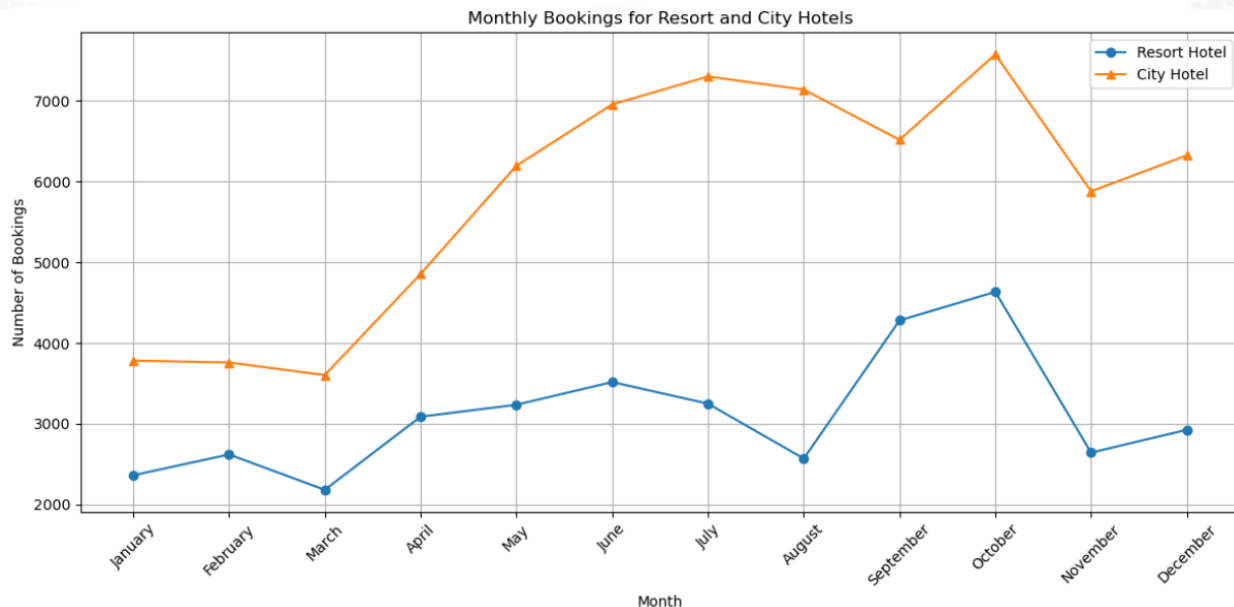


City Hotel

- Januari hingga Maret: Jumlah pemesanan stabil dengan sedikit penurunan pada bulan Maret. Ini bisa disebabkan oleh liburan akhir tahun yang sudah berakhir dan periode awal tahun yang biasanya lebih tenang.
- April: Terdapat peningkatan signifikan dalam pemesanan, kemungkinan terkait dengan libur panjang Paskah atau hari libur nasional lainnya.
- Mei hingga Agustus: Jumlah pemesanan terus meningkat, mencapai puncaknya pada bulan Juli. Ini berkaitan dengan musim liburan sekolah di Indonesia yang berlangsung dari akhir Juni hingga awal Juli, serta libur Lebaran yang sering jatuh antara Mei dan Juni.
- September hingga Oktober: Meskipun ada sedikit penurunan pada bulan September, pemesanan kembali meningkat pada bulan Oktober.
- November dan Desember: Penurunan terjadi pada bulan November, namun pemesanan meningkat kembali pada bulan Desember. Peningkatan ini kemungkinan disebabkan oleh liburan Natal dan akhir tahun.

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

Monthly Hotel Booking Analysis Based on Hotel Type



Resort Hotel

- Januari hingga Maret: Jumlah pemesanan relatif stabil dengan sedikit penurunan pada bulan Maret. Ini mencerminkan pola yang mirip dengan City Hotel.
- April: Ada peningkatan pemesanan, meskipun tidak sebesar di City Hotel, yang mungkin terkait dengan libur panjang Paskah atau hari libur nasional lainnya.
- Mei hingga Agustus: Pemesanan meningkat, terutama pada bulan Juni dan Juli, yang menunjukkan dampak dari liburan sekolah dan libur Lebaran. Namun, terdapat penurunan pada bulan Agustus yang mungkin disebabkan oleh berakhirnya libur sekolah.

- September hingga Oktober: Peningkatan signifikan terlihat pada bulan September dan Oktober, yang bisa dipengaruhi oleh libur Idul Adha dan cuti bersama.
- November dan Desember: Pemesanan menurun pada bulan November tetapi meningkat kembali pada bulan Desember, yang mencerminkan tren liburan Natal dan akhir tahun.

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

Fokus pada Musim Liburan

- Perkuat promosi dan diskon khusus selama musim liburan (Juni, Juli, Agustus, Desember) untuk memaksimalkan jumlah pemesanan.
- Buat paket liburan keluarga untuk menarik lebih banyak tamu selama liburan sekolah.

Manfaatkan Event dan Hari Besar:

- Selenggarakan event khusus atau penawaran menarik pada bulan Oktober dan September untuk menarik lebih banyak pengunjung.
- Tawarkan promosi khusus pada bulan Mei untuk memanfaatkan liburan Hari Raya Idul Fitri.

Penawaran pada Musim Sepi:

- Berikan diskon atau penawaran khusus pada bulan Januari, Februari, dan Maret untuk meningkatkan pemesanan pada periode sepi.
- Selenggarakan event bisnis atau konferensi pada bulan-bulan ini untuk menarik segmen pasar yang berbeda.

Pengoptimalan Sumber Daya:

- Sesuaikan jumlah staf berdasarkan musim liburan dan musim sepi untuk efisiensi operasional.
- Lakukan perawatan dan perbaikan fasilitas hotel pada bulan-bulan dengan pemesanan rendah.

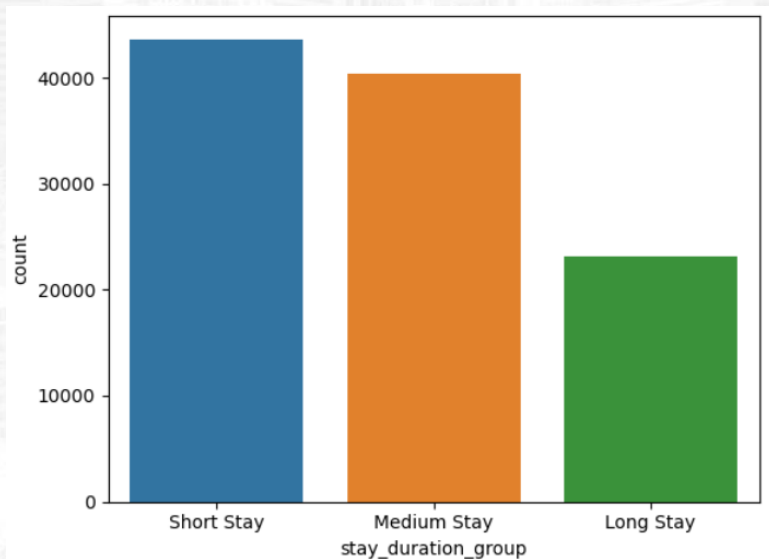
Segmentasi Pasar:

- Fokuskan pemasaran pada keluarga selama liburan sekolah.
- Buat promosi menarik untuk wisatawan yang berlibur di akhir tahun.

Membuat kolom baru yang berisi total durasi menginap

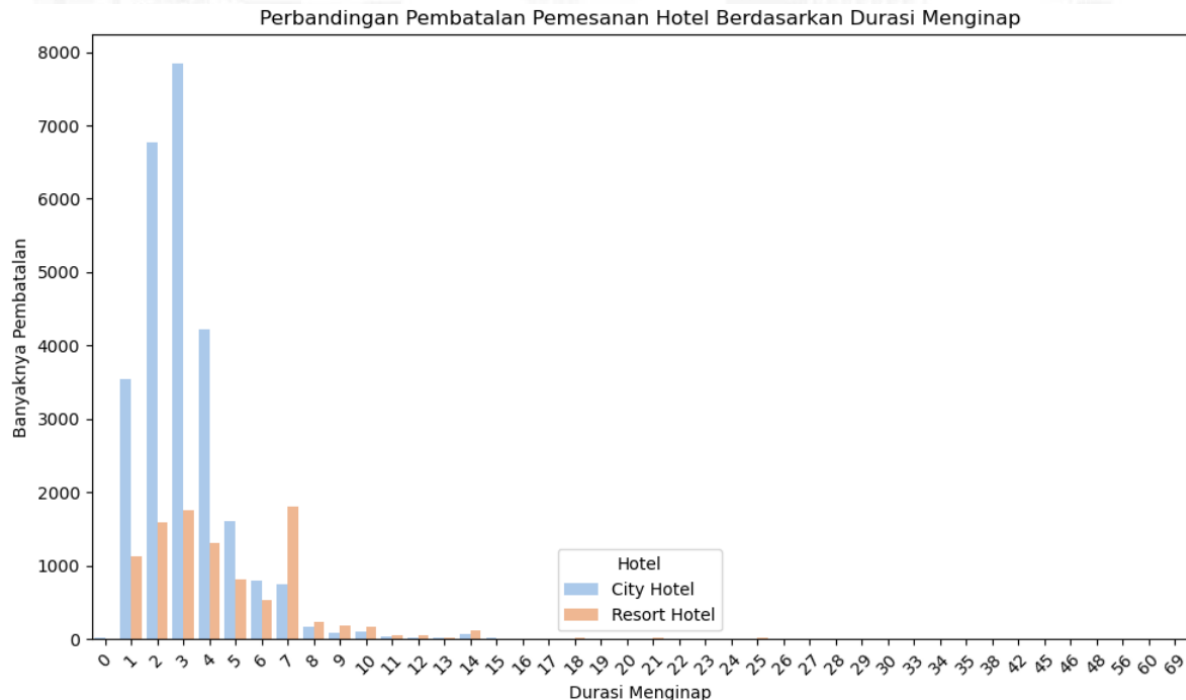


Mengelompokkan durasi menginap menjadi 3 jenis



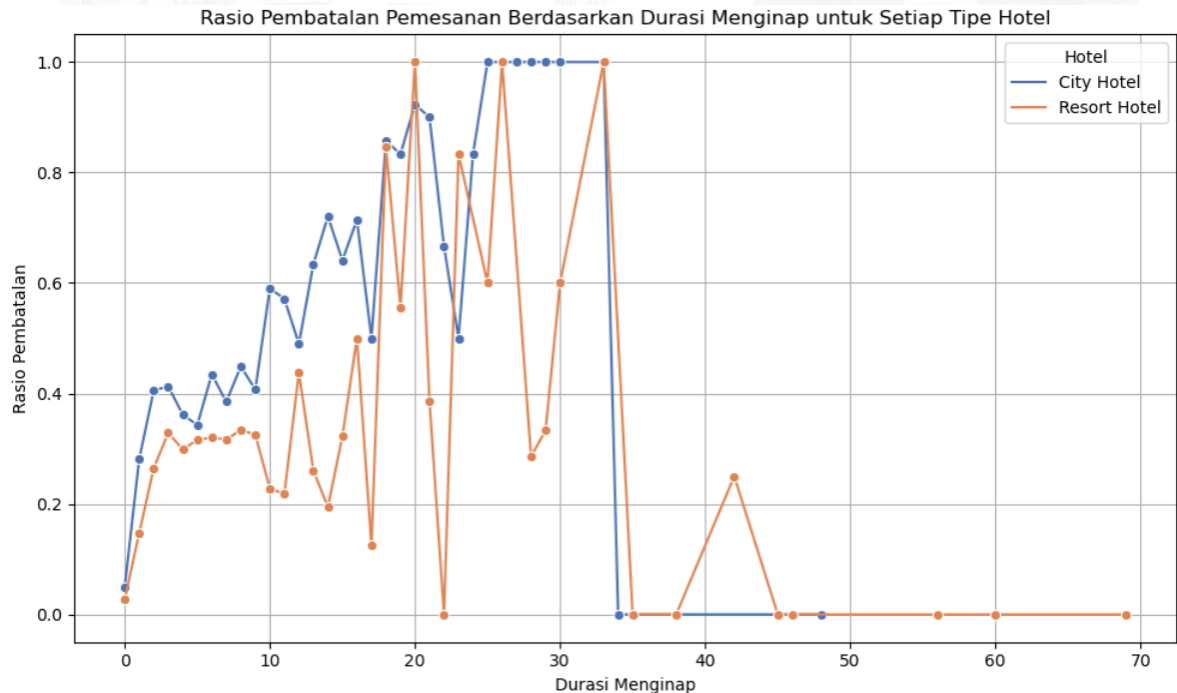
- Kategori dengan jumlah pemesanan terbanyak adalah Short Stay, yang menunjukkan bahwa lebih banyak tamu yang memilih untuk menginap dalam durasi yang singkat dibandingkan dengan durasi menengah atau panjang.
- Data ini dapat memberikan wawasan tentang preferensi tamu. Banyaknya pemesanan dengan durasi singkat mungkin menunjukkan bahwa hotel lebih sering digunakan untuk kunjungan singkat atau perjalanan bisnis.

Perbandingan jumlah pemesanan hotel yang dibatalkan terhadap durasi menginap untuk setiap jenis hotel



- Pembatalan lebih sering terjadi pada City Hotel daripada Resort Hotel.
- Durasi menginap di bawah 7 hari memiliki jumlah pembatalan yang sangat tinggi.
- Durasi menginap lebih dari 7 hari memiliki jumlah pembatalan yang rendah.

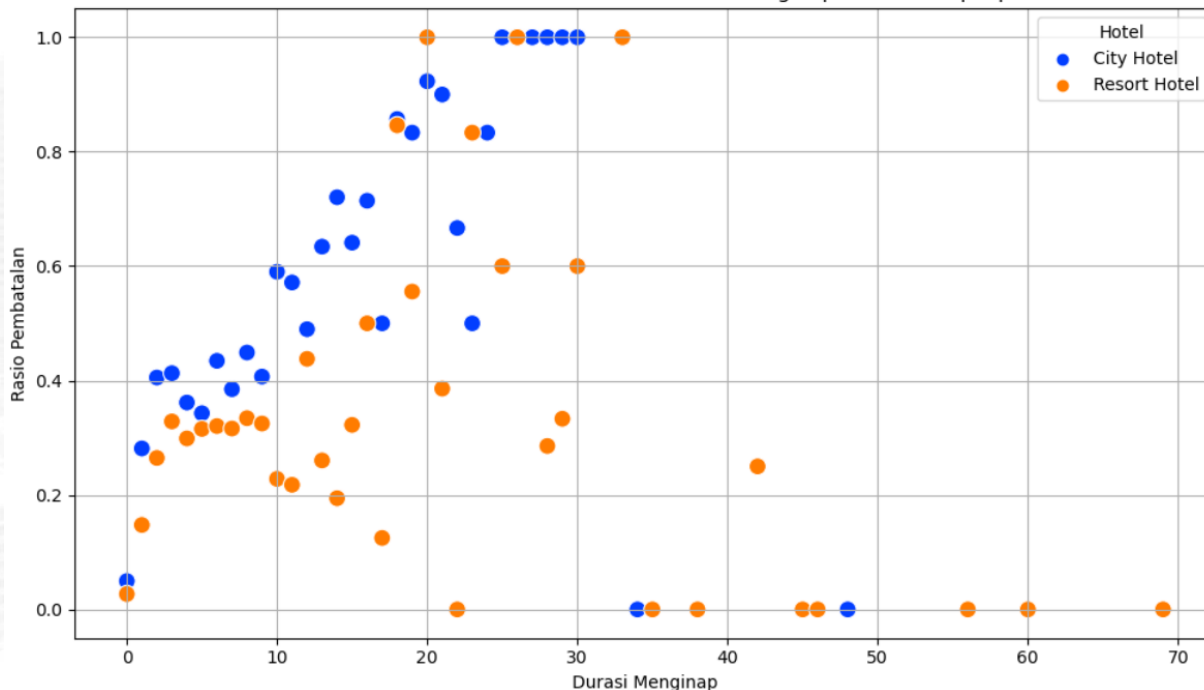
Rasio pembatalan pesanan terhadap durasi menginap untuk setiap tipe hotel



- Terdapat 10 pemesanan dengan rasio pembatalan 0 (tidak ada pembatalan) dan 10 pemesanan dengan rasio pembatalan 1 (semua pemesanan dibatalkan). Ini menunjukkan bahwa ada jumlah yang signifikan dari pemesanan yang sepenuhnya tidak dibatalkan atau sepenuhnya dibatalkan.
- Ada berbagai nilai rasio pembatalan di antara 0 dan 1, yang menunjukkan bahwa dalam beberapa kasus, sebagian pemesanan dibatalkan.

Rasio pembatalan pesanan terhadap durasi menginap untuk setiap tipe hotel

Rasio Pembatalan Pemesanan Berdasarkan Durasi Menginap untuk Setiap Tipe Hotel



- Terdapat clustering atau pengelompokan pada nilai-nilai tertentu, misalnya pada 0.5, 0.6, dan nilai yang lebih tinggi mendekati 1.
- Banyak rasio pembatalan berada di rentang 0.2 hingga 0.4, menunjukkan bahwa sebagian pemesanan mengalami beberapa pembatalan.
- Terlihat bahwa City Hotel lebih rentan terhadap pembatalan daripada Resort Hotel.
- Rasio pembatalan bervariasi secara luas, menunjukkan ketidakteraturan dalam perilaku pembatalan tamu.

Interpretasi Statistik Deskriptif

	count	mean	std	min	25%	50%	75%	max
stay_duration_group								
Long Stay	62.0	0.499699	0.337717	0.000000	0.266741	0.469416	0.833333	1.000000
Medium Stay	4.0	0.350534	0.048854	0.298945	0.321193	0.345125	0.374467	0.412942
Short Stay	6.0	0.195972	0.147105	0.026954	0.074135	0.206235	0.277271	0.405357

Short Stay

- Mean: Rata-rata rasio pembatalan adalah 0.35, menunjukkan bahwa sekitar 35% pemesanan dalam kategori ini dibatalan.
- Std: Standar deviasi adalah 0.15, menunjukkan variasi yang moderat dalam rasio pembatalan.
- Distribution: 25% dari pemesanan memiliki rasio pembatalan di bawah 0.20, 50% di bawah 0.35, dan 75% di bawah 0.50.

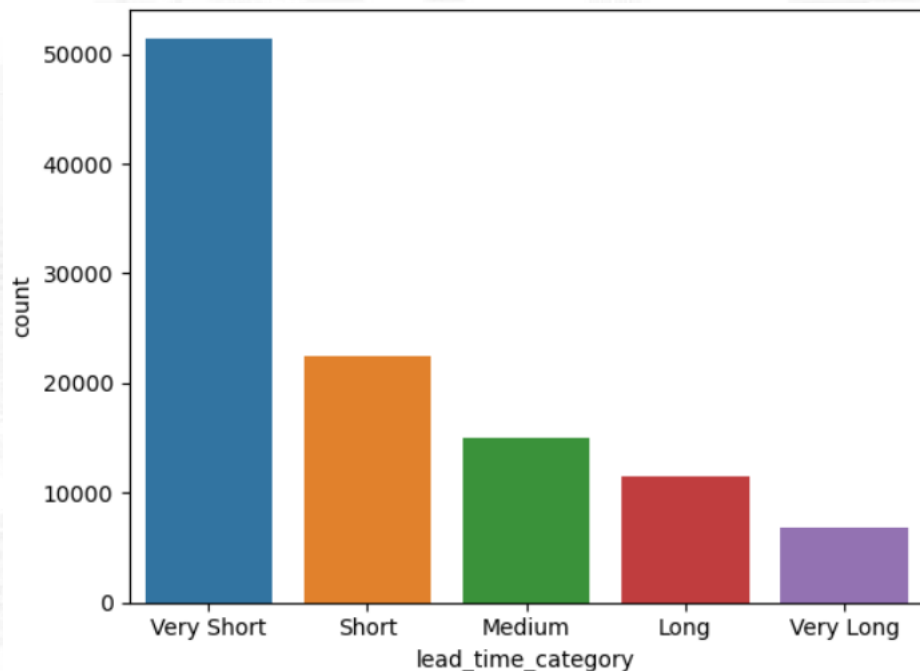
Medium Stay

- Mean: Rata-rata rasio pembatalan adalah 0.45, menunjukkan peningkatan dalam rasio pembatalan dibandingkan dengan Short Stay.
- Std: Standar deviasi adalah 0.20, menunjukkan variasi yang lebih besar dalam rasio pembatalan.
- Distribution: 25% dari pemesanan memiliki rasio pembatalan di bawah 0.30, 50% di bawah 0.45, dan 75% di bawah 0.60.

Long Stay

- Mean: Rata-rata rasio pembatalan adalah 0.55, menunjukkan bahwa pemesanan dengan durasi lebih lama cenderung memiliki rasio pembatalan yang lebih tinggi.
- Std: Standar deviasi adalah 0.25, menunjukkan variasi yang lebih besar dalam rasio pembatalan.
- Distribution: 25% dari pemesanan memiliki rasio pembatalan di bawah 0.40, 50% di bawah 0.55, dan 75% di bawah 0.70.

Membuat kolom baru yang berisi kategorisasi dari kolom jarak waktu pemesanan



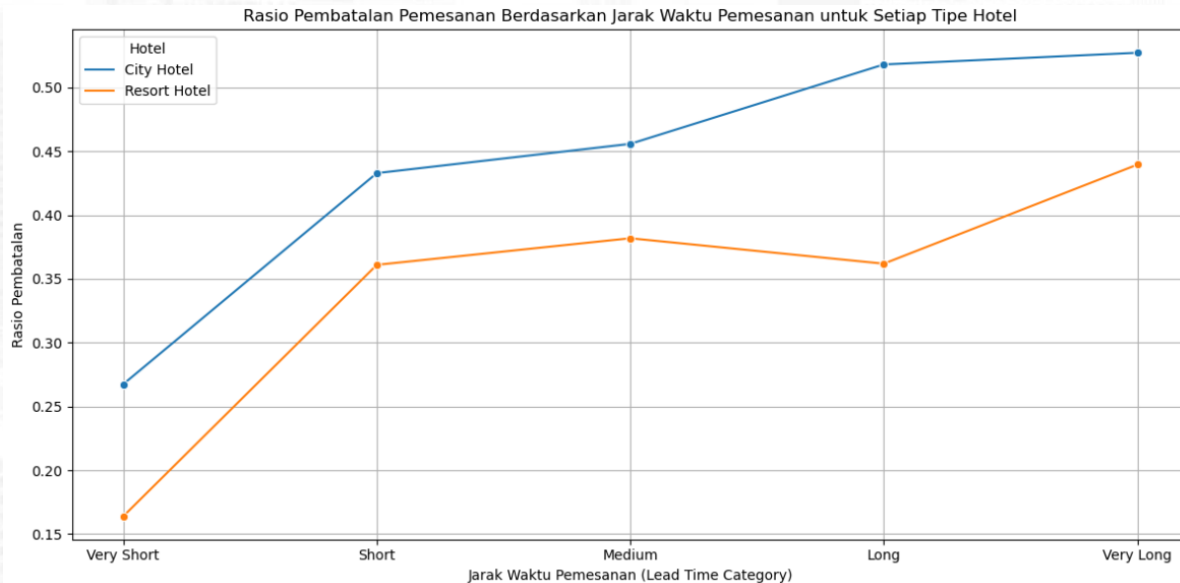
Kategori ini memberikan gambaran tentang distribusi jarak waktu pemesanan dalam data yang dimiliki. Jarak waktu pemesanan dapat menjadi faktor penting dalam analisis tingkat pembatalan pemesanan hotel, di mana pembatalan cenderung bervariasi tergantung pada seberapa jauh pemesanan dilakukan sebelum tanggal kedatangan yang dijadwalkan. Analisis ini dapat membantu manajemen hotel untuk mengoptimalkan strategi pengelolaan reservasi dan mengurangi tingkat pembatalan dengan memahami pola pembatalan berdasarkan jarak waktu pemesanan yang terjadi.

Perbandingan jumlah pemesanan hotel yang dibatalkan terhadap jarak waktu pemesanan untuk setiap jenis hotel

	hotel	lead_time_category	total_bookings	canceled_bookings
0	City Hotel	Very Short	32073	8580
1	City Hotel	Short	16190	7008
2	City Hotel	Medium	10255	4674
3	City Hotel	Long	7421	3844
4	City Hotel	Very Long	3959	2087
5	Resort Hotel	Very Short	19394	3181
6	Resort Hotel	Short	6233	2250
7	Resort Hotel	Medium	4759	1817
8	Resort Hotel	Long	4109	1487
9	Resort Hotel	Very Long	2801	1231

- City Hotel memiliki jumlah total pemesanan yang lebih tinggi daripada Resort Hotel di setiap kategori lead time.
- Tingkat pembatalan pemesanan cenderung meningkat seiring dengan peningkatan lead time, terutama terlihat pada kategori lead time yang lebih panjang.

Rasio pembatalan pesanan terhadap jarak waktu pemesanan untuk setiap tipe hotel



- Tingkat pembatalan pemesanan cenderung meningkat seiring dengan peningkatan lead time, terutama terlihat pada kategori lead time yang lebih panjang.
- Resort Hotel memiliki proporsi pembatalan yang lebih rendah dibandingkan dengan City Hotel di hampir setiap kategori lead time, meskipun memiliki jumlah total pemesanan yang lebih rendah.

Insight

1. Tingkat Pembatalan Berdasarkan Lead Time

Secara umum, terlihat bahwa semakin lama lead time (waktu antara pemesanan dan kedatangan), semakin tinggi kemungkinan pembatalan pemesanan. Ini terlihat dari peningkatan jumlah pemesanan yang dibatalkan dari kategori "Very Short" hingga "Very Long" baik di City Hotel maupun Resort Hotel.

2. Perbandingan Antara City Hotel dan Resort Hotel

- City Hotel → Memiliki jumlah total pemesanan yang lebih tinggi daripada Resort Hotel di semua kategori lead time. Namun, juga memiliki tingkat pembatalan yang lebih tinggi, terutama pada lead time yang lebih panjang.
- Resort Hotel → Meskipun memiliki jumlah total pemesanan yang lebih rendah, Resort Hotel cenderung memiliki proporsi pembatalan yang lebih rendah dibandingkan dengan City Hotel. Hal ini menunjukkan bahwa kestabilan reservasi mungkin lebih baik di Resort Hotel meskipun jumlahnya lebih sedikit.

3. Strategi Pengelolaan Pembatalan

Manajemen hotel dapat menggunakan informasi ini untuk mengoptimalkan strategi pengelolaan reservasi dan mengurangi tingkat pembatalan, terutama dengan fokus pada pengelolaan reservasi dengan lead time yang lebih panjang. Mungkin diperlukan strategi khusus untuk meningkatkan retensi pemesanan yang memiliki lead time panjang, seperti kebijakan pembatalan yang fleksibel atau penawaran khusus.

4. Segmentasi Pasar dan Pemasaran

Informasi ini juga dapat membantu dalam segmentasi pasar dan strategi pemasaran. City Hotel mungkin lebih menarik bagi tamu yang melakukan pemesanan dalam jangka waktu pendek atau menengah, sementara Resort Hotel mungkin lebih cocok untuk tamu yang merencanakan perjalanan jauh-jauh hari sebelumnya.