

Data Analytics for Business 2024

Project Based Learning

RedDoorz - Study Case

**A DATA DRIVEN APPROACH TO
HOSPITALITY EXCELLENCE**

CS 18 - 3 :

Nabila Putri Asy Syifa

KM-CS18165

Girinda Decalzgi Azade

KM-CS18270

Rendy Achmadiansyah Mukti

KM-CS18375

KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa atas limpahan rahmat-Nya, sehingga kami dapat menyelesaikan laporan Project-Based Learning (PBL) ini sebagai bagian dari Studi Independen di Bitlabs Academy pada Program Kampus Merdeka Batch 7. Laporan ini merupakan hasil kolaborasi tim:

1. Nabila Putri Asy Syifa (Universitas Singaperbangsa Karawang)
2. Girinda Decalzgi Azade (Universitas Jember)
3. Rendy Achmadiansyah Mukti (UPN "Veteran" Jawa Timur)

Kami mengucapkan terima kasih kepada Bitlabs Academy atas ilmu dan bimbingannya, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi atas program Kampus Merdeka, serta para mentor dan rekan tim atas dukungan dan kerja sama selama program ini.

Semoga laporan ini dapat bermanfaat dan menjadi inspirasi bagi pengembangan diri serta inovasi mahasiswa lainnya.

Karawang, Jember, dan Surabaya
08, Desember 2024

Tim Penyusun
Nabila Putri Asy Syifa
Girinda Decalzgi Azade
Rendy Achmadiansyah Mukti

DAFTAR ISI

KATA PENGANTAR.....	ii
DAFTAR ISI.....	iii
BAB I.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	2
BAB II	4
2.1 Data dan Sumber Data	4
2.2 Alur Pengerjaan Study Case	5
1. Tools yang digunakan dalam proses menganalisis data	5
2. Proses Analisis Data	7
BAB III.....	9
3.1 Deskripsi Data Hasil Analisis Bisnis	9
3.2 Hasil Analisis	12
3.3 Pembahasan.....	46
BAB IV	82
4.1 Simpulan	82
4.2 Saran	82
LAMPIRAN.....	84

BAB I

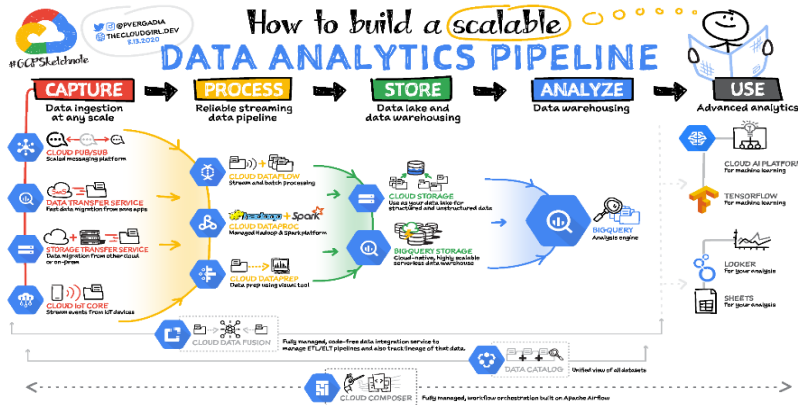
PENDAHULUAN

1.1 Latar Belakang

RedDoorz, didirikan pada tahun 2015, adalah jaringan perhotelan terkemuka yang menyediakan akomodasi berkualitas dengan harga terjangkau di seluruh Asia Tenggara. Dengan lebih dari 1.500 properti di negara-negara seperti Indonesia, Singapura, Vietnam, dan Filipina, RedDoorz bertujuan untuk menjadi pilihan utama para pelancong dengan memastikan pengalaman penginapan yang nyaman, bersih, dan aman. Memanfaatkan platform digital, RedDoorz membuat pemesanan lebih mudah dan fokus pada penggunaan teknologi untuk meningkatkan pengalaman pelanggan. Perusahaan juga menerapkan praktik berkelanjutan, yang bertujuan untuk mengurangi jejak karbon sambil mendukung masyarakat lokal.

Meskipun pertumbuhan dan kesuksesannya pesat, RedDoorz saat ini menghadapi tantangan dalam mengoptimalkan operasinya dan meningkatkan kepuasan pelanggan karena kesulitan dalam mengambil keputusan berdasarkan data. Kebutuhan akan pemahaman yang lebih mendalam mengenai kinerja bisnis, perilaku pelanggan, dan tren industri menjadi hal yang penting untuk mengatasi tantangan-tantangan ini. Oleh karena itu, sebagai bagian dari inisiatif ini, analisis data akan memainkan peran penting dalam mengungkap wawasan yang dapat ditindaklanjuti guna meningkatkan efisiensi operasional dan pengalaman pelanggan.

Studi kasus ini bertujuan untuk mengevaluasi keberhasilan merek dan properti RedDoorz dengan menganalisis kinerja bisnisnya. Hal ini juga akan fokus pada pemahaman pola perilaku pelanggan dan mengidentifikasi tren industri yang signifikan. Dengan menciptakan dashboard berbasis data, RedDoorz dapat mengambil keputusan yang tepat untuk mengoptimalkan operasional, melayani pelancong bisnis dan rekreasi secara efektif, dan pada akhirnya mencapai visinya untuk menjadi pilihan utama untuk akomodasi berkualitas dan terjangkau di Asia Tenggara.



Gambar 1. Data analytics pipeline

1.2 Identifikasi Masalah

RedDoorz menghadapi berbagai tantangan yang signifikan dalam upayanya untuk meningkatkan kualitas layanan dan mencapai tujuan bisnisnya. Berikut adalah masalah-masalah utama yang dihadapi oleh RedDoorz:

1. Kesulitan dalam Mengambil Keputusan Berbasis Data

RedDoorz menghadapi tantangan dalam melakukan optimasi operasional dan meningkatkan kepuasan pelanggan karena kesulitan dalam mengambil keputusan berbasis data. Meskipun memiliki akses terhadap data dari berbagai properti dan platform digital, penggunaan data tersebut untuk pengambilan keputusan strategis belum optimal.

2. Kurangnya Pemahaman Mendalam terhadap Performa Bisnis

RedDoorz memerlukan pemahaman yang lebih mendalam tentang performa bisnis, baik di tingkat merek maupun properti. Informasi terkait pendapatan, tingkat hunian, dan efisiensi operasional tidak sepenuhnya diolah menjadi wawasan yang dapat membantu perusahaan dalam mengidentifikasi area untuk perbaikan dan pertumbuhan.

3. Minimnya Analisis terhadap Perilaku Pelanggan

Kurangnya wawasan yang mendalam mengenai perilaku pelanggan, seperti preferensi, pola pemesanan, serta karakteristik pelanggan baru vs. pelanggan berulang, menyebabkan RedDoorz tidak dapat mengembangkan strategi yang lebih personal dalam memenuhi kebutuhan mereka. Hal ini berpotensi menghambat loyalitas pelanggan dan retensi pengguna.

4. Keterbatasan dalam Memahami Tren Industri Terkini

RedDoorz belum sepenuhnya memanfaatkan data untuk memahami tren industri yang sedang berkembang. Pengetahuan tentang tren dapat membantu perusahaan menyesuaikan layanan mereka agar lebih relevan dengan kebutuhan pasar, baik untuk wisatawan bisnis maupun rekreasi.

5. Tantangan dalam Optimalisasi Jaringan Properti

Dengan lebih dari 1.500 properti di berbagai negara, RedDoorz memerlukan strategi untuk mengoptimalkan jaringan properti yang mereka miliki. Saat ini, tidak adanya standar pengukuran performa yang konsisten untuk properti-properti tersebut menghambat perusahaan dalam mengambil tindakan yang tepat untuk memaksimalkan kinerja setiap properti.

6. Kurangnya Pemanfaatan Teknologi untuk Efisiensi

Operasional Meskipun RedDoorz telah menggunakan teknologi canggih dalam platform pemesanannya, potensi penggunaan teknologi untuk meningkatkan efisiensi operasional belum sepenuhnya dimanfaatkan. Ini termasuk penggunaan sistem manajemen properti (PMS) yang lebih terintegrasi, analitik data untuk prediksi permintaan, dan otomatisasi proses bisnis yang dapat membantu mengurangi biaya operasional serta meningkatkan kualitas layanan. Kurangnya pemanfaatan teknologi ini mengakibatkan operasi yang kurang efisien dan pemborosan sumber daya yang seharusnya dapat dihindari.

Dengan mengatasi permasalahan-permasalahan di RedDoorz dalam menghadapi tantangan dengan cara mengoptimalkan user journey pelanggan (*Optimasi User Journey Customer*) di seluruh tahapan, mulai dari discovery hingga post-stay. Proses yang belum sepenuhnya efisien menghambat kemampuan RedDoorz untuk memberikan pengalaman yang maksimal kepada pelanggan sehingga dapat berpengaruh terhadap kepuasan pelanggan, optimalisasi properti, serta penerapan teknologi yang lebih canggih dapat membantu RedDoorz dalam menghadapi tantangan yang ada dan mencapai visi mereka sebagai jaringan penginapan terkemuka di Asia Tenggara.

BAB II

METOLOGI

2.1 Data dan Sumber Data

Pada bagian ini dilaporkan jenis data dan sumber datanya dengan keterangan yang memadai. Jenis data yang digunakan dalam penelitian ini adalah data dalam format CSV yang diperoleh dari pihak Company RedDoorz. Data yang disediakan terdiri dari beberapa kategori utama, yaitu data transaksi, data pelanggan, dan data properti. Berikut adalah penjelasan terperinci:

1. Data Transaksi

Berisi informasi seperti `series_date`, `revenue`, `room/nights (RNs)`, `occupancy`, dan `Average Daily Rate`. Data ini memberikan gambaran tentang performa transaksi dari berbagai properti RedDoorz. Dengan data ini juga dapat menganalisis tren pendapatan, rata-rata tarif harian (ADR), serta perbandingan antara tarif dengan tingkat hunian. Dapat bisa melihat apakah ada hubungan antara tarif dan tingkat hunian dengan pendapatan yang dihasilkan. Memeriksa pola tingkat hunian selama periode waktu tertentu. Menganalisis apakah ada fluktuasi musiman atau hubungan dengan faktor lain (misalnya, kota atau jenis properti).

2. Data Pelanggan

Berisi informasi seperti pelanggan baru, pelanggan berulang, `gender`, tujuan perjalanan (bisnis/liburan), dan `usia`. Data ini digunakan untuk memahami perilaku dan karakteristik pelanggan RedDoorz. Dengan data ini juga dapat menganalisis perbedaan dalam perilaku antara pelanggan baru dan pelanggan yang kembali, termasuk frekuensi pemesanan, nilai transaksi, dan tingkat hunian. Menganalisis apakah terdapat perbedaan dalam pola pemesanan antara pelanggan yang bepergian untuk bisnis dan untuk rekreasi. Dan menganalisis bagaimana `usia` dan `gender` memengaruhi keputusan pemesanan serta jenis properti yang dipilih.

3. Data Properti

Terdiri dari informasi tentang lokasi properti (kota dan distrik), jumlah inventaris kamar, tanggal bergabungnya properti, dan tipe brand. Data ini membantu dalam mengevaluasi kinerja masing-masing properti dalam jaringan RedDoorz. Dengan data

ini juga dapat melihat distribusi konsentrasi properti di berbagai kota dan distrik, serta analisis terkait performa properti di wilayah tertentu. Menilai performa berdasarkan jenis merek, untuk melihat apakah ada perbedaan antara merek dalam hal tarif, pendapatan, dan tingkat hunian.

Dengan menggunakan data dari berbagai sumber ini, penelitian dapat memberikan analisis yang mendalam mengenai performa bisnis, perilaku pelanggan, serta efisiensi operasional dari properti-properti yang ada dalam jaringan RedDoorz.

2.2 Alur Pengerjaan Study Case

1. Tools yang digunakan dalam proses menganalisis data

Dalam proses analisis ini, berbagai alat telah digunakan untuk mendukung kebutuhan pengolahan dan visualisasi data secara efektif. Berikut macam-macam tools yang digunakan untuk memproses dan menganalisis data:

- **Draw.io**

Draw.io digunakan untuk pembuatan **Diagram BPMN** (Business Process Model and Notation) sebagai bagian dari pemodelan bisnis. Draw.io adalah alat online yang intuitif untuk membuat diagram alur proses yang membantu dalam pemahaman dan pengoptimalan alur kerja bisnis. Dengan menggunakan Draw.io, RedDoorz dapat menggambarkan proses bisnis yang kompleks dengan cara yang lebih terstruktur dan mudah dipahami oleh semua pihak yang berkepentingan.

- **Python**

Python digunakan sebagai alat utama dalam analisis data. Python merupakan bahasa pemrograman yang sangat populer dalam bidang data science karena kemampuannya yang luas dan pustaka yang lengkap seperti Pandas untuk manipulasi data, NumPy untuk perhitungan numerik, dan Matplotlib serta Seaborn untuk visualisasi data. Python memudahkan analisis data secara mendalam dan mendukung automasi proses pengolahan data, sehingga sangat efisien dalam menangani volume data yang besar, dimulai dari cleaning data hingga proses untuk mendapatkan insight hingga visualisasi di tools Google Collab.

- **DBeaver SQLite**

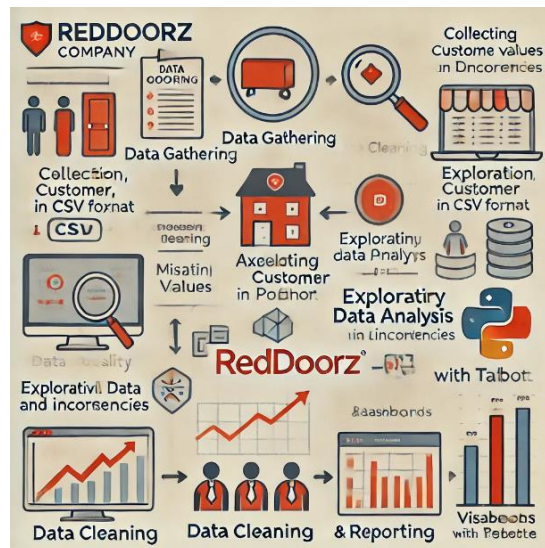
SQL digunakan untuk mengelola dan mengekstrak data dari database, sedangkan **Dbeaver SQLite** adalah alat yang digunakan untuk mempermudah akses dan manipulasi data dalam database. SQL (Structured Query Language) memungkinkan pengambilan data yang cepat dan efisien, sementara DBeaver adalah antarmuka yang membantu pengguna dalam melakukan query dan mengelola database dengan lebih mudah. Dengan kombinasi SQL dan DBeaver SQLite, RedDoorz dapat mengakses dan mengekstrak data yang relevan dari basis data untuk mendukung analisis yang lebih mendalam dengan metode query.

- **Tableau**

Tableau digunakan untuk membuat visualisasi data dan dashboard yang interaktif dan intuitif. Tableau adalah software berbasis visualisasi dan business intelligence yang memungkinkan pengguna untuk mengubah data mentah menjadi visualisasi yang mudah dipahami seperti grafik, diagram, dan dashboard. Tableau membantu dalam memvisualisasikan data secara efektif sehingga dapat memberikan wawasan mendalam yang mendukung pengambilan keputusan yang lebih baik.

Pemilihan alat-alat ini didasarkan pada kemampuan masing-masing untuk mengatasi permasalahan yang dihadapi, kemudahan integrasi dengan data yang tersedia, dan efisiensi dalam menghasilkan hasil analisis yang diinginkan. Kombinasi dari berbagai tools ini memungkinkan proses analisis data menjadi lebih efisien, dari pengolahan data mentah hingga pembuatan visualisasi yang menarik dan bermakna.

2. Proses Analisis Data



Gambar 2. Flowchart proses analisis data

Untuk menjelaskan proses analisis data yang dilakukan pada proyek RedDoorz, berikut adalah penjelasan terperinci yang dilengkapi dengan flowchart yang menggambarkan langkah-langkah yang dilakukan:

1. Pengumpulan Data (Data Gathering)

Langkah pertama dalam proses analisis adalah mengumpulkan data yang dibutuhkan. Pada kasus RedDoorz, data yang dikumpulkan mencakup dataset transaksi, pelanggan, dan properti, semuanya dalam format Excel. Data ini dikumpulkan atau didapatkan langsung dari perusahaan RedDoorz dan digunakan sebagai dasar untuk menjawab pertanyaan bisnis yang ada.

2. Penilaian Data (Assessing Data)

Setelah data terkumpul, tahap berikutnya adalah menilai kualitas dan struktur dari data tersebut. Proses penilaian data ini bertujuan untuk memastikan bahwa data yang digunakan berkualitas tinggi, lengkap, dan sesuai dengan kebutuhan analisis. Pada tahap ini, masalah seperti nilai yang hilang (*missing value*), data duplikat, inkonsisten tipe data, dan diidentifikasi outlier.

3. Pembersihan Data (Data Cleaning)

Setelah mengidentifikasi masalah pada tahap penilaian data, langkah berikutnya adalah pembersihan data. Proses ini bertujuan untuk memperbaiki

atau menghapus masalah yang ada, seperti menangani nilai yang hilang, menghilangkan duplikasi data, menangani outlier, mengubah tipe data yang tidak sesuai, dan memastikan konsistensi di seluruh dataset. Data yang telah dibersihkan ini kemudian siap untuk digunakan dalam analisis lebih lanjut.

4. Transformasi Data (Data Transformation)

Tahap ini mencakup proses transformasi data untuk mempersiapkan data agar dapat dianalisis lebih lanjut. Transformasi data meliputi normalisasi data, pengkodean variabel kualitatif, atau agregasi data. Langkah ini penting untuk memastikan bahwa data berada dalam format yang sesuai untuk keperluan analisis statistik atau pemodelan.

5. Eksplorasi Data (Exploratory Data Analysis - EDA)

Setelah data dibersihkan dan ditransformasi, tahap selanjutnya adalah eksplorasi data. Tujuan dari tahap ini adalah untuk memahami pola, tren, dan hubungan antar variabel dalam data. Analisis ini dilakukan menggunakan statistik deskriptif dan visualisasi data untuk menemukan wawasan awal yang dapat bermanfaat dalam menjawab pertanyaan bisnis.

6. Visualisasi dan Pelaporan (Visualization and Reporting)

Langkah terakhir dalam proses analisis adalah visualisasi dan pelaporan. Menggunakan Tableau, hasil dari analisis divisualisasikan dalam bentuk dashboard interaktif untuk memudahkan pemangku kepentingan memahami hasil analisis. Pelaporan ini bertujuan untuk menyampaikan wawasan yang diperoleh dan mendukung pengambilan keputusan bisnis.

BAB III

HASIL DAN PEMBAHASAN

3.1 Deskripsi Data Hasil Analisis Bisnis

Pada tahap ini, dilakukan analisis terhadap data yang telah dikumpulkan untuk memberikan gambaran menyeluruh mengenai kualitas dan keadaan data yang ada. Dataset yang digunakan mencakup beberapa variabel penting yang berkaitan dengan **transaksi**, **pelanggan**, dan **properti**. Dalam analisis ini, digunakan beberapa dataset yang masing-masing memiliki struktur yang berbeda, baik dalam jumlah kolom maupun baris. Berikut adalah deskripsi rinci mengenai masing-masing dataset yang digunakan:

1. Property Grade For Reference

Dataset ini terdiri dari 5 entri dengan 2 kolom, yaitu **Monthly Occupancy Percentage** dan **Grade**. Dimana dataset ini memberikan nilai tambah dengan menambahkan dimensi kontekstual ke data transaksi. Dengan menghubungkan informasi grade dan tingkat okupansi bulanan dengan data transaksi (booking, revenue, room nights), perusahaan dapat lebih memahami kinerja properti dan mengidentifikasi peluang peningkatan berdasarkan grade atau tingkat okupansi. Kedua kolom ini tidak memiliki nilai yang hilang dan tidak memiliki nilai duplikat. Kedua kolom berupa data kategorikal dengan tipe objek, sehingga statistik deskriptif numerik tidak diterapkan. Namun keduanya bertipe data objek. Informasi ini memberikan gambaran mengenai persentase okupansi bulanan dan grade yang diterima.

```
1. # Periksa Konsistensi Tipe Data Dataset Property Grade For Reference
2. property_grade_df.info()
3. # Hitung Null Value Dataset Property Grade For Reference
4. property_grade_df.isna().sum()
5. # Hitung duplikasi Dataset Property Grade For Reference
6. print('Jumlah duplikasi data:', property_grade_df.duplicated().sum())
7. # Hitung nilai statistik dataset Property Grade For Reference
8. property_grade_df.describe()
```

Kode Program 1. *Proses hasil analisis dataset property grade for reference pada python*

2. Bookings

Dataset ini lebih besar, dengan 100,000 entri dan 10 kolom. Beberapa kolom yang terdapat dalam dataset ini meliputi **Booking ID**, **User ID**, **Property Code**, **Booking Date**, dan **Revenue Dollar**. Semua kolom pada dataset ini memiliki nilai non-null, dan tidak memiliki nilai duplikat. Dataset **bookings** adalah bagian penting dari data transaksi karena mencatat setiap aktivitas pemesanan yang terjadi. Namun, tipe data bervariasi antara integer dan objek. Untuk nilai statistika deskriptif dimana: (**Rooms**: Rata-rata 2 kamar per pemesanan, dengan variasi dari 1 hingga 3 kamar), (**Nights**: Rata-rata 2 malam per pemesanan, dengan variasi dari 1 hingga 3 malam), (**Room Nights**: Rata-rata 4.01 kamar malam per pemesanan, dengan variasi yang lebih besar), dan (**Revenue Dollar**: Rata-rata pendapatan sekitar \$20.01 per pemesanan, dengan variasi dari \$10.00 hingga \$30.00). Dataset ini digunakan untuk menganalisis data pemesanan dan pendapatan.

```
1. # Periksa Konsistensi Tipe Data Dataset Bookings Table
2. bookings_df.info()
3. # Hitung Null Value Dataset Bookings Table
4. bookings_df.isna().sum()
5. # Hitung duplikasi Dataset Bookings Table
6. print('Jumlah duplikasi data:', bookings_df.duplicated().sum())
7. # Hitung nilai statistik dataset Bookings_Table
8. numerical_cols = bookings_df.select_dtypes(exclude=['datetime64[ns]']).columns
9. bookings_df[numerical_cols].describe()
```

Kode Program 2. Proses hasil analisis dataset bookings pada python

3. User

Dataset ini berisi 1,000 entri dengan 5 kolom, yang meliputi **User ID**, **User Type**, **User Gender**, **Travel Purpose**, dan **User Age**. Semua kolom memiliki nilai yang lengkap (non-null) dan tidak memiliki nilai duplikat dengan tipe data yang bervariasi antara integer dan objek. Dataset **user** merupakan bagian penting dari **data pelanggan** karena mencatat informasi mendetail tentang identitas, demografi, dan preferensi pelanggan. Untuk nilai statistika deskriptif dimana, **User Age**: Rata-rata usia pengguna adalah 39.83 tahun, dengan variasi antara 20 hingga 60 tahun. Dataset ini memberikan informasi terkait profil pengguna.

```
1. # Periksa Konsistensi Tipe Data Dataset User Table
2. user_df.info()
3. # Cek Null Value Dataset User Table
4. user_df.isna().sum()
5. # Cek duplikasi Dataset User Table
6. print('Jumlah duplikasi data:', user_df.duplicated().sum())
7. # Cek nilai statistik dataset User_Table
8. user_df.describe()
```

Kode Program 3. *Proses hasil analisis dataset user pada python*

4. Property

Dataset ini terdiri dari 500 entri dan 6 kolom, yang meliputi **Property Code**, **City**, **District**, **Cohort Date**, **Brand Type**, dan **Inventory**. Semua kolom dalam dataset ini juga tidak memiliki nilai yang hilang dan tidak memiliki data duplikat. Kolom **Inventory** bertipe integer, sementara yang lainnya bertipe objek. Untuk nilai statistika deskriptif dimana, **Inventory**: Rata-rata jumlah inventaris properti adalah 59.67, dengan variasi dari 20 hingga 100. Dataset **property**, yang mencakup informasi seperti **Property Code**, **City**, **District**, **Cohort Date**, **Brand Type**, dan **Inventory**, adalah inti dari **data properti** karena mencatat detail tentang karakteristik dan atribut properti yang tersedia dalam sistem. Dataset ini digunakan untuk analisis properti yang terdaftar.

```
1. # Periksa Konsistensi Tipe Data Dataset Property Table
2. property_df.info()
3. # Cek Null Value Dataset Property Table
4. property_df.isna().sum()
5. # Cek duplikasi Dataset Property Table
6. print('Jumlah duplikasi data:', property_df.duplicated().sum())
7. numerical_cols = property_df.select_dtypes(exclude=['datetime64[ns]']).columns
8. property_df[numerical_cols].describe()
```

Kode Program 4. *Proses hasil analisis dataset property pada python*

3.2 Hasil Analisis

A. Proses Data Cleaning Pada Python

Pada proses data cleaning yang dilakukan terhadap dataset yang diberikan dari company Reddoorz, terdiri dari beberapa dataset yang perlu untuk di cleaning, yakni dataset Bookings_Table, User_Table, dan Property_Table, beberapa langkah utama yang telah dilakukan antara lain:

1. Mengubah tipe data atribut dengan ID menjadi String

Tipe data pada kolom BOOKING_ID dan USER_ID dalam dataset Bookings_Table serta kolom USER_ID dalam dataset User_Table diubah menjadi tipe data string menggunakan fungsi .astype(str). Hal ini memastikan bahwa ID dalam bentuk string untuk konsistensi dan kemudahan dalam pengolahan data.

2. Mengubah tipe data atribut dengan waktu menjadi Date

Pada dataset Bookings_Table, kolom BOOKING_DATE, CHECK_IN_DATE, dan CHECK_OUT_DATE diubah menjadi tipe data datetime menggunakan fungsi pd.to_datetime(). Hal ini memastikan bahwa data yang berhubungan dengan waktu dapat dikelola dengan baik untuk analisis lebih lanjut seperti perhitungan durasi atau waktu tertentu. Begitu juga dengan kolom COHORT_DATE pada dataset Property_Table, yang diubah menjadi tipe data datetime untuk memastikan konsistensi format waktu.

3. Menyesuaikan format tanggal pada atribut yang berkaitan dengan waktu

Format tanggal pada kolom BOOKING_DATE, CHECK_IN_DATE, dan CHECK_OUT_DATE pada dataset Bookings_Table disesuaikan menggunakan parameter format='%d-%m-%Y' untuk memastikan format yang konsisten dalam analisis lebih lanjut.

Setelah melalui proses data cleaning, berikut adalah hasil yang didapatkan:

- **Data Konsistensi**

Semua kolom dengan ID, seperti BOOKING_ID dan USER_ID, telah berhasil diubah menjadi tipe string, yang memastikan tidak ada permasalahan dalam pemrosesan data berdasarkan ID tersebut.

- **Data Tanggal**

Semua kolom tanggal yang sebelumnya berformat objek atau string, seperti `BOOKING_DATE`, `CHECK_IN_DATE`, `CHECK_OUT_DATE`, dan `COHORT_DATE`, telah berhasil diubah menjadi tipe `datetime64`, yang memungkinkan analisis lebih lanjut terhadap aspek waktu.

- **Format Tanggal yang Konsisten**

Dengan mengubah format tanggal menjadi `'%d-%m-%Y'`, analisis data waktu kini dapat dilakukan dengan cara yang lebih terstruktur dan tidak ada kesalahan terkait format tanggal.

Melalui proses ini, data kini lebih siap untuk dianalisis lebih lanjut sesuai dengan kebutuhan bisnis, seperti analisis tren pemesanan, perilaku pelanggan, atau penilaian performa properti.

```
1. #Cleaning dataset Bookings_Table
2. # Mengubah tipe data atribut dengan ID menjadi String
3. bookings_df['BOOKING_ID'] = bookings_df['BOOKING_ID'].astype(str)
4. bookings_df['USER_ID'] = bookings_df['USER_ID'].astype(str)
5. # Mengubah tipe data atribut dengan waktu menjadi date
6. bookings_df['BOOKING_DATE'] = pd.to_datetime(bookings_df['BOOKING_DATE'],
    errors='coerce')
7. bookings_df['CHECK_IN_DATE'] = pd.to_datetime(bookings_df['CHECK_IN_DATE'],
    errors='coerce')
8. bookings_df['CHECK_OUT_DATE'] =
    pd.to_datetime(bookings_df['CHECK_OUT_DATE'], errors='coerce')
9. # Mengubah format tanggal pada atribut dengan waktu
10. bookings_df['BOOKING_DATE'] = pd.to_datetime(bookings_df['BOOKING_DATE'],
    format='%d-%m-%Y')
11. bookings_df['CHECK_IN_DATE'] = pd.to_datetime(bookings_df['CHECK_IN_DATE'],
    format='%d-%m-%Y')
12. bookings_df['CHECK_OUT_DATE'] =
    pd.to_datetime(bookings_df['CHECK_OUT_DATE'], format='%d-%m-%Y')
13.
```



```
14. #Cleaning dataset User_Table
15. # Mengubah tipe data atribut dengan ID menjadi String
16. user_df['USER_ID'] = user_df['USER_ID'].astype(str)
17.
18. #Cleaning dataset Property_Table
19. # Mengubah tipe data atribut dengan waktu menjadi date
20. property_df['COHORT_DATE'] = pd.to_datetime(property_df['COHORT_DATE'],
    errors='coerce')
```

Kode Program 5. *Proses data cleaning pada python*

B. Insight Booking Stay Duration and Booking Lead Time

Dalam tahap eksplorasi data ini, dua fitur baru telah diekstraksi dari data pemesanan terkait **durasi menginap** dan **waktu tunggu pemesanan**. Langkah-langkah yang dilakukan adalah sebagai berikut:

1. Durasi Menginap (Stay Duration)

Durasi menginap dihitung dengan mengurangi tanggal check-out (CHECK_OUT_DATE) dengan tanggal check-in (CHECK_IN_DATE). Hasilnya adalah durasi menginap dalam satuan hari untuk setiap pemesanan.

2. Waktu Tunggu Pemesanan (Lead Time)

Waktu tunggu pemesanan dihitung dengan mengurangi tanggal pemesanan (BOOKING_DATE) dengan tanggal check-in (CHECK_IN_DATE). Hasilnya menunjukkan berapa hari sebelum check-in pemesanan dilakukan.

Berdasarkan analisis yang dilakukan pada durasi menginap dan waktu tunggu pemesanan:

- **Durasi Menginap**

Rata-rata durasi menginap adalah 2 hari, dengan sebagian besar pelanggan menginap antara 1 hingga 3 hari. Ini menunjukkan bahwa kebanyakan tamu cenderung memilih masa menginap yang singkat.

- **Waktu Tunggu Pemesanan**

Rata-rata waktu tunggu pemesanan adalah 2 hari, dengan sebagian besar pemesanan dilakukan dalam waktu 0 hingga 1 hari sebelum check-in. Hal ini

menunjukkan bahwa pelanggan cenderung membuat pemesanan mendekati tanggal check-in mereka.

Analisis ini memberikan wawasan yang berguna bagi perusahaan dalam merencanakan kapasitas dan strategi pemasaran, terutama untuk mengelola ekspektasi terkait jumlah pemesanan yang dilakukan pada saat-saat terakhir dan durasi menginap yang relatif singkat.

```
1. #Explatory Data Analysis Terhadap Insight: Booking Duration and Booking Lead Time
2. #Membuat fitur baru Stay Duration: Menghitung Durasi Pemesanan
3. bookings_df['STAY_DURATION'] = (bookings_df['CHECK_OUT_DATE'] -
    bookings_df['CHECK_IN_DATE']).dt.days
4. #Membuat fitur baru Lead Time : Menghitung Waktu Tunggu Pemesanan
5. bookings_df['LEAD_TIME'] = (bookings_df['CHECK_IN_DATE'] -
    bookings_df['BOOKING_DATE']).dt.days
6.
7. # Validasi kolom baru
8. display("Preview of Stay Duration and Lead Time columns:")
9. display(bookings_df[['BOOKING_ID', 'BOOKING_DATE', 'CHECK_IN_DATE',
    'CHECK_OUT_DATE', 'STAY_DURATION', 'LEAD_TIME']].head())
10.
11. # Menghitung jumlah pemesanan customer dari berbagai variasi stay_duration
12. total_stay_duration = bookings_df['STAY_DURATION'].value_counts().sort_index()
13. display("Jumlah Pemesanan Variasi Stay Duration")
14. display(total_stay_duration)
15.
16. # Menghitung proporsi pemesanan dengan durasi
17. # Total pemesanan dengan durasi 1 hari
18. total_1_days = bookings_df[bookings_df['STAY_DURATION'] ==
    1]['STAY_DURATION'].count()
19. # Proporsi pemesanan dengan durasi 1 hari
20. proportion_1_days = (total_1_days / bookings_df['STAY_DURATION'].count()) * 100
21. # Total pemesanan dengan durasi 2 hari
22. total_2_days = bookings_df[bookings_df['STAY_DURATION'] ==
    2]['STAY_DURATION'].count()
23. # Proporsi pemesanan dengan durasi 2 hari
24. proportion_2_days = (total_2_days / bookings_df['STAY_DURATION'].count()) * 100
25. # Total pemesanan dengan durasi 3 hari
```

```

26. total_3_days = bookings_df[bookings_df['STAY_DURATION'] ==
    3]['STAY_DURATION'].count()
27. # Proporsi pemesanan dengan durasi 3 hari
28. proportion_3_days = (total_3_days / bookings_df['STAY_DURATION'].count()) * 100
29. # Menampilkan hasil
30.
31. # Analisis Stay Duration
32. display("\nStay Duration Statistics:")
33. display(bookings_df['STAY_DURATION'].describe())
34.
35. # Visualization distribution of Stay Duration
36. plt.figure(figsize=(8, 5))
37. sns.histplot(bookings_df['STAY_DURATION'], bins=10, kde=True, color='blue')
38. plt.title('Distribution of Stay Duration')
39. plt.xlabel('Stay Duration (Days)')
40. plt.ylabel('Frequency')
41. plt.grid(axis='y')
42. plt.show()
43.
44. # Menghitung jumlah pemesanan customer dari berbagai variasi lead_time
45. total_lead_time = bookings_df['LEAD_TIME'].value_counts().sort_index()
46. display("Jumlah Pemesanan Variasi Lead Time")
47. display(total_lead_time)
48.
49. # Analisis Lead Time
50. display("\nLead Time Statistics:")
51. display(bookings_df['LEAD_TIME'].describe())
52.
53. # Visualization distribution of Lead Time
54. plt.figure(figsize=(8, 5))
55. sns.histplot(bookings_df['LEAD_TIME'], bins=30, kde=True, color='green')
56. plt.title('Distribution of Lead Time')
57. plt.xlabel('Lead Time (Days)')
58. plt.ylabel('Frequency')
59. plt.grid(axis='y')
60. plt.show()

```

Kode Program 6. EDA untuk insight Stay Duration dan Lead Time pada python

C. Insight Revenue Per Night and Total Booking Revenue

Dalam tahapan eksplorasi data ini, dua fitur baru yang berkaitan dengan pendapatan dari pemesanan dihitung dan dianalisis, yaitu **Revenue per Night** dan **Total Booking Revenue**. Langkah-langkah yang dilakukan adalah sebagai berikut:

1. Revenue per Night (Pendapatan per Malam)

Fitur ini dihitung dengan membagi total pendapatan dari pemesanan (REVENUE_DOLLAR) dengan jumlah malam yang dipesan (ROOM_NIGHTS). Hasilnya, kita memperoleh pendapatan per malam untuk setiap pemesanan yang bisa digunakan untuk menganalisis seberapa besar pendapatan yang diperoleh per unit waktu (malam) yang dipesan.

2. Total Booking Revenue (Total Pendapatan Pemesanan)

Total pendapatan pemesanan dihitung dengan mengelompokkan data berdasarkan USER_ID dan menjumlahkan pendapatan untuk setiap pengguna. Ini memberikan gambaran tentang kontribusi masing-masing pelanggan terhadap total pendapatan yang diperoleh.

Berdasarkan analisis yang dilakukan, berikut adalah hasil yang diperoleh:

- **Revenue per Night**

Rata-rata pendapatan per malam adalah 7.47 USD, dengan nilai minimum 1.11 USD dan maksimum 30.00 USD. Sebagian besar pendapatan per malam berada dalam kisaran 3 hingga 10 USD, dengan distribusi yang menunjukkan puncak pada pendapatan yang lebih rendah, dan berkurang secara bertahap saat pendapatan per malam meningkat.

- **Total Booking Revenue**

Setelah mengelompokkan dan menjumlahkan pendapatan berdasarkan USER_ID, ditemukan bahwa pelanggan dengan total pendapatan tertinggi berkontribusi lebih besar terhadap pendapatan keseluruhan. Pengguna dengan USER_ID 6534621 tercatat sebagai pelanggan dengan pendapatan terbesar, yakni 2708.86 USD.

Dengan visualisasi distribusi total pendapatan pemesanan dan top 10 pelanggan berdasarkan pendapatan tersebut, kita bisa mendapatkan wawasan yang lebih mendalam tentang bagaimana pendapatan terdistribusi dan siapa pelanggan yang paling menguntungkan.

```

1. # Explatory Data Analysis Terhadap Insight: Revenue per Night and Booking Revenue Features
2. # Membuat fitur baru Revenue per Night: Menghitung Pendapatan per Malam (Revenue per Night)
3. bookings_df['REVENUE_PER_NIGHT'] = bookings_df['REVENUE_DOLLAR'] / bookings_df['ROOM_NIGHTS']
4.
5. # Analisis Revenue per Night
6. display("\nRevenue per Night Statistics:")
7. display(bookings_df['REVENUE_PER_NIGHT'].describe())
8.
9. # Visualisasi: Distribusi Revenue per Night
10. plt.figure(figsize=(8, 5))
11. plt.hist(bookings_df['REVENUE_PER_NIGHT'], bins=30, color='skyblue', edgecolor='black')
12. plt.title('Distribution of Revenue per Night')
13. plt.xlabel('Revenue per Night (USD)')
14. plt.ylabel('Frequency')
15. plt.grid(True)
16. plt.show()
17.
18. # Group by USER_ID to calculate Total Booking Revenue
19. Total_Booking_Revenue =
    bookings_df.groupby('USER_ID')['REVENUE_DOLLAR'].sum().reset_index()
20. Total_Booking_Revenue.rename(columns={'REVENUE_DOLLAR':
    'TOTAL_BOOKING_REVENUE'}, inplace=True)
21. # Menggabungkan Total Booking Revenue ke dalam DataFrame Utama (bookings_df)
22. bookings_df = pd.merge(bookings_df, TOTAL_BOOKING_REVENUE_df, on='USER_ID',
    how='left')
23.
24. # Analisis Total Booking Revenue
25. display("\nTotal Booking Revenue Statistics:")
26. display(TOTAL_BOOKING_REVENUE_df['TOTAL_BOOKING_REVENUE'].describe())
27.
28. # Menampilkan 10 pelanggan teratas dalam hal Total Pendapatan Pemesanan

```

```

29. top_10_customers =
    TOTAL_BOOKING_REVENUE_df.sort_values(by='TOTAL_BOOKING_REVENUE',
    ascending=False).head(10)
30. display(top_10_customers)
31.
32. # Visualize the top 10 customers
33. plt.figure(figsize=(7, 4))
34. sns.barplot(data=top_10_customers, x='USER_ID', y='TOTAL_BOOKING_REVENUE',
    palette='viridis')
35. plt.title('Top 10 Customers by Total Booking Revenue')
36. plt.xlabel('User ID')
37. plt.ylabel('Total Booking Revenue ($)')
38. plt.xticks(rotation=45)
39. plt.grid(axis='y')
40. plt.show()

```

Kode Program 7. EDA untuk insight Revenue Per Night and Total Booking Revenue pada python

D. Insight Property Performance and Occupancy Rate

Dalam tahapan eksplorasi data ini, perhitungan dan analisis dilakukan pada dua fitur penting terkait dengan Property Occupancy Rate dan Total Revenue per Property. Salah satu langkah kunci dalam perhitungan Days Active adalah penerapan teknik filter untuk menghindari hasil yang tidak valid, seperti nilai negatif atau nol. Langkah-langkah yang dilakukan adalah sebagai berikut:

1. Occupancy Rate (Tingkat Hunian)

Fitur **Occupancy Rate** dihitung dengan rumus. Dimana rumus **Occupancy Rate** digunakan untuk mengukur **efisiensi penggunaan kapasitas kamar**, membandingkan performa antar properti dalam mengisi kamar yang tersedia, dan menentukan strategi penyesuaian harga dan promosi untuk meningkatkan tingkat hunian.

➤ Memahami Komponen Rumus

- **Total Room Nights Booked:**

Jumlah total malam kamar yang dipesan untuk suatu properti selama periode tertentu. Data ini biasanya berasal dari jumlah pemesanan (booking) yang dilakukan oleh pelanggan.

- **Inventory:**

Jumlah kamar yang tersedia di properti tersebut. Ini adalah kapasitas total kamar yang dapat disewakan pada satu waktu.

- **Days Active:**

Periode waktu (dalam hari) selama properti tersebut aktif atau tersedia untuk disewakan. Ini dihitung sebagai selisih antara tanggal check-in pertama (**CHECK_IN_DATE**) dan tanggal properti mulai aktif (**COHORT_DATE**).

$$\text{Occupancy Rate} = \frac{\text{Total Room Nights Booked}}{\text{Inventory} \times \text{Days Active}}$$

Di mana **Days Active** adalah selisih antara **CHECK_IN_DATE** dan **COHORT_DATE** (dari tabel properti). Untuk memastikan perhitungan yang valid, langkah pertama yang dilakukan adalah memfilter nilai **Days Active** yang lebih kecil dari atau sama dengan nol. Pernyataan filter yang digunakan adalah sebagai berikut:

```
merged_df = merged_df[merged_df['DAYS_ACTIVE'] > 0]
```

Langkah ini penting untuk menghindari baris yang mengandung nilai **Days Active** negatif atau nol, yang tidak masuk akal dalam konteks analisis tingkat okupansi dan pendapatan. Setelah penerapan filter, analisis dapat dilakukan dengan data yang lebih bersih dan valid. Di mana **Days Active** adalah hasil perhitungan yang telah difilter, untuk memastikan hanya nilai yang valid yang digunakan dalam menghitung tingkat hunian.

Fitur Days Active ini memberikan gambaran tentang efisiensi penggunaan kamar di setiap properti, yaitu berapa banyak malam yang dipesan dibandingkan dengan jumlah malam yang tersedia untuk setiap periode.

2. Total Revenue per Property (Pendapatan Total per Properti)

Pendapatan total per properti dihitung berdasarkan **Total Room Nights Booked** dan **Occupancy Rate**, yang memberikan gambaran bagaimana performa pendapatan setiap properti dalam kaitannya dengan tingkat okupansi dan jumlah malam yang dipesan.

Berdasarkan analisis yang dilakukan, berikut adalah hasil yang diperoleh:

- **Days Active:**

Setelah menerapkan filter terhadap nilai **Days Active** yang lebih kecil dari atau sama dengan nol, diperoleh **48225 data valid**. Rata-rata **Days Active** adalah **590.45 hari**, dengan **minimum 1 hari** dan **maksimum 1718 hari**. Distribusi data menunjukkan bahwa sebagian besar properti aktif dalam rentang **240 hingga 885 hari**, dengan beberapa properti yang memiliki periode aktivitas lebih panjang.

- **Total Room Nights Booked:**

Statistik **Total Room Nights Booked** menunjukkan rata-rata **801.28 malam** per properti. Dengan **minimum 658 malam** dan **maksimum 998 malam**, properti secara umum memiliki performa yang cukup baik dalam hal jumlah kamar yang dipesan.

- **Total Revenue:**

Statistik **Total Revenue** menunjukkan rata-rata **4002.26 USD** per properti. Dengan nilai minimum **3255.23 USD** dan maksimum **4928.83 USD**, data ini menunjukkan distribusi yang cukup stabil meskipun ada beberapa properti dengan pendapatan yang lebih tinggi.

- **Top 10 Properties by Occupancy Rate:**

Properti dengan **Occupancy Rate** tertinggi adalah **JAK0253** dengan tingkat hunian sebesar **4014.29%**, diikuti oleh **JAK0052** (3804.76%) dan **BAN0428** (3560.00%). Properti-properti ini menunjukkan tingkat hunian yang sangat tinggi, yang dapat dijadikan sebagai benchmark untuk properti lainnya.

- **Korelasi antara Occupancy Rate dan Total Revenue:**

Analisis korelasi antara **Occupancy Rate** dan **Total Revenue** menunjukkan nilai **0.01**, yang menandakan hubungan yang sangat lemah antara kedua variabel ini. Hal ini mengindikasikan bahwa tingkat hunian tidak secara signifikan mempengaruhi pendapatan total per properti.


```

1. #EDA Terhadap Insight: Property Performance and Occupancy Rate
2. # Perhitungan Days Active
3. # Menggabungkan bookings_df dan property_df menjadi merged_df
4. merged_df = pd.merge(bookings_df, property_df, on='PROPERTY_CODE',
    how='left')
5.
6. # Step 1: Menghitung Days Active (Hitung selisih antara CHECK_IN_DATE dan
    COHORT_DATE untuk mendapatkan DAYS_ACTIVE)
7. merged_df['DAYS_ACTIVE'] = (merged_df['COHORT_DATE'] -
    merged_df['CHECK_IN_DATE']).dt.days
8.
9. # Filter out invalid DAYS_ACTIVE values (e.g., negative or zero days)
10. merged_df = merged_df[merged_df['DAYS_ACTIVE'] > 0]
11.
12. # UBAH DAYS ACTIVE KE TIPE DATA INT
13. merged_df['DAYS_ACTIVE'] = merged_df['DAYS_ACTIVE'].astype(int)
14.
15. # Menampilkan hasil
16. merged_df[['BOOKING_ID', 'PROPERTY_CODE', 'CHECK_IN_DATE',
    'COHORT_DATE', 'DAYS_ACTIVE']].head()
17.
18. # Analisis Days Active
19. display("Days Active Statistics:")
20. display(merged_df['DAYS_ACTIVE'].describe())
21.
22. # Distribusi Days Active Menggunakan Histogram
23. plt.figure(figsize=(10, 5))
24. sns.histplot(merged_df['DAYS_ACTIVE'], bins=30, kde=True, color='purple')
25. plt.title('Distribution of Days Active')
26. plt.xlabel('Days Active')
27. plt.ylabel('Frequency')
28. plt.grid(axis='y')
29. plt.show()
30.
31. # Step 2: HITUNG TOTAL MALAM KAMAR YANG DIPESAN dan
    PENDAPATAN TOTAL PER PROPERTI
32. property_performance = bookings_df.groupby('PROPERTY_CODE').agg(
33.     TOTAL_ROOM_NIGHTS_BOOKED=('ROOM_NIGHTS', 'sum'),
34.     TOTAL_REVENUE=('REVENUE_DOLLAR', 'sum')

```

```

35. ).reset_index()
36.
37. # menampilkan hasil
38. display("Preview of new features 'TOTAL_ROOM_NIGHTS_BOOKED', 'and
    'TOTAL_REVENUE':")
39. display(property_performance.head())
40. # Properti yang memiliki TOTAL_REVENUE tertinggi
41. top_properties = property_performance.sort_values(by='TOTAL_REVENUE',
    ascending=False).head(10)
42. # menampilkan hasil
43. display("Top 10 Properties by Total Revenue:")
44. display(top_properties)
45.
46. # Properti yang memiliki total room nights booked tertinggi
47. top_properties =
    property_performance.sort_values(by='TOTAL_ROOM_NIGHTS_BOOKED',
    ascending=False).head(10)
48. # menampilkan hasil
49. display("Top 10 Properties by Total Room Nights Booked:")
50. display(top_properties)
51.
52. # ANALISIS TOTAL ROOM NIGHTS BOOKED
53. display("Total Room Nights Booked Statistics:")
54. display(property_performance['TOTAL_ROOM_NIGHTS_BOOKED'].describe())
55. # ANALISIS TOTAL REVENUE
56. display("Total Revenue Statistics:")
57. display(property_performance['TOTAL_REVENUE'].describe())
58.
59. # Distribusi TOTAL_ROOM_NIGHTS_BOOKED menggunakan histogram
60. plt.figure(figsize=(10, 5))
61. sns.histplot(property_performance['TOTAL_ROOM_NIGHTS_BOOKED'], bins=30,
    kde=True, color='orange')
62. plt.title('Distribution of Total Room Nights Booked')
63. plt.xlabel('Total Room Nights Booked')
64. plt.ylabel('Frequency')
65. plt.grid(axis='y')
66. plt.show()
67.
68. # Distribusi TOTAL_REVENUE menggunakan histogram

```

```

69. plt.figure(figsize=(10, 5))
70. sns.histplot(property_performance['TOTAL_REVENUE'], bins=30, kde=True,
    color='blue')
71. plt.title('Distribution of Total Revenue')
72. plt.xlabel('Total Revenue')
73. plt.ylabel('Frequency')
74. plt.grid(axis='y')
75. plt.show()
76.
77. # Variasi dalam ukuran properti (inventory)
78. plt.figure(figsize=(10, 5))
79. sns.histplot(property_df['INVENTORY'], bins=30, kde=True, color='red')
80. plt.title('Distribution of Inventory')
81. plt.xlabel('Inventory')
82. plt.ylabel('Frequency')
83. plt.grid(axis='y')
84. plt.show()
85.
86. # Step 3: Merge Data for Occupancy Rate Calculation
87. merged_df = merged_df.merge(property_performance, on='PROPERTY_CODE',
    how='left')
88. # menampilkan hasil
89. display("Preview of merged DataFrame with new features:")
90. display(merged_df.head())
91.
92. # Step 4: HITUNG Occupancy Rate
93. merged_df['OCCUPANCY_RATE'] =
    merged_df['TOTAL_ROOM_NIGHTS_BOOKED'] / ((
94.     merged_df['INVENTORY'] * merged_df['DAYS_ACTIVE']
95. ))*100
96. # FORMAT OCCUPANCY RATE DESIMAL DUA BELAKANG KOMA
97. pd.options.display.float_format = '{:.2f}'.format
98. # menampilkan hasil
99. display("Preview of new feature 'OCCUPANCY_RATE':")
100. display(merged_df[['PROPERTY_CODE', 'TOTAL_ROOM_NIGHTS_BOOKED',
    'INVENTORY', 'DAYS_ACTIVE', 'OCCUPANCY_RATE']].head())
101.
102. # ANALISIS OCCUPANCY RATE
103. display("Occupancy Rate Statistics:")

```

```

104.display(merged_df['OCCUPANCY_RATE'].describe())
105.
106.# Distribusi OCCUPANCY_RATE menggunakan histogram
107.plt.figure(figsize=(10, 5))
108.sns.histplot(merged_df['OCCUPANCY_RATE'], bins=30, kde=True, color='blue')
109.plt.title('Distribution of Occupancy Rate')
110.plt.xlabel('Occupancy Rate')
111.plt.ylabel('Frequency')
112.plt.grid(axis='y')
113.plt.show()
114.g
115.# Step 5: Identify Top 10 Properties by Occupancy Rate
116.top_10_properties = merged_df.sort_values(by='OCCUPANCY_RATE',
      ascending=False).head(10)
117.# menampilkan hasil hanya kolom occupancy rate dan property code
118.display("Top 10 Properties by Occupancy Rate:")
119.display(top_10_properties[['PROPERTY_CODE', 'OCCUPANCY_RATE']])
120.
121.# Step 6: Visualizations
122.# Visualize Top 10 Properties by Occupancy Rate
123.plt.figure(figsize=(7, 5))
124.sns.barplot(data=top_10_properties, x='PROPERTY_CODE',
      y='OCCUPANCY_RATE', palette='coolwarm')
125.plt.title("Top 10 Properties by Occupancy Rate")
126.plt.xlabel('Property Code')
127.plt.ylabel('Occupancy Rate')
128.plt.xticks(rotation=45)
129.plt.grid(axis='y')
130.plt.show()
131.
132.# Step 7: Correlation Between Occupancy Rate and Total Revenue
133.correlation =
      merged_df['OCCUPANCY_RATE'].corr(merged_df['TOTAL_REVENUE'])
134.# dibulatkan hasilnya
135.correlation = round(correlation, 2)
136.# menampilkan hasil
137.display("Correlation between Occupancy Rate and Total Revenue:")
138.display(correlation)
139.

```

```
140.# Step 8: Scatter Plot for Occupancy Rate vs Total Revenue per Property
141.plt.figure(figsize=(7, 5))
142.sns.scatterplot(data=merged_df, x='OCCUPANCY_RATE', y='TOTAL_REVENUE',
                    hue='OCCUPANCY_RATE', palette='viridis')
143.plt.title('Correlation Between Occupancy Rate and Total Revenue')
144.plt.xlabel('Occupancy Rate')
145.plt.ylabel('Total Revenue ($)')
146.plt.grid()
147.plt.show()
```

Kode Program 8. EDA untuk insight Property Performance and Occupancy Rate

E. Proses Analisis Data Tingkat Lanjut Menggunakan SQL

Tahap analisis data tingkat lanjut dilakukan setelah proses pembersihan data menggunakan Python selesai. Dataset hasil pembersihan tersebut digunakan untuk analisis lanjutan dengan SQL.

Pada tahap ini, akan dilakukan beberapa analisis utama seperti menghitung total Room Nights (RNs), pendapatan (Revenue), dan Average Daily Rate (ADR) setiap kota di tahun 2024, mengidentifikasi 10 properti teratas dalam hal pertumbuhan RN saat membandingkan tahun-ke-tahun (YTD) 2023 dengan YTD 2024, menentukan segmentasi pelanggan berdasarkan perilaku pembelian, membandingkan harga musim normal dengan harga musim puncak pada tahun 2023 dan 2024, serta menganalisis kualitas properti. Sebelum memulai analisis, beberapa langkah persiapan perlu dilakukan, sebagai berikut :

Mengimport Dataset

Dataset yang telah disiapkan diimpor ke dalam tabel sesuai dengan struktur yang telah dibuat sebelumnya. Proses ini memastikan bahwa data tersedia untuk dianalisis secara optimal. Beberapa dataset yang diimpor antara lain: bookings_df, property_data, property_df, property_grade_df, dan user_df.

Setelah semua persiapan selesai, tahap selanjutnya adalah melanjutkan ke tahap utama, yaitu analisis data. Pada tahap ini, serangkaian analisis lanjutan akan dilakukan untuk mengeksplorasi dan menggali wawasan yang lebih mendalam dari dataset yang telah dipersiapkan. Melalui analisis ini, berbagai metrik penting akan dihitung, seperti total Room Nights (RNs), pendapatan (Revenue), dan Average Daily Rate (ADR) untuk setiap kota di tahun 2024. Selain itu, akan diidentifikasi 10 properti teratas berdasarkan pertumbuhan Room Nights (RN) antara tahun 2023 dan 2024, segmentasi pelanggan akan ditentukan berdasarkan perilaku pembelian, perbandingan harga musim normal dan musim puncak pada tahun 2023 dan 2024 akan dilakukan, serta kualitas properti akan dianalisis. Semua analisis ini bertujuan untuk memberikan wawasan yang lebih jelas dan mendalam yang dapat digunakan untuk mendukung pengambilan keputusan strategis yang lebih efektif.

F. Menghitung Total RNs, Revenue dan ADR Setiap Kota di Tahun 2024

Analisis ini berguna untuk mengetahui perbandingan Total RNs, Revenue dan ADR di setiap kota pada tahun 2024.

- Total RNs dihitung dengan menggunakan fungsi `SUM(b.ROOM_NIGHTS)` yang menjumlahkan jumlah Room Nights (malam yang dipesan) untuk setiap kota pada setiap bulan di tahun 2024.
- Total Revenue dihitung dengan menggunakan fungsi `SUM(b.REVENUE_DOLLAR)`, yang menjumlahkan total pendapatan yang dihasilkan dari pemesanan di setiap kota dan bulan pada tahun 2024.
- ADR (Average Daily Rate) dihitung dengan membagi total pendapatan (`SUM(b.REVENUE_DOLLAR)`) dengan total Room Nights (`SUM(b.ROOM_NIGHTS)`).

Hasil dari analisis ini berguna untuk memberikan wawasan mengenai kinerja operasional dan keuangan hotel di berbagai kota sepanjang tahun 2024.

```

1. SELECT
2.   p.CITY,
3.   strftime('%Y-%m', b.CHECK_IN_DATE) AS MONTH_YEAR,
4.   SUM(b.ROOM_NIGHTS) AS Total_RNs,
5.   SUM(b.REVENUE_DOLLAR) AS Total_Revenue,
6.   CASE
7.     WHEN SUM(b.ROOM_NIGHTS) = 0 THEN 0
8.     ELSE SUM(b.REVENUE_DOLLAR) / SUM(b.ROOM_NIGHTS)
9.   END AS ADR
10. FROM
11.   bookings b
12. JOIN
13.   property p ON b.PROPERTY_CODE = p.PROPERTY_CODE
14. WHERE
15.   strftime('%Y', b.CHECK_IN_DATE) = '2024'
16. GROUP BY
17.   p.CITY, strftime('%Y-%m', b.CHECK_IN_DATE)
18. ORDER BY
19.   p.CITY, MONTH_YEAR;

```

Kode Program 9. Menghitung Total RNs, Revenue dan ADR Setiap Kota di Tahun 2024

G. Mengidentifikasi 10 Properti Teratas Dalam Hal Pertumbuhan RN Saat Membandingkan Tahun-ke-Tahun (YTD) 2023 Dengan YTD 2024

Analisis ini bertujuan untuk menentukan properti dengan pertumbuhan terbaik dalam hal Room Nights (RN) dari tahun ke tahun (YTD) dengan membandingkan performa tahun 2024 dengan 2023.

- **Growth_Calculation CTE**

CTE ini menggabungkan data tahun 2023 dan 2024 untuk setiap properti menggunakan FULL OUTER JOIN pada PROPERTY_CODE. Di sini, dihitung selisih total RNs antara kedua tahun tersebut (RN_s_Growth) dan persentase pertumbuhannya (Growth_Percentage).

- **Growth Percentage**

Growth Percentage mengukur perubahan relatif jumlah Room Nights (RN) antara dua periode tahun yang berbeda, yaitu 2023 dan 2024.

$$\text{Growth Percentage} = \left(\frac{\text{RNs 2024} - \text{RNs 2023}}{\text{RNs 2023}} \right) \times 100$$

Rumus perhitungannya adalah selisih antara RNs tahun 2024 dan RNs tahun 2023, dibagi dengan RNs tahun 2023, kemudian dikalikan 100 untuk menghasilkan persentase. Rumus ini menunjukkan seberapa besar peningkatan atau penurunan jumlah pemesanan kamar yang terjadi pada tahun 2024 dibandingkan dengan tahun sebelumnya.

- **Growth Category**

Digunakan untuk mengklasifikasikan properti berdasarkan persentase pertumbuhan Room Nights (RNs) dari tahun 2023 ke 2024. Berdasarkan nilai Growth Percentage, properti akan dikelompokkan ke dalam tiga kategori:

- **High Growth:** Jika persentase pertumbuhan lebih dari 50%, properti akan digolongkan dalam kategori ini. Ini menunjukkan bahwa properti mengalami pertumbuhan yang sangat signifikan dalam jumlah pemesanan kamar dibandingkan dengan tahun sebelumnya.
- **Moderate Growth:** Jika persentase pertumbuhan berada antara 10% dan 50%, properti akan masuk dalam kategori ini. Ini menunjukkan adanya pertumbuhan yang moderat, yang masih cukup signifikan namun tidak sebesar kategori pertama.
- **Low/No Growth:** Jika persentase pertumbuhan kurang dari 10% atau bahkan negatif, properti akan digolongkan dalam kategori ini. Hal ini menunjukkan bahwa properti mengalami sedikit atau bahkan tidak ada pertumbuhan dalam pemesanan kamar, yang bisa mengindikasikan stagnasi atau penurunan.

Kategori ini membantu dalam memberikan gambaran yang lebih jelas tentang performa relatif setiap properti dalam hal pertumbuhan Room Nights. Dengan menganalisis data ini, manajer atau analis dapat mengetahui properti yang mengalami pertumbuhan signifikan dalam hal permintaan kamar dari tahun ke tahun dan membuat keputusan yang lebih baik terkait pengelolaan atau pemasaran properti.


```

1. WITH RNs_YTD AS (
2.     SELECT
3.         PROPERTY_CODE,
4.         strftime('%Y', CHECK_IN_DATE) AS Year,
5.         SUM(ROOM_NIGHTS) AS Total_RNs
6.     FROM
7.         bookings
8.     WHERE
9.         strftime('%m-%d', CHECK_IN_DATE) <= '11-30'
10.    GROUP BY
11.        PROPERTY_CODE, strftime('%Y', CHECK_IN_DATE)
12. ),
13. Growth_Calculation AS (
14.     SELECT
15.         COALESCE(a.PROPERTY_CODE, b.PROPERTY_CODE) AS
PROPERTY_CODE,
16.         COALESCE(b.Total_RNs, 0) AS RNs_2023,
17.         COALESCE(a.Total_RNs, 0) AS RNs_2024,
18.         COALESCE(a.Total_RNs, 0) - COALESCE(b.Total_RNs, 0) AS
RNs_Growth,
19.         CASE
20.             WHEN COALESCE(b.Total_RNs, 0) = 0 THEN NULL
21.             ELSE ((COALESCE(a.Total_RNs, 0) - COALESCE(b.Total_RNs, 0))
* 100.0) / COALESCE(b.Total_RNs, 0)
22.         END AS Growth_Percentage
23.     FROM
24.         RNs_YTD a
25.     FULL OUTER JOIN
26.         RNs_YTD b
27.         ON a.PROPERTY_CODE = b.PROPERTY_CODE
28.         AND b.Year = '2023'
29.         AND a.Year = '2024'
30. )
31. SELECT
32.     g.PROPERTY_CODE,
33.     g.RNs_2023,
34.     g.RNs_2024,
35.     g.RNs_Growth,
36.     ROUND(g.Growth_Percentage, 2) AS Growth_Percentage,
37.     p.CITY,
38.     CASE
39.         WHEN ROUND(g.Growth_Percentage, 2) > 50 THEN 'High Growth'
40.         WHEN ROUND(g.Growth_Percentage, 2) BETWEEN 10 AND 50
THEN 'Moderate Growth'
41.         ELSE 'Low/No Growth'
42.     END AS Growth_Category
43. FROM
44.     Growth_Calculation g
45. LEFT JOIN
46.     property p ON g.PROPERTY_CODE = p.PROPERTY_CODE
47. ORDER BY
48.     g.Growth_Percentage DESC NULLS LAST, -- Properti dengan data
pertumbuhan valid diprioritaskan
49.     g.RNs_Growth DESC
50. LIMIT 10;

```

Kode Program 10. Mengidentifikasi 10 properti teratas

H. Menentukan Segmentasi Pelanggan Berdasarkan Perilaku Pembelian

Analisis ini menggunakan pendekatan RFM (Recency, Frequency, Monetary) untuk mengidentifikasi dan mengklasifikasikan pelanggan berdasarkan perilaku pembelian mereka. RFM mengukur tiga faktor penting yang mempengaruhi perilaku pelanggan:

- **Recency:** Seberapa baru pelanggan melakukan pembelian. Pelanggan yang baru saja melakukan pembelian lebih mungkin untuk melakukan pembelian ulang.
- **Frequency:** Seberapa sering pelanggan melakukan pembelian. Pelanggan yang sering melakukan pembelian lebih cenderung loyal.
- **Monetary:** Seberapa banyak uang yang dibelanjakan oleh pelanggan. Pelanggan yang menghabiskan lebih banyak cenderung memberikan lebih banyak nilai pada bisnis.

Pada **RFM Scoring**, setiap pelanggan diberi skor untuk masing-masing faktor:

- Recency: Skor 5 untuk pelanggan yang baru membeli dalam 50 hari terakhir, dan skor menurun hingga 1 untuk pelanggan yang tidak membeli lebih dari 200 hari.
- Frequency: Skor 5 untuk pelanggan yang melakukan lebih dari 150 pembelian, dengan skor menurun sesuai dengan frekuensi pembelian.
- Monetary: Skor 5 untuk pelanggan yang menghabiskan lebih dari \$2000, dengan skor menurun berdasarkan total pengeluaran mereka.

Skor total RFM dihitung dengan **menjumlahkan skor untuk Recency, Frequency, dan Monetary**.

$$\text{RFM Score} = \text{Recency Score} + \text{Frequency Score} + \text{Monetary Score}$$

Berdasarkan total skor ini, pelanggan dibagi ke dalam empat segmen:

- **Platinum:** Total skor 13 atau lebih. Pelanggan dengan skor tertinggi, menunjukkan pelanggan yang sangat aktif dan menguntungkan.
- **Gold:** Total skor antara 9 dan 12. Pelanggan yang sering membeli dan menghabiskan banyak, tetapi tidak seaktif pelanggan Platinum.
- **Silver:** Total skor antara 6 dan 8. Pelanggan dengan skor moderat.

- **Bronze:** Total skor kurang dari 6. Pelanggan dengan skor rendah, menunjukkan mereka jarang melakukan pembelian dan menghabiskan sedikit.

Hasil analisis RFM ini membantu dalam pengambilan keputusan yang lebih terarah, meningkatkan efisiensi pemasaran, dan memperkuat strategi retensi dan loyalitas pelanggan.

```

1. WITH user_behavior AS (
2.     SELECT
3.         b.USER_ID,
4.         MAX(b.BOOKING_DATE) AS last_booking_date, -- Tanggal terakhir pemesanan
5.         COUNT(b.BOOKING_ID) AS total_bookings, -- Total pemesanan (Frequency)
6.         SUM(b.REVENUE_DOLLAR) AS total_revenue -- Total pendapatan (Monetary)
7.     FROM bookings b
8.     GROUP BY b.USER_ID
9. ),
10. current_date AS (
11.     SELECT DATE('now') AS today -- Tanggal hari ini
12. ),
13. rfm_scoring AS (
14.     SELECT
15.         ub.USER_ID,
16.         julianday(cd.today) - julianday(ub.last_booking_date) AS recency, -- Hitung
            Recency dalam hari
17.         ub.total_bookings AS frequency, -- Total pemesanan
18.         ub.total_revenue AS monetary, -- Total pendapatan
19.         -- Scoring Recency
20.         CASE
21.             WHEN julianday(cd.today) - julianday(ub.last_booking_date) <= 50 THEN 5
22.             WHEN julianday(cd.today) - julianday(ub.last_booking_date) BETWEEN 51
            AND 100 THEN 4
23.             WHEN julianday(cd.today) - julianday(ub.last_booking_date) BETWEEN 101
            AND 150 THEN 3
24.             WHEN julianday(cd.today) - julianday(ub.last_booking_date) BETWEEN 151
            AND 200 THEN 2
25.             ELSE 1
26.         END AS recency_score,
27.         -- Scoring Frequency
28.         CASE
29.             WHEN ub.total_bookings >= 150 THEN 5
30.             WHEN ub.total_bookings BETWEEN 100 AND 149 THEN 4
31.             WHEN ub.total_bookings BETWEEN 50 AND 99 THEN 3
32.             WHEN ub.total_bookings BETWEEN 20 AND 49 THEN 2
33.             ELSE 1
34.         END AS frequency_score,
35.         -- Scoring Monetary
36.         CASE
37.             WHEN ub.total_revenue >= 2000 THEN 5
38.             WHEN ub.total_revenue BETWEEN 1000 AND 1999 THEN 4
39.             WHEN ub.total_revenue BETWEEN 500 AND 999 THEN 3
40.             WHEN ub.total_revenue BETWEEN 200 AND 499 THEN 2
41.             ELSE 1
42.         END AS monetary_score
43.     FROM user_behavior ub
44.     CROSS JOIN current_date cd
45. ),
46. rfm_segments AS (
47.     SELECT
48.         r.USER_ID,
49.         r.recency,
50.         r.frequency,
51.         r.monetary,
52.         r.recency_score,
53.         r.frequency_score,
54.         r.monetary_score,

```

```

55.      -- Total RFM Score
56.      (r.recency_score + r.frequency_score + r.monetary_score) AS rfm_score,
57.      CASE
58.          WHEN (r.recency_score + r.frequency_score + r.monetary_score) >= 13 THEN
        'Platinum'
59.          WHEN (r.recency_score + r.frequency_score + r.monetary_score) BETWEEN 9
        AND 12 THEN 'Gold'
60.          WHEN (r.recency_score + r.frequency_score + r.monetary_score) BETWEEN 6
        AND 8 THEN 'Silver'
61.          ELSE 'Bronze'
62.      END AS customer_segment
63.  FROM rfm_scoring r
64. ),
65.  final_segments AS (
66.      SELECT
67.          rs.USER_ID,
68.          rs.recency,
69.          rs.frequency,
70.          rs.monetary,
71.          rs.recency_score,
72.          rs.frequency_score,
73.          rs.monetary_score,
74.          rs.rfm_score,
75.          rs.customer_segment,
76.          u.USER_TYPE,
77.          u.USER_GENDER,
78.          u.TRAVEL_PURPOSE,
79.          u.USER_AGE
80.      FROM rfm_segments rs
81.      LEFT JOIN user u ON rs.USER_ID = u.USER_ID
82.  )
83.  SELECT
84.      USER_ID,
85.      recency,
86.      frequency,
87.      monetary,
88.      recency_score,
89.      frequency_score,
90.      monetary_score,
91.      rfm_score,
92.      customer_segment,
93.      USER_TYPE,
94.      USER_GENDER,
95.      TRAVEL_PURPOSE,
96.      USER_AGE
97.  FROM final_segments
98.  ORDER BY rfm_score DESC, customer_segment;

```

Kode Program 11. *Menentukan segmentasi pelanggan*

I. Membandingkan Harga Musim Normal Dengan Harga Musim Puncak Pada Tahun 2023 dan 2024

Analisis ini bertujuan untuk membandingkan harga antara musim normal dan musim puncak pada tahun 2023 dan 2024, dengan fokus pada pemahaman perbedaan pendapatan dan pola harga berdasarkan periode musim.

- **Total Bookings**

Total pemesanan dihitung dengan **COUNT(BOOKING_ID)** untuk setiap kombinasi tahun dan musim. Ini memberikan gambaran tentang **jumlah pemesanan** yang diterima untuk masing-masing musim (normal atau puncak) pada setiap tahun.

- **Total Revenue**

Total pendapatan dihitung dengan **SUM(REVENUE_DOLLAR)**, yang menunjukkan **total pendapatan yang dihasilkan** dari semua pemesanan dalam periode tersebut.

- **Total Room Nights**

Jumlah malam yang dipesan dihitung dengan **SUM(ROOM_NIGHTS)**, memberikan informasi tentang **jumlah total malam kamar yang dipesan** dalam periode tersebut.

- **Harga Rata-rata per Malam (Avg Revenue per Night):**

Harga rata-rata per malam dihitung dengan rumus:

$$\text{Avg Revenue per Night} = \frac{\text{Total Revenue}}{\text{Total Room Nights}}$$

Rumus ini memberikan gambaran tentang **pendapatan rata-rata yang dihasilkan per malam yang dipesan**. Ini adalah metrik penting untuk membandingkan apakah harga kamar cenderung lebih tinggi atau lebih rendah selama musim puncak dibandingkan dengan musim normal.

- **Harga Rata-rata per Pemesanan (Avg Revenue per Booking):**

Harga rata-rata per pemesanan dihitung dengan rumus:

$$\text{Avg Revenue per Booking} = \frac{\text{Total Revenue}}{\text{Total Bookings}}$$

Rumus ini memberikan gambaran tentang **pendapatan rata-rata per pemesanan** yang diterima, yang dapat digunakan untuk menilai seberapa besar kontribusi setiap pemesanan terhadap pendapatan total.

Analisis ini sangat berguna untuk menilai perbedaan tarif dan pendapatan antara musim puncak dan normal. Dengan melihat harga rata-rata per malam, hotel dapat mengetahui apakah tarif mereka lebih tinggi pada musim puncak dan menyesuaikan strategi harga mereka untuk memaksimalkan pendapatan.

```

1.  WITH booking_data AS (
2.      SELECT
3.          b.BOOKING_ID,
4.          b.REVENUE_DOLLAR,
5.          b.ROOM_NIGHTS,
6.          strftime('%Y', b.BOOKING_DATE) AS year,
7.          CASE
8.              WHEN strftime('%m', b.BOOKING_DATE) IN ('06', '07', '12', '01') THEN 'Peak
Season'
9.              ELSE 'Normal Season'
10.         END AS season
11.     FROM
12.         bookings b
13.     WHERE
14.         strftime('%Y', b.BOOKING_DATE) IN ('2023', '2024')
15. ),
16. seasonal_pricing AS (
17.     SELECT
18.         year,
19.         season,
20.         COUNT(BOOKING_ID) AS total_bookings, -- Total jumlah pemesanan
21.         SUM(REVENUE_DOLLAR) AS total_revenue, -- Total pendapatan
22.         SUM(ROOM_NIGHTS) AS total_room_nights, -- Total malam yang dipesan
23.         SUM(REVENUE_DOLLAR) / SUM(ROOM_NIGHTS) AS
avg_revenue_per_night, -- Harga rata-rata per malam
24.         SUM(REVENUE_DOLLAR) / COUNT(BOOKING_ID) AS
avg_revenue_per_booking -- Harga rata-rata per pemesanan
25.     FROM
26.         booking_data
27.     GROUP BY
28.         year,
29.         season
30. )
31. SELECT
32.     year,
33.     season,
34.     total_bookings,
35.     total_revenue,
36.     total_room_nights,
37.     ROUND(avg_revenue_per_night, 2) AS avg_revenue_per_night,
38.     ROUND(avg_revenue_per_booking, 2) AS avg_revenue_per_booking
39. FROM
40.     seasonal_pricing
41. ORDER BY
42.     year,
43.     season;

```

Kode Program 12. Membandingkan harga musim normal

J. Menganalisis Kualitas Properti

Analisis ini bertujuan untuk menganalisis kualitas properti berdasarkan tingkat occupancy rate dan grade properti. Berikut penjelasan detail:

1. Occupancy Rate

Tingkat hunian properti dihitung dengan **ROUND(pd.OCCUPANCY_RATE, 2)**, yang memberikan informasi tentang **persentase tingkat hunian properti** pada periode tertentu. Nilai ini menunjukkan seberapa banyak kamar yang terisi dibandingkan dengan total kamar yang tersedia.

2. Pengelompokan Berdasarkan Occupancy Rate

Properti dikelompokkan ke dalam kategori berdasarkan tingkat occupancy rate-nya menggunakan pernyataan **CASE**. Berikut pembagian kategori occupancy rate:

- **Monthly_occupancy > 80%**: Properti dengan tingkat hunian di atas 80% dianggap memiliki kinerja yang sangat baik.
- **Monthly_occupancy between 70% & < 80%**: Properti dengan tingkat hunian antara 70% hingga 80% memiliki kinerja yang baik.
- **Monthly_occupancy between 40% & < 70%**: Properti dengan tingkat hunian antara 40% hingga 70% dianggap memiliki kinerja rata-rata.
- **Monthly_occupancy between 20% & < 40%**: Properti dengan tingkat hunian antara 20% hingga 40% menunjukkan kinerja yang kurang baik.
- **Monthly_occupancy < 20%**: Properti dengan tingkat hunian di bawah 20% dianggap memiliki kinerja yang sangat buruk.

Analisis ini sangat berguna untuk menilai kinerja dan kualitas properti berdasarkan tingkat hunian. Dengan melihat hubungan antara occupancy rate dan grade, pihak manajemen dapat mengidentifikasi properti yang berperforma sangat baik dan yang perlu perbaikan. Properti dengan occupancy rate yang rendah bisa menunjukkan adanya masalah yang perlu ditangani, seperti harga yang terlalu tinggi, layanan yang kurang memadai, atau strategi pemasaran yang tidak efektif.


```

1. WITH property_quality AS (
2.   SELECT
3.     pd.PROPERTY_CODE,
4.     ROUND(pd.OCCUPANCY_RATE, 2) AS OCCUPANCY_RATE,
5.     pg.Grade
6.   FROM
7.     property_data pd
8.   JOIN
9.     property_grade pg
10.  ON
11.    CASE
12.      WHEN pd.OCCUPANCY_RATE > 80 THEN pg.Monthly_Occupancy_Percentage =
'Monthly_occupancy > 80%'
13.      WHEN pd.OCCUPANCY_RATE BETWEEN 70 AND 80 THEN
pg.Monthly_Occupancy_Percentage = 'Monthly_occupancy between 70% > & < 80%'
14.      WHEN pd.OCCUPANCY_RATE BETWEEN 40 AND 70 THEN
pg.Monthly_Occupancy_Percentage = 'Monthly_occupancy between 40% > & < 70%'
15.      WHEN pd.OCCUPANCY_RATE BETWEEN 20 AND 40 THEN
pg.Monthly_Occupancy_Percentage = 'Monthly_occupancy between 20% > & < 40%'
16.      ELSE pg.Monthly_Occupancy_Percentage = 'Monthly_occupancy < 20%'
17.    END
18. )
19. SELECT
20.   PROPERTY_CODE,
21.   Grade,
22.   OCCUPANCY_RATE
23. FROM
24.   property_quality
25. ORDER BY
26.   Grade ASC, OCCUPANCY_RATE DESC;

```

Kode Program 13. *Menganalisis kualitas properti*

K. SQL View Untuk Laporan Metrik Utama

Analisis ini menggunakan SQL VIEW untuk menyimpan hasil perhitungan metrik utama yang berkaitan dengan performa properti dalam berbagai kota selama enam bulan terakhir. Metrik utama yang dihitung adalah Total Room Nights (RNs), Total Revenue, dan Average Daily Rate (ADR).

- **Filtered Bookings**

Pada bagian pertama, subquery **filtered_bookings** digunakan untuk **menyaring data pemesanan** yang terjadi dalam enam bulan terakhir. Dengan menggunakan fungsi `strftime('%Y-%m', b.BOOKING_DATE)`, data pemesanan dikategorikan berdasarkan **bulan dan tahun** (contohnya: '2024-06'). Hal ini memungkinkan analisis untuk melihat performa di tiap bulan selama enam bulan terakhir. **ROOM_NIGHTS** dan **REVENUE_DOLLAR** disaring untuk mendapatkan total malam yang dipesan dan pendapatan yang dihasilkan.

- **City Metrics**

Subquery **city_metrics** melakukan agregasi data berdasarkan kota (CITY) dan bulan pemesanan (booking_month). Di sini, tiga metrik utama dihitung:

- **Total Room Nights (RNs):** Merupakan jumlah keseluruhan malam yang dipesan pada tiap kota dan bulan. Dihitung dengan SUM(fb.ROOM_NIGHTS).
- **Total Revenue:** Total pendapatan yang dihasilkan oleh pemesanan di tiap kota dan bulan. Dihitung dengan SUM(fb.REVENUE_DOLLAR).
- **Average Daily Rate (ADR):** Harga rata-rata per malam, dihitung dengan AVG(fb.REVENUE_DOLLAR / fb.ROOM_NIGHTS), yang memberikan informasi tentang pendapatan rata-rata per malam kamar yang terjual.

- **Pembuatan View**

Setelah perhitungan dilakukan, hasilnya disimpan dalam **view** yang bernama **city_metrics_last_6_months**, sehingga laporan metrik utama untuk setiap kota dalam enam bulan terakhir dapat diakses dengan mudah.

Manfaat analisis ini adalah untuk memantau kinerja properti dalam berbagai kota selama enam bulan terakhir, dengan fokus pada total room nights, total pendapatan, dan rata-rata harga per malam (ADR). Analisis ini memberikan wawasan yang berguna bagi manajer properti untuk mengevaluasi tren musiman, melihat bulan-bulan dengan performa terbaik atau terburuk, serta mengidentifikasi potensi peningkatan pendapatan. Dengan informasi ini, strategi harga dan pemasaran dapat disesuaikan untuk memaksimalkan kinerja properti di setiap kota, serta merespons perubahan permintaan dan kondisi pasar yang dapat mempengaruhi performa.

```
1. CREATE VIEW city_metrics_last_6_months AS
2. WITH filtered_bookings AS (
3.   SELECT
4.     b.BOOKING_DATE,
5.     b.ROOM_NIGHTS,
6.     b.REVENUE_DOLLAR,
7.     b.PROPERTY_CODE,
8.     strftime('%Y-%m', b.BOOKING_DATE) AS booking_month
9.   FROM
10.    bookings b
11.  WHERE
```

```

12.     b.BOOKING_DATE >= date('now', '-6 months')
13. ),
14. city_metrics AS (
15.     SELECT
16.         p.CITY,
17.         fb.booking_month,
18.         SUM(fb.ROOM_NIGHTS) AS total_RNs,
19.         SUM(fb.REVENUE_DOLLAR) AS total_revenue,
20.         AVG(fb.REVENUE_DOLLAR / fb.ROOM_NIGHTS) AS avg_ADR
21.     FROM
22.         filtered_bookings fb
23.     JOIN
24.         property p
25.     ON
26.         fb.PROPERTY_CODE = p.PROPERTY_CODE
27.     GROUP BY
28.         p.CITY, fb.booking_month
29. )
30. SELECT
31.     CITY,
32.     booking_month,
33.     total_RNs,
34.     total_revenue,
35.     avg_ADR
36. FROM
37.     city_metrics;

38. SELECT * FROM city_metrics_last_6_months;

```

Kode Program 14. *SQL View untuk Laporan Metrik Utama*

SQL View dirancang untuk menyederhanakan proses analisis data dengan menyimpan hasil perhitungan metrik utama dalam struktur yang terorganisir. Fokus utama dari SQL View adalah pada Total Room Nights (RNs), Total Revenue, dan Average Daily Rate (ADR) berdasarkan performa properti dalam enam bulan terakhir. Berikut adalah manfaat dari SQL View :

1. **Efisiensi Analisis:** Dengan SQL View, analisis data dapat dilakukan lebih cepat tanpa perlu menghitung ulang metrik utama setiap kali diperlukan.
2. **Monitoring Performa:** Data yang disimpan dalam view memberikan wawasan performa properti berdasarkan kota dan bulan. Hal ini memungkinkan manajer untuk memantau tren musiman, bulan dengan performa terbaik, dan potensi area untuk peningkatan.

Dukungan Pengambilan Keputusan: Informasi seperti fluktuasi ADR, total revenue, dan volume pemesanan membantu dalam merancang strategi pemasaran, promosi, dan penyesuaian harga.

L. Revenue Trend by City and Brand Type

Visualisasi dari “Revenue Trend by City and Brand Type” dibuat untuk memahami tren pendapatan dari berbagai kota dan tipe mereknya (RedDoorz, Koolkost, RedPartner) selama periode 2020-2024. Dengan menggunakan grafik garis, dapat diperoleh hasil untuk mengidentifikasi fluktuasi pendapatan, kota yang mendominasi pendapatan, serta kontribusi setiap merek terhadap total pendapatan. Tujuan utamanya adalah untuk mengidentifikasi area yang membutuhkan perhatian lebih, seperti kota dengan tren pendapatan menurun, sehingga dapat dibuat strategi promosi dan pricing yang sesuai. Oleh karena itu, dibuat visualisasi pada tableau dengan langkah pengerjaan sebagai berikut:

1. Data Preparation

Dataset dimuat ke tableau, dengan kolom Booking Date diatur dalam hierarki Year-Month agar analisis dapat dilakukan pada tingkat tahunan maupun bulanan. Kolom City digunakan untuk membedakan data berdasarkan kotanya, sedangkan Brand Type digunakan untuk menampilkan kontribusi merek secara spesifik. Kolom Revenue Dollar dipilih sebagai metrik utama yang menunjukkan total pendapatan per kota dan merek. Filter waktu diterapkan untuk analisis dari Januari 2020 hingga September 2024 yang dapat diatur rentangnya untuk menampilkan berdasarkan periode yang diinginkan.

2. Visualisasi

Grafik garis dibuat untuk menunjukkan tren pendapatan per tahun dan bulan di setiap kota. Dimensi City digunakan untuk memisahkan garis untuk masing-masing kota, dan pemberian warna garis digunakan untuk membedakan merek (RedDoorz, Koolkost, RedPartner). Sumbu vertikal menunjukkan total pendapatan, sedangkan sumbu horizontal menampilkan rentang waktu. Filter tambahan memungkinkan untuk memilih kota atau merek tertentu untuk fokus analisis lanjutan. Hasil visualisasi akan menunjukkan tren pendapatan sesuai periode yang ditentukan dengan memperhatikan masing-masing tren dari merek (RedDoorz, Koolkost, RedPartner) di setiap kota.

M. Customer Segmentation and Behavior Pattern

Visualisasi ini bertujuan untuk memahami perilaku pelanggan berdasarkan frekuensi kunjungan, pengeluaran total, usia, dan jenis kelamin. Diagram batang dan lingkaran digunakan untuk menganalisis distribusi pendapatan berdasarkan segmen pelanggan, termasuk kelompok usia, jenis kelamin, dan tipe pengguna (baru/ulang). Tujuannya adalah untuk mengidentifikasi segmen pelanggan yang paling menguntungkan dan merancang strategi pemasaran yang lebih terfokus. Oleh karena itu, dibuat visualisasi pada tableau dengan langkah pengerjaan sebagai berikut:

1. Data Preparation

Langkah pertama dalam data preparation melibatkan penghitungan total pemesanan per pelanggan menggunakan formula { FIXED [User Id] : COUNT([Booking Id]) }. Dengan formula ini, setiap pelanggan dihitung jumlah total pemesanannya secara unik berdasarkan ID pengguna. Selanjutnya, pelanggan dikelompokkan ke dalam kategori Customer Frequency Segmentation berdasarkan jumlah pemesanan mereka, dengan klasifikasi sebagai berikut.

- Moderate Stayer : Total pemesanan ≤ 50
- Frequent Stayer : Total pemesanan yaitu antara 51 hingga 100
- Super Frequent Stayer : Total pemesanan > 100

Selain itu, total pengeluaran per pelanggan dihitung menggunakan formula { FIXED [User Id] : SUM([Revenue Dollar]) } yang kemudian dikelompokkan ke dalam Total Spend Segment dengan klasifikasi sebagai berikut:

- High Spender : Total pengeluaran $> \$2000$
- Medium Spender : Total pengeluaran yaitu antara \$1000-\$2000
- Low Spender : Total pengeluaran $\leq \$1000$

Untuk menganalisis pengaruh usia terhadap pendapatan, pelanggan dikelompokkan ke dalam User Age Segment menggunakan kriteria sebagai berikut:

- Young Adults : Usia ≤ 29 tahun
- Young Professionals : Usia diantara 30-39 tahun
- Middle-aged Adults : Usia diantara 40-49 tahun
- Older Adults : Usia diantara 50-59 tahun)
- Seniors : Usia ≥ 60 tahun

2. Visualisasi

Visualisasi dibuat untuk menggambarkan pola perilaku pelanggan dari berbagai segmen. Untuk melihat pola perilaku pelanggan, terdapat tiga visualisasi untuk mengelompokkan pelanggan berdasarkan total pemesanan, total pengeluaran, dan usia. Berikut merupakan masing-masing visualisasi yang dibuat:

- 1) Diagram batang segmented digunakan untuk menunjukkan kontribusi pendapatan dari setiap segmen pengeluaran (High, Medium, Low Spender) dan frekuensi kunjungan (Moderate, Frequent, dan Super Frequent Stayer). Visualisasi ini memberikan gambaran tentang pelanggan mana yang paling berkontribusi terhadap total revenue.
- 2) Diagram lingkaran digunakan untuk memvisualisasikan distribusi pendapatan berdasarkan jenis kelamin pelanggan (pria dan wanita) dan tipe pengguna (*new user* atau *repeat user*). Visualisasi ini menunjukkan kontribusi pendapatan dari masing-masing gender dan tipe pengguna serta dapat difilter berdasarkan *travel purpose* mereka untuk analisis lebih lanjut.
- 3) Diagram batang berdasarkan kelompok usia digunakan untuk menganalisis kontribusi pendapatan dari pelanggan berdasarkan kategori usia, sehingga mempermudah identifikasi segmen usia yang paling menguntungkan. Kombinasi visualisasi ini memungkinkan analisis menyeluruh terhadap perilaku pelanggan dan pola pengeluaran mereka.
- 4) Diagram garis yang menunjukkan tren tahunan pendapatan pelanggan berdasarkan User Gender (jenis kelamin) dan User Type (pengguna baru atau ulang) selama periode 2020 hingga 2024. Dengan membandingkan kontribusi pendapatan antara pria dan wanita, serta antara pengguna baru dan pengguna ulang, visualisasi ini membantu memahami pola perilaku pelanggan dari waktu ke waktu. Tujuannya adalah untuk mengidentifikasi segmen pelanggan yang memiliki tren positif atau negatif dalam kontribusi pendapatan, sehingga dapat dirancang strategi yang sesuai untuk mempertahankan atau meningkatkan kontribusi mereka.

N. The Top 10 Properties With The Highest Room Nights With The Brand Type Details

Visualisasi ini dirancang untuk mengidentifikasi properti dengan performa terbaik berdasarkan jumlah Room Nights (RNs) yang mencerminkan tingkat pemanfaatan

kamar di setiap properti. Tujuan utama dari visualisasi ini adalah untuk menentukan properti-properti unggulan yang dapat dijadikan model dalam strategi pengembangan lebih lanjut, seperti akuisisi properti baru atau peningkatan properti di lokasi lain. Selain itu, visualisasi ini juga mengevaluasi kontribusi setiap merek (Brand Type) terhadap performa properti, memberikan wawasan tentang merek yang paling mendominasi dalam segmen properti berkinerja tinggi. Oleh karena itu, dibuat visualisasi pada tableau dengan langkah pengerjaan sebagai berikut:

1. Data Preparation

Untuk mempersiapkan data, langkah pertama adalah menghitung peringkat (Rank Calculation) menggunakan formula `index()` untuk memberi nomor urut berdasarkan jumlah total RNs di setiap properti. Properti kemudian diurutkan secara descending berdasarkan nilai RNs, dan filter diterapkan untuk menampilkan hanya 10 properti dengan jumlah RNs tertinggi. Total RNs dihitung untuk setiap properti dengan nilai agregat dari kolom Room Nights (RNs). Dimensi Brand Type ditambahkan untuk menunjukkan merek yang berkontribusi pada performa setiap properti, sehingga memungkinkan evaluasi lebih lanjut terhadap kontribusi masing-masing merek.

2. Visualisasi

Visualisasi dibuat dalam bentuk grafik batang, di mana jumlah room nights untuk setiap properti diurutkan berdasarkan rank. Properti dengan room nights tertinggi ditempatkan di bagian paling kiri, dan warna batang menunjukkan merek (Brand Type) yang terkait dengan setiap properti, seperti RedDoorz, Koolkost, dan RedPartner. Label jumlah room nights ditampilkan di atas setiap batang untuk memberikan informasi kuantitatif yang jelas. Visualisasi ini hanya menampilkan 10 properti dengan performa terbaik, memberikan fokus pada properti yang paling berkontribusi terhadap perusahaan berdasarkan jumlah room nights.

O. RedDoorz Business Insight Dashboard

Dashboard ini dirancang untuk memberikan gambaran menyeluruh tentang kinerja bisnis RedDoorz dengan fokus pada tren pendapatan berdasarkan kota dan merek, pola perilaku pelanggan, serta properti dengan performa terbaik berdasarkan Room Nights

(RNs). Tujuan utama dari dashboard ini adalah untuk mengidentifikasi area yang memerlukan perbaikan dalam penjualan, harga, dan akuisisi properti, serta memberikan rekomendasi untuk kampanye pemasaran yang lebih efektif dan terfokus. Dengan visualisasi yang interaktif, pengguna dapat memahami bagaimana setiap faktor berkontribusi terhadap total pendapatan sebesar \$2,001,130 dan memanfaatkannya untuk pengambilan keputusan strategis. Oleh karena itu, dibuat visualisasi pada tableau dengan langkah pengerjaan sebagai berikut.

1. Data Preparation

Semua visualisasi yang telah dibuat dicek kembali agar tidak terjadi kesalahan saat menaruhnya ke dalam dashboard. Kemudian, menambahkan informasi tambahan seperti total booking, total kamar, dan total pendapatan untuk memperkaya informasi yang ada pada dashboard.

2. Visualisasi

Setelah visualisasi siap, langkah berikutnya adalah membuat berbagai visualisasi yang akan dimasukkan ke dalam dashboard. Revenue Trends by City and Brand Type dibuat menggunakan grafik garis, memvisualisasikan pendapatan berdasarkan kota dan merek dengan hierarki waktu (Year-Month) pada sumbu horizontal. Visualisasi Customer Segmentation by Frequency and Spending Patterns dibuat menggunakan diagram batang segmented, yang menunjukkan kontribusi pendapatan dari pelanggan berdasarkan frekuensi kunjungan dan segmen pengeluaran. Diagram lingkaran digunakan untuk Revenue Distribution by User Type and Travel Purpose, menggambarkan distribusi pendapatan berdasarkan tujuan perjalanan dan jenis pengguna. Untuk Gender-Based Revenue Distribution for New and Repeat Users dan Age-Based Revenue Analysis Across New and Repeat Users, diagram lingkaran dan diagram batang digunakan untuk menunjukkan pola pendapatan berdasarkan jenis kelamin dan kelompok usia. Terakhir, The Top 10 Properties With the Highest RNs With Brand Type Details dibuat menggunakan grafik batang, di mana properti diurutkan berdasarkan total Room Nights (RNs) dengan warna batang mencerminkan Brand Type. Semua visualisasi diintegrasikan ke dalam dashboard interaktif, dengan filter untuk Booking Date, City, dan Travel

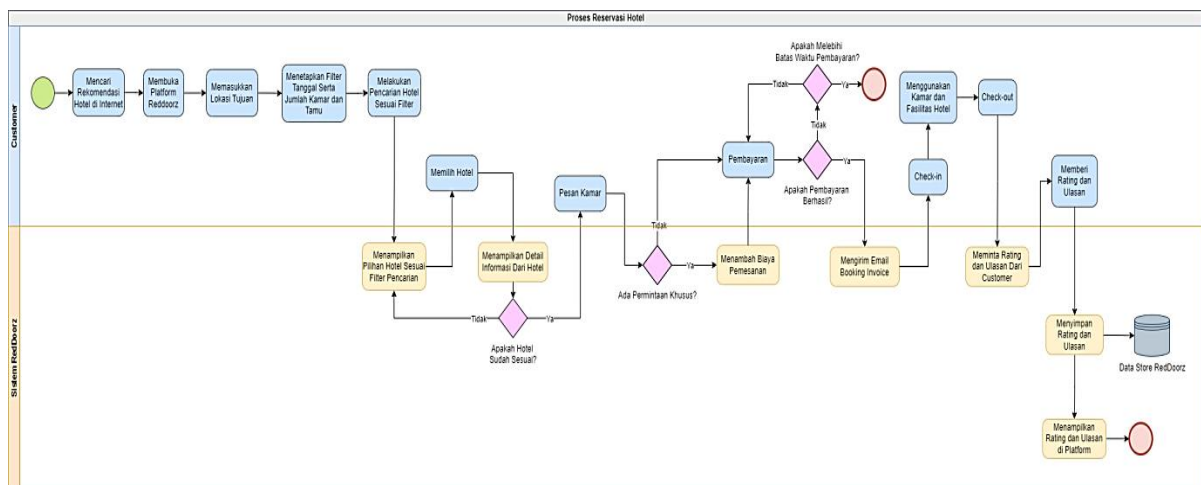
Purpose diterapkan secara menyeluruh untuk memastikan visualisasi saling terhubung dan dinamis.

3.3 Pembahasan

➤ Business Modeling

Optimasi User Journey Customer

Analisis business modeling untuk mengidentifikasi lebih lanjut mengenai tahapan user journey pelanggan disajikan dalam diagram BPMN sebagai berikut:



Gambar 3. Diagram BPMN User Journey Customer Dalam Proses Reservasi Hotel RedDoorz

Diagram BPMN menggambarkan alur *user journey* pelanggan RedDoorz yang terdiri dari lima tahap utama: *discovery*, *booking*, *pre-stay*, *during stay*, dan *post-stay*. Berikut adalah analisis keseluruhan untuk setiap tahap:

1. Discovery (Pencarian Properti)

Pada tahap ini, pelanggan memulai dengan mencari informasi hotel di internet, kemudian mereka membuka platform RedDoorz, dan menetapkan filter seperti lokasi, tanggal check-in dan check-out, serta jumlah kamar dan tamu untuk menemukan hotel yang sesuai dengan kebutuhan. Dengan demikian, pelanggan dapat mencari hotel sesuai dengan preferensi yang mereka inginkan.

Sistem harus memastikan untuk menampilkan pilihan hotel yang akurat sesuai dengan filter yang ditetapkan oleh pelanggan. Hal ini akan memberikan

pengalaman yang baik kepada pelanggan. Dengan demikian, kepuasan pelanggan terhadap platform RedDoorz dapat meningkat.

2. Booking (Pemesanan Kamar)

Setelah memilih hotel, pelanggan melanjutkan ke tahap pemesanan dengan memasukkan detail reservasi yaitu memilih permintaan khusus, dan melakukan pembayaran. Sistem secara otomatis menambah biaya untuk permintaan khusus, namun opsi permintaan khusus yang tersedia masih terbatas antara lain yaitu, early check-in, late check-out, dan add breakfast. Pembayaran harus dilakukan dalam batas waktu tertentu, dan jika gagal, reservasi dibatalkan.

RedDoorz dapat bekerjasama dengan hotel partnernya untuk meningkatkan fleksibilitas pada opsi permintaan khusus. Selain itu, sistem RedDoorz dapat menyediakan waktu yang cukup untuk melakukan pembayaran dan memastikan pengingat otomatis mendekati batas waktu pembayaran untuk mencegah pembatalan.

3. Pre-Stay (Sebelum Menginap)

Setelah pembayaran selesai, sistem mengirimkan email konfirmasi kepada pelanggan yang berisi detail reservasi. Dengan demikian, pelanggan dapat mengetahui dengan jelas semua informasi penting mengenai reservasi yang telah dibayar sebelumnya. Pelanggan juga dapat menghubungi *Contact Person* yang tersedia untuk berkomunikasi lebih lanjut yang mungkin memerlukan penyesuaian tambahan pada reservasi.

RedDoorz dapat mengoptimalkan komunikasi dengan pelanggan melalui pengingat otomatis ketika menjelang check-in. Selain itu, RedDoorz juga dapat menambahkan opsi untuk menyampaikan kebutuhan spesifik tambahan oleh pelanggan seperti melalui pengisian formulir digital yang akan memudahkan pelanggan. Hal ini juga harus diikuti oleh koordinasi yang baik oleh RedDoorz dengan hotel partnernya agar permintaan khusus pelanggan dapat dieksekusi dengan baik. Upaya ini dapat meningkatkan kepercayaan pelanggan untuk melanjutkan proses reservasi ke tahap menginap nanti.

4. During Stay (Selama Menginap)

Tahapan ini melibatkan proses check-in di hotel, penggunaan fasilitas hotel oleh pelanggan selama menginap, hingga pelanggan melakukan check-out. Alur pada tahapan ini cukup sederhana, namun koordinasi RedDoorz dengan hotel partnernya sangat penting selama proses menginap. Hotel partner harus menangani dengan baik jika ada keluhan dari pelanggan selama menginap dan RedDoorz juga harus memastikan bahwa keluhan tersebut apakah sudah ditangani oleh layanan hotel atau ada kendala tertentu yang menghambat penanganan masalah.

RedDoorz dapat memberikan email kepada pelanggan yang menanyakan apakah pelanggan memiliki keluhan dan bagaimana pihak hotel mengatasi keluhan tersebut. Dengan demikian, RedDoorz dapat memantau pengalaman pelanggan selama menginap sehingga akan meminimalisir kemungkinan adanya ulasan buruk dari pelanggan yang dapat menurunkan rating hotel dan layanan RedDoorz.

5. Post-Stay (Setelah Menginap)

Setelah check-out, sistem meminta ulasan atau feedback dari pelanggan dalam bentuk rating. Rating dan ulasan yang diberikan oleh pelanggan akan disimpan oleh sistem RedDoorz dan akan ditampilkan pada detail informasi hotel terkait. Hal ini sangat berguna untuk evaluasi RedDoorz dan hotel partnernya. Namun, tidak semua pelanggan mau memberikan rating dan ulasannya setelah menginap. Permintaan pemberian rating dan ulasan harus dibuat lebih menarik agar pelanggan mau memberikan rating dan ulasannya.

RedDoorz dapat memberikan insentif seperti voucher diskon untuk pemesanan berikutnya atau hal lainnya yang dapat meningkatkan partisipasi pelanggan dalam pengisian ulasan. Upaya ini dapat meningkatkan kepercayaan pelanggan lain yang akan menginap karena mereka dapat membaca lebih banyak ulasan dari pelanggan yang sudah pernah menginap. Selain itu, data feedback dapat digunakan untuk analisis operasional yang lebih baik sehingga meningkatkan layanan RedDoorz dalam proses reservasi dari tahap pencarian hotel hingga pelanggan selesai menginap.

Secara keseluruhan, alur *user journey* pelanggan dalam proses reservasi hotel RedDoorz sudah cukup lengkap dan efektif sehingga memudahkan pelanggan dalam melakukan reservasi. Namun, beberapa hal di setiap tahapannya perlu ditingkatkan. Mulai dari hasil pencarian yang harus lebih akurat, permintaan khusus yang lebih fleksibel, dan pembayaran yang didukung oleh pengingat otomatis. Selain itu, perlu adanya komunikasi aktif dengan pelanggan sebelum menginap, dan merespons dengan cepat terhadap keluhan mereka, serta memberikan insentif tertentu untuk mendorong mereka memberi rating dan ulasan. Beberapa hal tersebut merupakan rekomendasi upaya untuk dapat memaksimalkan pengalaman pelanggan di setiap tahapan reservasinya sehingga dapat meningkatkan kepuasan mereka.

➤ **Data Cleaning Pada Python**

Kolom `USER_ID` dan `BOOKING_ID` idealnya direpresentasikan sebagai string karena berisi nilai-nilai unik yang mengidentifikasi pengguna dan pemesanan. Mengubahnya menjadi tipe data string memastikan bahwa nilai-nilai tersebut diperlakukan sebagai data kategorikal dan tidak digunakan dalam perhitungan numerik yang tidak relevan. Hal ini juga memfasilitasi operasi seperti penggabungan (`merge`) dan pengelompokan (`groupby`) berdasarkan ID.

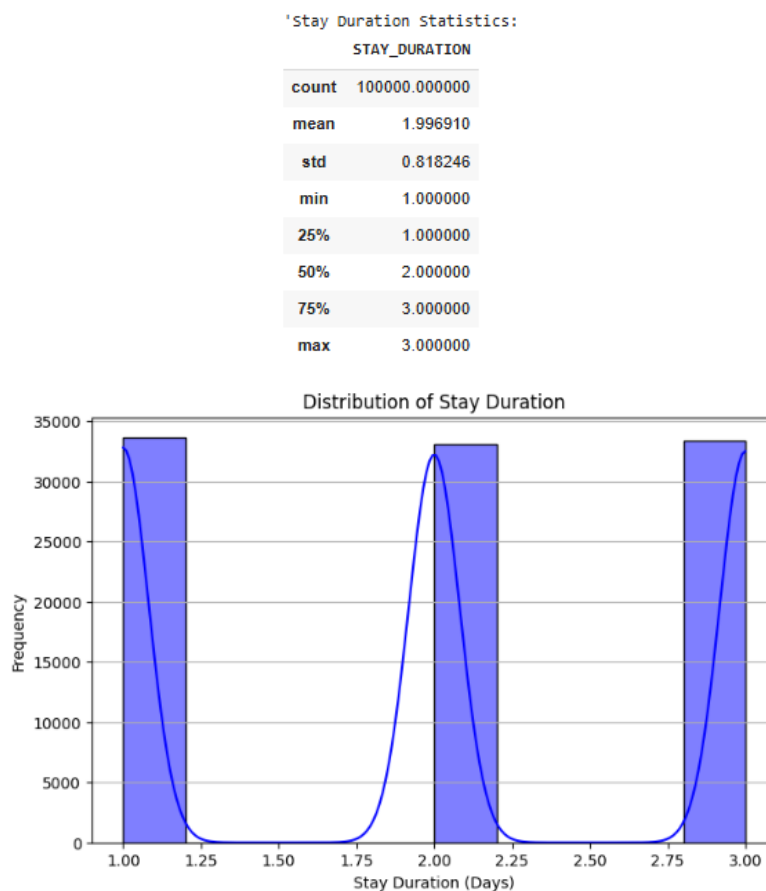
Kolom-kolom `BOOKING_DATE` `CHECK_IN_DATE` `CHECK_OUT_DATE` ini berisi informasi tanggal dan waktu, sehingga mengubahnya menjadi tipe data *datetime* memungkinkan kita untuk melakukan operasi dan analisis yang spesifik untuk data waktu, seperti menghitung selisih waktu (*stay duration*, *lead time*), mengekstrak komponen tanggal (hari, bulan, tahun), dan melakukan pengurutan berdasarkan waktu. Menggunakan tipe data *datetime* juga memastikan konsistensi dan akurasi dalam representasi data waktu.

Perubahan format tanggal ini dilakukan untuk memastikan bahwa data tanggal di kolom-kolom `BOOKING_DATE` `CHECK_IN_DATE` `CHECK_OUT_DATE` tersebut konsisten dan sesuai dengan format yang diinginkan (tanggal(d)-bulan(m)-tahun(y)). Format yang konsisten penting untuk menghindari kesalahan dalam

interpretasi dan analisis data, serta untuk memfasilitasi operasi seperti penggabungan dan pengelompokan berdasarkan tanggal.

Kolom `COHOR_DATE` kemungkinan berisi informasi tanggal, sehingga mengubahnya menjadi tipe data `datetime` memberikan manfaat yang sama seperti yang dijelaskan sebelumnya untuk kolom-kolom tanggal di dataset `bookings_df`. Hal ini memungkinkan kita untuk melakukan analisis dan manipulasi data waktu dengan lebih mudah dan akurat.

➤ Insight Booking Stay Duration and Booking Lead Time



Gambar 4. *Insight Distribusi Stay Duration*

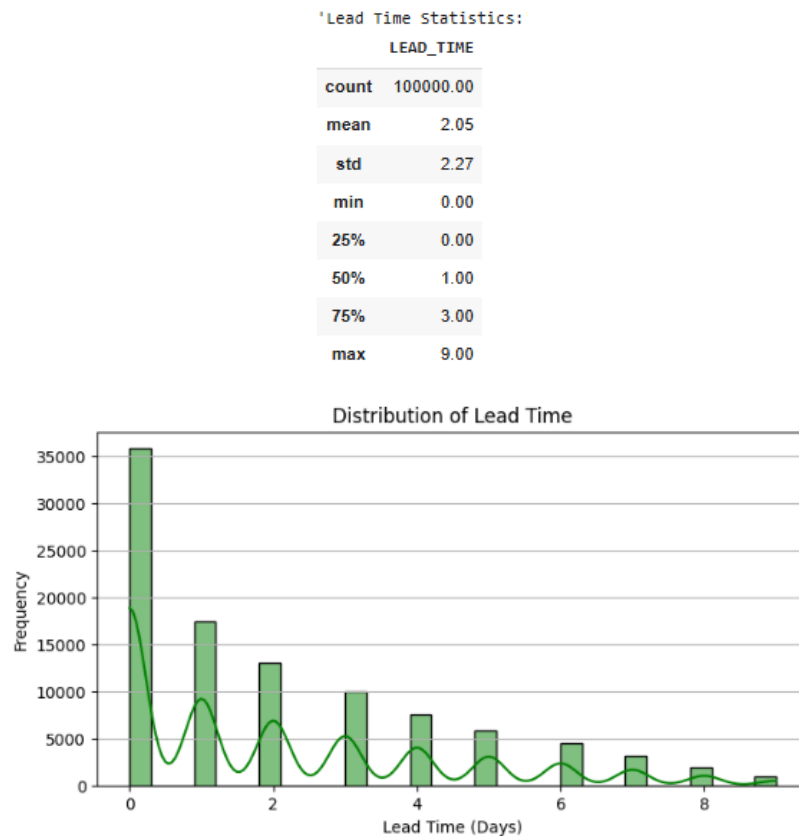
Berdasarkan hasil analisis terhadap **Stay Duration** (Durasi Menginap), disini mendapatkan beberapa informasi penting mengenai pola menginap pelanggan di **RedDoorz** yang dapat memberikan insight mendalam untuk strategi bisnis yang lebih efektif. Stay Duration dihitung dengan mengambil selisih antara tanggal **CHECK_OUT_DATE** dan **CHECK_IN_DATE**. Ini memberikan sebuah

informasi tentang berapa lama tamu menginap di properti tersebut dalam satuan hari. Hasil analisis menunjukkan bahwa berdasarkan **statistik deskriptif** yang telah dihitung durasi menginap sebagian besar pelanggan menginap selama **2 hari**, dengan sedikit variasi untuk durasi yang lebih pendek **1 hari** dan lebih lama **3 hari**, yang bisa menunjukkan bahwa banyak pelanggan yang memilih menginap selama dua malam sebagai durasi yang optimal untuk perjalanan singkat.

Dari data yang ada, durasi menginap **1 hari** (33,631 pemesanan), **2 hari** (33,047 pemesanan), dan **3 hari** (33,322 pemesanan) menunjukkan distribusi yang hampir **seimbang**, dengan ketiga durasi tersebut memiliki jumlah pemesanan yang hampir setara. **Variasi durasi menginap yang tidak terlalu signifikan** ini menunjukkan bahwa pelanggan RedDoorz memiliki kecenderungan untuk memilih **durasi menginap yang cukup fleksibel**, tanpa dominasi durasi tertentu yang jelas. Bahwa meskipun durasi menginap 3 hari cukup populer dengan jumlah pemesanan yang signifikan (33.3%), mayoritas pelanggan RedDoorz (66.7%) cenderung memilih durasi menginap yang lebih singkat, yaitu 1 hingga 2 hari. Hal ini menunjukkan bahwa RedDoorz perlu memperhatikan kebutuhan pelanggan yang menginap dalam waktu singkat dan menawarkan layanan atau promosi yang sesuai dan hal ini juga dapat memberikan peluang bagi RedDoorz untuk mengeksplorasi penawaran atau paket harga bagi pelanggan yang menginap lebih lama. Maka, terlihat menggambarkan bahwa mayoritas pemesanan berfokus pada masa tinggal yang lebih singkat seperti perjalanan bisnis atau liburan akhir pekan.

Ini memberikan wawasan strategi bisnis yang penting bagi manajemen properti bahwa mayoritas pelanggan lebih cenderung memilih masa tinggal yang singkat. Untuk meningkatkan pendapatan per malam, RedDoorz bisa menawarkan **penawaran untuk menginap lebih lama** atau **paket durasi panjang** bagi pelanggan yang menginap selama 3 hari atau lebih. Misalnya diskon untuk **3 malam berturut-turut** atau layanan tambahan untuk mendorong pelanggan memilih durasi menginap lebih panjang. Penawaran ini bisa menarik bagi pelanggan yang mencari akomodasi untuk liburan lebih lama atau perjalanan bisnis yang membutuhkan waktu lebih lama. Dengan mengetahui bahwa mayoritas pelanggan menginap selama **1 hingga 2 hari**, RedDoorz bisa **mengoptimalkan**

pengelolaan ketersediaan kamar dengan lebih fleksibel untuk durasi menginap tersebut. Misalnya, menawarkan **penawaran harga lebih rendah** untuk menginap lebih pendek atau menyesuaikan tarif untuk pelanggan yang menginap lebih lama, dan bisa juga dengan memastikan harga yang kompetitif pada masa-masa puncak dengan durasi menginap lebih pendek, seperti akhir pekan.



Gambar 5. *Insight Distribusi Lead Time*

Berdasarkan analisis dan visualisasi distribusi **Lead Time** (waktu yang diperlukan pelanggan untuk melakukan pemesanan sebelum check-in). Insight ini akan membantu RedDoorz memahami pola pemesanan pelanggan dan bagaimana strategi pemasaran serta penawaran dapat disesuaikan. Lead time dihitung sebagai selisih antara **BOOKING_DATE** dan **CHECK_IN_DATE**, menunjukkan berapa lama sebelum tanggal check-in pemesanan dilakukan. Berdasarkan statistik deskriptif bahwa **Rata-rata (mean) Lead Time** adalah **2.05 hari**, yang menunjukkan bahwa pelanggan **cenderung memesan sekitar 2 hari sebelum check-in**. Ini berarti banyak pelanggan melakukan pemesanan dalam waktu yang cukup dekat dengan tanggal kedatangan mereka, yang berarti ada permintaan yang tinggi untuk **last-minute booking** menunjukkan adanya kebutuhan untuk

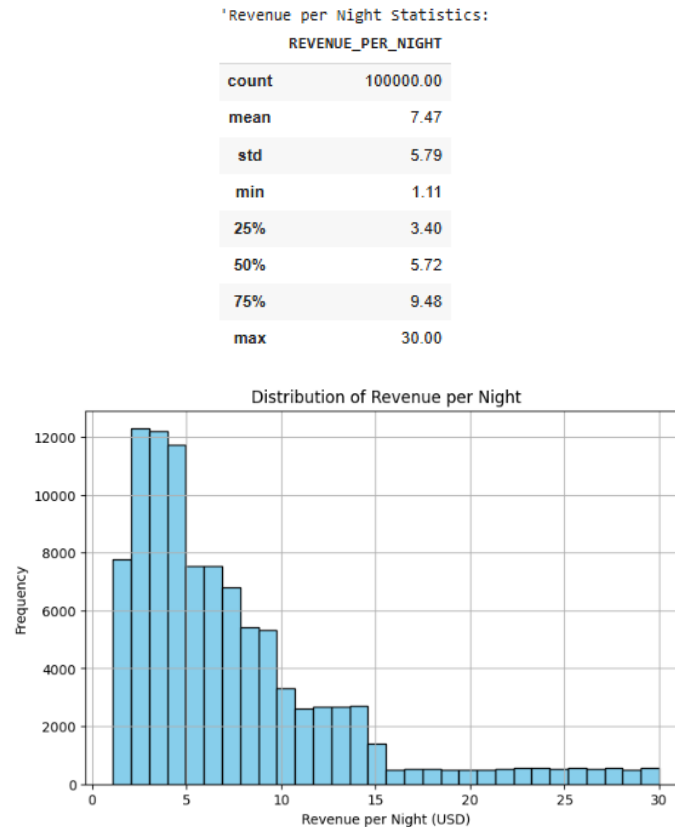
fleksibilitas yang lebih tinggi dalam sistem pemesanan dan harga. Banyak tamu mungkin mencari akomodasi dengan segera, mungkin untuk perjalanan mendadak atau karena adanya perubahan rencana.

Berdasarkan grafik distribusi Lead Time dapat dilihat bahwa ada variasi yang cukup signifikan dalam **Lead Time**, bahwa **Lead Time 0 hari** (pemesanan yang dilakukan pada hari yang sama) adalah yang paling dominan, dengan jumlah pemesanan yang sangat tinggi, yaitu **35,793 pemesanan**. Setelah itu, **Lead Time 1 hari** juga cukup tinggi, mencapai **17,462 pemesanan**, yang menunjukkan bahwa banyak pelanggan yang cenderung memesan mendekati hari check-in.

Meskipun setelah **Lead Time 2 hari** frekuensi pemesanan mulai menurun secara signifikan, masih ada segmen pelanggan yang memilih untuk memesan dengan lead time lebih panjang namun frekuensi nya juga menurun (3 hingga 9 hari).

Sehingga ini memberikan wawasan bahwa **Mayoritas pemesanan terjadi pada Lead Time 0 hingga 1 hari**, yang menunjukkan pentingnya memiliki strategi penjualan yang fleksibel dan responsif terhadap pemesanan mendadak, RedDoorz dapat meningkatkan strategi pemasaran untuk **penawaran last-minute** atau **flash sales** dengan memberikan harga lebih murah untuk pelanggan yang memesan beberapa hari sebelum check-in. Ini dapat membantu mengisi kamar yang kosong dengan lebih efisien, terutama selama periode low-season atau mendekati waktu keberangkatan. Dengan mengetahui bahwa banyak pemesanan dilakukan dalam waktu singkat, penting untuk memiliki **sistem pemesanan yang mudah dan cepat**. Hal ini juga mempengaruhi manajemen kapasitas di mana RedDoorz bisa mempertimbangkan untuk menggunakan data ini untuk lebih mengelola ketersediaan kamar dan membuat prediksi yang lebih tepat mengenai tingkat hunian dalam waktu singkat.

➤ **Insight Revenue Per Night and Total Booking Revenue**



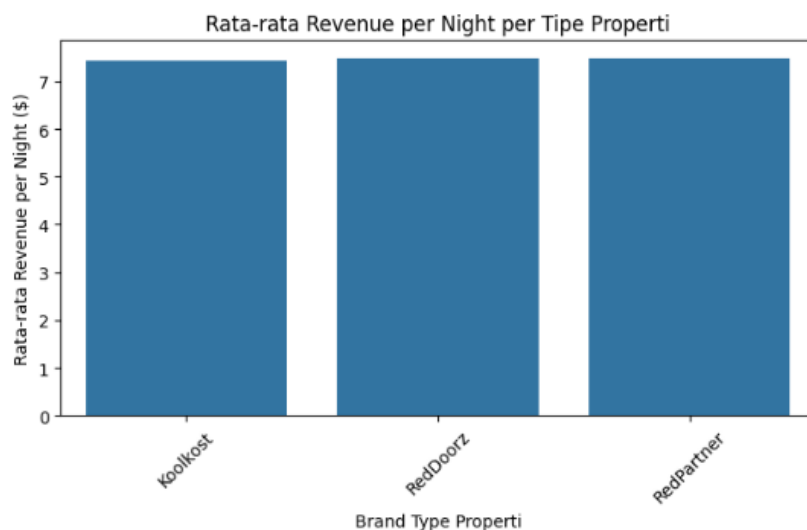
Gambar 6. *Insight Distribusi Revenue Per Night*

Rata-rata **Revenue per Night** sebesar **7.47 USD** memberikan gambaran tentang **harga yang dibayarkan pelanggan** untuk menginap di RedDoorz. Ini menunjukkan bahwa RedDoorz berada pada kisaran harga yang relatif terjangkau, yang menarik untuk **segmen pasar budget-conscious**. Dengan harga rata-rata ini, RedDoorz memposisikan dirinya sebagai pilihan **akomodasi yang terjangkau** bagi banyak pelanggan, baik wisatawan lokal maupun internasional yang mencari penginapan yang lebih ekonomis, namun tetap menawarkan kualitas penginapan yang memadai.

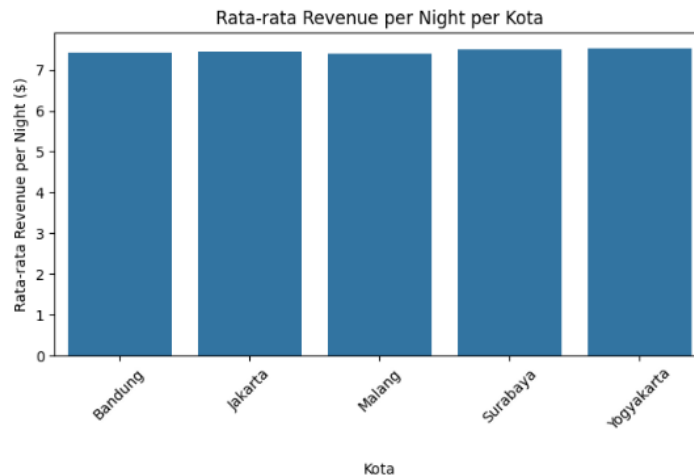
Variasi yang cukup signifikan dalam **Revenue per Night**, dengan standar deviasi **5.79**, menunjukkan bahwa harga penginapan di RedDoorz sangat bervariasi. Rentang harga dimulai dari **1.11 USD** hingga **30.00 USD**. Hal ini mengindikasikan bahwa **sejumlah pemesanan memiliki harga yang jauh lebih rendah** dengan penginapan dengan fasilitas standar, sementara yang lain seperti **properti premium** atau **pemesanan untuk grup** dapat dikenakan berada pada harga yang

jauh lebih tinggi. RedDoorz mungkin menggunakan **strategi harga dinamis** yang memungkinkan harga bervariasi berdasarkan waktu, lokasi, jenis properti, atau **promosi tertentu** yang sedang berlangsung. Hal ini memungkinkan RedDoorz untuk **menarik berbagai segmen pelanggan** yang memiliki tingkat pengeluaran berbeda.

Mayoritas properti menargetkan pasar ekonomis memiliki pendapatan per malam di kisaran **\$3 hingga \$5 USD**, menunjukkan fokus utama RedDoorz pada pasar **budget-friendly**. Distribusi **positively skewed** dengan sebagian kecil properti memiliki pendapatan per malam tinggi (di atas **\$15 USD**) hingga **\$30 USD**, mengindikasikan segmen properti premium atau di lokasi strategis. Properti dengan pendapatan rendah (<\$3 USD) memiliki potensi untuk meningkatkan pendapatan melalui **strategi dynamic pricing** atau peningkatan fasilitas. Properti dengan pendapatan tinggi menawarkan peluang untuk diperluas atau dipromosikan lebih lanjut kepada pelanggan dengan daya beli lebih besar. RedDoorz didominasi oleh properti yang menargetkan pasar ekonomis, namun terdapat peluang signifikan untuk mengoptimalkan pendapatan melalui strategi harga dinamis dan promosi pada properti premium. Diversifikasi strategi dapat meningkatkan pendapatan secara keseluruhan tanpa mengesampingkan pasar utama.



Gambar 7. *Insight Average Revenue per Night per brand type*



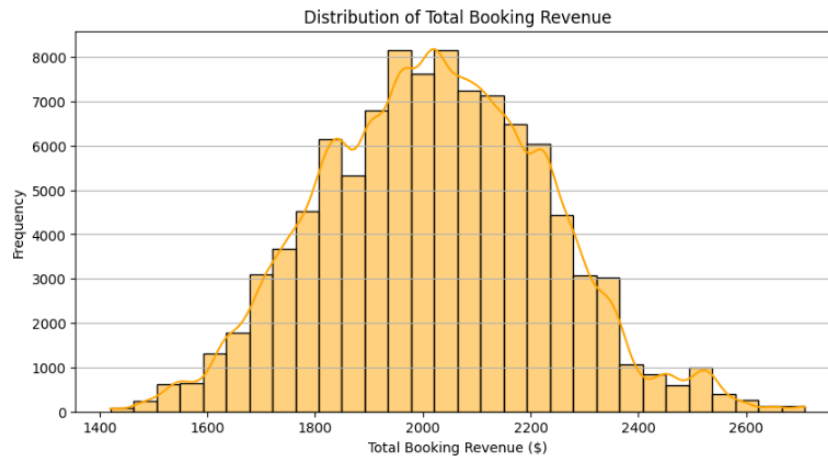
Gambar 8. *Insight Average Revenue per Night per city*

Harga per malam dipengaruhi oleh beberapa faktor penting, terutama **tipe properti** dan **lokasi**. Berdasarkan analisis, tidak ada perbedaan harga yang signifikan antara **RedDoorz**, **RedPartner**, dan **Koolkost** (semua berada di kisaran harga rata-rata yang hampir sama, yaitu sekitar 7.47 USD). Namun, **lokasi** menunjukkan adanya sedikit variasi, dengan kota-kota yang lebih populer seperti **Yogyakarta** dan **Surabaya** memiliki sedikit harga lebih tinggi dibandingkan kota-kota lain. Hal ini bisa disebabkan oleh tingginya **tingkat okupansi** di kota-kota tersebut, yang memungkinkan properti di daerah tersebut untuk menawarkan harga sedikit lebih mahal. Untuk kota dengan permintaan lebih rendah, seperti **Malang**, RedDoorz dapat menawarkan **penawaran khusus** atau harga lebih rendah untuk menarik pelanggan. Meskipun harga antar tipe properti terlihat seragam, ada kemungkinan bahwa **penawaran premium** atau **fasilitas tambahan** yang ditawarkan di beberapa properti dapat meningkatkan **revenue per night** jika harga dapat disesuaikan dengan jenis layanan yang disediakan.

```

Total Booking Revenue Statistics:
TOTAL_BOOKING_REVENUE
count      1000.00
mean       2001.13
std        209.08
min        1419.79
25%        1848.58
50%        2000.57
75%        2146.23
max        2708.86
dtype: float64

```

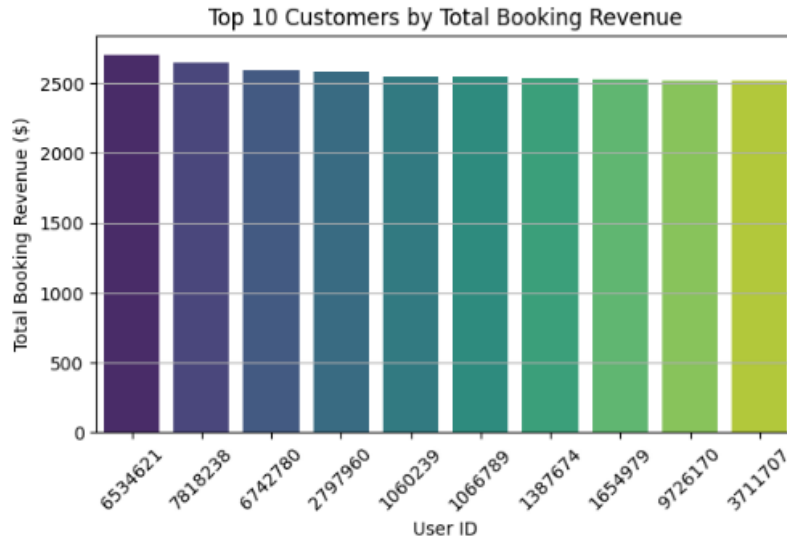


Gambar 9. *Insight analisis distribusi total booking revenue*

Distribusi Pendapatan Pelanggan, dengan rata-rata total pendapatan pemesanan sekitar **2000 USD**, sebagian besar pelanggan menghasilkan pendapatan di kisaran **1,800 hingga 2,200 USD**. Grafik distribusi menunjukkan bentuk **distribusi normal** untuk total pendapatan pemesanan, dengan puncak di sekitar **2,000 USD**, yang menunjukkan bahwa mayoritas pelanggan menghasilkan pendapatan pemesanan di kisaran tersebut. Ini menunjukkan bahwa pelanggan dengan pendapatan yang lebih tinggi (seperti pemesanan untuk durasi lama atau lebih banyak kamar) berkontribusi secara signifikan terhadap total pendapatan. Berdasarkan analisis, pelanggan dengan total pendapatan pemesanan tertinggi adalah mereka yang menghabiskan lebih banyak uang untuk pemesanan, yang mencakup lebih banyak malam atau kamar yang lebih premium. Ini memberikan gambaran bahwa strategi pemesanan yang berfokus pada pelanggan ini dapat meningkatkan pendapatan lebih lanjut.

top_10_customers in terms of Total Booking Revenue :

USER_ID	TOTAL_BOOKING_REVENUE	
766	6534621	2708.86
839	7818238	2651.86
774	6742780	2596.37
501	2797960	2584.47
58	1060239	2548.25
63	1066789	2545.37
317	1387674	2544.78
429	1654979	2530.24
982	9726170	2524.71
560	3711707	2523.20



Gambar 10. Insight analisis Top 10 Customers by Total Booking Revenue

Berdasarkan hasil visualisasi tersebut, terlihat bahwa beberapa pelanggan dengan ID tertentu memiliki kontribusi yang sangat besar terhadap total pendapatan pemesanan. Disini pelanggan dengan ID **6534621** berkontribusi lebih dari **2,700 USD**. Ini menunjukkan bahwa pelanggan-pelanggan tertentu yang melakukan pemesanan dengan nilai lebih tinggi atau lebih sering memiliki dampak yang signifikan terhadap pendapatan perusahaan. Dengan mengetahui siapa pelanggan dengan total pendapatan terbesar, RedDoorz dapat menciptakan strategi yang lebih terfokus untuk meningkatkan loyalitas pelanggan ini. Misalnya, dengan memberikan **reward khusus** atau **penawaran eksklusif** bagi pelanggan yang menghasilkan pendapatan terbesar untuk mendorong pemesanan lebih lanjut.

➤ Insight Property Performance and Occupancy Rate

'Top 10 Properties by Total Revenue:'

	PROPERTY_CODE	TOTAL_ROOM_NIGHTS_BOOKED	TOTAL_REVENUE
431	YOG0221	924	4928.83
496	YOG0490	961	4785.12
110	JAK0031	974	4757.58
343	SUR0214	900	4731.78
215	MAL0079	915	4705.37
245	MAL0202	937	4680.02
483	YOG0438	930	4668.22
0	BAN0002	924	4657.61
281	MAL0400	996	4643.54
365	SUR0326	968	4633.20

'Total Revenue Statistics:

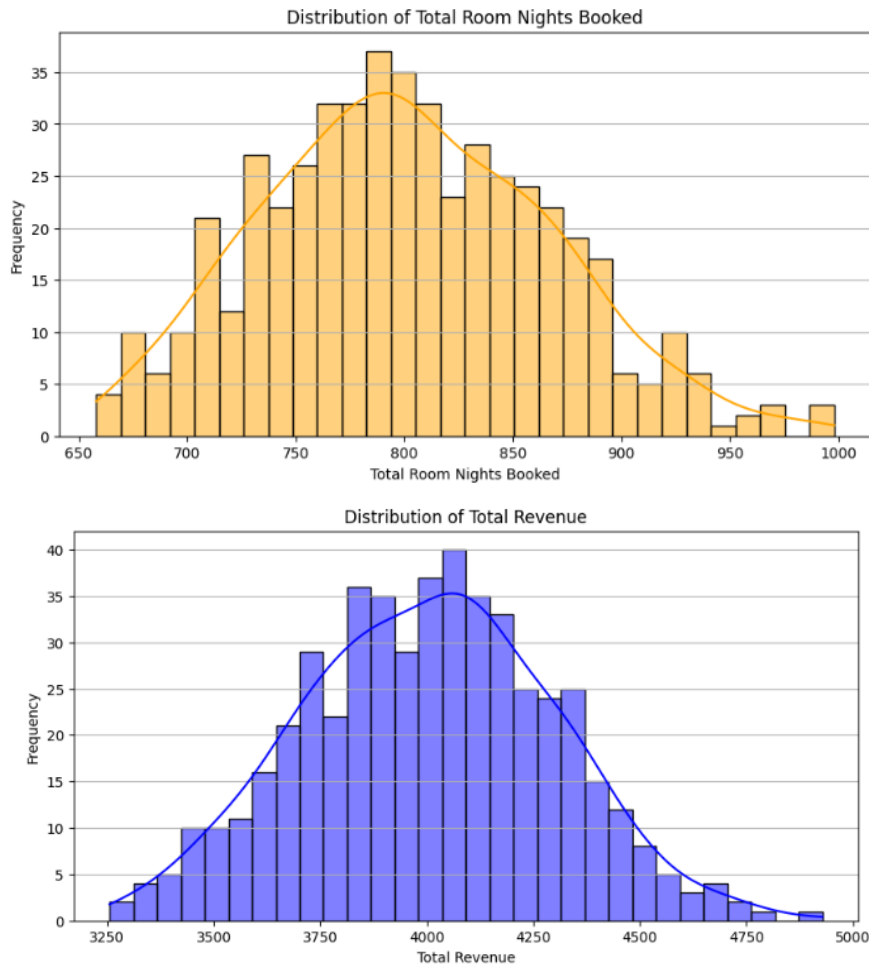
TOTAL_REVENUE	
count	500.00
mean	4002.26
std	295.18
min	3255.23
25%	3797.96
50%	4016.23
75%	4202.58
max	4928.83

'Top 10 Properties by Total Room Nights Booked:'

	PROPERTY_CODE	TOTAL_ROOM_NIGHTS_BOOKED	TOTAL_REVENUE
476	YOG0421	998	4550.25
281	MAL0400	996	4643.54
369	SUR0344	990	4553.02
110	JAK0031	974	4757.58
313	SUR0056	970	4494.59
365	SUR0326	968	4633.20
336	SUR0191	961	4405.23
496	YOG0490	961	4785.12
78	BAN0370	951	4315.71
74	BAN0336	940	4220.36

'Total Room Nights Booked Statistics:

TOTAL_ROOM_NIGHTS_BOOKED	
count	500.00
mean	801.28
std	65.88
min	658.00
25%	753.75
50%	796.50
75%	848.25
max	998.00



Gambar 11. *property performance*

Berdasarkan data terkait property performance bahwa properti yang memiliki **Total Room Nights Booked** tertinggi adalah **YOG0421**, dengan **998 malam** yang dipesan. Properti ini menunjukkan tingkat hunian yang sangat tinggi, mengindikasikan popularitas atau kapasitas yang sangat baik dalam memenuhi permintaan pelanggan. Properti ini berada di posisi teratas dalam hal volume pemesanan. Ini menunjukkan wawasan bahwa properti dengan **Total Room Nights Booked** tertinggi menunjukkan bahwa volume pemesanan yang tinggi tidak hanya bergantung pada harga atau fasilitas, tetapi juga pada faktor lain seperti **lokasi**, **ketersediaan fasilitas**, atau **musim** yang mempengaruhi jumlah pemesanan. Properti seperti ini bisa menjadi model untuk properti lain dalam mengoptimalkan kapasitas dan meningkatkan pemesanan dengan strategi yang lebih baik.

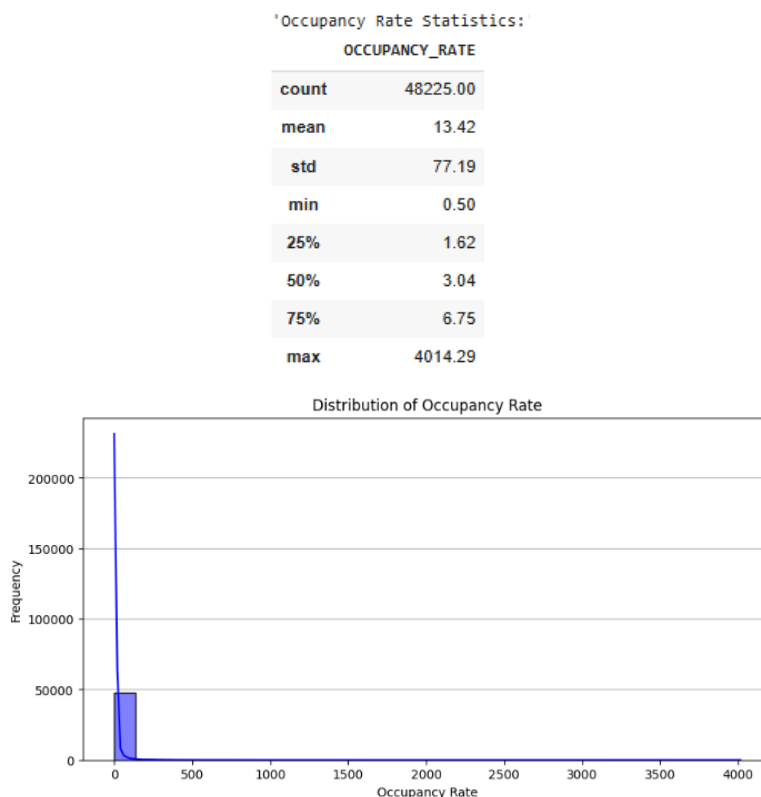
Properti yang memiliki **Total Revenue** tertinggi adalah **YOG0221**, dengan pendapatan sebesar **4928.83 USD**. Pendapatan yang lebih tinggi ini dapat disebabkan oleh faktor-faktor seperti **harga per malam** yang lebih tinggi, jumlah pemesanan yang cukup banyak, atau jenis layanan premium yang ditawarkan. Ini menunjukkan wawasan bahwa pencapaian **Total Revenue** tertinggi tidak selalu terkait dengan jumlah malam yang dipesan. Properti seperti **YOG0221** dengan harga premium atau layanan tambahan bisa mendapatkan pendapatan lebih tinggi meskipun jumlah pemesanan kamar relatif lebih rendah. Oleh karena itu, strategi untuk meningkatkan **revenue** dapat melibatkan **penyesuaian harga** atau **peningkatan layanan premium** di properti lainnya.

Terdapat variasi yang signifikan antara **Total Room Nights Booked** dan **Total Revenue**. Meskipun beberapa properti memiliki pemesanan yang tinggi, seperti **YOG0421**, properti lainnya dengan jumlah malam dipesan yang lebih rendah dapat menghasilkan pendapatan yang lebih tinggi, seperti **YOG0221**. Ini menunjukkan wawasan bahwa variasi ini menunjukkan bahwa performa properti tidak hanya bergantung pada jumlah malam yang dipesan, tetapi juga dipengaruhi oleh **faktor harga, fasilitas, dan jenis layanan** yang ditawarkan. Properti dengan pendapatan tinggi sering kali menawarkan layanan premium atau tarif yang lebih tinggi, yang memungkinkan mereka menghasilkan lebih banyak meskipun dengan volume pemesanan yang lebih sedikit. Oleh karena itu, properti dengan volume pemesanan lebih rendah dapat fokus pada peningkatan kualitas layanan atau strategi penetapan harga yang lebih tinggi.

Property performance (termasuk **Total Room Nights Booked** dan **Total Revenue**) dipengaruhi oleh berbagai faktor, seperti **lokasi, harga per malam, fasilitas, dan tipe properti**. Properti dengan fasilitas yang lebih lengkap atau harga yang lebih tinggi cenderung memiliki performa yang lebih baik dalam hal **pendapatan**, meskipun volumenya mungkin lebih kecil. Ini menunjukkan wawasan bahwa faktor-faktor seperti **lokasi yang strategis** dan **tipe properti** yang menawarkan pengalaman unik dapat meningkatkan **performansi properti** secara signifikan. Properti di lokasi populer atau yang menawarkan pengalaman eksklusif bisa menghasilkan pendapatan lebih tinggi meskipun dengan volume pemesanan

yang lebih rendah. Oleh karena itu, pengelola properti harus menyesuaikan layanan dan harga dengan **kebutuhan pasar** dan **preferensi pelanggan** untuk meningkatkan performa keseluruhan.

Berdasarkan grafik distribusi **Total Room Nights Booked** dan **Total Revenue** menunjukkan pola distribusi yang cenderung **normal** dengan sedikit skew ke kanan. Sebagian besar properti memiliki performa yang konsisten dengan **jumlah malam yang dipesan** sekitar **750 hingga 850 malam**, dan **pendapatan** antara **3500 hingga 4500 USD**. Ada beberapa properti yang jauh lebih tinggi dalam pendapatan dan jumlah malam yang dipesan. Ini menunjukkan wawasan terkait pola distribusi ini mengindikasikan bahwa sebagian besar properti memiliki performa yang **seragam** dalam hal pemesanan dan pendapatan. Namun, ada beberapa properti dengan performa jauh lebih tinggi, yang menunjukkan bahwa ada **peluang besar untuk optimasi**. Properti dengan performa rendah bisa mengadopsi strategi dari properti dengan kinerja tinggi, seperti **penyesuaian harga, peningkatan fasilitas, atau kampanye pemasaran yang lebih baik**.



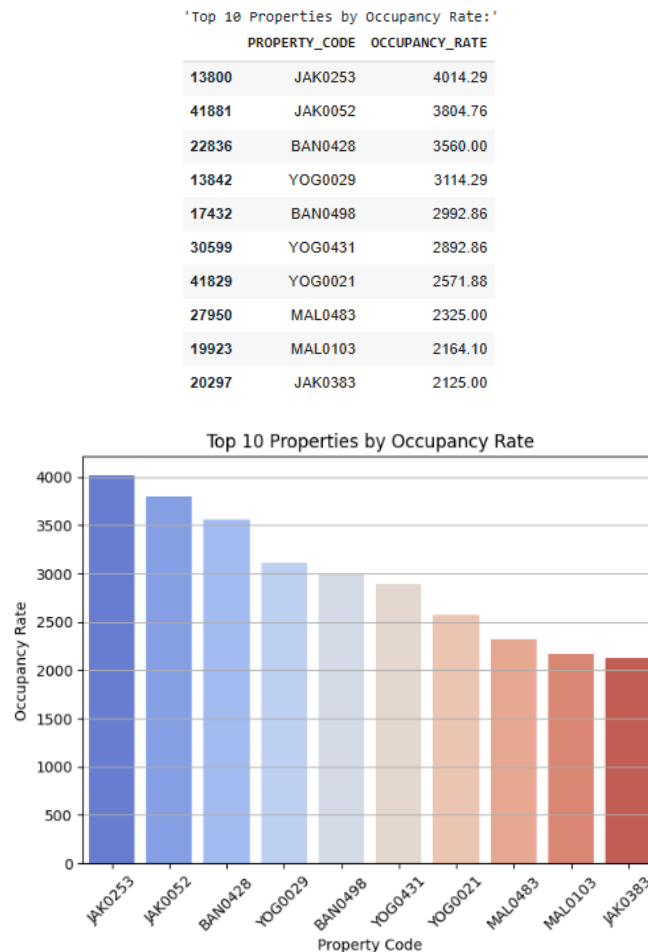
Gambar 12. *Occupancy rate*

Rata-rata **Occupancy Rate** di **RedDoorz** adalah **13.42%**, dengan nilai minimum **0.50%** dan maksimum **4014.29%**. Hal ini menunjukkan adanya distribusi yang sangat bervariasi dalam tingkat hunian antar properti. Sebagian besar properti memiliki **Occupancy Rate** yang rendah, namun ada beberapa properti yang sangat efisien dalam mengisi kamar mereka, dengan tingkat hunian yang sangat tinggi. Ini menunjukkan wawasan bahwa rata-rata **Occupancy Rate** yang rendah menandakan bahwa **banyak properti** yang memiliki kapasitas tidak terisi secara optimal. Properti yang memiliki **Occupancy Rate** rendah mungkin bisa memanfaatkan strategi promosi atau **penyesuaian harga** untuk meningkatkan tingkat hunian. Sebaliknya, properti dengan **Occupancy Rate** tinggi menunjukkan bahwa ada potensi **efisiensi yang lebih tinggi** dalam pengelolaan kapasitas.

Beberapa faktor utama yang mempengaruhi **Occupancy Rate** diantaranya adalah Tipe Properti biasanya properti dengan layanan premium atau yang menawarkan pengalaman eksklusif cenderung memiliki tingkat hunian yang lebih tinggi. Lokasi dimana properti yang terletak di daerah strategis atau dekat dengan tujuan wisata, pusat bisnis, atau transportasi umum biasanya memiliki tingkat hunian lebih tinggi. Penetapan harga yang kompetitif dan penyesuaian harga selama musim sepi dapat meningkatkan tingkat hunian. Tingkat hunian lebih tinggi biasanya terjadi pada musim liburan atau akhir pekan. Ini menunjukkan wawasan bahwa faktor-faktor ini memberikan petunjuk penting tentang bagaimana meningkatkan **Occupancy Rate**. Untuk properti dengan **Occupancy Rate** rendah, strategi yang bisa diterapkan adalah **menyesuaikan harga** sesuai dengan permintaan pasar, **menawarkan diskon musiman**, dan **meningkatkan kualitas layanan**. Properti yang memiliki **Occupancy Rate** tinggi perlu mempertahankan kualitas layanan dan terus mengoptimalkan harga agar tetap kompetitif.

Pada grafik terlihat bawah distribusi **Occupancy Rate** menunjukkan **skewness** yang sangat besar, dengan sebagian besar properti memiliki tingkat hunian yang sangat rendah, dan beberapa properti memiliki tingkat hunian yang sangat tinggi (seperti **JAK0253**). Grafik distribusi menunjukkan puncak yang tajam pada nilai yang sangat rendah, dengan ekor panjang yang menuju nilai yang lebih tinggi. Ini menunjukkan wawasan bahwa Distribusi yang sangat miring ini menunjukkan

bahwa banyak properti belum mengisi kapasitas mereka secara optimal, meskipun ada beberapa properti yang sangat efisien dalam mengelola tingkat hunian. Ini mengindikasikan bahwa ada **peluang besar untuk optimasi**. Properti dengan **Occupancy Rate** rendah dapat meningkatkan tingkat hunian dengan strategi pemasaran yang lebih baik, seperti penawaran musiman, promosi harga, dan peningkatan eksposur ke pasar yang lebih luas.



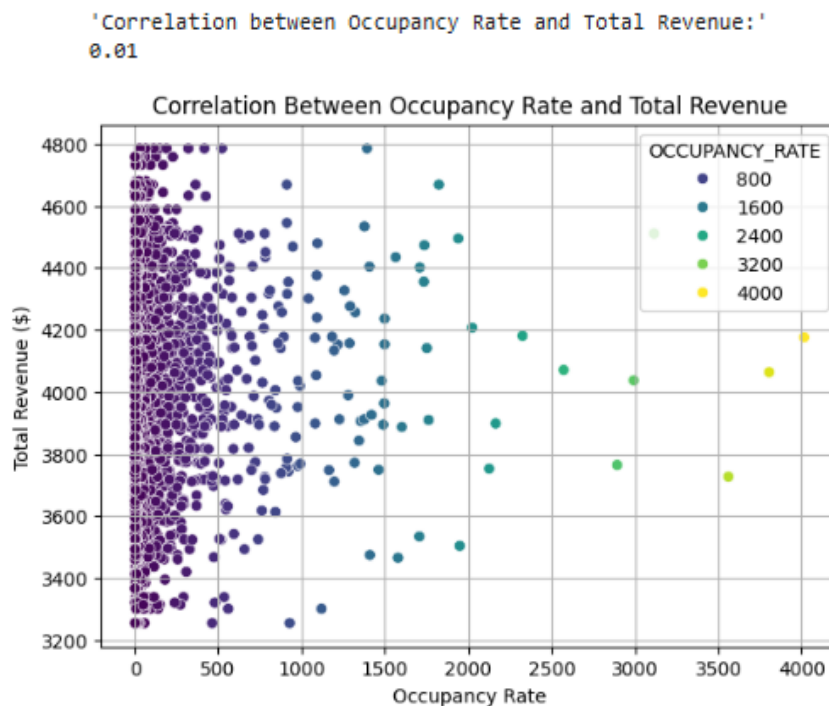
Gambar 13. Top 10 Properti berdasarkan Occupancy rate

JAK0253 memiliki **Occupancy Rate** tertinggi, yaitu **4014.29%**. Hal ini menunjukkan bahwa properti ini sangat efisien dalam mengisi kamar mereka, bahkan mungkin jauh melampaui kapasitas yang tersedia, yang menunjukkan popularitas atau penetapan harga yang sangat optimal.

Ini menunjukkan wawasan bahwa properti dengan **Occupancy Rate** sangat tinggi seperti **JAK0253** dapat menjadi model untuk properti lainnya. Faktor yang mungkin berkontribusi pada hasil ini termasuk **lokasi strategis**, **penawaran**

layanan premium, atau **strategi harga dinamis**. Properti dengan performa seperti ini harus dianalisis lebih lanjut untuk mengidentifikasi praktik terbaik yang dapat diadopsi oleh properti lainnya.

Terdapat variasi yang signifikan dalam **Occupancy Rate** antar properti. Beberapa properti seperti **JAK0253** memiliki tingkat hunian yang sangat tinggi, sedangkan properti lain seperti **JAK0383** memiliki **Occupancy Rate** yang jauh lebih rendah, hanya **2125.00%**. Ini menunjukkan wawasan bahwa variasi besar dalam **Occupancy Rate** ini menunjukkan bahwa ada beberapa properti yang sangat sukses dalam mengisi kamar mereka, sementara yang lain tidak begitu optimal. Hal ini dapat disebabkan oleh berbagai faktor seperti **lokasi**, **tipe properti**, **penetapan harga**, atau bahkan **musim**. Properti dengan **Occupancy Rate** rendah perlu melakukan analisis untuk mengetahui penyebabnya dan menerapkan strategi yang lebih efektif untuk meningkatkan tingkat hunian.



Gambar 14. Korelasi antara occupancy rate dan total revenue per property

Hasil analisis menunjukkan korelasi yang sangat rendah, yaitu **0.01**, antara **Occupancy Rate** dan **Total Revenue**. Hal ini mengindikasikan bahwa tingkat hunian (Occupancy Rate) tidak memiliki hubungan yang signifikan dengan total pendapatan yang dihasilkan oleh properti. Meskipun beberapa properti dengan

Occupancy Rate tinggi dapat menghasilkan pendapatan yang lebih besar, hasil ini tidak cukup konsisten untuk mengonfirmasi hubungan yang kuat antara keduanya.

Ini menunjukkan wawasan bahwa korelasi yang rendah ini menunjukkan bahwa meskipun **Occupancy Rate** dapat mempengaruhi pendapatan total, **faktor-faktor lain**, seperti **harga per malam**, **lokasi**, **jenis layanan**, dan **musim**, lebih berperan dalam menentukan total pendapatan properti. Oleh karena itu, **tingkat hunian** bukan satu-satunya indikator kesuksesan dalam mengoptimalkan pendapatan. Pengelolaan harga dan peningkatan layanan akan memberikan dampak yang lebih besar terhadap **Total Revenue** dibandingkan dengan hanya bergantung pada tingkat hunian.

Catatan:

Meskipun **Occupancy Rate** adalah metrik penting untuk mengukur efisiensi hunian properti, hubungan langsungnya dengan **Total Revenue** sangat lemah karena **faktor-faktor lain** seperti harga per malam, layanan tambahan, lokasi, dan kapasitas properti yang lebih berperan dalam menentukan pendapatan total. Sehingga, kami tidak bisa menindaklanjuti analisis pada EDA Property Performance and Occupancy Rate, dikarenakan juga dengan isi informasi dataset yang mungkin kurang cukup mendukung dan juga efek dari perhitungan informasi days active yang minus.

➤ **Insight Total Room Nights (RNs), Revenue, dan ADR Setiap Kota Tahun 2024**

	A-Z CITY	A-Z MONTH_YEAR	123 Total_RNs	123 Total_Revenue	123 ADR
1	Bandung	2024-01	1,513	7,553.43	4.9923529412
2	Bandung	2024-02	1,440	7,077.76	4.9151111111
3	Bandung	2024-03	1,507	7,257.78	4.8160451228
4	Bandung	2024-04	1,557	7,711.03	4.9524919717
5	Bandung	2024-05	1,533	7,410.05	4.833692107
6	Bandung	2024-06	1,404	6,764.83	4.8182549858
7	Bandung	2024-07	1,484	7,235.43	4.8756266846
8	Bandung	2024-08	1,442	7,348.3	5.0959084605
9	Bandung	2024-09	1,360	6,897.85	5.0719485294
10	Bandung	2024-10	101	477.46	4.7273267327
11	Jakarta	2024-01	1,426	7,049.2	4.9433380084
12	Jakarta	2024-02	1,432	6,826.67	4.7672276536
13	Jakarta	2024-03	1,329	6,642.23	4.9979157261
14	Jakarta	2024-04	1,396	6,723.21	4.8160530086
15	Jakarta	2024-05	1,563	7,458.56	4.7719513756
16	Jakarta	2024-06	1,259	6,302.15	5.0056791104
17	Jakarta	2024-07	1,357	6,806.52	5.0158585114
18	Jakarta	2024-08	1,313	6,584.74	5.0150342727
19	Jakarta	2024-09	1,450	7,116.78	4.9081241379
20	Jakarta	2024-10	120	618.69	5.15575
21	Malang	2024-01	1,275	6,017.29	4.7194431373
22	Malang	2024-02	1,372	6,698.02	4.8819387755
23	Malang	2024-03	1,355	6,982.83	5.1533800738
24	Malang	2024-04	1,398	6,513.73	4.6593204578
25	Malang	2024-05	1,376	6,572.03	4.776184593
26	Malang	2024-06	1,254	5,986.55	4.7739633174
27	Malang	2024-07	1,467	6,871.65	4.6841513292
28	Malang	2024-08	1,346	6,690.63	4.9707503715

Gambar 15. *Insight Total Room Nights (RNs), Revenue, dan ADR*

Insight Total Room Nights (RNs), Revenue, dan ADR Setiap Kota Tahun 2024

memberikan pemahaman yang mendalam tentang performa kota-kota utama berdasarkan tiga metrik utama: **Total Room Nights (RNs)**, **Total Revenue**, dan **Average Daily Rate (ADR)**. Analisis ini mengungkapkan bahwa terdapat variasi signifikan dalam kinerja masing-masing kota, mencerminkan segmentasi pasar yang berbeda.

Bandung mencatat **Total RNs yang stabil** sepanjang tahun, dengan rata-rata mencapai **1.500 malam per bulan**, menegaskan bahwa kota ini memiliki permintaan yang konsisten di pasar. Namun, **ADR Bandung sebesar \$5.0 USD** menunjukkan bahwa kota ini lebih fokus pada pasar menengah ke bawah, menarik pelanggan yang sensitif terhadap harga. Stabilitas ini menjadikan Bandung sebagai pasar ekonomis dengan volume tinggi. **Jakarta**, di sisi lain, memiliki **ADR tertinggi sebesar \$7.8 USD**, menargetkan segmen premium. Hal ini menunjukkan bahwa Jakarta memiliki daya tarik sebagai destinasi bagi pelanggan yang mencari akomodasi berkualitas tinggi, dengan fokus pada pendapatan per malam yang lebih besar. Meskipun Jakarta memiliki Total RNs yang sedikit lebih rendah dibanding

Bandung, kontribusi revenue per malamnya jauh lebih signifikan. **Surabaya**, meskipun memiliki **ADR yang lebih rendah**, berkontribusi pada volume pasar yang tinggi dengan strategi harga yang kompetitif. Kota ini menargetkan pelanggan yang mencari akomodasi dengan nilai terbaik, mengindikasikan potensi untuk meningkatkan revenue dengan optimalisasi harga.

Distribusi ADR yang tidak merata di antara kota-kota tersebut mengindikasikan adanya peluang untuk menerapkan strategi harga yang lebih spesifik. Dengan menyesuaikan harga sesuai dengan karakteristik masing-masing segmen pasar di setiap kota, perusahaan dapat meningkatkan pendapatan secara signifikan sambil tetap mempertahankan volume pemesanan yang stabil.

➤ **Insight 10 Properti Teratas Berdasarkan Pertumbuhan Room Nights (2023 vs 2024)**

	PROPERTY_CODE	123 RNs_2023	123 RNs_2024	123 RNs_Growth	123 Growth_Percentage	CITY	Growth_Category
1	YOG0342	119	217	98	82.35	Yogyakarta	High Growth
2	YOG0297	72	129	57	79.17	Yogyakarta	High Growth
3	BAN0251	113	197	84	74.34	Bandung	High Growth
4	JAK0197	93	159	66	70.97	Jakarta	High Growth
5	BAN0306	100	166	66	66	Bandung	High Growth
6	SUR0414	105	174	69	65.71	Surabaya	High Growth
7	JAK0311	103	164	61	59.22	Jakarta	High Growth
8	JAK0160	95	151	56	58.95	Jakarta	High Growth
9	SUR0174	128	199	71	55.47	Surabaya	High Growth
10	YOG0023	119	185	66	55.46	Yogyakarta	High Growth

Gambar 16. Insight 10 Properti

Insight 10 Properti Teratas Berdasarkan Pertumbuhan Room Nights (2023 vs 2024) memberikan gambaran tentang properti dengan peningkatan performa paling signifikan dalam hal Room Nights (RNs) dari tahun 2023 ke 2024. Analisis ini mengidentifikasi properti yang berhasil memanfaatkan strategi pemasaran, harga, atau peningkatan layanan untuk mencapai pertumbuhan tinggi.

Properti dengan Pertumbuhan Tinggi (High Growth) seperti **YOG0342** di Yogyakarta dan **BAN0251** di Bandung mencatat **pertumbuhan lebih dari 70% dalam Room Nights (RNs)** dibandingkan tahun sebelumnya. Pertumbuhan ini menegaskan keberhasilan strategi pemasaran lokal yang intensif, promosi diskon musiman, dan lokasi properti yang strategis. **Properti BAN0251** menunjukkan bahwa kota seperti Bandung, dengan kombinasi volume tinggi dan lokasi strategis, memiliki potensi pertumbuhan yang kuat. Keberhasilan properti ini dapat dijadikan

contoh untuk memperluas strategi serupa ke properti lain. **Properti seperti BAN0251**, dengan pertumbuhan 74.34%, menunjukkan bahwa lokasi di Bandung tidak hanya menarik volume pemesanan tetapi juga mempertahankan pertumbuhan stabil. Strategi tambahan, seperti promosi grup atau penawaran diskon musiman, dapat membantu meningkatkan properti dalam kategori Moderate dan Low Growth.

➤ Insight Segmentasi Pelanggan Berdasarkan Perilaku Pembelian

	USER_ID	recency	frequency	monetary	recency_s	frequency_s	monetary_s	rfm_so	customer_seg	USER_TYP	USER_G	TRAVEL_PUF	USER_AGE
1	1005054	70	103	2,009.35	4	4	5	13	Platinum	Repeat User	Female	Leisure	47
2	10103753	84	107	2,234.25	4	4	5	13	Platinum	New User	Male	Business	20
3	10122208	83	104	2,073.19	4	4	5	13	Platinum	New User	Male	Business	33
4	10133837	69	104	2,072.74	4	4	5	13	Platinum	Repeat User	Female	Business	22
5	10173311	70	105	2,070.36	4	4	5	13	Platinum	New User	Female	Business	51
6	10178220	83	112	2,030.93	4	4	5	13	Platinum	New User	Male	Business	24
7	10179740	83	123	2,376.69	4	4	5	13	Platinum	New User	Female	Business	30
8	10225748	72	113	2,182.98	4	4	5	13	Platinum	New User	Female	Business	60
9	10273904	72	102	2,055.5	4	4	5	13	Platinum	Repeat User	Female	Leisure	48
10	10284483	78	107	2,053.22	4	4	5	13	Platinum	New User	Male	Leisure	49
11	10404052	71	100	2,047.27	4	4	5	13	Platinum	New User	Male	Leisure	28
12	10411891	69	125	2,515.5	4	4	5	13	Platinum	New User	Female	Leisure	54
13	10466003	76	101	2,093.3	4	4	5	13	Platinum	Repeat User	Female	Business	57
14	10497091	74	102	2,010.47	4	4	5	13	Platinum	New User	Male	Leisure	24
15	10509541	91	109	2,289.59	4	4	5	13	Platinum	New User	Male	Business	26
16	10514509	97	115	2,366.8	4	4	5	13	Platinum	Repeat User	Male	Leisure	47
17	10522158	84	103	2,012.65	4	4	5	13	Platinum	New User	Male	Leisure	26
18	10544569	71	104	2,094.52	4	4	5	13	Platinum	Repeat User	Male	Business	29
19	10549245	84	108	2,135.07	4	4	5	13	Platinum	Repeat User	Male	Business	29
20	10556405	95	109	2,271.16	4	4	5	13	Platinum	Repeat User	Male	Leisure	34

Gambar 17. Insight Segmentasi Pelanggan

Insight Segmentasi Pelanggan Berdasarkan Perilaku Pembelian bertujuan untuk memahami perilaku pelanggan melalui RFM Analysis (Recency, Frequency, Monetary). Analisis ini membagi pelanggan menjadi beberapa segmen berdasarkan seberapa baru mereka melakukan pembelian, seberapa sering mereka bertransaksi, dan seberapa besar nilai transaksi mereka. Pendekatan ini memberikan wawasan strategis untuk meningkatkan loyalitas pelanggan dan pendapatan perusahaan.

Platinum Customers, Segmen ini mencakup pelanggan dengan skor RFM tertinggi (≥ 13), yang menyumbang lebih dari 50% total revenue meskipun hanya terdiri dari 25% total pelanggan. Pelanggan ini sangat loyal, melakukan pembelian dengan frekuensi tinggi, dan memiliki pengeluaran yang besar. Strategi eksklusif seperti program loyalitas VIP, penawaran personalisasi, atau layanan prioritas direkomendasikan untuk mempertahankan mereka. **Gold Customers** Pelanggan ini memiliki skor RFM antara 9–12, memberikan kontribusi yang signifikan terhadap pendapatan, meskipun tidak seaktif segmen Platinum. Promosi insentif berbasis volume pembelian atau diskon khusus dapat membantu meningkatkan

keterlibatan mereka. **Bronze Customers:** Segmen ini memiliki skor RFM rendah (<6), dengan kontribusi revenue yang kecil. Pelanggan ini umumnya jarang bertransaksi dan memiliki pengeluaran rendah. Namun, mereka memiliki potensi besar jika diaktifkan melalui kampanye diskon pembelian pertama, referral program, atau bundling penawaran yang menarik.

Distribusi Revenue menunjukkan pola yang tidak merata, di mana sebagian besar pendapatan berasal dari pelanggan Platinum dan Gold. Hal ini menegaskan pentingnya fokus pada pelanggan bernilai tinggi, sementara tetap memanfaatkan peluang untuk meningkatkan keterlibatan pelanggan di segmen Bronze.

➤ **Insight Perbandingan Harga Musim Normal dan Musim Puncak (2023 & 2024)**

	Az year	Az season	i23 total_bookings	i23 total_revenue	i23 total_room_nights	i23 avg_revenue_per_night	i23 avg_revenue_per_booking
1	2023	Normal Season	13,906	278,288.83	55,898	4.98	20.01
2	2023	Peak Season	7,139	142,177.64	28,350	5.02	19.92
3	2024	Normal Season	10,457	209,417.29	42,269	4.95	20.03
4	2024	Peak Season	5,120	101,832.18	20,922	4.87	19.89

Gambar 18. *Insight Perbandingan Harga Musim Normal dan Musim Puncak*

Rata-rata harga musim puncak mencapai \$5.02 USD per malam, sedangkan musim normal sedikit lebih rendah di \$4.98 USD. Penurunan ADR pada 2024 menunjukkan kemungkinan dampak dari penyesuaian harga yang tidak tepat atau penurunan daya beli pelanggan. Namun, jumlah total pemesanan di musim normal tetap lebih tinggi dibandingkan musim puncak, menandakan bahwa volume tetap menjadi penentu utama pendapatan di luar musim liburan. Musim puncak memiliki potensi optimalisasi melalui strategi premium pricing, sementara musim normal membutuhkan promosi volume untuk mempertahankan stabilitas kinerja. Diskon early bird dan bundling dapat meningkatkan volume pada kedua musim.

➤ **Insight Kualitas Properti Berdasarkan Occupancy Rate**

	A-z PROPERTY_CODE	A-z Grade	123 OCCUPANCY_RATE
1	JAK0052	A	3,804.76
2	SUR0036	A	1,216.67
3	MAL0012	A	992.86
4	JAK0059	A	819.54
5	SUR0191	A	565.29
6	BAN0423	A	333.71
7	MAL0250	A	236.34
8	JAK0197	A	211.31
9	SUR0402	A	206.61
10	BAN0213	A	193
11	MAL0218	A	143.07
12	BAN0370	A	135.47
13	YOG0060	A	135.09
14	JAK0258	A	124.24
15	YOG0490	A	122.89
16	JAK0458	A	115.25
17	MAL0478	A	114.15
18	BAN0146	A	107.76
19	MAL0084	A	101.36
20	JAK0184	A	98.89

Gambar 19. *Insight Kualitas Properti*

Insight Kualitas Properti Berdasarkan Occupancy Rate berfokus pada analisis tingkat hunian (occupancy rate) untuk mengevaluasi kinerja dan kualitas properti. Analisis ini mengelompokkan properti berdasarkan persentase hunian menjadi beberapa kategori, mulai dari sangat baik hingga sangat buruk. Pendekatan ini membantu mengidentifikasi properti yang berkinerja tinggi dan properti yang memerlukan perbaikan.

Grade A (>80%): Properti seperti **JAK0052** dan **SUR0036** menunjukkan tingkat hunian sangat tinggi, mendekati saturasi kapasitas. Properti dalam kategori ini biasanya terletak di lokasi strategis, memiliki layanan unggul, dan harga yang kompetitif. Strategi untuk properti ini difokuskan pada optimalisasi harga untuk meningkatkan pendapatan tanpa mengurangi tingkat hunian.

Grade C (40%-70%): Properti dalam kategori ini menunjukkan performa rata-rata. Tingkat hunian yang moderat mengindikasikan adanya peluang untuk meningkatkan kinerja, seperti melalui promosi diskon musiman, peningkatan layanan, atau revisi harga agar lebih kompetitif.

Grade D (20-40%): Properti dengan tingkat hunian rendah, terutama di lokasi sekunder, membutuhkan perhatian lebih. Masalah utama mungkin terkait dengan lokasi yang kurang strategis, harga yang tidak kompetitif, atau kualitas layanan yang perlu ditingkatkan. Strategi khusus seperti kampanye pemasaran lokal atau bundling penawaran dengan mitra strategis dapat membantu meningkatkan daya tarik properti ini.

Insight Utama:

- Properti Grade A menunjukkan performa yang sangat baik dan dapat dimaksimalkan melalui strategi premium pricing.
- Properti Grade C dan D memerlukan evaluasi menyeluruh untuk meningkatkan daya saing, termasuk analisis lokasi, layanan, dan strategi pemasaran.

➤ Laporan Metrik Utama (SQL View)

	Az CITY	Az booking_month	123 total_RNs	123 total_revenue	123 avg_ADR
1	Bandung	2024-06	972	4,958.92	7.5345544218
2	Bandung	2024-07	1,491	7,296.94	7.1805937118
3	Bandung	2024-08	1,452	7,385.79	7.4664087848
4	Bandung	2024-09	1,363	6,819.71	7.4133611931
5	Jakarta	2024-06	919	4,693.01	7.8949507736
6	Jakarta	2024-07	1,391	6,972.43	7.5280195618
7	Jakarta	2024-08	1,273	6,380.81	7.2654606838
8	Jakarta	2024-09	1,491	7,313.62	7.3112739726
9	Malang	2024-06	1,050	4,960.94	7.0586411111
10	Malang	2024-07	1,466	6,809.99	7.1412399936
11	Malang	2024-08	1,345	6,753.5	7.5488273453
12	Malang	2024-09	1,220	6,394.9	7.8497923323
13	Surabaya	2024-06	854	4,432.44	8.3750597907
14	Surabaya	2024-07	1,335	6,358.56	7.2945859053
15	Surabaya	2024-08	1,344	6,752.19	7.4122869551
16	Surabaya	2024-09	1,139	5,535.77	7.2612594584
17	Yogyakarta	2024-06	1,120	5,525.93	7.276918086
18	Yogyakarta	2024-07	1,498	7,112.17	7.0936628615
19	Yogyakarta	2024-08	1,583	7,992.58	7.551526046
20	Yogyakarta	2024-09	1,513	7,354.77	6.9927137325

Gambar 20. Insight Total Room Nights (RNs), Revenue, dan ADR

SQL View adalah alat yang **efektif** untuk menyederhanakan proses analisis data, khususnya dalam skala besar. Dalam konteks ini, SQL View `city_metrics_last_6_months` dirancang untuk menghitung metrik utama seperti **Total Room Nights (RNs)**, **Total Revenue**, dan **Average Daily Rate (ADR)** berdasarkan performa properti selama enam bulan terakhir. SQL View ini menyimpan data yang telah disegmentasi berdasarkan **kota** dan **bulan**,

memberikan wawasan tentang **fluktuasi musiman**, perbandingan performa antar kota, serta potensi untuk mengidentifikasi peluang pertumbuhan. Dengan **Total RNs yang tinggi di Bandung** dan **ADR premium di Jakarta**, SQL View ini memaparkan perbedaan segmentasi pasar yang mencerminkan kebutuhan pelanggan di setiap lokasi. **Bandung menunjukkan stabilitas volume pemesanan**, sementara **Jakarta mendominasi pasar premium** dengan pendapatan per malam yang lebih tinggi.

Melalui SQL View, perusahaan dapat lebih **efisien** dalam menganalisis tren performa tanpa perlu menghitung ulang data setiap kali diperlukan. Informasi yang diperoleh membantu mengidentifikasi **periode musiman terbaik, kota dengan kinerja tinggi**, dan **properti yang memerlukan perhatian lebih**. Selain itu, SQL View mendukung pengambilan keputusan strategis seperti **penyesuaian harga, promosi musiman**, dan **alokasi sumber daya** untuk meningkatkan pendapatan dan efisiensi operasional. Dengan gambaran yang **konsisten** dan **akurat**, SQL View ini menjadi alat penting untuk mendukung manajer dalam merancang strategi yang lebih **responsif** terhadap dinamika pasar.

➤ **Revenue Trend by City and Brand Type**

Gambar berikut menunjukkan tren pendapatan dan bulanan dari tahun 2020 hingga 2024 berdasarkan kota (Bandung, Jakarta, Malang, Surabaya, dan Yogyakarta) dan Brand Type (RedDoorz, Koolkost, RedPartner). Grafik ini memberikan wawasan tentang kontribusi pendapatan dari masing-masing kota dan merek, serta perubahan tren selama lima tahun.



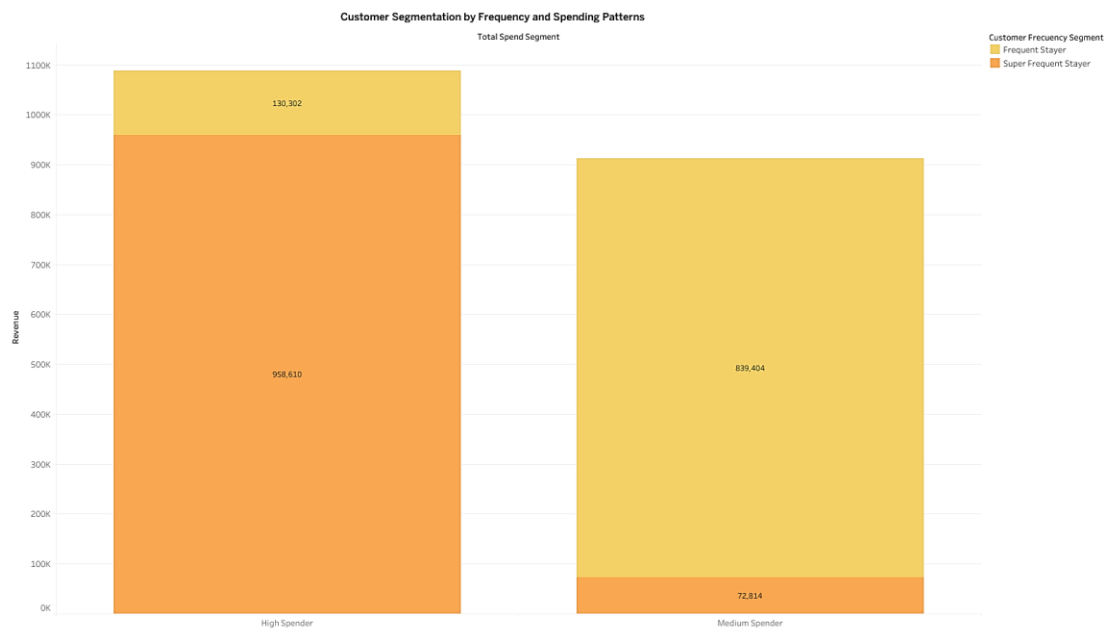
Gambar 21. Line Chart Revenue Trend by City and Brand Type

Dari visualisasi ini, kita dapat memahami bagaimana performa kota dan merek memengaruhi pendapatan secara keseluruhan yang dapat menjadi dasar untuk strategi pengembangan dan pemasaran. Tren pendapatan menunjukkan bahwa kota Yogyakarta dan Surabaya memiliki performa paling stabil dengan puncak pendapatan di Surabaya sebesar \$34,903 pada tahun 2023. Sebaliknya, kota seperti Jakarta dan Malang menunjukkan tren penurunan signifikan, dengan pendapatan Jakarta turun ke \$18,726 pada tahun 2024. Dari sisi merek, RedDoorz menjadi kontributor utama pendapatan di hampir semua kota, sementara Koolkost dan RedPartner memberikan kontribusi yang lebih kecil. Penurunan pendapatan di kota-kota tertentu menunjukkan perlunya strategi peningkatan, seperti promosi atau pengembangan properti baru, sementara performa stabil di Surabaya dan Yogyakarta menunjukkan potensi pengembangan lebih lanjut di wilayah ini.

➤ Customer Segmentation by Frequency and Spending Patterns

Gambar berikut menunjukkan kontribusi pendapatan berdasarkan frekuensi kunjungan pelanggan (Moderate Stayer, Frequent Stayer, Super Frequent Stayer) dan kategori pengeluaran (High Spender, Medium Spender, Low Spender).

Visualisasi ini dirancang untuk memahami perilaku pelanggan berdasarkan pola kunjungan dan pengeluaran mereka.

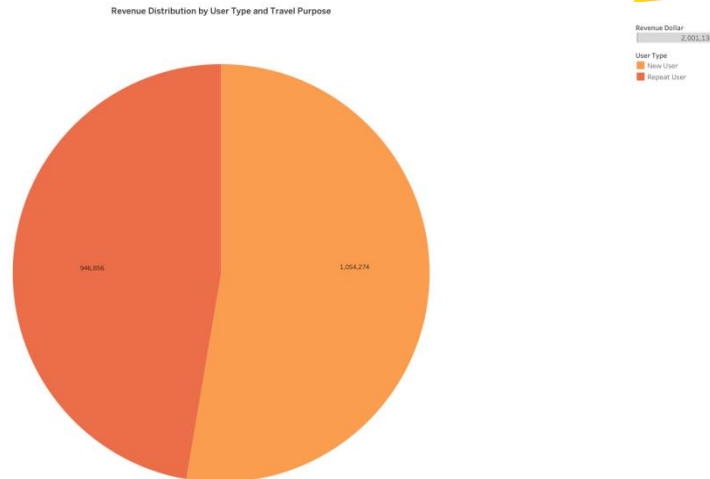


Gambar 22. Bar Chart Customer Segmentation by Frequency and Spending Patterns

Dari visualisasi ini, kita dapat mengetahui segmen pelanggan yang memberikan kontribusi pendapatan terbesar, sehingga strategi loyalitas dapat difokuskan pada segmen tersebut. Pelanggan dalam kategori Super Frequent Stayer memberikan kontribusi pendapatan tertinggi sebesar \$958,610, yang sebagian besar berasal dari kategori High Spender. Sementara itu, Frequent Stayer memberikan kontribusi sebesar \$839,404, dengan distribusi yang lebih merata di antara kategori pengeluaran. Moderate Stayer memberikan kontribusi paling kecil, menunjukkan potensi peningkatan jika mereka dapat didorong menjadi Frequent atau Super Frequent Stayer. Strategi loyalitas dan promosi harus difokuskan pada mempertahankan pelanggan dalam kategori High Spender dan meningkatkan engagement dengan Frequent Stayer untuk meningkatkan kontribusi mereka.

➤ Revenue Distribution by User Type and Travel Purpose

Gambar berikut menunjukkan distribusi pendapatan berdasarkan User Type (New User dan Repeat User) serta tujuan perjalanan (Business dan Leisure). Visualisasi ini dirancang untuk memahami kontribusi pendapatan dari berbagai tipe pengguna dan tujuan perjalanan.

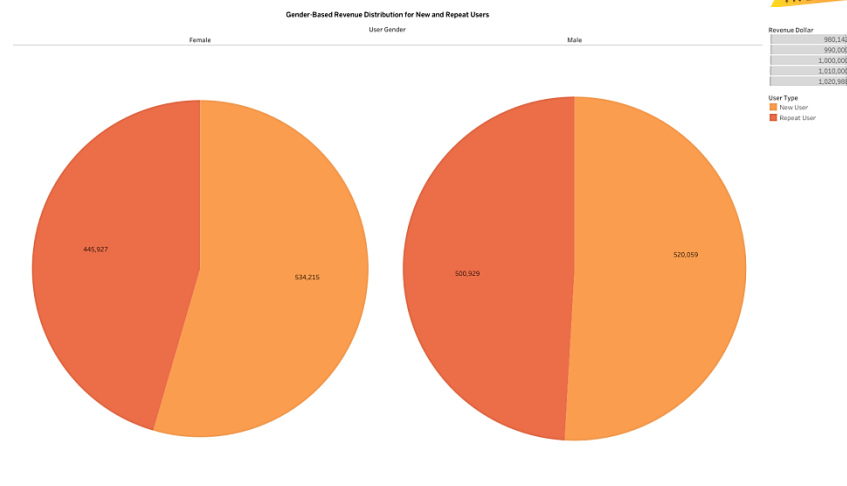


Gambar 23. *Pie Chart Revenue Distribution by User Type and Travel Purpose*

Dari visualisasi ini, kita dapat mengetahui pola pendapatan dari tujuan perjalanan yang berbeda, yang membantu merancang strategi pemasaran yang sesuai. Pengguna ulang (Repeat User) memberikan kontribusi pendapatan lebih besar sebesar \$1,054,274 dibandingkan pengguna baru yang menghasilkan \$946,856. Tujuan perjalanan bisnis dan leisure memiliki kontribusi yang hampir seimbang, menunjukkan bahwa kedua segmen ini penting untuk dijaga. Kampanye pemasaran dapat disesuaikan untuk menarik lebih banyak pengguna baru dan memperkuat hubungan dengan pelanggan ulang, misalnya dengan program loyalty yang menawarkan insentif untuk perjalanan bisnis atau leisure.

➤ **Gender-Based Revenue Distribution for New and Repeat Users**

Gambar berikut menunjukkan distribusi pendapatan berdasarkan jenis kelamin pelanggan (pria dan wanita) serta User Type (New User dan Repeat User). Visualisasi ini memberikan gambaran tentang kontribusi pendapatan dari segmen gender yang berbeda.

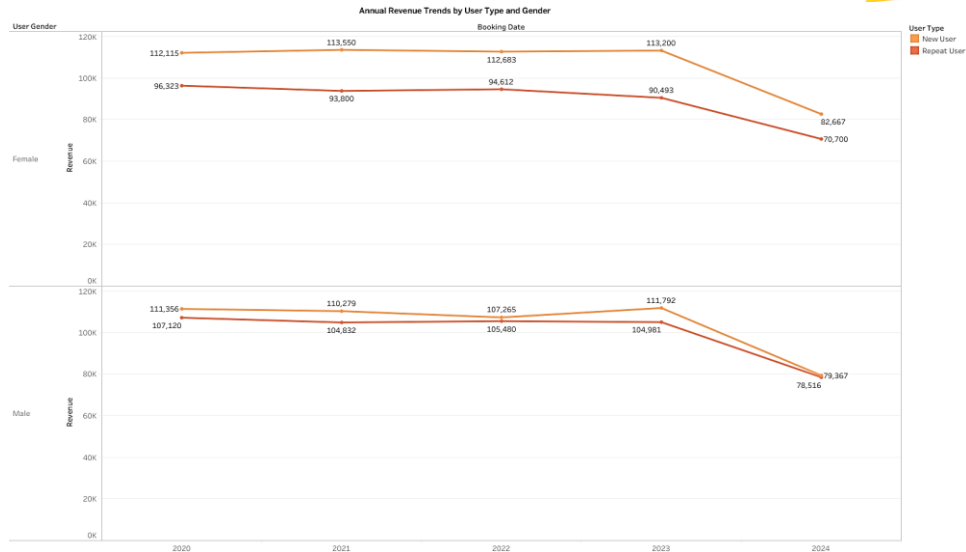


Gambar 24. *Pie Chart Gender-Based Revenue Distribution for New and Repeat Users*

Dari visualisasi ini, kita dapat melihat perbedaan kontribusi pendapatan antara pria dan wanita, serta strategi yang dapat diterapkan untuk masing-masing segmen. Pelanggan pria memberikan kontribusi pendapatan yang lebih besar sebesar \$1,020,988, dibandingkan dengan wanita yang menyumbang \$980,142. Namun, pengguna baru wanita memberikan kontribusi yang lebih besar dibandingkan pria baru, menunjukkan potensi pasar yang dapat ditingkatkan melalui kampanye pemasaran yang menarik bagi segmen ini. Strategi loyalitas harus difokuskan pada mempertahankan pelanggan pria yang sudah ada, sementara pengguna wanita baru dapat dijadikan target utama untuk kampanye digital.

➤ **Annual Revenue Trends by User Type and Gender**

Gambar berikut menunjukkan visualisasi tren tahunan pendapatan dari tahun 2020 hingga 2024, dipisahkan berdasarkan User Gender (pria dan wanita) dan User Type (pengguna baru dan ulang). Grafik ini dirancang untuk memberikan wawasan tentang kontribusi pendapatan masing-masing segmen pelanggan serta perubahan tren selama lima tahun terakhir.



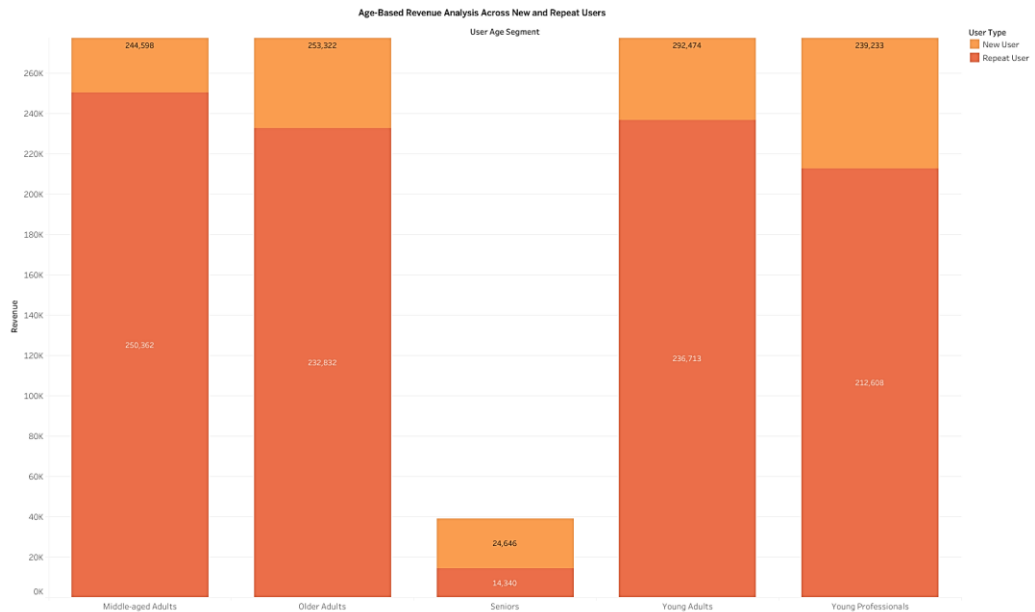
Gambar 25.Line Chart Annual Revenue Trends by User Type and Gender

Dari visualisasi ini, kita dapat memahami pola perilaku pelanggan dan perbedaan kontribusi pendapatan dari masing-masing segmen, yang memberikan dasar untuk strategi peningkatan pendapatan di masa mendatang. Visualisasi ini menunjukkan bahwa pengguna ulang (Repeat User) memberikan kontribusi pendapatan yang lebih besar dibandingkan pengguna baru (New User) pada kedua jenis kelamin. Pengguna wanita ulang memiliki pendapatan tertinggi di tahun 2020 sebesar \$113,550, tetapi mengalami penurunan menjadi \$82,667 di tahun 2024. Pengguna baru wanita juga mengalami tren serupa, turun dari \$96,323 di tahun 2020 menjadi \$70,700 di tahun 2024.

Sementara itu, pengguna pria ulang menunjukkan tren penurunan yang lebih stabil, dari \$111,356 di tahun 2020 menjadi \$79,367 di tahun 2024. Pengguna baru pria memiliki kontribusi yang lebih stabil dibandingkan kelompok lainnya, dengan penurunan yang kecil dari \$107,120 di tahun 2020 menjadi \$78,516 di tahun 2024. Penurunan yang signifikan pada pengguna ulang, terutama wanita, menunjukkan perlunya strategi retensi pelanggan, seperti program loyalitas yang menarik atau penawaran eksklusif. Di sisi lain, pengguna baru wanita dapat menjadi target kampanye pemasaran yang lebih personal dan relevan untuk meningkatkan kontribusi pendapatan mereka. Strategi ini akan membantu perusahaan mempertahankan dan meningkatkan pendapatan dari pelanggan yang berbeda segmennya.

➤ Age-Based Revenue Analysis Across New and Repeat Users

Gambar berikut menunjukkan kontribusi pendapatan berdasarkan kelompok usia (Young Adults, Young Professionals, Middle-aged Adults, Older Adults, Seniors) dan User Type (New User dan Repeat User). Visualisasi ini bertujuan untuk memahami kontribusi pendapatan dari pelanggan berdasarkan usia mereka.



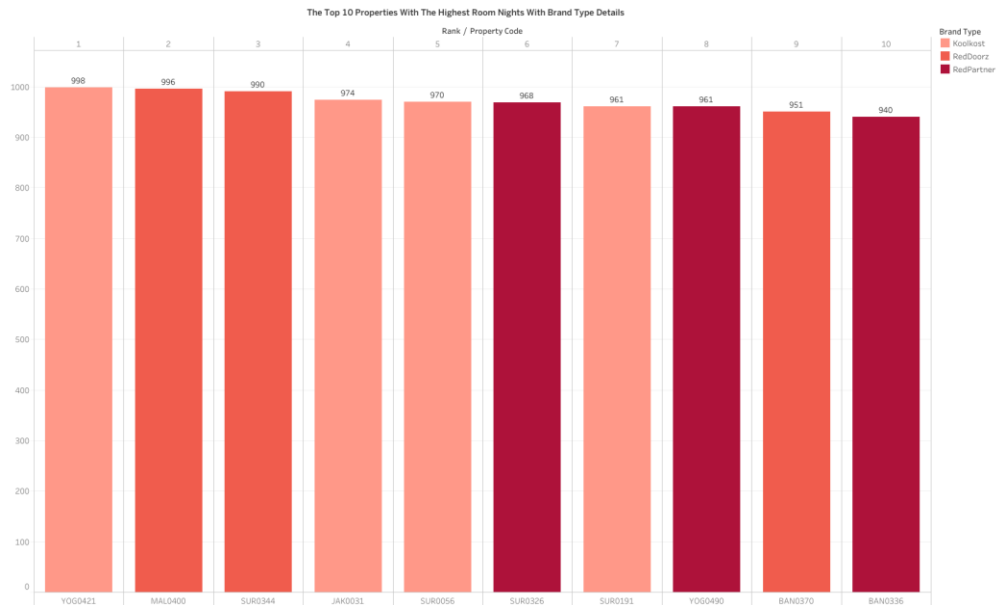
Gambar 26. Bar Chart Age-Based Revenue Analysis Across New and Repeat Users

Dari visualisasi ini, kita dapat mengetahui segmen usia yang paling memberikan kontribusi pendapatan, sehingga strategi pemasaran dapat difokuskan pada segmen tersebut. Segmen usia Young Professionals memberikan kontribusi pendapatan tertinggi sebesar \$292,474, diikuti oleh Young Adults dengan \$236,713. Segmen usia lainnya, seperti Middle-aged Adults dan Older Adults, memberikan kontribusi lebih kecil, sementara Seniors hanya menyumbang \$24,646. Strategi pemasaran harus difokuskan pada segmen usia muda, terutama Young Professionals, dengan kampanye yang relevan dan menarik untuk menjaga dan meningkatkan kontribusi pendapatan mereka.

➤ The Top 10 Properties With The Highest Room Nights With Brand Type Details

Gambar berikut menunjukkan 10 properti dengan jumlah Room Nights (RNs) tertinggi, serta kontribusi dari masing-masing merek (RedDoorz, Koolkost,

RedPartner). Visualisasi ini memberikan wawasan tentang properti dengan performa terbaik yang dapat dijadikan model pengembangan.

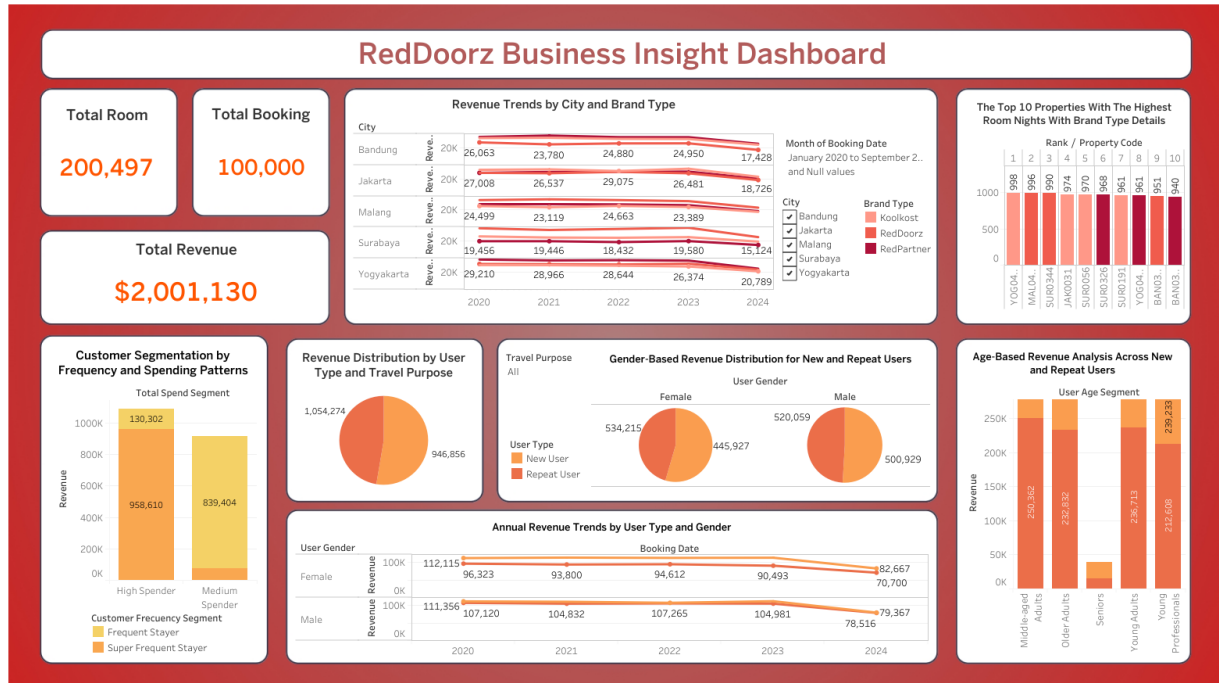


Gambar 7. Bar Chart The Top 10 Properties With The Highest Room Nights With Brand Type Details

Dari visualisasi ini, kita dapat memahami properti mana yang memberikan kontribusi terbesar dan bagaimana merek memengaruhi performa properti. Properti dengan jumlah Room Nights tertinggi adalah YOG0421 di Yogyakarta dengan 998 RNs, diikuti oleh MAL0400 di Malang dengan 996 RNs. Dari 10 properti teratas, merek RedDoorz mendominasi dengan kontribusi pada 6 properti, sementara Koolkost memiliki 3 properti dan RedPartner hanya 1. Total RNs dari 10 properti ini adalah 9,329, menunjukkan kontribusi signifikan terhadap pendapatan perusahaan. Properti ini dapat menjadi model untuk pengembangan lebih lanjut, baik melalui peningkatan fasilitas maupun akuisisi properti baru di lokasi dengan karakteristik serupa.

➤ **RedDoorz Business Insight Dashboard**

Gambar berikut menunjukkan hasil analisis yang memberikan wawasan tentang tren pendapatan, segmentasi pelanggan, dan performa properti berdasarkan data RedDoorz. Informasi ini digunakan untuk mengidentifikasi area yang memerlukan perbaikan dalam penjualan, harga, dan akuisisi properti, serta menyusun rekomendasi untuk kampanye pemasaran yang lebih efektif.



Gambar 28. RedDoorz Business Insight Dashboard

Dengan memanfaatkan dashboard ini, kita dapat menggalang lebih dalam mengenai peluang peningkatan pendapatan melalui pengelolaan properti dan kampanye pemasaran yang lebih terarah.

- **Areas of Improvement in Sales, Pricing, and Property Acquisition**

Berdasarkan visualisasi Revenue Trends by City and Brand Type, terlihat bahwa kota seperti Jakarta dan Malang mengalami penurunan pendapatan signifikan pada tahun-tahun terakhir. Penurunan ini menunjukkan perlunya strategi peningkatan penjualan, seperti menawarkan promosi musiman, diskon untuk masa inap jangka panjang, atau paket bundling yang dapat menarik pelanggan baru maupun ulang. Sebaliknya, kota seperti Yogyakarta dan Surabaya, yang memiliki performa stabil dengan total pendapatan tertinggi (contohnya Surabaya mencapai \$34,903 pada 2023), dapat dijadikan referensi untuk mengembangkan properti baru di lokasi strategis lainnya. Selain itu, peningkatan fasilitas atau layanan di properti yang berperforma baik dapat membantu mempertahankan keunggulan di wilayah ini.

- **Recommendations for Targeted Marketing Campaigns**

Dari hasil visualisasi Customer Segmentation and Behavior Pattern, terlihat bahwa pelanggan dalam kategori Super Frequent Stayer dan High Spender memberikan kontribusi pendapatan terbesar, dengan masing-masing menyumbang \$958,610 dan sebagian besar berasal dari pelanggan pria. Oleh karena itu, kampanye pemasaran harus menargetkan kelompok ini dengan program loyalitas yang menarik, seperti diskon eksklusif untuk masa inap ulang atau reward berbasis poin.

Segmen usia muda, seperti Young Professionals (30-39 tahun) dan Young Adults (≤ 29 tahun), yang menyumbang pendapatan tertinggi sebesar \$292,474 dan \$236,713, dapat dijangkau melalui kampanye digital yang dirancang untuk menarik perhatian mereka. Pengguna baru wanita juga menunjukkan potensi besar, dengan kontribusi pendapatan signifikan dibandingkan pengguna baru pria. Strategi pemasaran digital, seperti penawaran khusus di platform media sosial atau aplikasi, dapat meningkatkan keterlibatan pelanggan baru wanita.

Program loyalitas yang dipersonalisasi, seperti reward untuk perjalanan ulang atau diskon khusus untuk pria, juga dapat membantu mempertahankan pelanggan lama, yang menjadi kontributor utama pendapatan perusahaan. Dashboard ini tidak hanya membantu mengidentifikasi area yang membutuhkan perbaikan, tetapi juga memberikan wawasan strategis untuk meningkatkan pendapatan melalui pengelolaan properti, pemasaran yang terfokus, dan pengalaman pelanggan yang lebih baik.

BAB IV

KESIMPULAN DAN SARAN

4.1 Simpulan

Hasil analisis data RedDoorz memberikan wawasan mendalam mengenai kinerja properti, perilaku pelanggan, dan tren industri. Analisis ini mengidentifikasi bahwa properti dengan tingkat hunian tinggi tidak selalu menghasilkan pendapatan maksimal, karena faktor seperti harga, lokasi, dan layanan premium memainkan peran signifikan. Selain itu, pola pemesanan menunjukkan preferensi pelanggan terhadap durasi menginap yang singkat dan kebutuhan fleksibilitas dalam pemesanan mendadak. Segmentasi pelanggan berbasis RFM juga membantu memahami perilaku pelanggan untuk meningkatkan loyalitas. Secara keseluruhan, pemanfaatan teknologi analitik data memungkinkan optimasi operasional dan pengambilan keputusan berbasis data yang lebih efektif.

4.2 Saran

1. Optimalisasi Harga dan Layanan

Terapkan strategi harga dinamis untuk memaksimalkan pendapatan properti, terutama pada properti dengan tingkat hunian rendah. Tingkatkan layanan di properti dengan potensi tinggi untuk menarik pelanggan premium.

2. Strategi Promosi

Fokus pada promosi untuk durasi menginap lebih panjang guna meningkatkan pendapatan. Kembangkan penawaran khusus seperti diskon untuk pemesanan mendadak atau musim tertentu.

3. Segmentasi Pelanggan

Gunakan hasil segmentasi RFM untuk mengembangkan program loyalitas pelanggan yang lebih terarah. Berikan penawaran eksklusif kepada segmen pelanggan Platinum dan Gold.

4. Pengelolaan Properti

Evaluasi properti dengan performa rendah untuk menentukan area yang membutuhkan perbaikan, seperti fasilitas atau strategi pemasaran. Replikasi praktik terbaik dari properti dengan performa tinggi.

5. Pemanfaatan Teknologi

Tingkatkan integrasi teknologi seperti sistem manajemen properti (PMS) untuk efisiensi operasional. Gunakan analitik prediktif untuk mengidentifikasi tren pasar dan mengoptimalkan strategi pemasaran.

Pendekatan yang lebih strategis ini diharapkan dapat membantu RedDoorz meningkatkan efisiensi operasional, kepuasan pelanggan, dan performa bisnis secara keseluruhan.

LAMPIRAN

A. Online Diagram

https://drive.google.com/file/d/1P1faulk_om9aVa0rOGwkBSvPyepQn42m/view?usp=sharing

B. Python Code

https://colab.research.google.com/drive/1vXrMpUg-zkSCUJcma4XIgDECF7BBHit_?usp=sharing

C. Tableau Public

https://public.tableau.com/views/KELOMPOK3-CS18/RedDoorzDashboard?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

D. Link Presentasi Recording

<https://youtu.be/7NOGwkag8kw>

E. PPT

<https://docs.google.com/presentation/d/1hPCW3PcOgOkk55Mb5JroV69RlrYN-Y8B/edit?usp=sharing&ouid=109353194798525920179&rtpof=true&sd=true>

F. Notulen

https://docs.google.com/document/d/1a9zGw-uJG3v1dUh1w5GpUK_PqnZ1AunC/edit?usp=sharing&ouid=109353194798525920179&rtpof=true&sd=true