

Data Analytics for Business 2024

MID EXAM

Anggota Kelompok 8

- | | | |
|----|-------------------------|-----------------|
| 1. | Michael Luwi Pallea | CS05-KM-CS05278 |
| 2. | Keisya Akhmala Tazkiyah | CS05-KM-CS05509 |
| 3. | Rayhan Anugrah Ramadhan | CS05-KM-CS05488 |

BAB I

Pendahuluan

1.1 Latar belakang masalah

Dalam lingkungan perhotelan yang kompetitif, kepuasan pelanggan dan manajemen operasional yang efisien menjadi kunci untuk mempertahankan loyalitas pelanggan dan meningkatkan pendapatan. Namun, perusahaan hotel ini menghadapi berbagai tantangan dalam proses reservasi, yang berdampak negatif pada tingkat pembatalan reservasi serta menurunkan minat pelanggan untuk melakukan reservasi jangka panjang. Berdasarkan analisis terhadap dataset, beberapa masalah utama telah diidentifikasi sebagai faktor penyebab kendala operasional ini.

1.1.1 Penjelasan Rinci dari Masalah

Berikut adalah penjelasan lebih rinci dari masalah-masalah adalah:

1. Tingkat Pembatalan yang Tinggi

Berdasarkan data pada kolom `is_canceled`, tercatat bahwa persentase pembatalan reservasi cukup tinggi. Pelanggan sering kali membatalkan reservasi setelah melakukan pemesanan, meskipun kamar sudah dialokasikan untuk mereka. Hal ini menyebabkan kerugian operasional karena kamar yang sudah disiapkan menjadi kosong, serta merusak prediksi okupansi hotel. Tingginya pembatalan ini dapat disebabkan oleh berbagai faktor, termasuk ketidakpastian layanan atau ketidakpuasan pelanggan pada proses reservasi yang ada.

2. Lead Time yang Bervariasi

Kolom `lead_time` menunjukkan variasi yang signifikan dalam durasi antara tanggal reservasi dan tanggal kedatangan. Hal ini membuat pengelolaan sumber daya menjadi sulit bagi pihak hotel. Pada beberapa kasus, lead time yang sangat panjang diikuti dengan pembatalan dapat mengurangi peluang hotel untuk menerima reservasi dari pelanggan lain yang mungkin lebih membutuhkan kamar dalam

waktu dekat. Sebaliknya, lead time yang pendek bisa menyebabkan kekurangan kamar jika reservasi datang mendadak. Akibatnya, hotel berisiko mengalami overbooking atau underbooking yang berdampak pada kepuasan pelanggan dan stabilitas operasional

3. Permintaan Khusus Tidak Dikelola dengan Baik

Kolom `total_of_special_requests` menunjukkan bahwa pelanggan sering kali mengajukan permintaan khusus, seperti permintaan kamar tertentu atau layanan tambahan. Namun, tidak adanya mekanisme penanganan permintaan khusus yang efektif menyebabkan ketidakpuasan pelanggan. Ketika permintaan mereka tidak terpenuhi, risiko pembatalan meningkat. Oleh karena itu, diperlukan sistem yang mampu mengakomodasi permintaan khusus pelanggan untuk meningkatkan tingkat kepuasan dan loyalitas mereka.

4. Proses Pembayaran yang Tidak Efisien

Berdasarkan data pada kolom `adr` (Average Daily Rate) dan `reservation_status`, ditemukan bahwa proses pembayaran sering kali mengalami kendala. Beberapa pelanggan melaporkan kegagalan transaksi, yang membuat mereka ragu untuk melanjutkan reservasi. Proses pembayaran yang lambat atau rumit menyebabkan hilangnya kepercayaan pelanggan, yang pada akhirnya berpotensi menyebabkan pembatalan dan penurunan jumlah reservasi. Perusahaan perlu memperbaiki sistem pembayaran agar lebih efisien dan mudah digunakan.

5. Kurangnya Komunikasi Selama Proses Check-in dan Check-out

Analisis pada kolom `reservation_status` menunjukkan bahwa komunikasi selama proses check-in dan check-out sering kali tidak berjalan lancar. Pelanggan terkadang tidak menerima informasi yang tepat waktu, yang menyebabkan keterlambatan atau kebingungan pada saat check-in dan check-out. Ketidaktepatan komunikasi ini tidak hanya menyebabkan ketidaknyamanan bagi pelanggan, tetapi juga dapat mempengaruhi persepsi mereka terhadap kualitas layanan hotel secara keseluruhan.

1.1.2 Penyelesaian Masalah

Untuk mengatasi masalah-masalah tersebut, perusahaan akan melakukan beberapa tahapan sebagai berikut:

1. Analisis Data Pembatalan

Menganalisis pola pembatalan dengan melihat variabel-variabel terkait seperti `lead_time`, `deposit_type`, dan `reservation_status`. Tujuan dari analisis ini adalah untuk mengidentifikasi faktor-faktor utama yang mempengaruhi pembatalan, serta memberikan rekomendasi terkait kebijakan pembatalan yang lebih fleksibel atau insentif bagi pelanggan yang melakukan reservasi lebih awal.

2. Optimasi Lead Time dan Pengelolaan Sumber Daya

Mengidentifikasi pola `lead_time` untuk merencanakan alokasi kamar dan staf secara lebih efisien. Hotel dapat menerapkan strategi pricing atau diskon untuk pelanggan dengan lead time yang pendek guna meningkatkan pemanfaatan kamar yang mendadak.

3. Sistem Permintaan Khusus yang Terintegrasi

Mengembangkan atau memperbarui sistem reservasi agar dapat menerima dan mengelola permintaan khusus pelanggan secara otomatis. Dengan demikian, pihak hotel dapat memastikan bahwa permintaan ini diterima dan dipenuhi sesuai dengan harapan pelanggan.

4. Penyederhanaan Proses Pembayaran

Mengimplementasikan metode pembayaran yang lebih efisien dan aman. Misalnya, hotel dapat menambahkan opsi pembayaran digital atau integrasi dengan platform pembayaran pihak ketiga untuk meningkatkan kenyamanan pelanggan.

5. Peningkatan Komunikasi Proses Check-in dan Check-out

Memastikan bahwa setiap informasi terkait reservasi, kebijakan check-in, dan check-out dikomunikasikan dengan jelas kepada pelanggan. Pihak hotel dapat mempertimbangkan penggunaan pesan otomatis atau aplikasi seluler untuk memberikan pembaruan informasi kepada pelanggan.

1.1.3 Output

Output yang diharapkan antara lain:

1. Penurunan Tingkat Pembatalan

Dengan mengidentifikasi dan menangani penyebab utama pembatalan, diharapkan tingkat pembatalan reservasi akan menurun.

2. Efisiensi Manajemen Sumber Daya

Variasi dalam lead time dapat dioptimalkan sehingga alokasi kamar dan staf lebih efektif.

3. Peningkatan Kepuasan Pelanggan

Sistem permintaan khusus yang terintegrasi akan meningkatkan kepuasan pelanggan, terutama bagi mereka yang mengajukan permintaan tertentu.

4. Komunikasi yang Lebih Baik

Informasi yang tepat waktu dan akurat selama proses check-in dan check-out akan meningkatkan pengalaman pelanggan dan mengurangi kemungkinan keluhan.

BAB II

Business Process Analysis

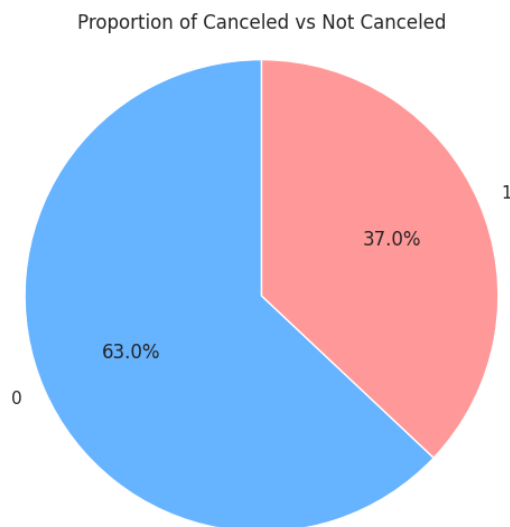
2.1 Identifikasi Masalah

Dalam proses reservasi hotel, beberapa tahapan utama yang umum ditemui pada proses reservasi adalah sebagai berikut

- 1) Pemilihan Kamar, Pelanggan memulai dengan memilih kamar yang diinginkan berdasarkan ketersediaan dan preferensi. Pada tahap ini, pelanggan bisa menambahkan permintaan khusus, seperti kebutuhan tertentu terkait fasilitas kamar.
- 2) Pemesanan, Setelah memilih kamar, pelanggan memasukkan detail pemesanan, seperti tanggal kedatangan, durasi menginap, dan jumlah tamu. Informasi ini kemudian disimpan dalam sistem untuk memastikan ketersediaan kamar yang diinginkan.
- 3) Konfirmasi Pemesanan, Setelah pemesanan selesai, hotel memberikan konfirmasi kepada pelanggan, biasanya melalui email atau SMS. Pada tahap ini, hotel juga akan memberikan informasi tentang kebijakan pembatalan dan detail lain terkait pemesanan.
- 4) Pra-Kedatangan, Beberapa hari sebelum kedatangan, hotel dapat mengirim pengingat atau informasi tambahan kepada tamu dapat berupa email atau SMS. Pada tahap ini, hotel bisa melakukan penyesuaian terhadap pemesanan jika ada perubahan dari pihak tamu.
- 5) Check-in, Pada hari kedatangan, tamu datang ke hotel untuk melakukan check-in. Staf front desk melakukan verifikasi data, menyerahkan kunci kamar, dan memastikan bahwa semua permintaan khusus telah diakomodasi sesuai kebutuhan pelanggan.
- 6) Selama Menginap, Selama menginap, pelanggan menikmati layanan hotel, termasuk kemungkinan penanganan tambahan untuk permintaan khusus yang diajukan.

- 7) Check-out, Pada akhir masa menginap, tamu melakukan check-out, menyelesaikan pembayaran, dan mengembalikan kunci kamar. Pelanggan juga diberi kesempatan untuk memberikan umpan balik mengenai pelayanan hotel. Proses pembayaran juga dapat mencakup biaya tambahan yang terkait dengan layanan tambahan yang digunakan selama menginap.

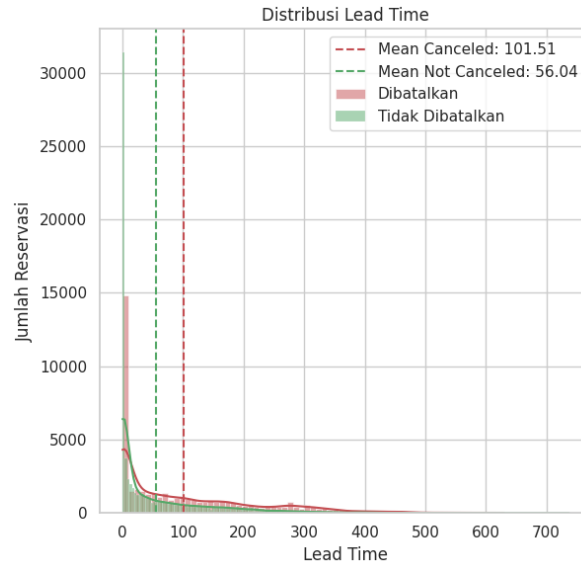
Pada proses tersebut pada praktiknya, dapat terjadi permasalahan yang menyebabkan dampak yang negatif terhadap hotel, seperti meningkatnya tingkat pembatalan. Berdasarkan analisis data yang dilakukan pada data *hotel data reservation* pada kolom *is_canceled* ditemukan tingginya tingkat pembatalan. Total pemesanan sebanyak 119389, 37% persen atau sebanyak 44224 dari total pemesanan tersebut dibatalkan pada periode dalam data yaitu tahun 2015-2017. Pembatalan disini adalah saat dimana pelanggan telah melakukan reservasi tetapi kemudian membatalkan pesanan meski kamar sudah dialokasikan untuk mereka. Proporsi pada antara pesanan yang dibatalkan dapat dilihat pada **Gambar 1**.



Gambar 1. Pie Chart Proporsi data Pembatalan

Pembatalan yang dilakukan pelanggan seringkali terjadi pada rata-rata *lead_time* sebesar 101,51 hari. Artinya, rata-rata pembatalan dilakukan sekitar 101 hari sebelum tanggal kedatangan. Meskipun demikian, Sebagian besar pembatalan juga terjadi dalam periode *lead_time* yang sangat singkat. Distribusi menunjukkan bahwa banyak pembatalan terjadi pada lead

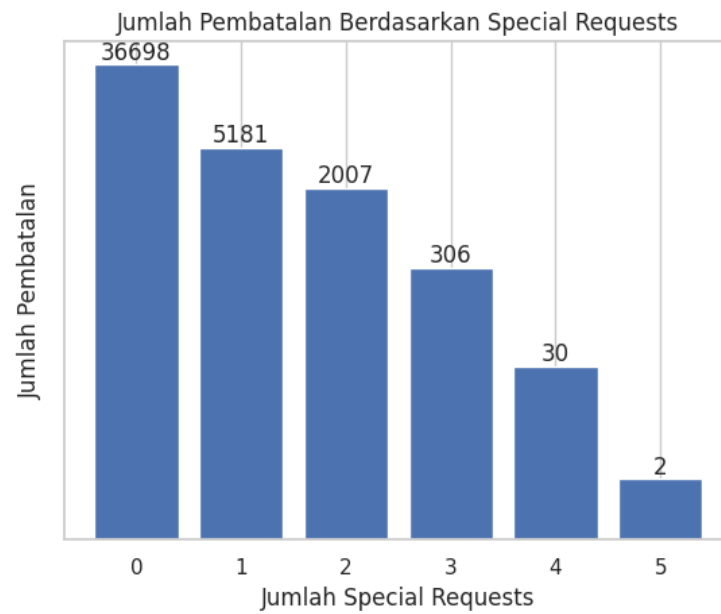
time mendekati tanggal check-in. Artinya, *lead_time* yang rendah juga sering berkorelasi dengan tingkat pembatalan yang tinggi, seperti pada **Gambar 2**.



Gambar 2. Histogram Distribusi Jumlah Reservasi Berdasarkan Lead Time

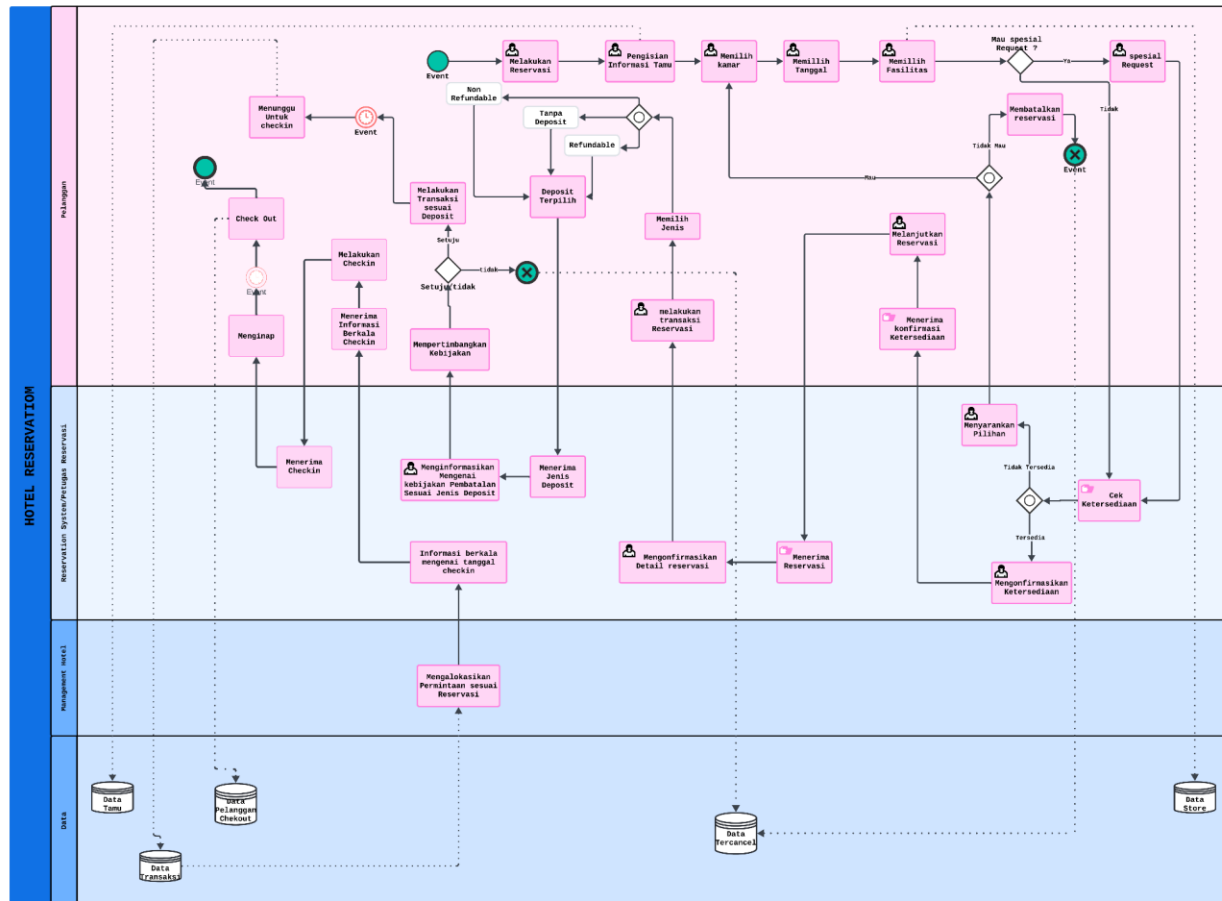
Oleh karena itu, variabilitas dalam *lead_time* menjadi faktor penting yang harus diperhatikan dalam manajemen risiko pembatalan. Pada satu sisi, pembatalan dengan *lead_time* yang panjang memungkinkan waktu untuk melakukan penyesuaian, tetapi pada sisi lain, tingginya frekuensi pembatalan pada *lead_time* yang rendah menuntut kebijakan yang lebih fleksibel atau insentif untuk mengurangi pembatalan mendadak.

Berdasarkan data, ditemukan pula distribusi jumlah pembatalan yang signifikan berdasarkan spesial request. Tidak optimalnya pemenuhan spesial request ini menjadi salah satu alasan utama tingginya angka pembatalan. Hal ini dapat dilihat pada distribusi pembatalan berdasarkan jenis spesial request yang diajukan oleh pelanggan **Gambar 3**.



Gambar 3. Histogram Jumlah Pembatalan Berdasarkan Special Request

2.2 Diagram BPMN



Gambar 4. Diagram BPMN Reservasi Hotel

Diagram pada **Gambar 4** diatas terdiri dari beberapa swimlanes yang membedakan peran dalam proses, yaitu: Pelanggan (Customer): Representasi dari aktivitas yang dilakukan oleh pelanggan. Reservation System/Front Desk (System or Reception): Berfokus pada proses sistem reservasi atau front desk yang menangani pemesanan. Management Hotel (Hotel Management): Berfokus pada langkah-langkah yang diambil oleh manajemen hotel. Data: Berisi data terkait, seperti data transaksi dan data tamu. Adapun tahapan proses yang terjadi adalah sebagai berikut:

- 1) Pelanggan mulai dengan melakukan reservasi melalui sistem.
- 2) Pengisian Informasi: Pelanggan mengisi informasi dasar yang dibutuhkan untuk pemesanan.

- 3) Memilih Kamar dan Fasilitas: Pelanggan memilih kamar yang sesuai dengan kebutuhan dan fasilitas tambahan.
- 4) Memilih Tanggal: Pelanggan menentukan tanggal check-in
- 5) Special Request: Pelanggan bisa menambahkan permintaan khusus, seperti fasilitas tambahan
- 6) Pemeriksaan Ketersediaan:
 - a) Cek Ketersediaan: Sistem atau front desk mengecek apakah kamar dan fasilitas yang diminta tersedia.
 - b) Jika tidak tersedia, pelanggan akan diberi saran pilihan lain, atau mereka dapat membatalkan pemesanan.
 - c) Jika tersedia, pelanggan melanjutkan ke tahap berikutnya
- 7) Transaksi dan Konfirmasi:
 - a) Memilih Jenis Deposit: Pelanggan memilih jenis deposit yang tersedia:
 - i) Non-Refundable: Tidak ada pengembalian dana
 - ii) Refundable: Ada opsi pengembalian dana tergantung kebijakan
 - iii) Tanpa deposit
 - b) Jika memilih refundable, pelanggan melakukan transaksi deposit.
 - c) Setelah deposit dipilih dan disetujui, sistem mengonfirmasi reservasi.
- 8) Pra-Kedatangan:
 - a) Informasi mengenai kebijakan pembatalan disampaikan kepada pelanggan berdasarkan jenis deposit yang dipilih.
 - b) Dua hari sebelum check-in, kamar disiapkan oleh manajemen hotel sesuai dengan reservasi.
- 9) Check-In dan Selama Menginap:
 - a) Check-In: Pelanggan datang pada tanggal yang telah ditentukan untuk melakukan check-in.
 - b) Informasi mengenai tanggal check-in diberikan kepada pelanggan sebelumnya.
 - c) Menginap: Pelanggan menginap sesuai dengan durasi yang telah disepakati.
- 10) Check-Out dan Penyimpanan Data:

- a) Check-Out: Pelanggan menyelesaikan proses check-out dan membayar biaya tambahan jika ada.
- b) Data transaksi disimpan dalam database hotel untuk kepentingan manajemen dan evaluasi.

Titik Potensi Masalah

Dalam diagram BPMN ini, ada beberapa titik potensi masalah yang bisa menyebabkan hambatan dalam proses, yaitu:

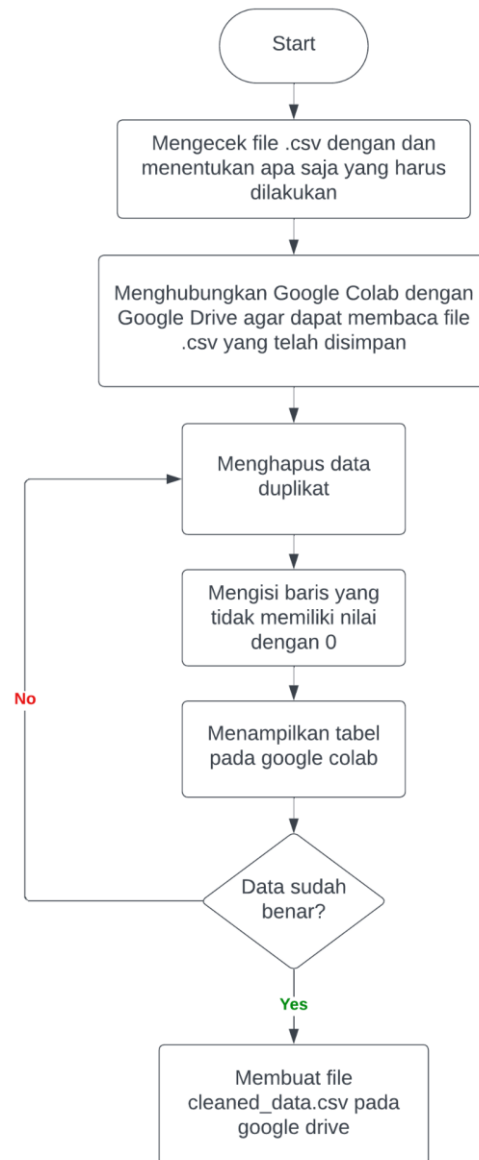
- 1. Spesial Request yang tidak terpenuhi bisa memicu pembatalan oleh pelanggan.
- 2. Kebijakan Pembatalan yang mungkin tidak fleksibel dapat meningkatkan tingkat pembatalan, terutama pada pemesanan refundable.
- 3. Pemeriksaan Ketersediaan yang tidak akurat bisa menyebabkan ketidakpuasan pelanggan.
- 4. Kurangnya pengingat pra-kedatangan dapat menyebabkan pelanggan lupa atau melakukan pembatalan mendadak.

BAB III

Data Preparation and Structured Query Language

3.1 Data Preparation

Awalnya, kami memeriksa file ``midterm_hotel_data.csv`` untuk melihat apakah ada data yang perlu dibersihkan. Kami menemukan bahwa beberapa kolom, misalnya ``lead_time``, memiliki banyak nilai kosong (null). Untuk memudahkan proses analisis data, kami memutuskan untuk mengisi nilai-nilai kosong tersebut dengan angka 0. Setelah itu, kami menghubungkan Google Colab dengan Google Drive agar bisa mengakses file CSV yang sudah kami simpan. Kemudian, kami melakukan data preparation, termasuk menghapus data yang duplikat dan mengisi nilai null dengan angka 0. Setelah memastikan bahwa data sudah dalam kondisi yang bersih dan siap diolah, kami menyimpan hasilnya ke dalam file ``cleaned_data.csv``. Untuk flowchart dapat dilihat pada Gambar 6.



Gambar 5. Flow data preparation yang dilakukan.

Kode Program 1. Proses *data cleansing* pada Python.

```

1. # Menghubungkan Google Drive dan membaca file midterm_hotel_data ke dalam DataFrame
2. from google.colab import drive
3. drive.mount('/content/drive')
4.
5. import pandas as pd
  
```

```
6.
7. file_path = '/content/drive/MyDrive/MSIB Batch 7/Hands on
   Assignment/MidTerm/midterm_hotel_data.csv'
8. df = pd.read_csv(file_path)
9. # Baris kode ini untuk menghapus semua duplikat
10. df.drop_duplicates()
11.
12. # Bagian kode ini untuk mengisi data yang awalnya kosong menjadi 0
13. df['lead_time'] = df['lead_time'].fillna(0)
14. df['stays_in_weekend_nights'] = df['stays_in_weekend_nights'].fillna(0)
15. df['adults'] = df['adults'].fillna(0)
16. df['agent'] = df['agent'].fillna(0)
17. df['company'] = df['company'].fillna(0)
18. df['adr'] = df['adr'].fillna(0)
19. df['total_of_special_requests'] = df['total_of_special_requests'].fillna(0)
20.
21. # Baris kode ini untuk menampilkan tabel
22. display(df)
23.
24. # Baris kode ini untuk membuat file baru bernama "cleaned_data" dimana datanya sudah
   bersih
25. df.to_csv('/content/drive/MyDrive/Studi Independen MSIB/Hands on
   Assignment/MidTerm/cleaned_data.csv', index=False)
```

Pada baris 1-10, kode ini digunakan untuk menghubungkan Google Drive, membaca file `midterm_hotel_data.csv`, serta mengimpor library pandas yang akan digunakan untuk analisis data. Baris 10 menjalankan perintah untuk menghapus semua duplikat dalam data, sehingga hanya data yang unik yang tersisa. Selanjutnya, pada baris 13-19, kode ini mengisi baris yang memiliki nilai kosong dengan angka 0, memastikan tidak ada nilai yang hilang dalam kolom tersebut. Baris 22 digunakan untuk menampilkan tabel yang sudah dibersihkan, sehingga kita dapat melihat hasilnya secara langsung. Terakhir, pada baris 25, data yang sudah dibersihkan disimpan kembali ke dalam file bernama `cleaned_data.csv` dan diletakkan di folder Google Drive yang sama, agar bisa diakses kapan saja.

3.2 Data Extraction

Anda dapat menjelaskan setiap nomor pada Module SQL pada bagian ini. Penggunaan sub-heading juga diperbolehkan jika memang dibutuhkan. Untuk lampiran *query* yang digunakan, bisa dilihat pada potongan Kode Program 2 berikut:

Kode Program 1. Pembatalan Reservasi setiap Jenis Hotel Berdasarkan Tahun

```
1. SELECT
2.   hotel,
3.   arrival_date_year,
4.   ROUND(SUM(CASE WHEN is_canceled = 1 THEN 1 ELSE 0 END) * 100.0 / COUNT(*), 2) AS
   cancellation_rate
5. FROM
6.   data_hotel
7. GROUP BY
8.   hotel,
9.   arrival_date_year
10. ORDER BY
11.   hotel,
12.   arrival_date_year;
```

Pada soal nomor 1 module 3 kami diminta menganalisis tingkat pembatalan reservasi, kita perlu menghitung persentase pembatalan untuk setiap jenis hotel berdasarkan tahun kedatangan. Query ini akan menampilkan hasil dalam bentuk persentase pembatalan terhadap total reservasi per tahun dan per hotel. Dari baris 1-4 ini memilih kolom hotel dan arrival_date_year, lalu menghitung tingkat pembatalan dengan menggunakan SUM untuk menjumlahkan nilai 1 ketika is_canceled adalah 1 (reservasi dibatalkan), membaginya dengan total reservasi yang dihitung menggunakan COUNT(*), dan mengalikan hasilnya dengan 100 untuk mendapatkan persentase, kemudian membulatkannya hingga dua desimal dengan ROUND. Lalu baris ke 5-6 menentukan tabel data_hotel sebagai sumber data dari mana query akan menarik informasi. Selanjutnya baris ke 7-9 kita mengelompokkan data berdasarkan kolom hotel dan arrival_date_year, sehingga perhitungan pembatalan dilakukan untuk setiap kombinasi hotel dan tahun kedatangan secara terpisah. Dan baris ke 10-12 kita akan

mengurutkan hasil akhir query berdasarkan kolom hotel dan arrival_date_year, menampilkan data dalam urutan yang terstruktur menurut jenis hotel dan tahun.

Berikut adalah output dari query yang sudah kita jalankan

data_hotel 1 ×

SELECT hotel, arrival_date_year, F Enter a SQL expression to filter results (use C

	A-Z hotel	123 arrival_date_year	123 cancellation_rate
1	City Hotel	2,015	43.88
2	City Hotel	2,016	40.4
3	City Hotel	2,017	42.5
4	Resort Hotel	2,015	25.72
5	Resort Hotel	2,016	26.55
6	Resort Hotel	2,017	30.76

Gambar 6. Pembatalan Reservasi Setiap Jenis Hotel Setiap Tahun

Pada output tersebut menunjukkan bahwa kolom tersebut menjelaskan antara lain:

1. hotel: Menyebutkan jenis hotel yang tercatat dalam data, yaitu *City Hotel* dan *Resort Hotel*.
2. arrival_date_year: Menyatakan tahun kedatangan pelanggan, mulai dari 2015 hingga 2017.
3. cancellation_rate: Menunjukkan persentase pembatalan reservasi untuk masing-masing jenis hotel di setiap tahun tertentu. Persentase ini menunjukkan seberapa sering reservasi dibatalkan

Interpretasi dari data tersebut adalah:

1. *City Hotel* memiliki tingkat pembatalan yang lebih tinggi dibandingkan *Resort Hotel* untuk setiap tahunnya
2. Tingkat pembatalan *City Hotel* cenderung fluktuatif, dengan sedikit penurunan pada tahun 2016 dan kenaikan lagi pada tahun 2017
3. *Resort Hotel* menunjukkan tren kenaikan tingkat pembatalan dari tahun ke tahun, meskipun tidak sebesar kenaikan pada *City Hotel*

Kode Program 2. Rata-Rata Waktu Tunggu Sebelum Check-in

```
1. WITH avg_lead_time_cte AS (  
2.     SELECT AVG(lead_time) AS overall_avg_lead_time  
3.     FROM data_hotel  
4. )  
5. SELECT  
6.     total_of_special_requests,  
7.     ROUND(CAST(AVG(lead_time) AS NUMERIC), 2) AS average_lead_time  
8. FROM  
9.     data_hotel  
10. WHERE  
11.     lead_time > (SELECT overall_avg_lead_time FROM avg_lead_time_cte)  
12. GROUP BY  
13.     total_of_special_requests  
14. ORDER BY  
15.     total_of_special_requests;
```

Pada soal nomor 2 module 3 kami diminta untuk menghitung rata-rata waktu tunggu sebelum check-in (lead_time) berdasarkan kelompok jumlah permintaan khusus (total_of_special_requests), hanya data dengan lead time di atas rata-rata keseluruhan yang akan disertakan dalam perhitungan. Pada baris 1-4 kita menggunakan CTE ini menghitung rata-rata keseluruhan dari kolom lead_time di tabel data_hotel dan menyimpannya sebagai overall_avg_lead_time. CTE ini digunakan untuk memberikan referensi rata-rata waktu lead yang dapat dipakai pada bagian query utama. Selanjutnya pada query utama yaitu baris ke 5- 15 Query ini memilih kolom total_of_special_requests dan menghitung rata-rata lead_time (dalam format numerik yang dibulatkan hingga dua desimal) untuk setiap jumlah permintaan khusus dengan mengambil data dari tabel data_hotel, memfilter hanya reservasi yang memiliki lead_time lebih besar dari rata-rata keseluruhan yang dihitung sebelumnya dalam CTE, mengelompokkan hasil berdasarkan jumlah permintaan khusus, dan mengurutkan hasil berdasarkan nilai total_of_special_requests agar lebih mudah dibaca.

Berikut adalah output dari query yang sudah kita jalankan

data_hotel 1 X

WITH avg_lead_time_cte AS (SEL | Enter a SQL expression to filter results (use Ctrl+Space)

	123 total_of_special_requests	123 average_lead_time	
1	0	193.78	
2	1	172.72	
3	2	166.56	
4	3	163.26	
5	4	182.83	
6	5	171.55	

Gambar 7. Rata-Rata Waktu Lead Time Berdasarkan Jumlah Special Request

Berikut adalah penjelasan dari setiap kolomnya:

1. total_of_special_requests: Menunjukkan jumlah permintaan khusus yang dibuat oleh pelanggan untuk setiap reservasi, dari 0 (tanpa permintaan khusus) hingga 5 permintaan khusus
2. average_lead_time: Menyatakan waktu rata-rata antara tanggal pemesanan dan tanggal kedatangan (lead time) dalam hari, berdasarkan jumlah permintaan khusus

Dari data ini dapat dilihat bahwa Lead time cenderung menurun seiring dengan bertambahnya jumlah permintaan khusus. Pelanggan dengan lebih banyak permintaan khusus memiliki waktu pemesanan yang lebih dekat dengan tanggal kedatangan

Kode Program 3. Tanggal Reservasi dengan Pendapatan Tertinggi

```

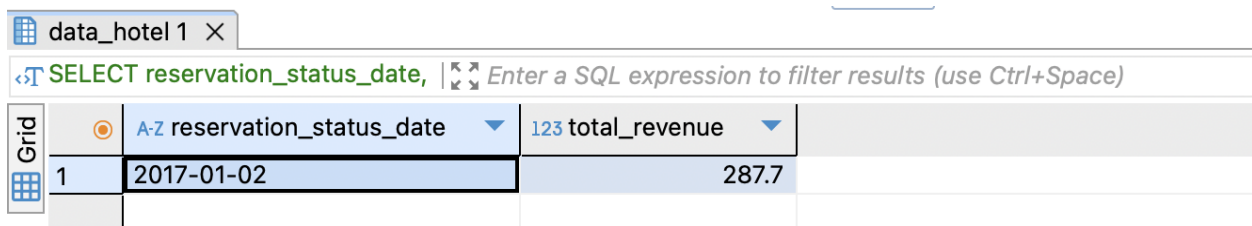
1. SELECT
2.     reservation_status_date,
3.     ROUND(AVG(adr), 2) AS total_revenue
4. FROM
5.     data_hotel
6. WHERE
7.     reservation_status = 'Check-Out'
8.     AND total_of_special_requests > 2
9. GROUP BY
10.    reservation_status_date
11. ORDER BY

```

```
12.     total_revenue DESC
13. LIMIT 1;
```

Pada soal nomor 3 module 3 kami kami untuk menemukan tanggal reservasi dengan pendapatan rata-rata harian tertinggi dari reservasi yang telah selesai (status "Check-Out") dan memiliki lebih dari dua permintaan khusus. Pendapatan dihitung sebagai rata-rata tarif harian (ADR) per tanggal reservasi untuk memberikan gambaran tanggal dengan performa terbaik berdasarkan kriteria ini. Pada baris 1-3 kita memilih tanggal reservasi (reservation_status_date) dan menghitung rata-rata tarif harian (adr) yang dibulatkan hingga dua desimal sebagai total pendapatan harian (total_revenue). Selanjutnya baris 4-5 kita mengambil data dari tabel data_hotel. Lalu baris 6-8 kita memfilter hanya untuk reservasi yang berstatus "Check-Out" dan memiliki lebih dari dua permintaan khusus (total_of_special_requests). Baris 9-12 kita mengelompokkan hasil berdasarkan tanggal reservasi untuk mendapatkan rata-rata tarif harian per tanggal dan mengurutkan hasil dari pendapatan tertinggi ke terendah. Dan pada baris ke 13 kita ambil hanya satu hasil teratas, yaitu tanggal dengan pendapatan rata-rata tertinggi.

Berikut adalah output dari query yang sudah kita jalankan



	A-Z reservation_status_date	123 total_revenue
1	2017-01-02	287.7

Gambar 8. Tanggal Reservasi Dengan Total Revenue Tertinggi

Berikut adalah penjelasan dari setiap kolomnya:

1. reservation_status_date: Kolom ini menunjukkan tanggal ketika status reservasi terakhir diperbarui. Dalam contoh ini, hanya ada satu tanggal yang ditampilkan, yang tertinggi yaitu 2017-01-02.
2. total_revenue: Kolom ini menunjukkan total pendapatan yang dihasilkan pada tanggal reservasi tersebut. Dalam kasus ini, total pendapatan pada tanggal 2017-01-02 adalah 287.7

Jadi berdasarkan data yang ditampilkan, kita dapat menyimpulkan bahwa pada tanggal 2017-01-02 , hotel tersebut menghasilkan pendapatan tertinggi sebesar 287.7

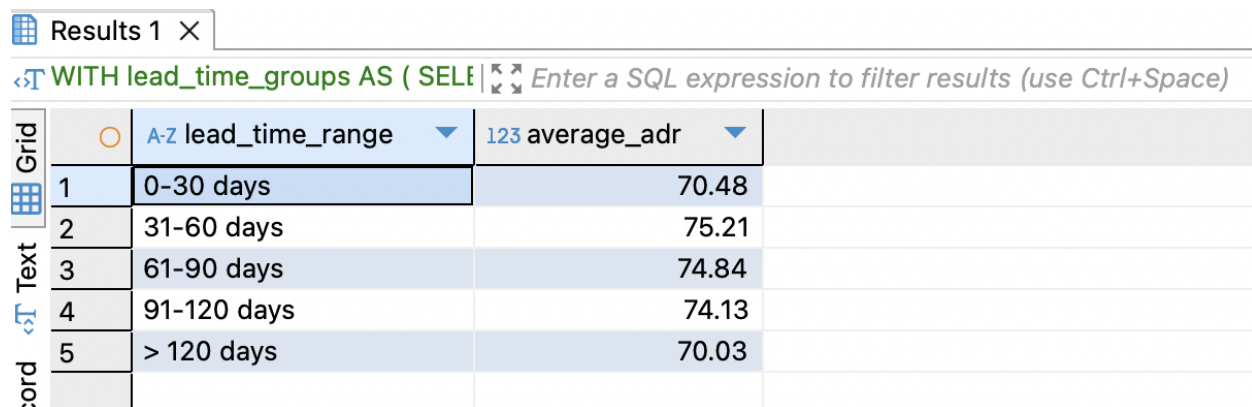
Kode Program 4. Hubungan Antara lead-time dan Keberhasilan Pembayaran

```
1. WITH lead_time_groups AS (  
2.     SELECT  
3.         CASE  
4.             WHEN lead_time BETWEEN 0 AND 30 THEN '0-30 days'  
5.             WHEN lead_time BETWEEN 31 AND 60 THEN '31-60 days'  
6.             WHEN lead_time BETWEEN 61 AND 90 THEN '61-90 days'  
7.             WHEN lead_time BETWEEN 91 AND 120 THEN '91-120 days'  
8.             WHEN lead_time > 120 THEN '> 120 days'  
9.             ELSE 'Unknown'  
10.        END AS lead_time_range,  
11.        adr  
12.    FROM data_hotel  
13. )  
14. SELECT  
15.     lead_time_range,  
16.     ROUND(AVG(adr), 2) AS average_adr  
17. FROM  
18.     lead_time_groups  
19. GROUP BY  
20.     lead_time_range  
21. ORDER BY  
22.     lead_time_range;
```

Pada soal nomor 4 module 3 kita diminta untuk menganalisis hubungan antara lead time dan keberhasilan pembayaran, kita mengelompokkan data dalam rentang lead time tertentu dan hitung rata-rata nilai pembayaran (adr) untuk setiap rentang. Analisis ini akan membantu mengidentifikasi apakah lead time mempengaruhi tingkat pembayaran yang berhasil diterima. Pada baris 1-9 kita menggunakan CTE lead_time_groups membuat tabel sementara yang mengelompokkan nilai lead_time dari tabel data_hotel ke dalam rentang waktu tertentu: 0-30 hari, 31-60 hari, 61-90 hari, 91-120 hari, dan lebih dari 120 hari, sementara nilai yang tidak sesuai kategori diberi label 'Unknown'. Pada baris 10-22

menjelaskan mengenai kolom `adr` (average daily rate) diambil dari tabel `data_hotel`, dimana bagian utama query ini memilih dua kolom—`lead_time_range`, yang berisi rentang waktu pemesanan yang telah ditentukan, dan `ROUND(AVG(adr), 2) AS average_adr`, yang menghitung rata-rata nilai `adr` untuk setiap rentang waktu pemesanan dan membulatkannya hingga dua desimal menggunakan fungsi `AVG` dan `ROUND`; selanjutnya, query ini menggunakan pernyataan `GROUP BY` untuk mengelompokkan hasil berdasarkan `lead_time_range`, sehingga semua nilai `adr` dalam rentang waktu yang sama akan digabungkan untuk menghasilkan rata-rata yang sesuai, dan hasil akhirnya diurutkan menggunakan `ORDER BY lead_time_range`, yang menyusun hasil berdasarkan urutan dari rentang waktu yang telah ditentukan untuk memudahkan pembacaan dan analisis.

Berikut adalah output dari query yang sudah kita jalankan



The screenshot shows a database query results window titled "Results 1". The query is: `WITH lead_time_groups AS (SELECT ...)`. The results are displayed in a table with two columns: `lead_time_range` and `average_adr`. The table has 5 rows of data.

	lead_time_range	average_adr
1	0-30 days	70.48
2	31-60 days	75.21
3	61-90 days	74.84
4	91-120 days	74.13
5	> 120 days	70.03

Gambar 9. Hubungan Antara Rata-Rata Lead Time dengan Rata-Rata ADR

Berikut adalah penjelasan dari setiap kolomnya:

1. `lead_time_range`: Kolom ini adalah jangka waktu antara saat sebuah pesanan atau reservasi dilakukan hingga saat layanan atau produk tersebut benar-benar diberikan. Dalam konteks hotel, ini bisa diartikan sebagai waktu antara saat tamu melakukan pemesanan hingga tanggal kedatangan
2. `average_adr`: Ini adalah rata-rata pendapatan per kamar yang tersedia setiap harinya

Interpretasi datanya adalah:

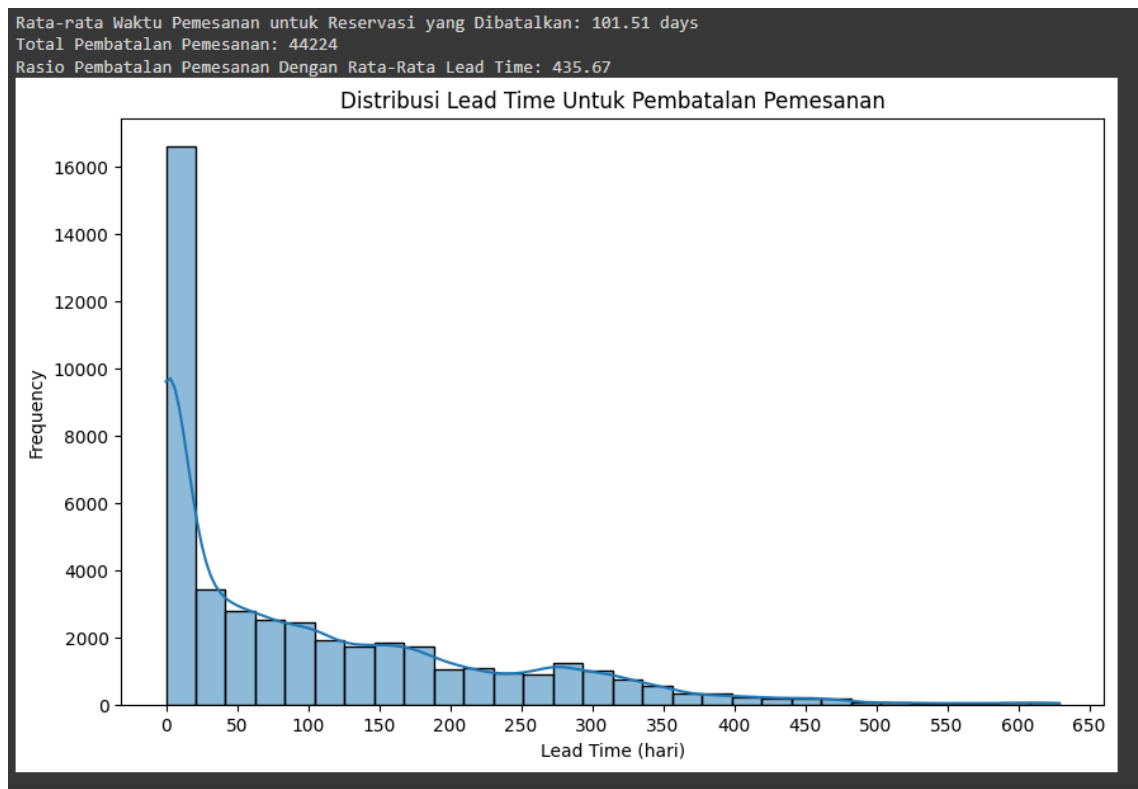
1. Rentang waktu 31-60 hari memiliki ADR tertinggi. Ini bisa mengindikasikan bahwa tamu yang memesan kamar dengan jangka waktu pemesanan sekitar 1-2 bulan cenderung membayar harga yang lebih tinggi
2. Tamu yang memesan kamar dengan lead time sangat pendek (0-30 hari) cenderung membayar harga yang sedikit lebih rendah. Ini bisa terjadi karena mereka mungkin melakukan pemesanan mendadak atau memanfaatkan diskon last minute
3. Tamu yang memesan kamar jauh-jauh hari (lebih dari 120 hari) juga cenderung membayar harga yang sedikit lebih rendah

BAB IV

Python Programming

4.1 Exploratory Data Analysis

Pada soal nomor 1 module 4 kami diminta untuk menunjukkan rasio antara lead_time dengan cancellation.



Gambar 10. Distribusi Lead Time Untuk Pembatalan Pemesanan

Grafik di atas menunjukkan distribusi waktu pemesanan (lead time) untuk reservasi yang dibatalkan pada sebuah hotel. Sumbu x menggambarkan jumlah hari antara tanggal pemesanan dan tanggal kedatangan (lead time), sementara sumbu y menunjukkan frekuensi pembatalan untuk setiap rentang lead time tersebut. Mayoritas pembatalan terjadi pada lead time yang sangat pendek, terutama di bawah 50 hari, menunjukkan bahwa banyak tamu yang membatalkan pemesanan mereka dalam jangka waktu dekat sebelum tanggal kedatangan. Frekuensi pembatalan menurun drastis seiring bertambahnya lead time, namun tetap ada

beberapa pembatalan yang terjadi bahkan hingga lebih dari 600 hari sebelum tanggal kedatangan. Rata-rata lead time untuk pembatalan adalah sekitar 101,51 hari, dan total pembatalan yang terjadi adalah 44.224. Rasio pembatalan dengan rata-rata lead time mencapai sekitar 435,67, yang mungkin menunjukkan korelasi antara semakin jauh lead time, semakin kecil peluang pembatalan.

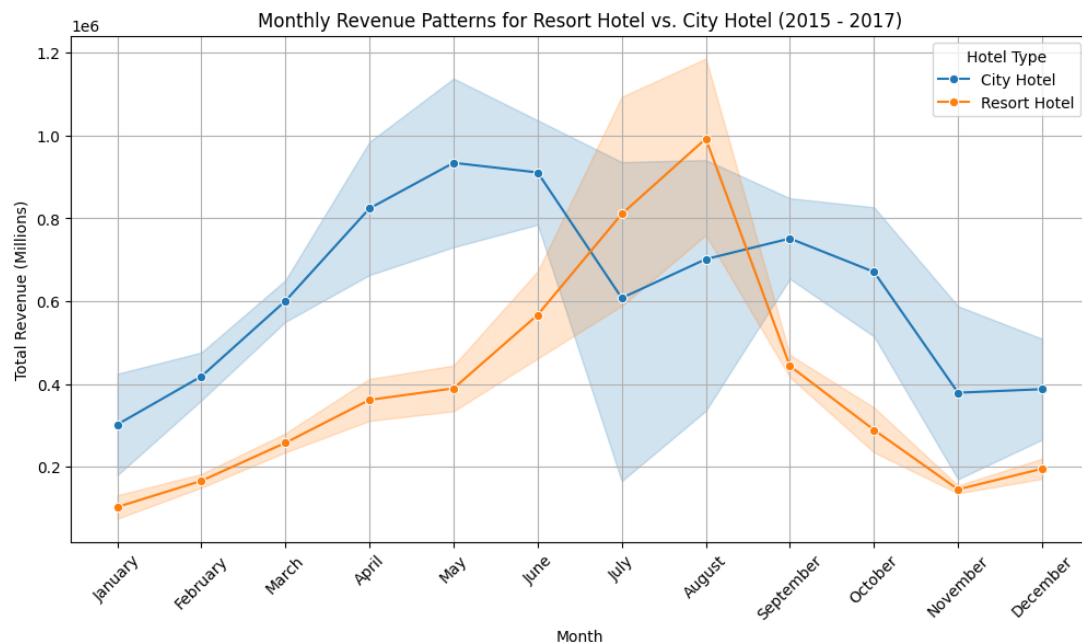
Kode Program 6. Distribusi Lead Time Untuk Pembatalan Pemesanan.

```
1. # Nomor 1
2. canceled_reservations = df[df['is_canceled'] == 1]
3.
4. average_lead_time_canceled = canceled_reservations['lead_time'].mean()
5. total_canceled = canceled_reservations['lead_time'].count()
6.
7. ratio = total_canceled / average_lead_time_canceled
8.
9. print(f"Rata-rata Waktu Pemesanan untuk Reservasi yang Dibatalkan:
    {average_lead_time_canceled:.2f} days")
10. print(f"Total Pembatalan Pemesanan: {total_canceled}")
11. print(f"Rasio Pembatalan Pemesanan Dengan Rata-Rata Lead Time: {ratio:.2f}")
12.
13. plt.figure(figsize=(10, 6))
14. sns.histplot(canceled_reservations['lead_time'], bins=30, kde=True)
15. plt.title('Distribusi Lead Time Untuk Pembatalan d Pemesanan')
16. plt.xlabel('Lead Time (hari)')
17. plt.ylabel('Frequency')
18.
19. plt.xticks(range(0, int(canceled_reservations['lead_time'].max()) + 50, 50))
20.
21. plt.show()
```

Pada baris 2, kode memfilter data untuk mendapatkan semua pemesanan yang dibatalkan dengan mengacu pada kolom `is_canceled`. Pada baris 4, kode menghitung rata-rata waktu pemesanan (lead time) dari reservasi yang dibatalkan menggunakan method `.mean()`, dan pada baris 5, menghitung total jumlah pembatalan menggunakan `.count()`. Baris 7 menghitung rasio antara total pembatalan dan rata-rata lead time dengan membagi jumlah total pembatalan dengan rata-rata lead time tersebut. Pada baris 9-11, kode menampilkan hasil analisis berupa rata-rata waktu pemesanan, total pembatalan, dan rasio

pembatalan menggunakan f-string. Selanjutnya, baris 13-17 membuat visualisasi distribusi waktu pemesanan menggunakan histogram dengan 30 bins, dilengkapi dengan kurva KDE (Kernel Density Estimate) untuk melihat pola distribusi secara lebih halus. Label sumbu x diatur sebagai 'Lead Time (hari)' dan sumbu y sebagai 'Frequency'. Pada baris 19, label sumbu x disesuaikan dengan interval 50 hari untuk meningkatkan keterbacaan grafik, dan grafik ditampilkan dengan `plt.show()` pada baris 21.

Pada soal nomor 2 module 4 kami diminta untuk menentukan apakah ada perbedaan pola pendapatan antara resort hotel dan city hotel dalam bulan tertentu?



Gambar 11. Pola Pendapatan Resort Hotel dan City Hotel dari 2015 hingga 2017

Berdasarkan grafik tersebut, terdapat pola pendapatan yang berbeda secara signifikan antara hotel resort dan hotel kota, dengan tingkat variabilitas yang ditunjukkan oleh area bayangan (confidence interval) di sekitar garis tren. City Hotel mencapai pendapatan tertinggi pada Mei-Juni (sekitar 900 ribu), dengan area bayangan yang relatif sempit menunjukkan konsistensi pendapatan yang baik, sementara Resort Hotel memuncak di bulan Agustus (hampir 1 juta) dengan variabilitas yang lebih tinggi ditandai oleh area bayangan yang lebih lebar. Perbedaan paling kontras terlihat di awal tahun (Januari-April) dimana City Hotel

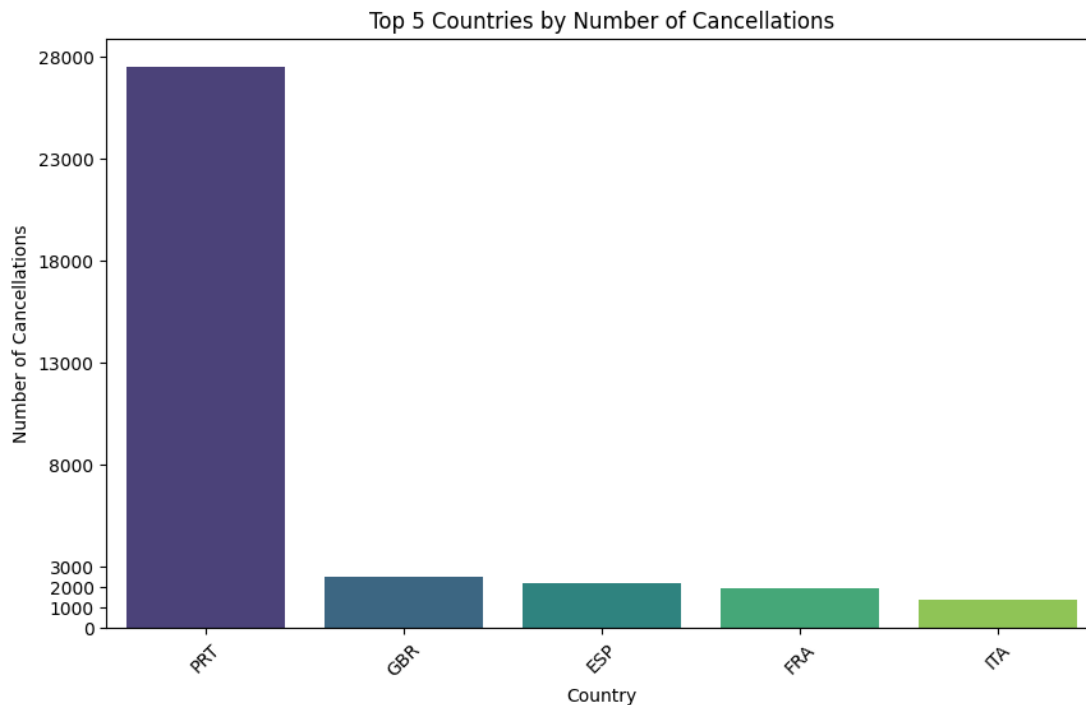
memiliki pendapatan yang lebih tinggi dibandingkan Resort Hotel. Situasi berbalik di bulan Agustus, dimana Resort Hotel mengalami lonjakan pendapatan signifikan dengan tingkat variabilitas yang meningkat, sementara City Hotel justru mengalami penurunan namun dengan konsistensi yang tetap terjaga seperti ditunjukkan oleh area bayangan yang lebih sempit. Di akhir tahun (November-Desember), kedua jenis hotel mengalami penurunan pendapatan dengan tingkat variabilitas yang relatif rendah, namun City Hotel tetap mempertahankan pendapatan yang lebih tinggi dibandingkan hotel resort, menunjukkan pola musiman yang lebih stabil untuk City Hotel sepanjang tahun.

Kode Program 7. Pola Pendapatan Resort Hotel dan City Hotel dari 2015 hingga 2017

```
1. # Nomor 2
2. # Membuat kolom baru untuk total revenue per reservation
3. df['total_revenue'] = df['adr'] * (df['stays_in_weekend_nights'] +
   df['stays_in_week_nights'])
4.
5. # Group by year, month, dan hotel untuk menghitung total revenue per bulan
6. monthly_revenue = df.groupby(['arrival_date_year', 'arrival_date_month',
   'hotel'])['total_revenue'].sum().reset_index()
7.
8. # Mengurutkan Bulan
9. month_order = ['January', 'February', 'March', 'April', 'May', 'June',
10.               'July', 'August', 'September', 'October', 'November', 'December']
11. monthly_revenue['arrival_date_month'] =
   pd.Categorical(monthly_revenue['arrival_date_month'], categories=month_order,
   ordered=True)
   a.
12. # Visualisasi
13. plt.figure(figsize=(12, 6))
14. sns.lineplot(data=monthly_revenue, x='arrival_date_month', y='total_revenue',
   hue='hotel', marker='o')
15. plt.title('Monthly Revenue Patterns for Resort Hotel vs. City Hotel')
16. plt.xlabel('Month')
17. plt.ylabel('Total Revenue (Millions)')
18. plt.xticks(rotation=45)
19. plt.grid()
20. plt.legend(title='Hotel Type')
21. plt.show()
```

Pada baris 3, kode membuat kolom baru 'total_revenue' dengan mengalikan kolom 'adr' dengan total durasi menginap (penjumlahan 'stays_in_weekend_nights' dan 'stays_in_week_nights'). Selanjutnya, pada baris 6, data dikelompokkan menggunakan method groupby dari pandas berdasarkan tiga kolom ('arrival_date_year', 'arrival_date_month', 'hotel') dan menghitung total pendapatan bulanan dengan fungsi sum(), kemudian hasilnya direset indeksnya menjadi dataframe biasa. Pada baris 9-11, kode membuat list urutan bulan dan mengkonversi kolom 'arrival_date_month' menjadi tipe data Categorical dengan urutan yang telah ditentukan untuk memastikan visualisasi menampilkan bulan secara berurutan. Baris 14-15 memulai pembuatan visualisasi dengan mengatur ukuran figure menjadi 12x6 inci dan membuat line plot menggunakan seaborn dengan parameter data=monthly_revenue, sumbu x menampilkan bulan, sumbu y menampilkan total revenue dalam satuan millions atau juta, dan hue untuk membedakan tipe hotel. Terakhir, pada baris 16-22, kode menyempurnakan visualisasi dengan menambahkan judul grafik, label sumbu, merotasi label bulan 45 derajat, menambahkan grid dan legend, serta menampilkan hasilnya dengan plt.show().

Pada soal nomor 3 module 4 kami diminta untuk menentukan turis dari negara mana yang paling sering melakukan reservation cancellation?



Gambar 12. Urutan Asal Negara Turis yang Paling Sering Membatalkan Pemesanan

Bar chart ini menunjukkan urutan negara yang paling sering membatalkan reservasi hotel. Turis dari PRT (Portugal) berada di posisi pertama dengan 27.519 pembatalan. Berikutnya adalah GBR (Britania Raya) dengan 2.453 pembatalan, diikuti oleh ESP (Spanyol) dengan 2.177 pembatalan. Peringkat keempat diduduki oleh FRA (Prancis) dengan 1.934 pembatalan, dan posisi kelima ditempati oleh ITA (Italia) dengan 1.333 pembatalan.

Kode Program 8. Urutan Asal Negara Turis yang Paling Sering Membatalkan Pemesanan

```

1. # Nomor 3
2. # Mengambil data pembatalan
3. canceled_reservations = df[df['is_canceled'] == 1]
4.
5. # Mengelompokkan berdasarkan negara dan menghitung jumlah pembatalan
6. cancellation_by_country =
   canceled_reservations['country'].value_counts().reset_index()
7.
8. # Menamai kolom agar lebih jelas
9. cancellation_by_country.columns = ['Country', 'Number of Cancellations']
10.

```

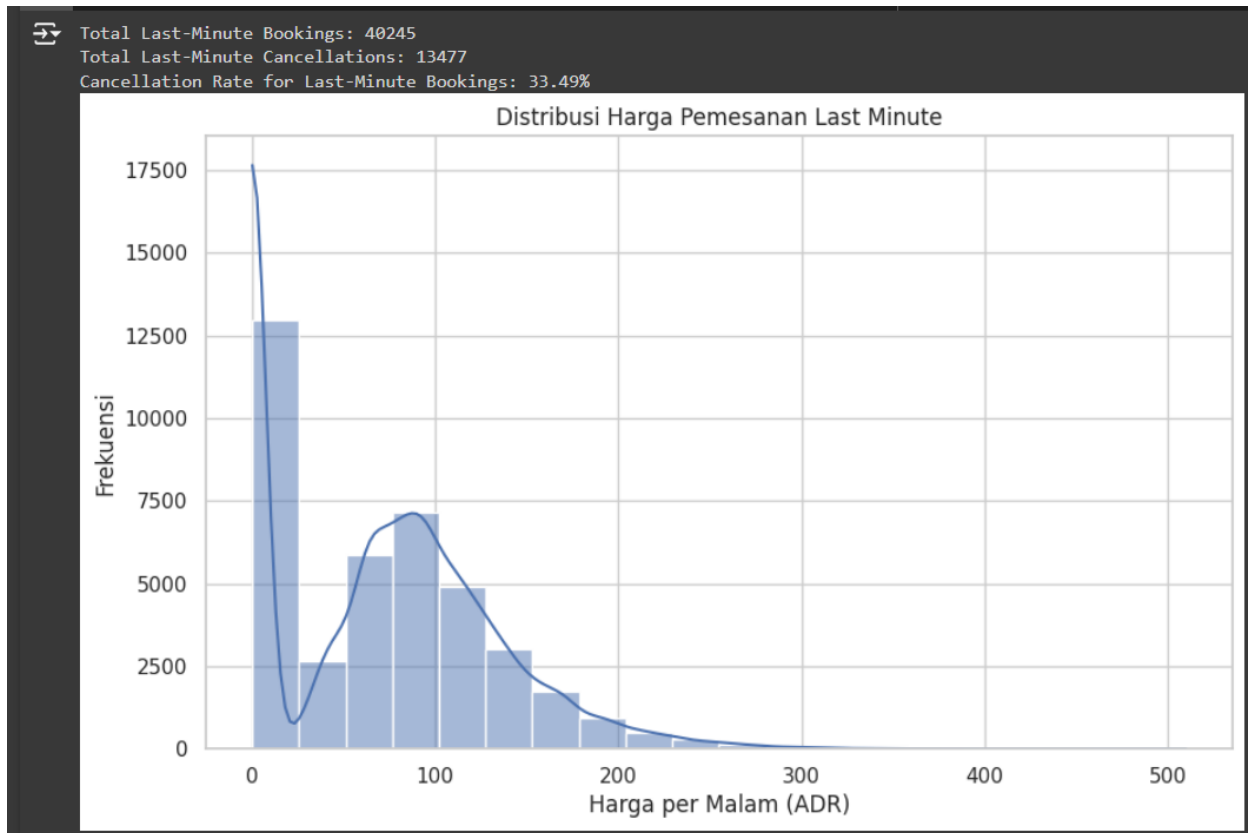
```

11. # Mengambil 10 negara dengan jumlah pembatalan terbanyak
12. top_5_countries = cancellation_by_country.head(5)
13.
14. # Mengatur ukuran grafik
15. plt.figure(figsize=(10, 6))
16.
17. # Membuat bar chart
18. sns.barplot(data=top_5_countries, x='Country', y='Number of Cancellations',
    palette='viridis')
19.
20. # Menambahkan judul dan label
21. plt.title('Top 5 Countries by Number of Cancellations')
22. plt.xlabel('Country')
23. plt.ylabel('Number of Cancellations')
24.
25. # Memutar label di sumbu x untuk memudahkan pembacaan
26. plt.xticks(rotation=45)
27.
28. max_value = top_5_countries['Number of Cancellations'].max()
29. yticks = list(range(0, 3000, 1000)) + list(range(3000, max_value + 5000, 5000))
30.
31. # Mengatur rentang pada sumbu y
32. plt.yticks(yticks)
33.
34. # Menampilkan grafik
35. plt.show()

```

Pada baris 3, kode mengambil data pemesanan yang dibatalkan dengan memfilter berdasarkan kolom `is_canceled` yang bernilai 1. Pada baris 6-9, data dikelompokkan berdasarkan negara menggunakan `value_counts()` untuk menghitung jumlah pembatalan per negara, kemudian hasilnya diubah menjadi DataFrame dengan kolom `Country` dan `Number of Cancellations`. Pada baris 12, kode mengambil 5 negara dengan jumlah pembatalan terbanyak menggunakan `head(5)` dan mengatur ukuran grafik dengan `plt.figure()`. Pada baris 18-23, kode membuat bar chart untuk memvisualisasikan jumlah pembatalan, mengatur palet warna menjadi 'viridis', dan menambahkan judul serta label sumbu x dan y. Pada baris 26, sumbu x diputar 45 derajat untuk memudahkan pembacaan. Pada baris 29, kode mengatur rentang sumbu y, menampilkan angka dengan interval per 1000 hingga 3000 dan setelahnya menggunakan interval per 5000 hingga mencapai nilai maksimum. Terakhir, pada baris 35, kode menggunakan `plt.show()` untuk menampilkan grafik tersebut.

Pada nomor 4 module 4 soal meminta kami untuk menentukan bagaimana perilaku pemesanan last minute?



Gambar 13. Distribusi Harga Pemesanan (ADR) Last Minute

Grafik ini menunjukkan distribusi harga kamar (ADR - Average Daily Rate) untuk pemesanan last-minute, dengan total 44.113 pemesanan dan tingkat pembatalan 31.42% (13.868 pembatalan). Distribusi harga menunjukkan pola yang menarik dimana mayoritas pemesanan last-minute terjadi di rentang harga sangat rendah (sekitar 0-50), dengan frekuensi mencapai sekitar 14.000 pemesanan. Terdapat pola kedua yang menonjol di rentang harga 75-150 dengan puncak sekitar 7.500 pemesanan. Setelah rentang harga 200, frekuensi pemesanan menurun drastis dan terus mengalami penurunan hingga harga 500. Pola ini mengindikasikan bahwa pemesanan last-minute cenderung didominasi oleh kamar-kamar dengan harga ekonomis, mungkin sebagai strategi hotel untuk mengisi kamar kosong atau strategi tamu untuk mendapatkan harga yang lebih terjangkau.

Kode Program 9. Distribusi Harga Pemesanan (ADR) Last Minute

```
1. # Nomor 4
2. # Tentukan threshold untuk pemesanan last minute
3. last_minute_threshold = 0 # dalam hari
4.
5. # Filter untuk pemesanan last minute
6. last_minute_bookings = df[df['lead_time'] <= last_minute_threshold]
7.
8. # Hitung total pemesanan last minute
9. total_last_minute_bookings = last_minute_bookings['lead_time'].count()
10.
11. # Hitung total pembatalan pemesanan last minute
12. last_minute_cancellations = last_minute_bookings[last_minute_bookings['is_canceled']
    == 1]
13. total_last_minute_cancellations = last_minute_cancellations['lead_time'].count()
14.
15. # Hitung tingkat pembatalan pemesanan last minute
16. if total_last_minute_bookings > 0:
17.     cancellation_rate = (total_last_minute_cancellations /
        total_last_minute_bookings) * 100
18. else:
19.     cancellation_rate = 0
20.
21. # Cetak hasil analisis
22. print(f"Total Last-Minute Bookings: {total_last_minute_bookings}")
23. print(f"Total Last-Minute Cancellations: {total_last_minute_cancellations}")
24. print(f"Cancellation Rate for Last-Minute Bookings: {cancellation_rate:.2f}%")
25.
26. # Analisis karakteristik pemesanan last minute
27. plt.figure(figsize=(10, 6))
28. sns.histplot(last_minute_bookings['adr'], bins=20, kde=True)
29. plt.title('Distribusi Harga Pemesanan Last Minute')
30. plt.xlabel('Harga per Malam (ADR)')
31. plt.ylabel('Frekuensi')
32. plt.show()
```

Pada baris 3-6, kode menetapkan threshold 0 hari untuk menentukan kriteria pemesanan last minute dan memfilter data menggunakan boolean indexing untuk mendapatkan pemesanan yang sesuai kriteria. Selanjutnya, pada baris 9-13, kode menghitung total pemesanan last minute menggunakan method count() dan memfilter kembali untuk mendapatkan jumlah pembatalan dengan kondisi 'is_canceled' sama dengan 1. Pada baris 16-19, kode menghitung

tingkat pembatalan dengan membagi jumlah pembatalan dengan total pemesanan dan mengalikan dengan 100 untuk mendapatkan persentase, dengan pengecualian jika total pemesanan 0. Baris 22-24 mencetak hasil analisis menggunakan f-string untuk menampilkan total pemesanan, total pembatalan, dan tingkat pembatalan dengan format yang rapi. Terakhir, pada baris 27-32, kode membuat visualisasi menggunakan seaborn histplot dengan ukuran figure 10x6, menampilkan distribusi harga ADR untuk pemesanan last minute dengan 20 bins dan kurva density, serta dilengkapi dengan judul, label sumbu x dan y, dan perintah `plt.show()` untuk menampilkan grafik.

4.2 A/B Testing

Berdasarkan hasil pengujian A/B testing yang telah dilakukan untuk membandingkan rata-rata ADR (Average Daily Rate) antara City Hotel dan Resort Hotel, kami menggunakan kode Python yang memisahkan data menjadi dua kelompok dan melakukan uji statistik menggunakan independent t-test dengan ``equal_var=False``. Dari hasil pengujian, kami mendapatkan nilai T-statistic sebesar 17.86 dan P-value yang sangat kecil yaitu 0.0000. Dengan tingkat signifikansi 5% (0.05), p-value yang jauh lebih kecil ini memberikan bukti statistik yang sangat kuat bahwa terdapat perbedaan yang signifikan antara rata-rata ADR kedua jenis hotel, sehingga kita menolak hipotesis nol (H_0) dan menerima hipotesis alternatif (H_1). Kesimpulan ini didukung oleh nilai T-statistic yang positif dan besar, yang mengindikasikan perbedaan substansial antara rata-rata kedua kelompok, menunjukkan bahwa perbedaan harga antara City Hotel dan Resort Hotel bukanlah hasil dari kebetulan atau variasi acak.

```
Hasil A/B Testing:
-----
Statistik Uji (T-statistic): 17.86
Nilai-P (P-value): 0.0000

Keputusan:
-----
Karena nilai-p < 0.05, maka MENOLAK  $H_0$ 
Kesimpulan: Terdapat perbedaan yang signifikan pada rata-rata ADR antara City Hotel dan Resort Hotel
Hipotesis yang diterima:  $H_1$ 
```

Gambar 14. Hasil A/B Testing

Kode Program 10. A/B Testing

```
1. # Memisahkan data ADR untuk kedua kelompok hotel
2. # Kelompok A: City Hotel
3. # Kelompok B: Resort Hotel
4. adr_hotel_city = df[df['hotel'] == 'City Hotel']['adr']
5. adr_hotel_resort = df[df['hotel'] == 'Resort Hotel']['adr']
6.
7. # Melakukan A/B Testing menggunakan independent t-test
8. nilai_t, nilai_p = stats.ttest_ind(adr_hotel_city, adr_hotel_resort,
    equal_var=False)
9.
10. # Mencetak hasil pengujian
11. print("Hasil A/B Testing:")
12. print("-"*50)
13. print(f"Statistik Uji (T-statistic): {nilai_t:.2f}")
14. print(f"Nilai-P (P-value): {nilai_p:.4f}")
15.
16. # Menetapkan tingkat signifikansi
17. alpha = 0.05
18.
19. # Membuat keputusan
20. print("\nKeputusan:")
21. print("-"*50)
22. if nilai_p < alpha:
23.     print("Karena nilai-p < 0.05, maka MENOLAK H0")
24.     print("Kesimpulan: Terdapat perbedaan yang signifikan pada rata-rata ADR antara
    City Hotel dan Resort Hotel")
25.     print("Hipotesis yang diterima: H1")
26. else:
27.     print("Karena nilai-p >= 0.05, maka GAGAL MENOLAK H0")
28.     print("Kesimpulan: Tidak ada perbedaan yang signifikan pada rata-rata ADR antara
    City Hotel dan Resort Hotel")
```

Pada baris 4-5, kode ini memisahkan data ADR (Average Daily Rate) menjadi dua kelompok menggunakan pandas, yaitu `adr_hotel_city` untuk kelompok A (City Hotel) dan `adr_hotel_resort` untuk kelompok B (Resort Hotel). Selanjutnya, pada baris 8, dilakukan pengujian A/B testing menggunakan independent t-test dari library scipy.stats dengan parameter `equal_var=False` untuk menghasilkan nilai statistik uji (nilai_t) dan nilai-p (nilai_p). Pada baris 11-14, kode mencetak hasil pengujian dengan format yang rapi menggunakan garis pemisah berupa karakter "-" sebanyak 50 kali dan menampilkan nilai T-statistic serta P-value yang dibulatkan. Baris 17 menetapkan tingkat signifikansi (alpha)

sebesar 0.05. Terakhir, pada baris 20-28, kode membuat keputusan berdasarkan perbandingan nilai_p dengan alpha - jika $\text{nilai_p} < 0.05$ maka menolak H_0 dan menerima H_1 yang berarti ada perbedaan signifikan antara rata-rata ADR kedua hotel.

BAB V

Kesimpulan & Saran

A. Kesimpulan utama

Berdasarkan analisis yang dilakukan pada data hotel dari tahun 2015 hingga 2017, ditemukan bahwa tingkat pembatalan reservasi merupakan tantangan utama yang dihadapi oleh pihak hotel. Dari total 119.389 pemesanan, 37% atau 44.224 di antaranya dibatalkan, dengan rata-rata lead time pembatalan sekitar 101 hari sebelum kedatangan. Pembatalan lebih sering terjadi pada lead time pendek, yang mengindikasikan bahwa tamu cenderung membatalkan mendekati tanggal check-in, sehingga menyulitkan manajemen dalam mengatur ketersediaan kamar. Selain itu, analisis terhadap pola pendapatan menunjukkan adanya perbedaan signifikan antara City Hotel dan Resort Hotel. City Hotel cenderung memiliki pendapatan yang lebih konsisten sepanjang tahun, sedangkan Resort Hotel mengalami puncak pendapatan di bulan Agustus. Variasi ini menunjukkan adanya tren musiman yang perlu diperhatikan dalam penyusunan strategi bisnis. Selain itu, permintaan khusus yang tidak dikelola dengan baik dan sistem pembayaran yang kurang efisien turut berkontribusi pada tingkat pembatalan yang tinggi. Faktor-faktor ini menunjukkan bahwa peningkatan dalam manajemen reservasi dan pelayanan sangat diperlukan untuk meningkatkan kinerja hotel.

B. Saran

Untuk mengurangi tingkat pembatalan yang tinggi, disarankan bagi pihak hotel untuk menerapkan kebijakan reservasi yang lebih fleksibel, seperti memberikan insentif untuk pemesanan jauh hari atau penawaran khusus bagi tamu yang tidak membatalkan. Optimalisasi sistem untuk menangani permintaan khusus secara otomatis dapat meningkatkan kepuasan pelanggan dan mengurangi risiko pembatalan. Peningkatan efisiensi dalam proses pembayaran, misalnya melalui integrasi dengan platform digital atau pembayaran yang lebih cepat, juga dapat meningkatkan kepercayaan pelanggan. Selain itu,

pemantauan tren musiman yang lebih mendalam memungkinkan hotel untuk menyesuaikan strategi pemasaran dan harga sesuai dengan periode tertentu, baik untuk City Hotel maupun Resort Hotel. Penelitian lebih lanjut dapat berfokus pada analisis faktor eksternal seperti kondisi ekonomi atau tren pariwisata yang memengaruhi pola pembatalan, serta pada pengembangan sistem otomatis untuk mengoptimalkan pengelolaan ketersediaan kamar secara real-time.

BAB VI

Lampiran

A. Online Diagram

https://lucid.app/lucidchart/bb35d4c5-4967-4363-91d1-bce22577b4ff/edit?viewport_loc=-1823%2C-815%2C6148%2C3376%2CNKXkErc2Hqk3&invitationId=inv_72169178-efc4-404f-8c71-ddeb3df9a9ca

B. Python Code

https://drive.google.com/drive/u/0/folders/10_RyqvS90kLvL5dbrnaNrIEvlRvhK9RL

C. Recording

<https://youtu.be/ezH0GXvSBgc?si=wPQrEXORQeVc-q09>