



HOTEL BOOKING DATA ANALYTICS

L U ' L U ' A H Z A K I Y A H





INTRODUCTION

Saya, Lu'Lu'Ah Zakiyah,
Perkenalkan saya seorang Data Analyst Enthusiast yang saat ini sedang menjalani pendidikan di DIBIMBING. Saya memiliki minat yang besar dalam mengolah data, mendapatkan insight yang diperlukan dan mengambil keputusan berdasarkan data. Dengan latar belakang ini, saya akan terus mengembangkan keterampilan saya. Saya percaya bahwa data adalah sebuah kunci untuk mendorong inovasi dan efisiensi dalam berbagai industri.

[HTTPS://WWW.LINKEDIN.COM/IN/LU-LU-AH-ZAKIYAH/](https://www.linkedin.com/in/lu-lu-ah-zakiyah/)



O 1

PROJECT OVERVIEW





ABOUT THE PROJECT



Proyek ini berfokus pada analisis data pemesanan hotel untuk mengungkap pola dan wawasan yang berarti yang dapat mendukung pengambilan keputusan bisnis. Tujuan utamanya adalah melakukan pembersihan data, analisis data eksplorasi, dan evaluasi statistik untuk memahami perilaku pelanggan, tren pemesanan, dan peluang pendapatan potensial. Melalui eksplorasi metrik seperti Rata-rata Tarif Harian (ADR), waktu tunggu, tingkat pembatalan, jenis pelanggan, dan segmen pasar, proyek ini bertujuan untuk memberikan wawasan yang dapat ditindaklanjuti untuk tim manajemen hotel.

Hasil dari proyek ini diharapkan dapat membantu hotel:

- Meningkatkan pendapatan melalui strategi penetapan harga dinamis.
- Mengoptimalkan inventaris kamar berdasarkan komposisi tamu dan perilaku pemesanan.
- Mengurangi kerugian finansial dengan memahami pola pembatalan dan menyesuaikan kebijakan.

Proyek ini penting karena industri perhotelan sangat bergantung pada interpretasi data yang akurat untuk tetap kompetitif. Dengan memanfaatkan analisis data, hotel dapat meningkatkan pengalaman pelanggan, meningkatkan profitabilitas, dan menyederhanakan strategi operasional. Penerima manfaat langsung dari proyek ini termasuk manajer hotel, analis pendapatan, dan tim pemasaran.



O 2

BUSINESS PROBLEM





BUSINESS PROBLEM



> Masalah Bisnis

- **Ketidakpastian dalam Penetapan Harga Kamar:** Kesulitan dalam menetapkan harga kamar bulanan yang optimal (ADR) karena kurangnya perkiraan berbasis data.
- **Promosi yang Tidak Efektif:** Promosi dan diskon seringkali tidak ditargetkan karena pola musiman dan perilaku pelanggan tidak dipahami dengan baik.
- **Peluang Pendapatan yang Terlewatkan:** Tanpa prediksi ADR yang akurat, hotel berisiko kehilangan pendapatan selama permintaan tinggi dan menghadapi tingkat hunian rendah selama periode sepi.

> Tujuan Bisnis

- **Meningkatkan Pendapatan Hotel:** Optimalisasi ADR, berdasarkan prediksi, dapat memaksimalkan pendapatan.
- **Penetapan Harga dan Promosi Proaktif:** Tren musiman dan permintaan membenarkan perlunya strategi proaktif.
- **Meningkatkan Perencanaan:** Distribusi prediktif dan tren historis memberikan dasar untuk perencanaan operasional dan keuangan yang lebih baik.



O 3

DATA UNDERSTANDING

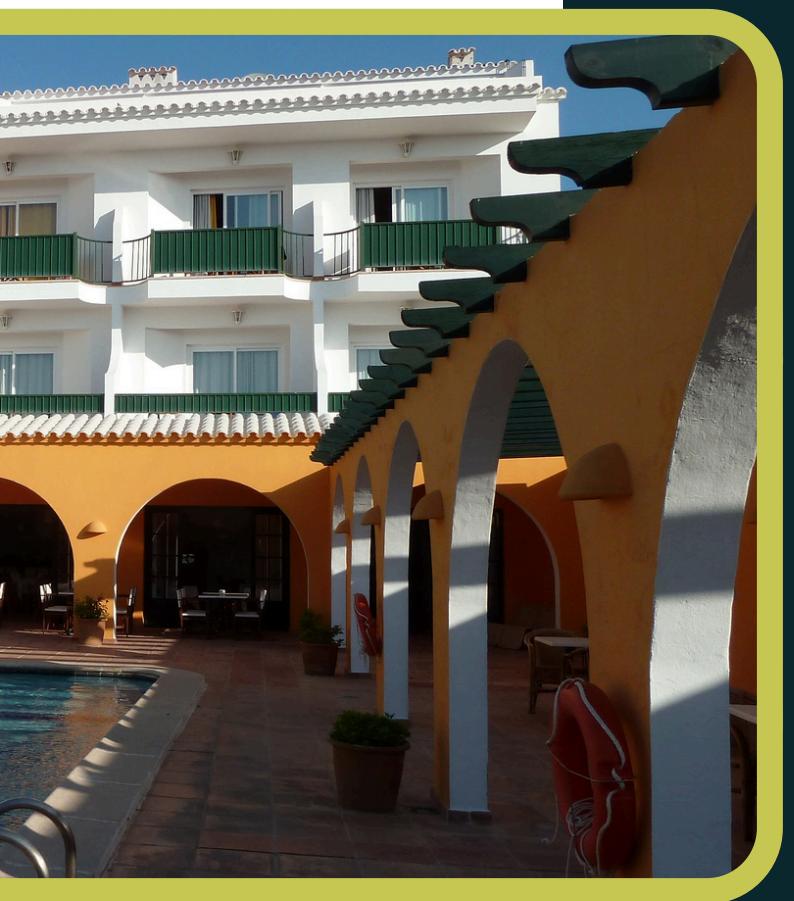


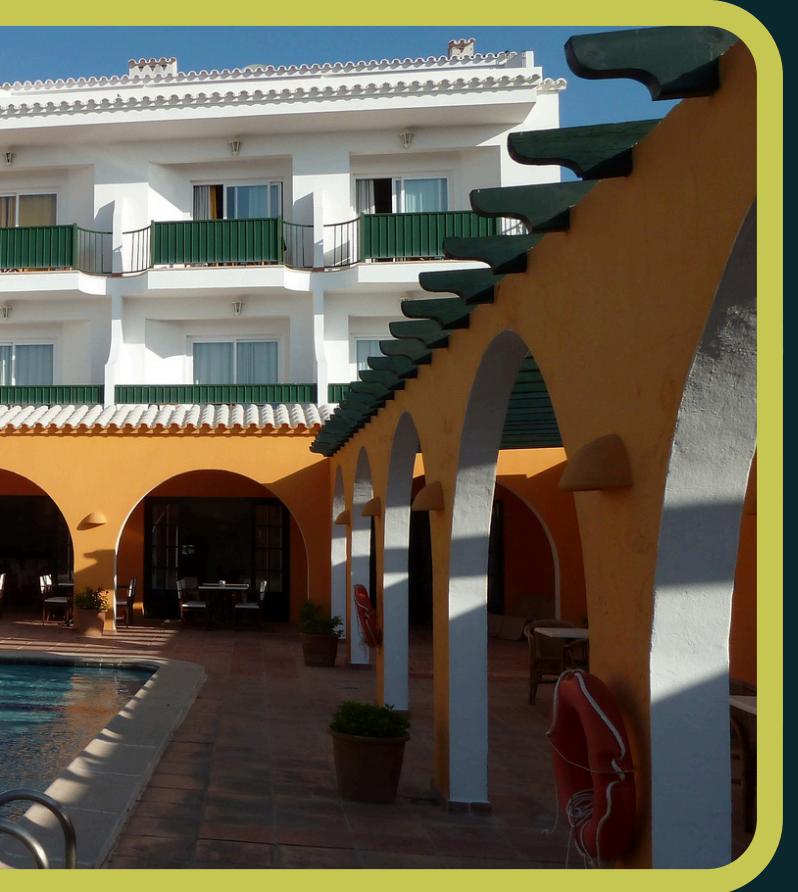
DATASET OVERVIEW

Dataset berisi **83.293** entri (mewakili pemesanan individual) dan mencakup **33** kolom dengan berbagai jenis data. Berikut adalah rincian singkat struktur dataset:

Kolom Kunci:

- **Informasi Hotel:**
 - hotel: Jenis hotel (misalnya, Resort, City Hotel).
- **Informasi Pelanggan:**
 - country: Negara asal tamu.
 - is_repeated_guest: Menunjukkan apakah tamu adalah pelanggan berulang (1 = ya, 0 = tidak).
 - previous_cancellations: Jumlah pembatalan sebelumnya oleh tamu.
 - previous_bookings_not_canceled: Jumlah pemesanan sebelumnya oleh tamu yang tidak dibatalkan.





DATASET OVERVIEW

Informasi Kolom Pemesanan:

- is_canceled: Menunjukkan apakah pemesanan dibatalkan (1 = dibatalkan, 0 = tidak dibatalkan).
- lead_time: Waktu (dalam hari) antara pemesanan dan kedatangan.
- arrival_date_year, arrival_date_month, arrival_date_day_of_month: Kolom terkait tanggal untuk melacak kapan pemesanan dijadwalkan.
- stays_in_weekend_nights, stays_in_week_nights: Durasi menginap (dalam malam).
- adults, children, babies: Jumlah tamu (dewasa, anak-anak, bayi) dalam pemesanan.
- meal: Jenis makanan yang dipilih (misalnya, bed & breakfast, half board).
- market_segment: Kategori pelanggan (misalnya, agen perjalanan online, pemesanan langsung).
- distribution_channel: Saluran tempat pemesanan dilakukan.

Informasi Kolom Keuangan:

- adr (Average Daily Rate): Rata-rata tarif harian kamar yang dipesan.
- days_in_waiting_list: Jumlah hari pemesanan berada dalam daftar tunggu.



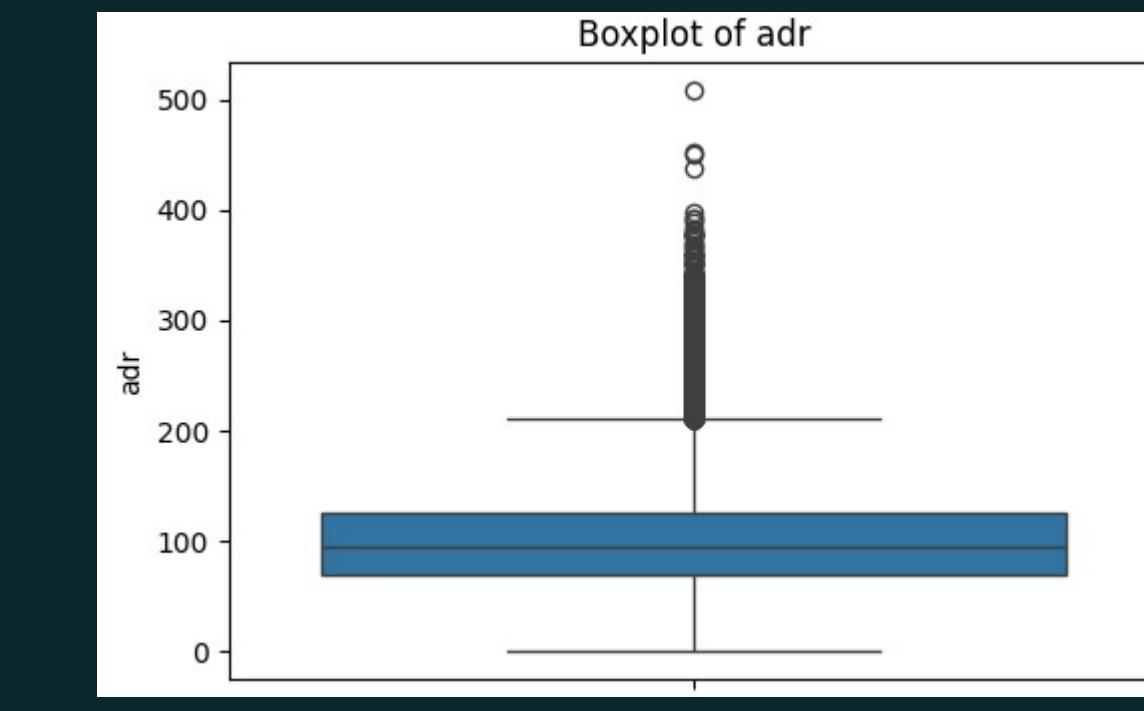
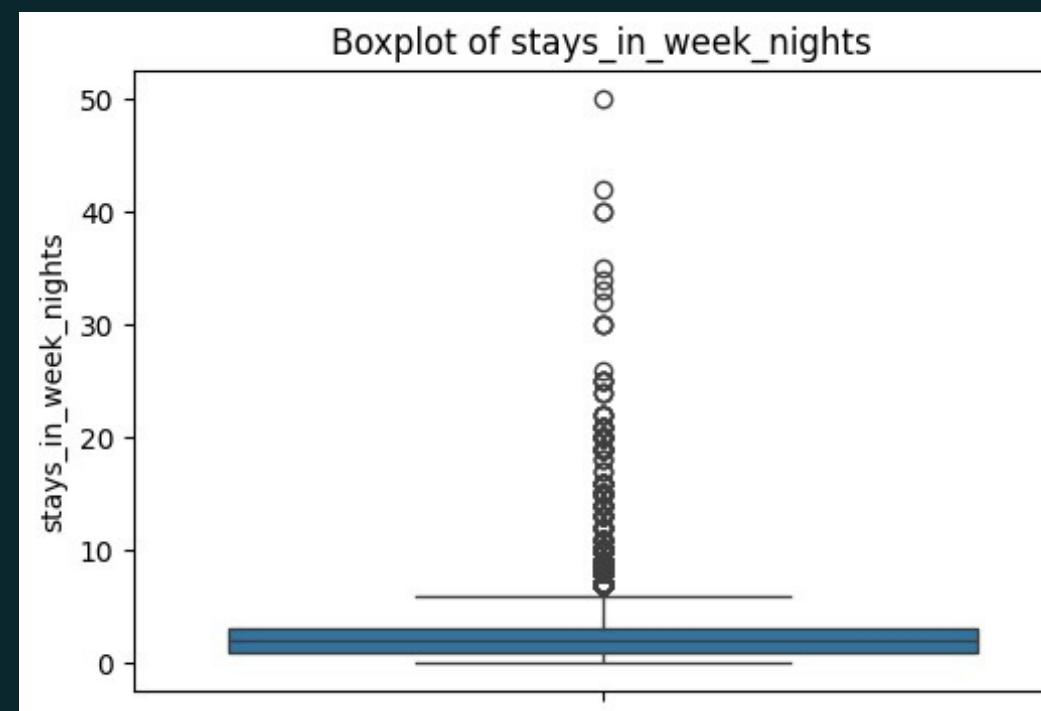
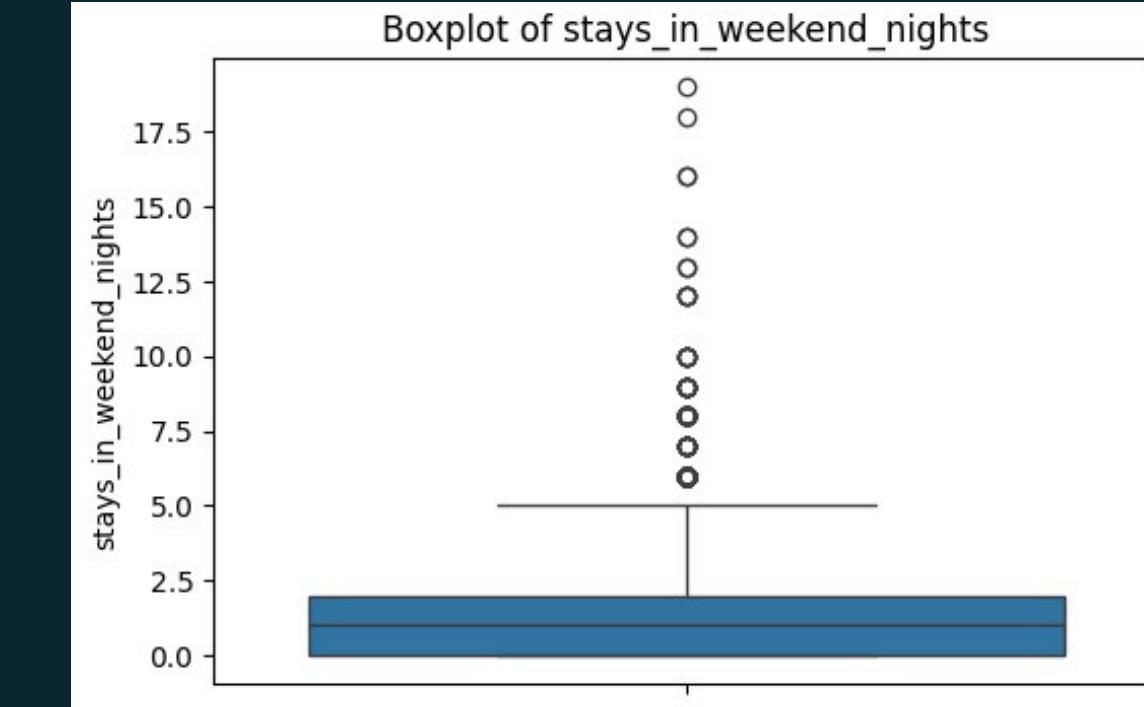
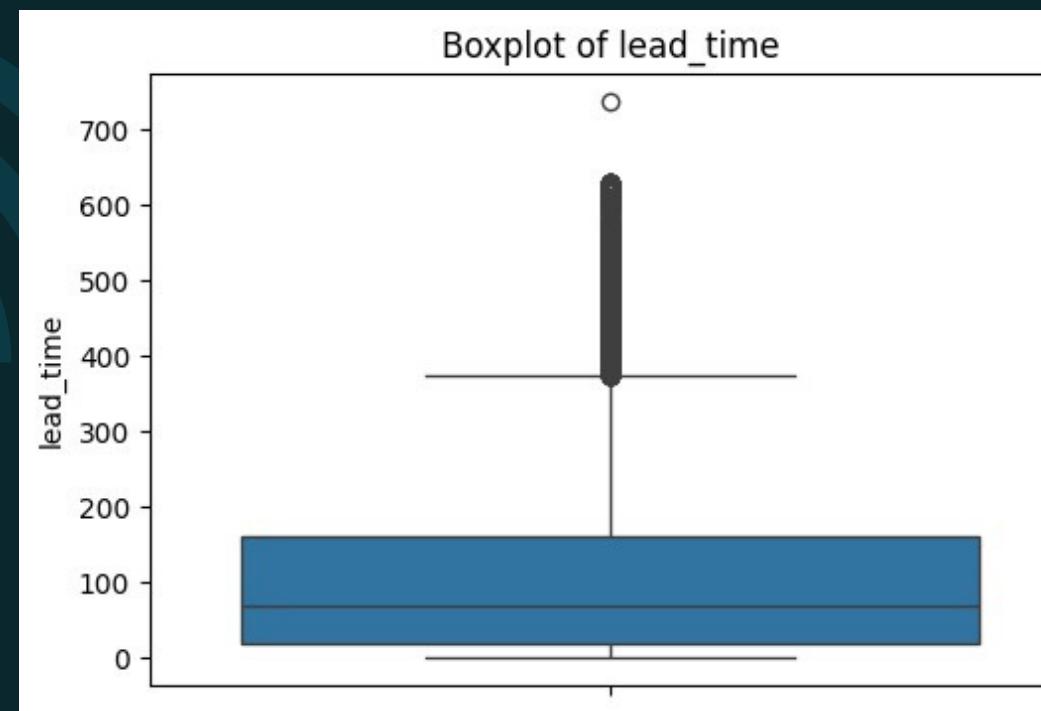
DATA CLEANING

- **Periksa Data Duplikat:** Dataset aman dari data duplikat.
- **Periksa Nilai Hilang:** Ada nilai hilang pada kolom '**Country**', '**Children**', '**Agent**', dan '**Company**'. Persentase nilai hilang pada setiap kolom, kecuali 'Company' yang memiliki 94,32%.
 - Kolom 'Agent' diisi dengan 0, menunjukkan pemesanan tanpa agen atau agen tidak dikenal.
 - Kolom 'Country' diisi dengan 'Unknown' untuk mempertahankan semua catatan untuk analisis.
 - Kolom 'Children' diisi dengan 0, mengasumsikan hilang berarti tidak ada anak yang hadir dalam pemesanan.



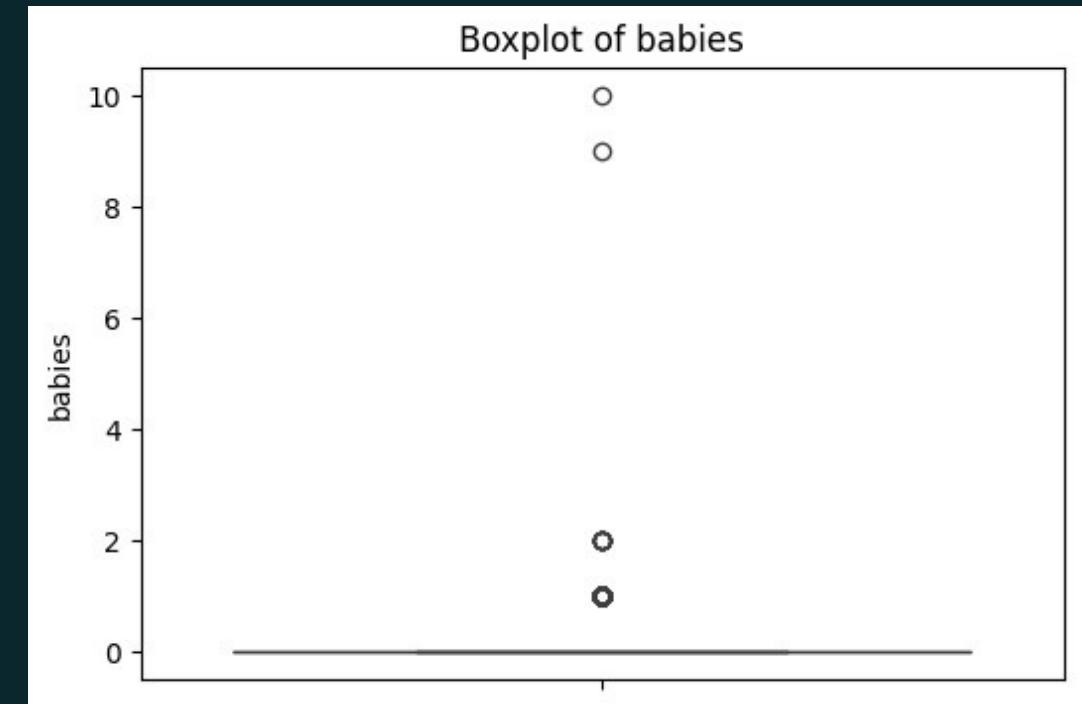


OUTLIER





OUTLIER



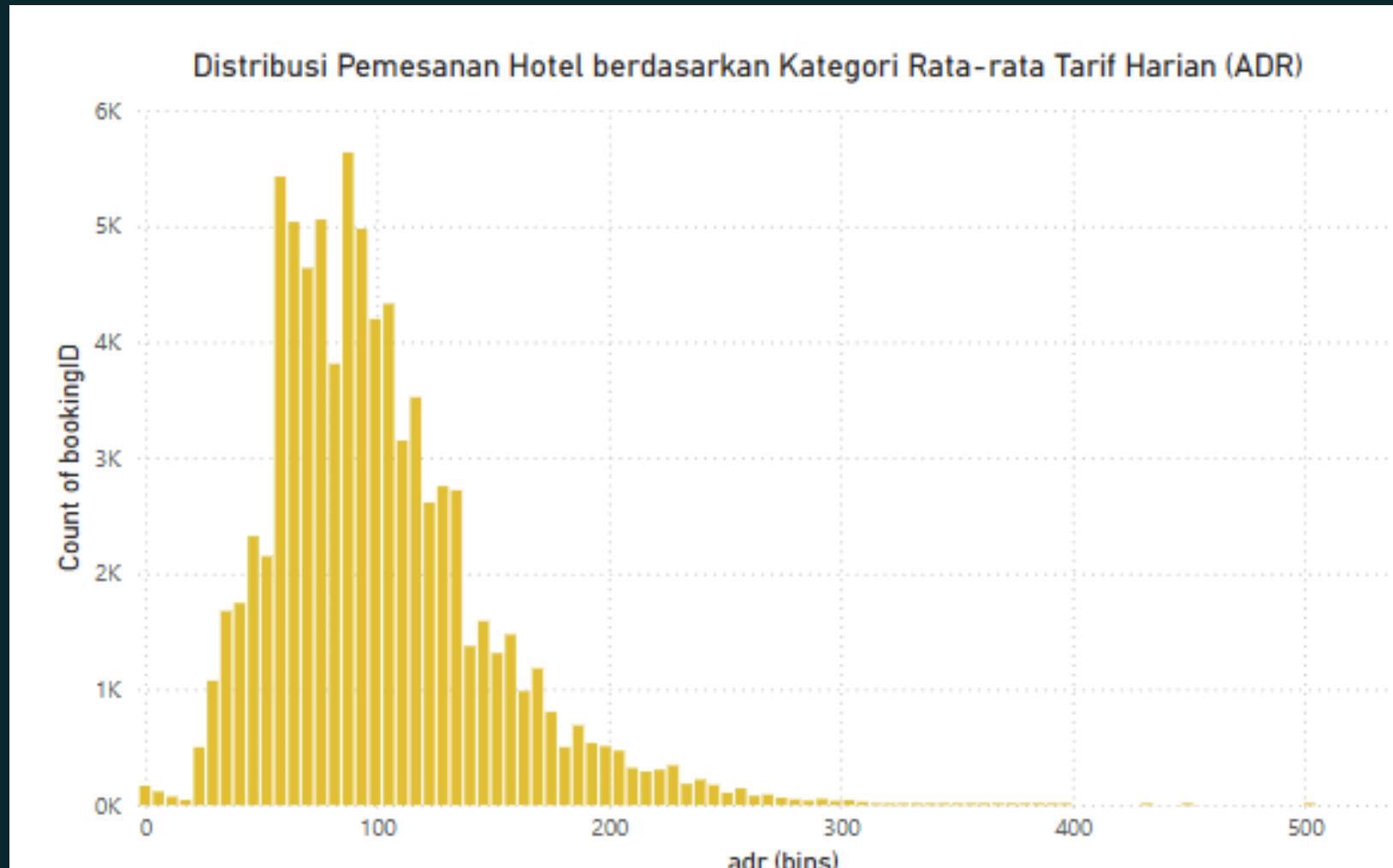


O 4

EXPLANATORY DATA ANALYSIS AND VISUALISATIONS



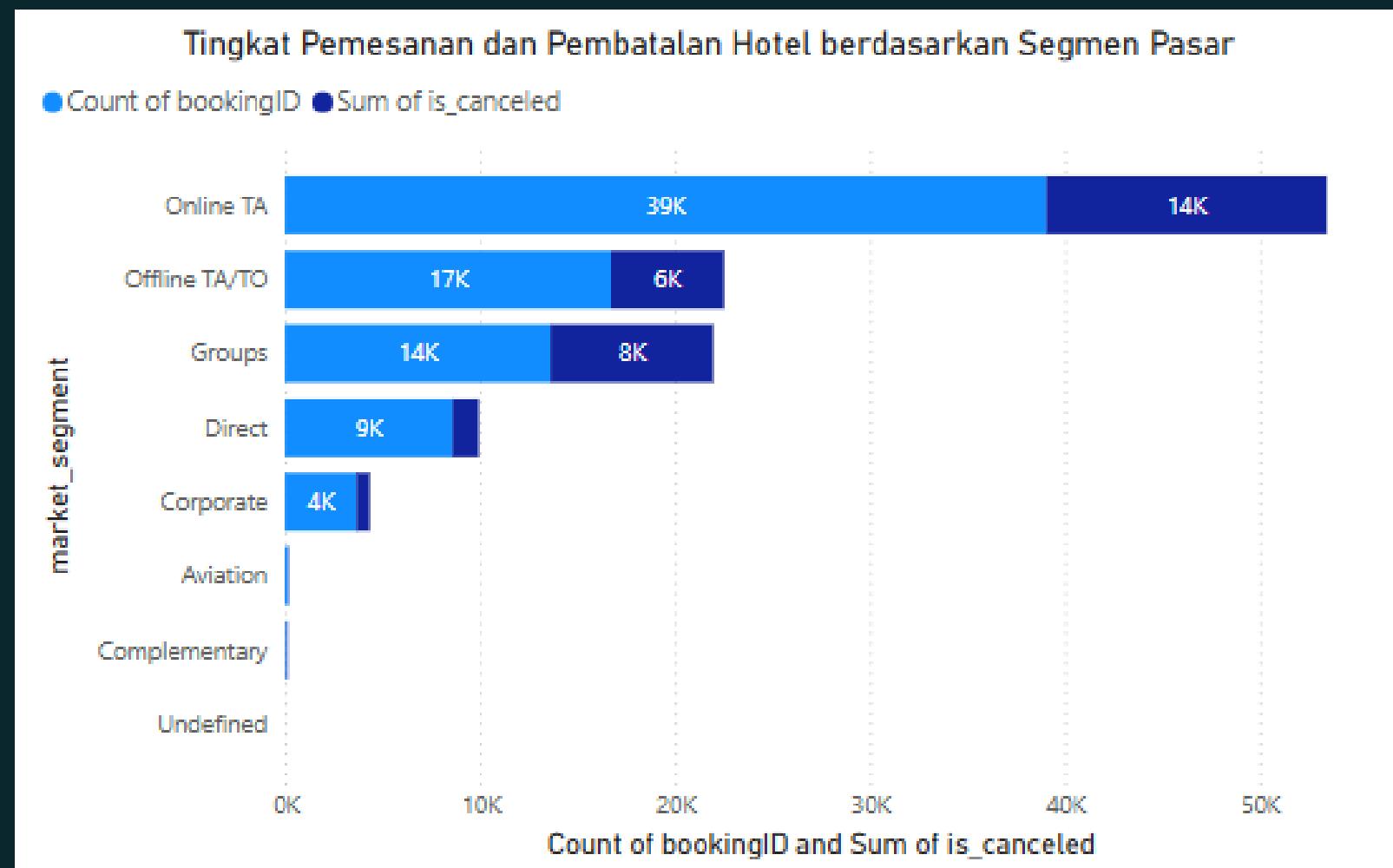
DISTRIBUSI PEMESANAN HOTEL BERDASARKAN RATA-RATA TARIF HARIAN (ADR)



- Berdasarkan visualisasi, sebagian besar pemesanan hotel termasuk dalam ADR (Average Daily Rate). Ini menunjukkan bahwa mayoritas pelanggan lebih memilih **harga menengah**, menjadikannya segmen yang paling menarik dan kompetitif bagi hotel. Untuk pertumbuhan bisnis, hotel harus fokus pada **pengoptimalan penawaran dan strategi pemasaran dalam kisaran ADR \$50-\$149** untuk menarik pangsa tamu terbesar, sambil juga mengembangkan paket premium yang ditargetkan untuk secara bertahap meningkatkan pemesanan dalam kategori ADR yang lebih tinggi.



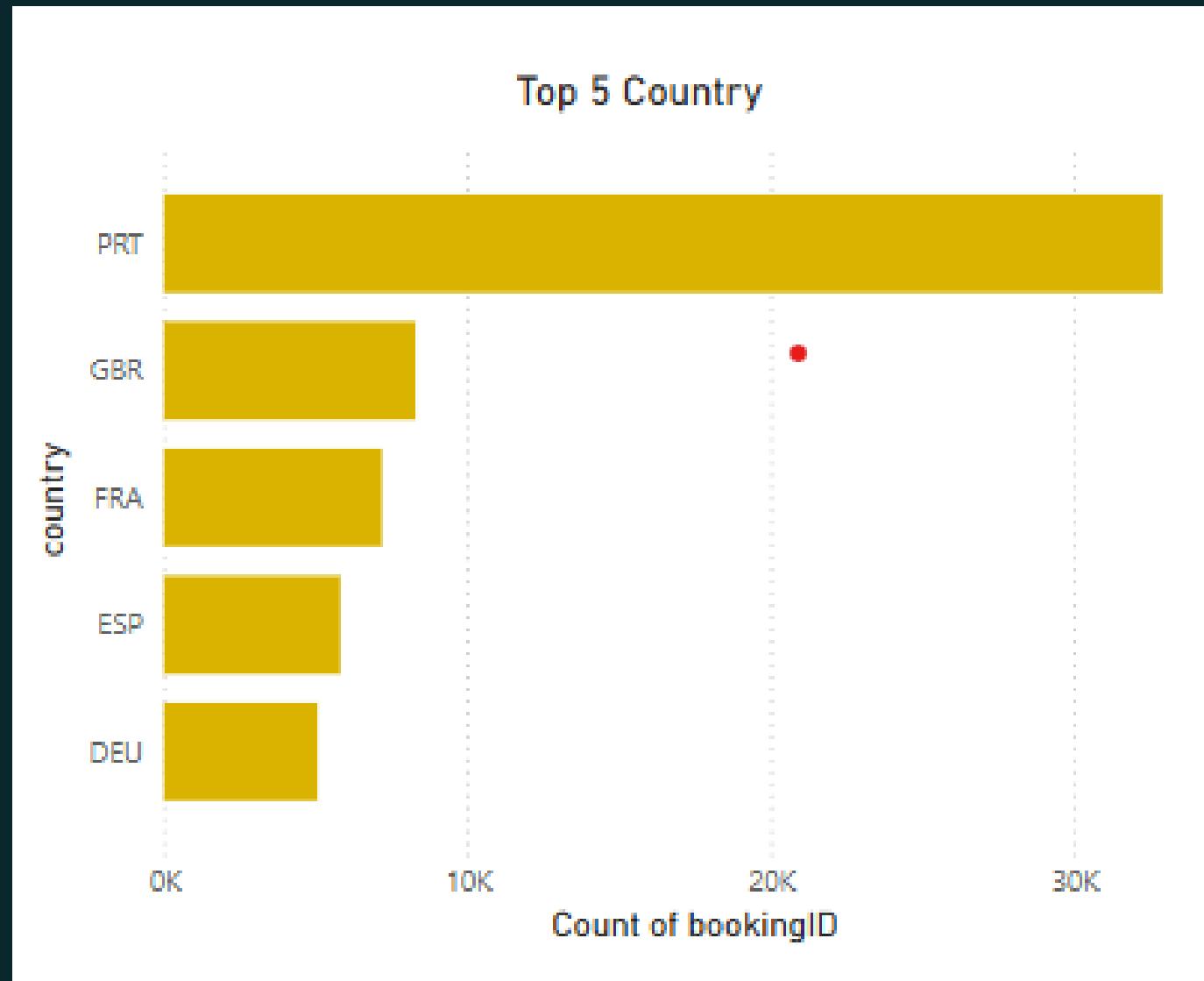
TINGKAT PEMESANAN DAN PEMBATALAN HOTEL BERDASARKAN SEGMENT PASAR



Berdasarkan visualisasi, segmen Agen Perjalanan Online (Online TA) menghasilkan jumlah pemesanan hotel tertinggi yaitu 39.000, tetapi juga memiliki jumlah pembatalan terbesar yaitu 14.000, mewakili tingkat pembatalan sekitar 37%. Offline TA/TO merupakan segmen pasar kedua tertinggi menghasilkan jumlah pemesanan yaitu 17.000 dengan jumlah pembatalan di 6.000. Grup dengan 14.000 pemesanan, dan jumlah pembatalan 8.000 (60,7%). Pemesanan langsung jauh lebih rendah yaitu 9.000, dengan jumlah pembatalan yang jauh lebih rendah yaitu 1.000.



KOTA YANG BANYAK PEMINATNYA



Berdasarkan visualisasi, **Portugal (PRT)** sejauh ini merupakan sumber pemesanan hotel terbesar, menyumbang **32.930** reservasi, hampir empat kali lebih banyak dari negara tertinggi berikutnya, **Inggris Raya (GBR)** dengan **8.310** pemesanan. **Prancis (FRA)**, **Spanyol (ESP)**, dan **Jerman (DEU)** mengikuti dengan masing-masing **7.190**, **5.830**, dan **5.070** pemesanan. Ini menunjukkan bahwa permintaan domestik dari Portugal adalah pendorong utama bisnis hotel.



05

MACHINE LEARNING PROCESS



MACHINE LEARNING

Model	R2(Train)	MAE(Train)	RMSE(Train)
Random Forest	0.9726	4.58	7.95
LightGBM	0.7752	16.45	22.76
Decision Tree	0.9999	0	0.35
XGBoost	0.8288	14.15	19.86
Linear Regression	0.5837	23.34	30.98

Model	R2(Validation)	MAE(Validation)	RMSE(Validation)
Random Forest	0.798	12.46	21.54
LightGBM	0.7514	17.16	23.9
Decision Tree	0.6314	15.64	29.1
XGBoost	0.7785	15.7	22.56
Linear Regression	0.5746	23.46	31.26

Model	R2(Test)	MAE(Test)	RMSE(Test)
Random Forest	0.8042	12.29	21.17
LightGBM	0.768	16.75	23.05
Decision Tree	0.6443	15.25	28.54
XGBoost	0.7961	15.24	21.6
Linear Regression	0.5868	23.31	30.76



MACHINE LEARNING

- Random Forest adalah model dengan kinerja terbaik, menunjukkan nilai R² tertinggi, serta MAE (Mean Absolute Error) dan RMSE (Root Mean Squared Error) terendah di seluruh dataset (Latih, Validasi, dan Uji). Ini menjadikannya pilihan terbaik untuk prediksi ini.
- XGBoost adalah pesaing kuat, dengan nilai R² yang sebanding dengan Random Forest, namun memiliki MAE dan RMSE yang sedikit lebih tinggi. Meskipun demikian, ini tetap merupakan model yang baik.
- LightGBM memberikan hasil yang cukup baik, tetapi tidak mengungguli Random Forest atau XGBoost.
- Decision Tree tampaknya mengalami overfitting, menunjukkan hasil yang sempurna pada set pelatihan tetapi berkinerja buruk pada set pengujian.
- Linear Regression adalah model terlemah, menunjukkan kinerja yang buruk di semua metrik. Model ini kemungkinan besar tidak cocok untuk tugas ini.



05
RECOMMENDATION





RECOMENDATION

- Berikan diskon maupun iklan yang lebih menarik untuk pemesanan secara online, dikarenakan banyaknya pemesanan pada online dan berikan juga konsekuensi tidak membalikkan DP agar pelanggan tidak akan membatalkan pemesanan kamar
- Untuk pertumbuhan bisnis, hotel harus fokus pada **pengoptimalan penawaran dan strategi pemasaran dalam kisaran ADR \$50–\$149** untuk menarik pangsa tamu terbesar, sambil juga mengembangkan paket premium yang ditargetkan untuk secara bertahap meningkatkan pemesanan dalam kategori ADR yang lebih tinggi.
- Dengan letak hotel yang berada pada banyak negara, bisa diprioritaskan kepada 5 negara yang memiliki banyak dampak pada pemesanan kamar yaitu negara Portugal, Inggris Raya, Prancis, Spanyol dan Jerman
- Machine Learning yang paling baik digunakan untuk model ini adalah Random Forest menunjukkan nilai R2 tertinggi, serta MAE (Mean Absolute Error) dan RMSE (Root Mean Squared Error) terendah di seluruh dataset (Train, Validation, dan Test)



THANK
YOU

