

# PROJECT PRESENTATION

PRESENTED DIMAS HARDIANTO

28 - May - 2025

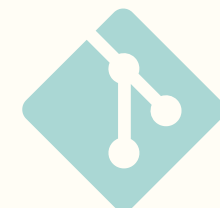


# About Me

Hi, I'm Dimas Hardianto, with 5+ years of experience in banking and fintech, focusing on customer-facing and data-driven roles like Relationship Manager and Marketing Specialist. Motivated by a strong interest in data-informed decisions, I recently completed a Data Science bootcamp where I built projects in churn prediction, customer segmentation (RFM & clustering), and sentiment analysis.

I'm skilled in Python (Pandas, Scikit-Learn), SQL, Power BI, and Streamlit, and enjoy translating raw data into actionable business insights.

One project I'm proud of was analyzing TransJakarta transaction data to segment users and support service improvement. Now, I'm eager to contribute as a Data Scientist and drive real business impact through data.



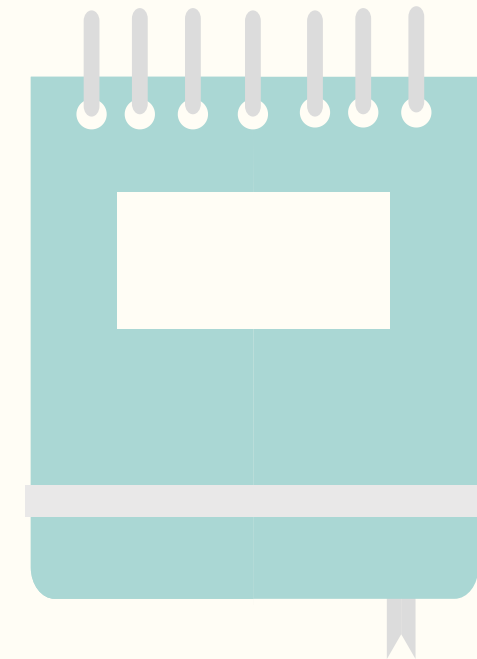
git

---

# Table of Contents

---

- Introduction
- Overview Project
- Background Project
- Problem Statement
- Data Understanding
- Eksploratif ( EDA )
- Analisis Statistik & Business Insights
- Information Dashboard
- Recommendation





# OVERVIEW PROJECT



# Early Detection Of Cancellation Risk

## Tujuan Utama

Memahami pola booking hotel dan mengidentifikasi faktor-faktor kunci yang mendorong pembatalan reservasi, guna merancang strategi pencegahan yang lebih efektif.



*“Dengan memahami siapa, kapan, dan kenapa pelanggan membatalkan pemesanan, bisnis dapat mengambil tindakan preventif yang lebih presisi dan meningkatkan efisiensi operasional.”*



- **Exploratory Data Analysis (EDA) Mendalam.**

Menganalisis karakteristik booking, tipe customer, perilaku cancel, dan pola harga untuk menemukan insight yang tersembunyi.



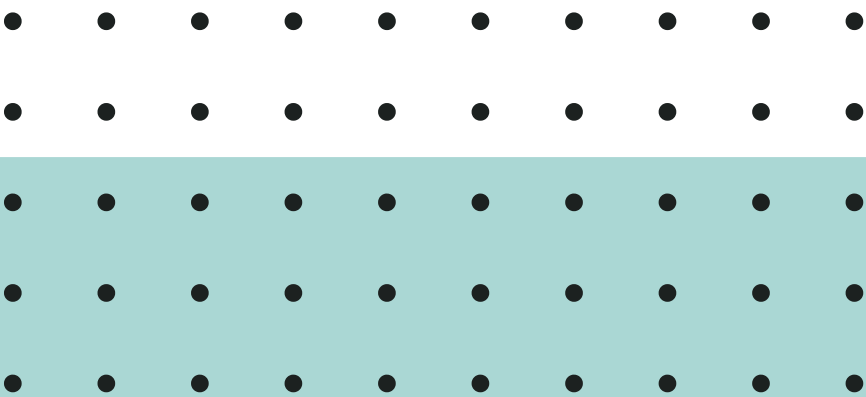
- **Identifikasi Faktor Risiko Pembatalan**

Menyusun visualisasi dan statistik deskriptif untuk memahami kombinasi variabel (seperti ADR, durasi inap, lead time) yang memengaruhi risiko pembatalan.

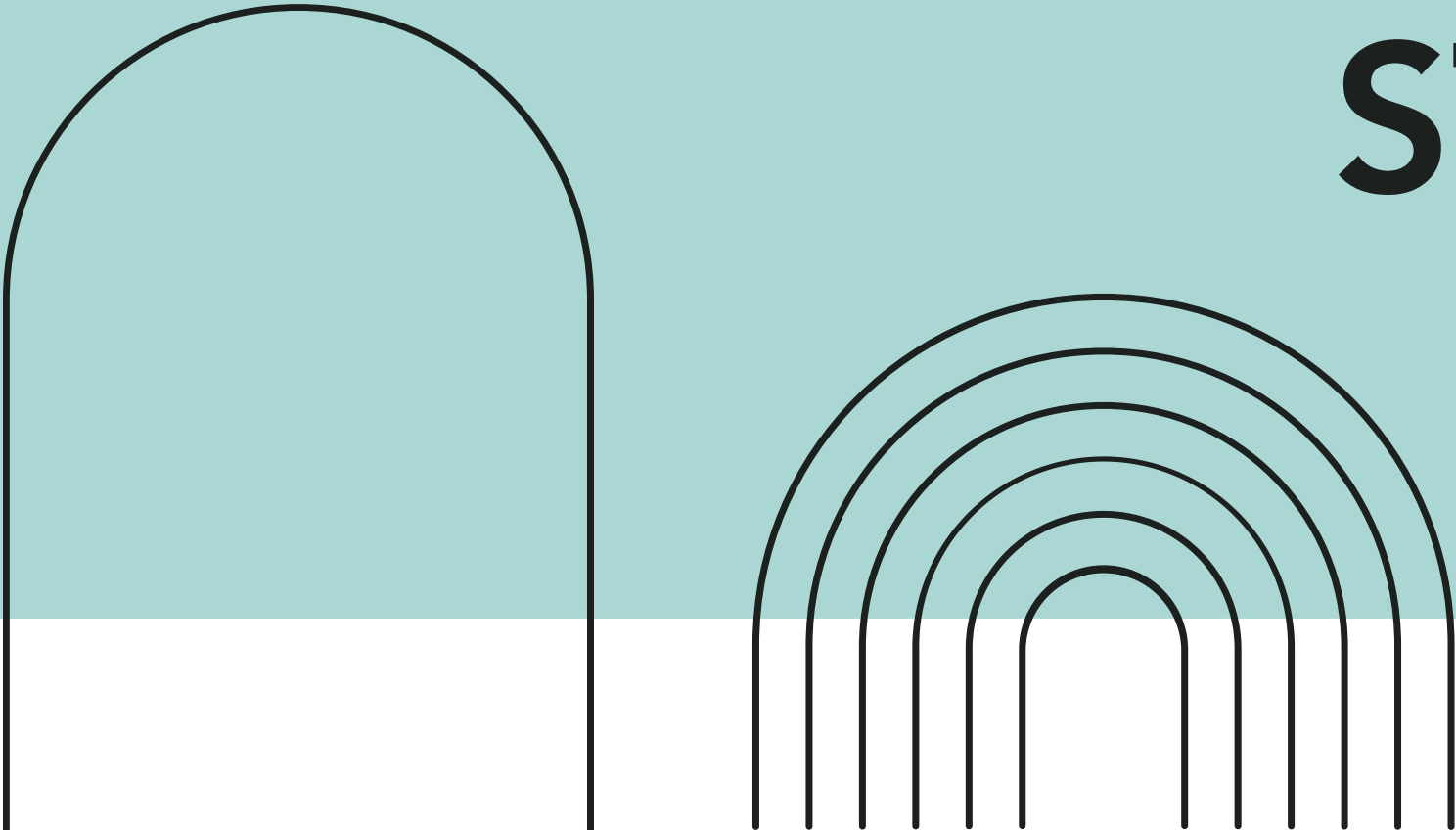


- **Dasar Strategi Bisnis Preventif**

Menyediakan insight data-driven sebagai dasar rekomendasi bisnis, seperti strategi harga adaptif, kebijakan deposit, dan segmentasi pelanggan.



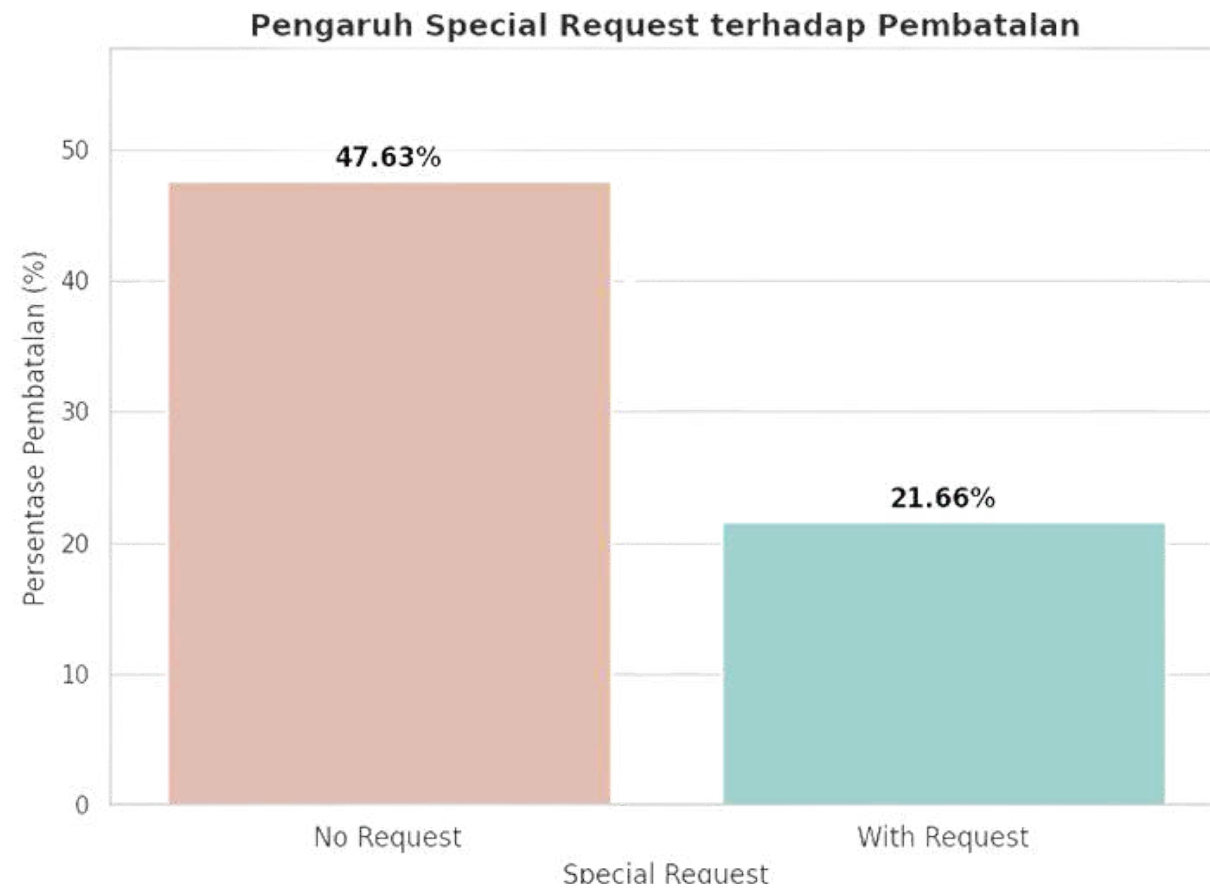
# ANALISIS POLA BOOKING & PREDIKSI PEMBATALAN HOTEL UNTUK REKOMENDASI STRATEGI BISNIS



All scripts, data, and other technical part of the project can be accessed in this [link](#)



# Background



1

Pelanggan dengan permintaan khusus memiliki tingkat pembatalan hanya 21%, jauh lebih rendah dibandingkan 37% pada pelanggan tanpa permintaan.

2

Permintaan khusus mencerminkan keseriusan & keterlibatan pelanggan, sehingga dapat dijadikan indikator loyalitas dan dasar strategi pencegahan pembatalan.

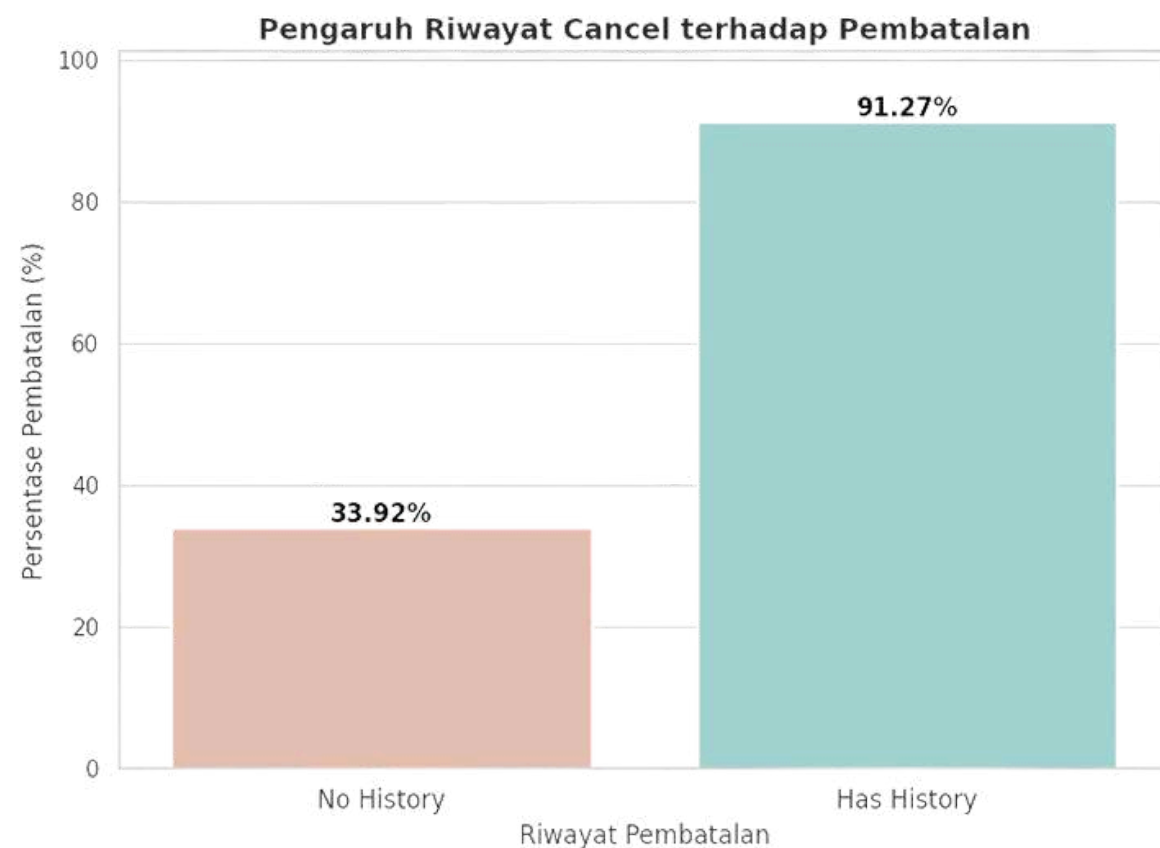


1

Pelanggan dengan riwayat pembatalan memiliki peluang pembatalan ulang sebesar 74%, dibandingkan hanya 24% untuk pelanggan tanpa riwayat.

2

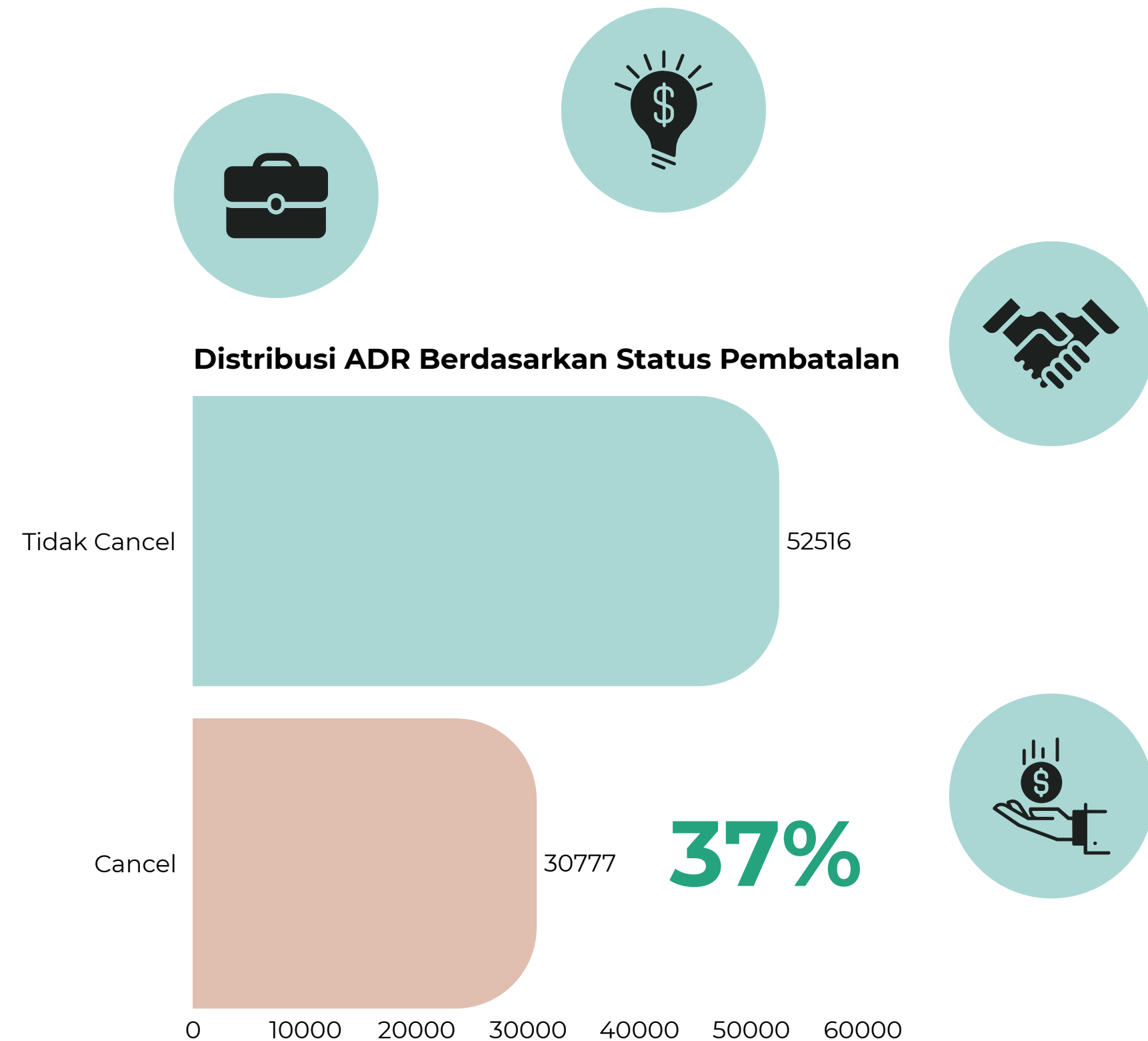
Riwayat pembatalan adalah indikator kuat risiko churn, mencerminkan perilaku tidak loyal atau perbandingan harga aktif antar platform.



# Problem Statement

Mengapa hampir 4 dari 10 reservasi dibatalkan, dan apa pola perilaku customer yang menyebabkan pembatalan?

- Dari total **83.293** pemesanan, terdapat sekitar **37%** pemesanan yang dibatalkan.
- Ini berarti *lebih dari 1 dari 3 tamu* membatalkan reservasinya, yang tentu berdampak langsung pada revenue dan okupansi hotel.
- Tingkat pembatalan yang tinggi ini bukan hanya angka, tapi potensi kerugian operasional yang signifikan.





# Metodelogi Analisis Data

## Initial Research

01

- Identifikasi topik: Analisis Pola Booking & Pembatalan Hotel.
- Penentuan Data Hotel



## Data Collection

02

- Periode: 1 Januari 2017 – 31 Desember 2019.
- Data: Memahami struktur, ukuran, dan tipe data awal (jumlah kolom, baris, tipe fitur).



## Data Processing

03

- Identifikasi nilai hilang, duplikasi, dan pencilan yang dapat memengaruhi analisis.



## Reporting

05

- Analisis Stastik dan Business Insight
- Pembuatan Dashboard





# DATA UNDERSTANDING

## Read Data

Dataset berisi **83.293** data booking hotel dengan **33 fitur** untuk analisis

## Data Cleaning

Kolom company dihapus karena **94.3%** missing, hanya muncul di **5.6%** data, dan berpotensi jadi noise dalam analisis.

## Proses EDA

Membuat pertanyaan yang dirancang untuk mendukung pengambilan keputusan berbasis data dalam meningkatkan efisiensi operasional dan profitabilitas bisnis.



## Check Data

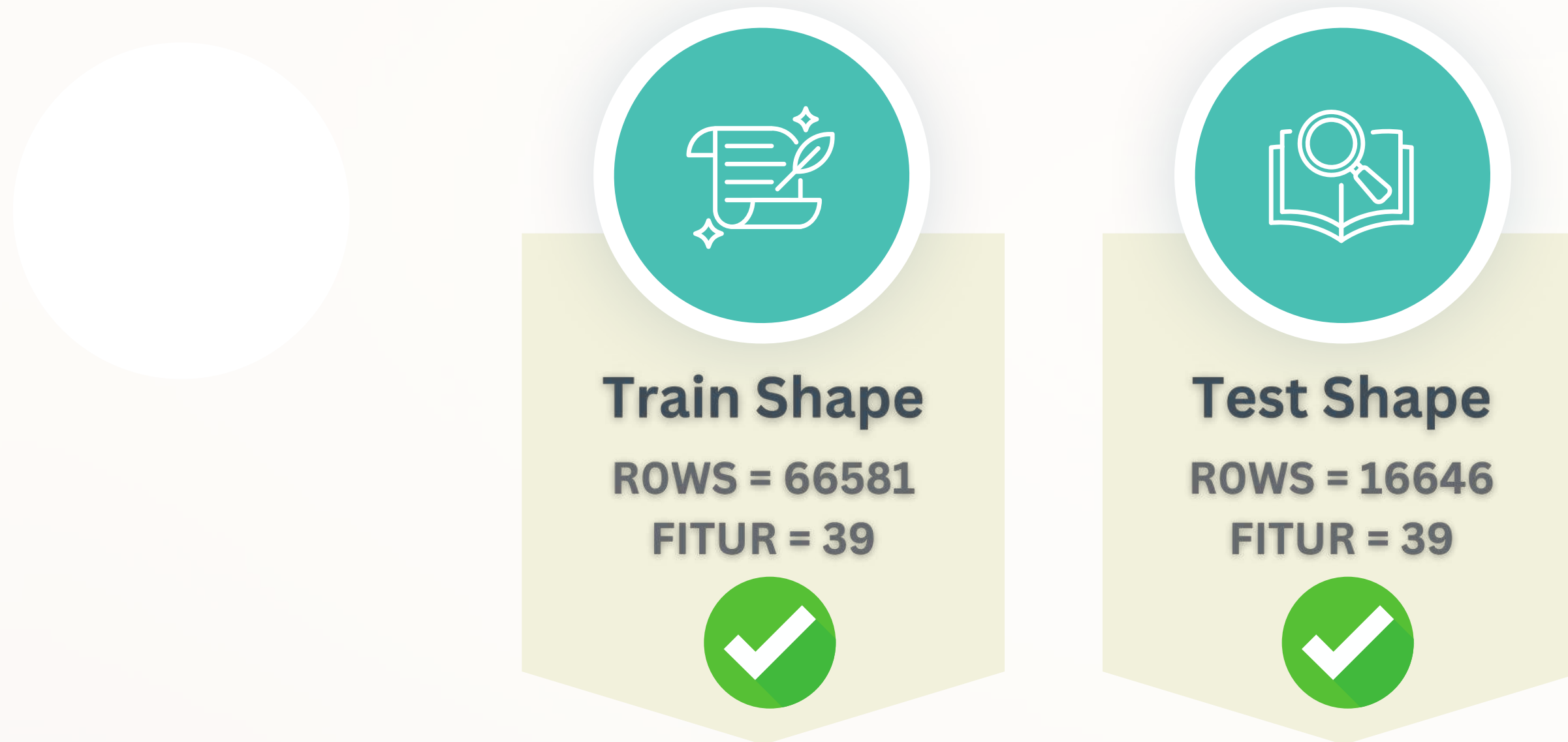
Terdapat missing values pada kolom children (**3 baris**), country (**346 baris**), agent (**11.404 baris**), dan company (78.599 baris). Kolom bookingID berfungsi sebagai ID unik tiap reservasi.

## Feature Engineering

Setelah proses feature engineering, terdapat 13 kolom baru, menjadikan total kolom meningkat dari **33 menjadi 46**. Kolom tambahan mencakup informasi seperti arrival\_day\_of\_week, arrival\_quarter, total\_nights, hingga flag tertentu (is\_family, room\_mismatch\_flag, dll) untuk mendukung analisis dan prediksi yang lebih akurat.



# MENYIAPKAN DATA UNTUK PEMODELAN: PEMBAGIAN TRAIN-TEST 80:20



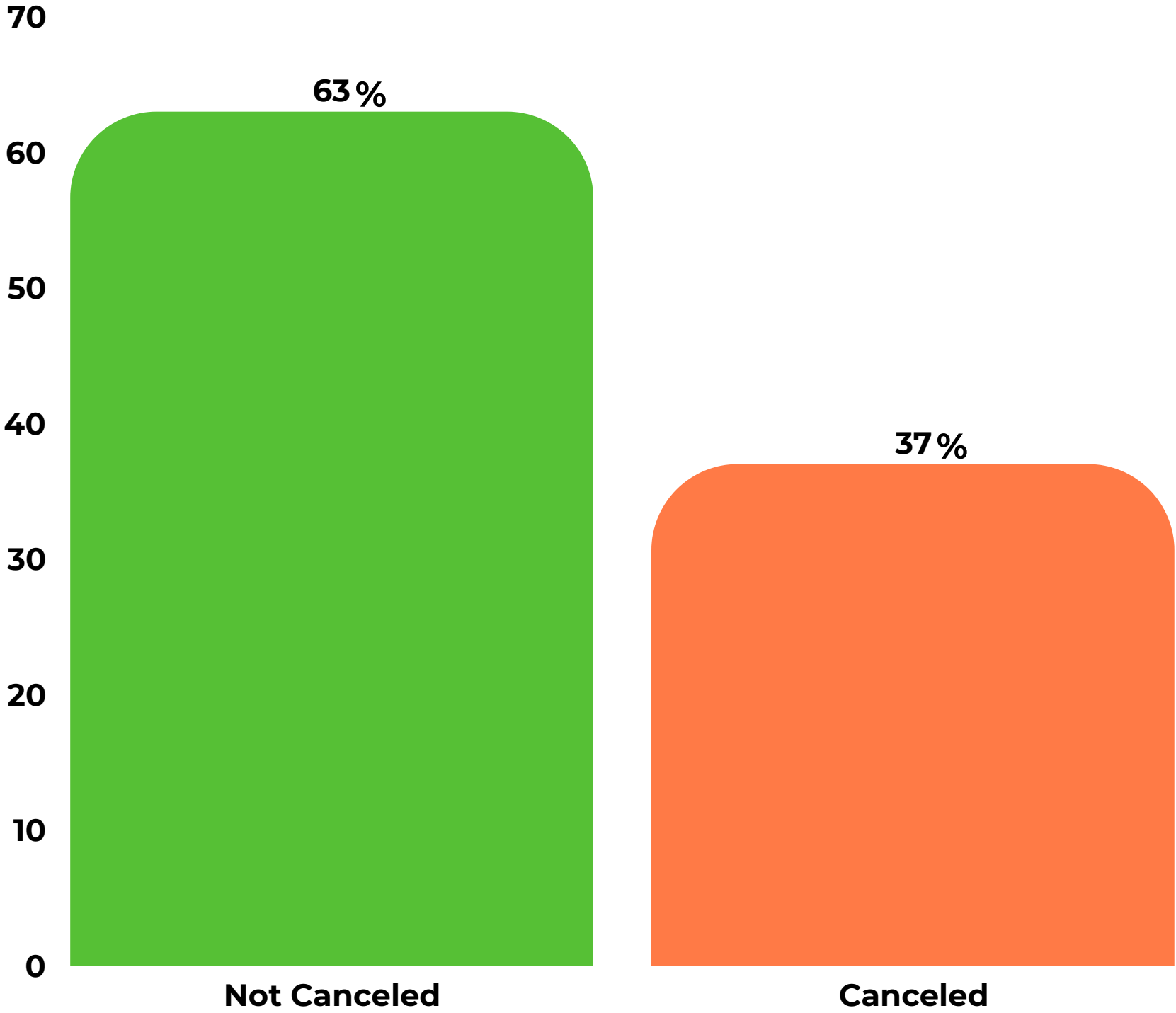
**Data is clean and ready for modeling**



# **EXPLORATORY DATA ANALYSIS**



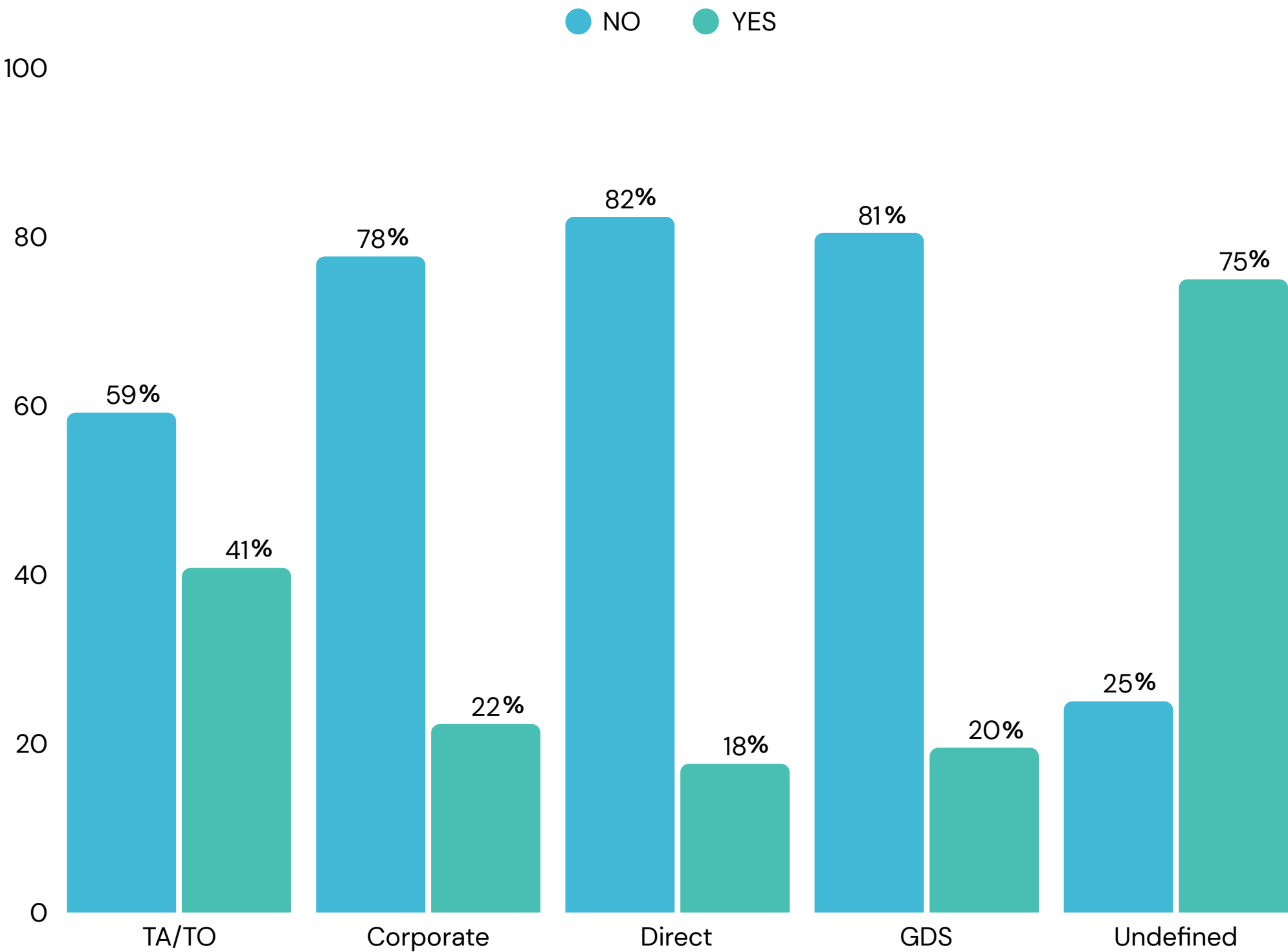
# Distribusi Pembatalan Booking



- Mayoritas booking tidak dibatalkan – Sebanyak **63.06%** reservasi berhasil dilanjutkan hingga selesai, menunjukkan potensi pendapatan yang masih dominan bisa diamankan.
- Risiko pembatalan signifikan – Sebanyak **36.94%** booking dibatalkan, yang merupakan angka cukup tinggi dan berpotensi menimbulkan kerugian operasional serta pendapatan yang hilang jika tidak ditangani secara proaktif.

# Distribusi Pembatalan Booking Berdasarkan Channel Penjualan

Distribusi channel vs is\_canceled



Channel TA/TO paling banyak digunakan Menyumbang jumlah reservasi tertinggi (68.2% dari total), namun memiliki tingkat pembatalan cukup tinggi (40.8%)

Channel Direct paling stabil. Memiliki tingkat pembatalan terendah (17.6%), menunjukkan bahwa pemesanan langsung lebih loyal dan dapat diandalkan.

Corporate booking cenderung tidak dibatalkan. Hanya 22.3% dari pemesanan via channel ini yang dibatalkan, menandakan potensi kerjasama B2B lebih aman secara revenue.

Channel Undefined dan GDS minim volume tapi rentan. Meski volume kecil, Undefined menunjukkan tingkat pembatalan tertinggi (75%), penting untuk ditindak lanjut validasi datanya.



# Bagaimana tren pembatalan booking per bulan dan per kuartal?

Tren Rata-rata Pembatalan Booking Hotel per Bulan



- Pembatalan Tertinggi Terjadi pada Kuartal 2 (April-Juni) **Rata-rata cancel rate > 40%** di bulan April dan Juni, menunjukkan lonjakan signifikan pada Q2.
- Tren Pembatalan Menurun di Kuartal 4 (Oktober-Desember) Tingkat pembatalan relatif lebih rendah (**sekitar 31-35%**), menunjukkan periode ini lebih stabil dalam pemesanan.



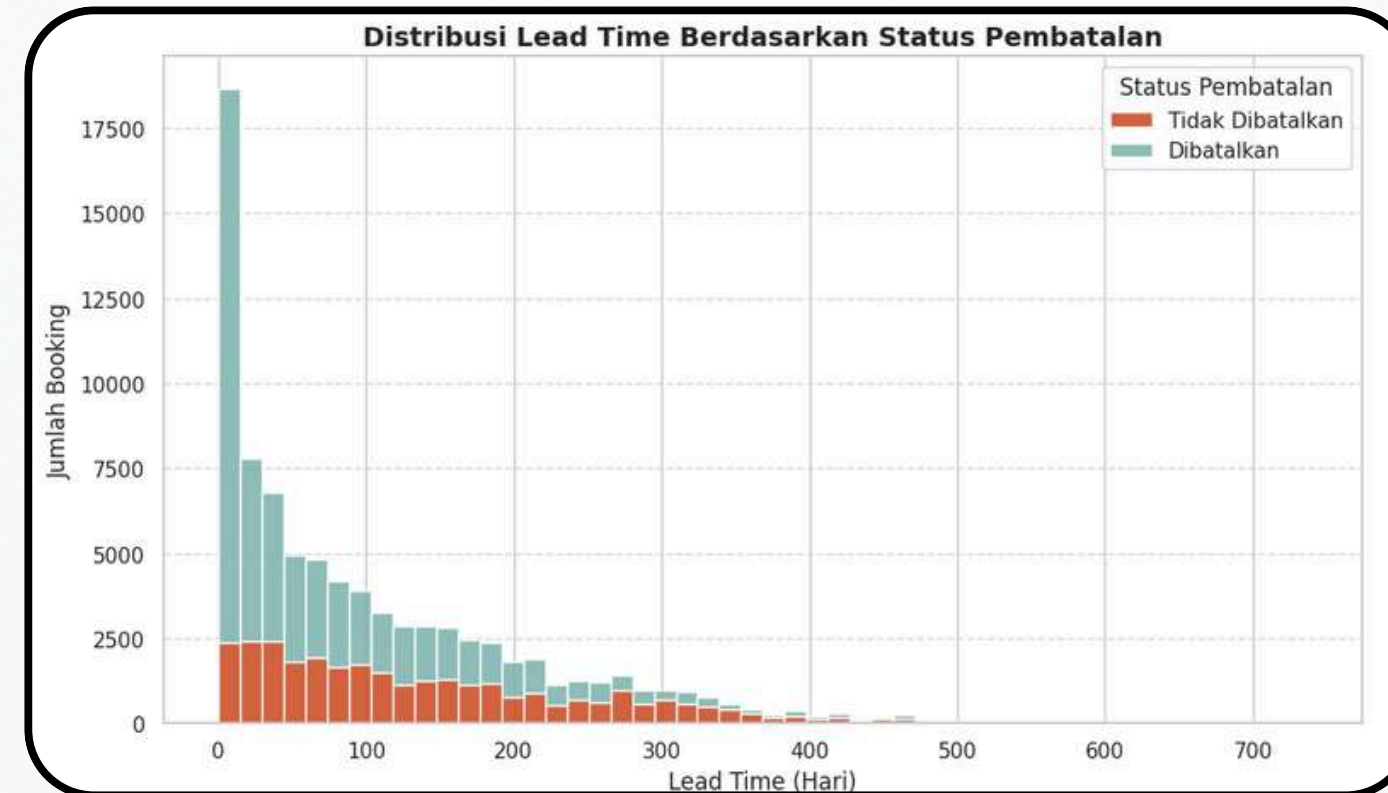
# APAKAH LEAD TIME (WAKTU ANTARA BOOKING DAN CHECK-IN) MEMENGARUHI PEMBATALAN?

Booking terbanyak terjadi pada lead time  
0–15 hari, menunjukkan mayoritas tamu  
memesan mendekati tanggal menginap.

Booking dengan lead time pendek (0–30 hari)  
mendominasi volume booking. Ini bisa  
menunjukkan perilaku pelanggan yang lebih  
spontan dalam memesan hotel.



Namun, pembatalan juga cenderung tinggi di  
lead time yang panjang (>90 hari). Hal ini  
masuk akal karena semakin jauh dari tanggal  
menginap, kemungkinan terjadi perubahan  
rencana pelanggan semakin besar.



Pada lead time menengah (30–90 hari),  
terdapat distribusi yang lebih seimbang antara  
booking yang dibatalkan dan tidak. Ini  
menunjukkan bahwa di fase ini pelanggan  
mulai lebih pasti, tapi risiko pembatalan masih  
ada.

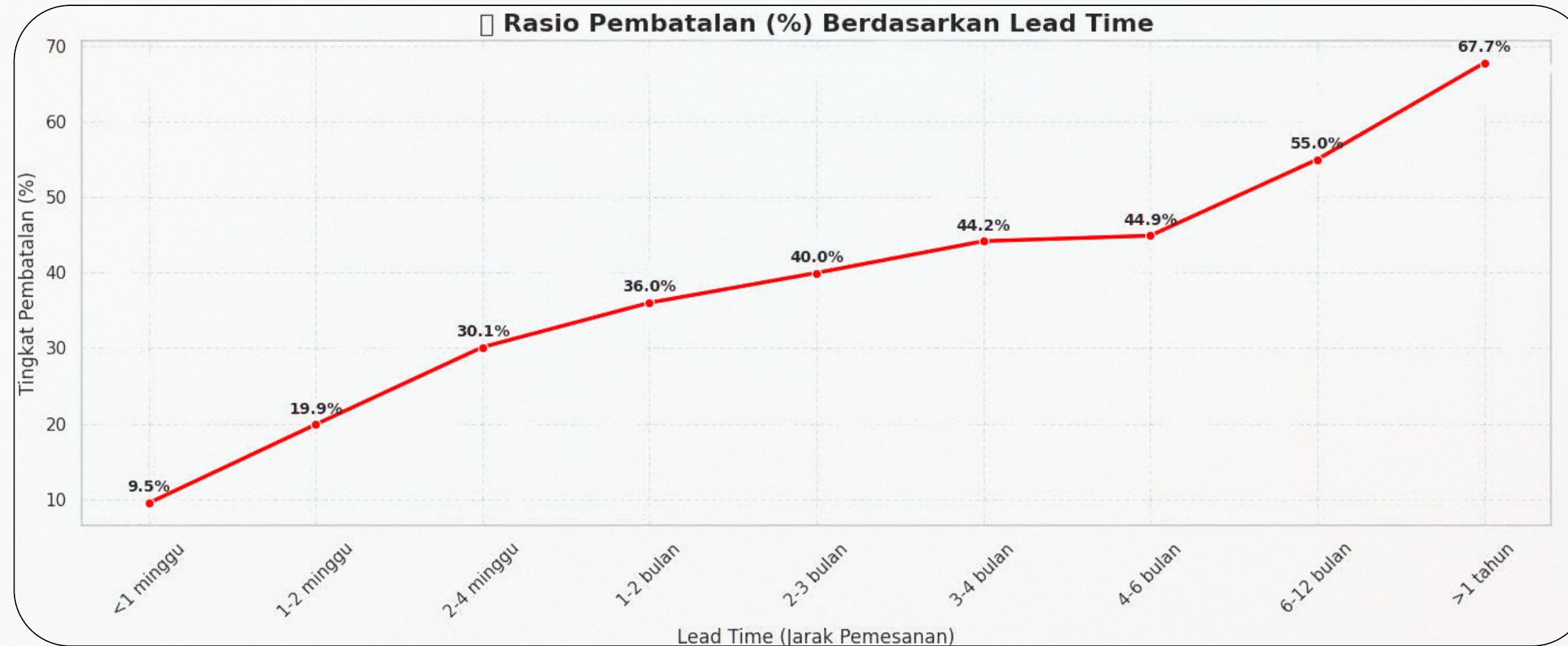


Booking dengan lead time ekstrem panjang  
(>200 hari) jarang terjadi, tapi proporsi  
pembatalannya terlihat cukup besar dibanding  
jumlah booking-nya. Ini bisa menjadi  
pertimbangan untuk kebijakan refund atau  
deposit untuk jangka waktu jauh.

**Kesimpulan: Lead time memiliki korelasi positif dengan pembatalan — makin jauh jarak  
antara booking dan check-in, makin besar risiko pembatalan.**



# SEMAKIN CEPAT DIPESAN, SEMAKIN RAWAN BATAL?



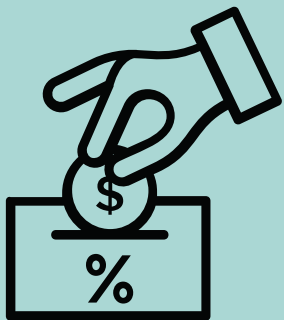
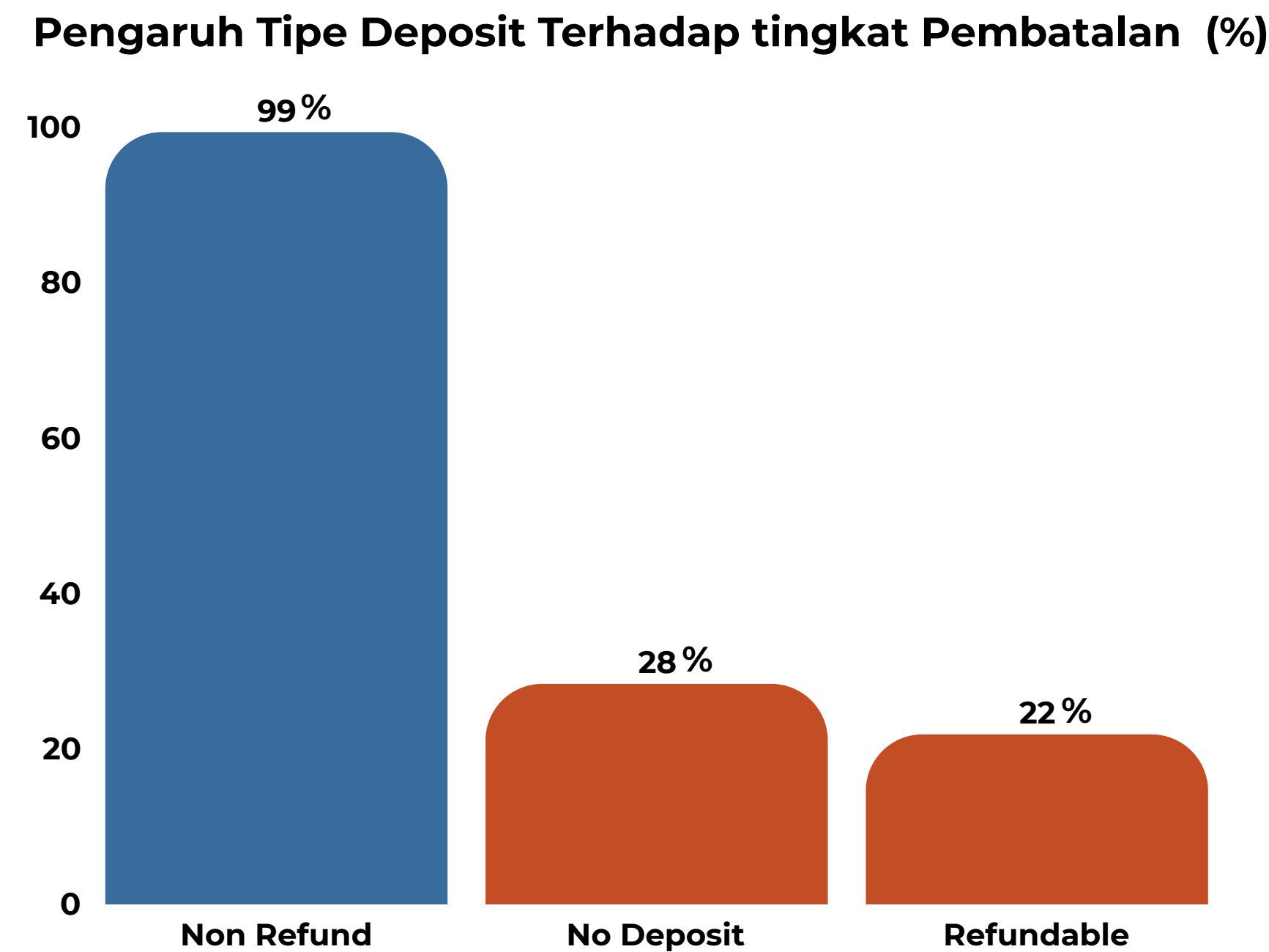
Semakin jauh jarak waktu pemesanan (lead time), semakin tinggi kemungkinan pembatalan — Rasio pembatalan meningkat dari 9.5% (< 1 minggu) hingga 67.7% (> 1 tahun).

Pemesanan mendekati tanggal menginap lebih stabil dan jarang dibatalkan — Karena tamu biasanya sudah yakin dan memiliki rencana yang pasti.

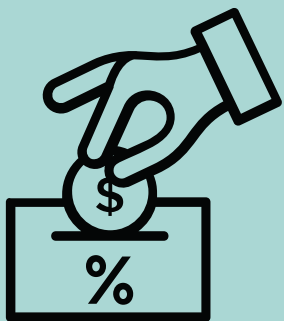


# Fleksibilitas vs Komitmen: Dampaknya pada Pembatalan

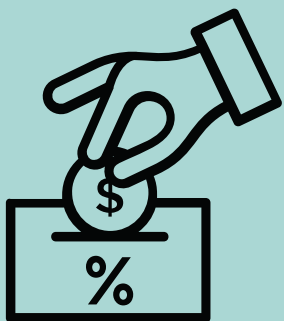
*Apa pengaruh tipe deposit terhadap kemungkinan pembatalan?*



**Not Refund = 99.3% dibatalkan**  
→ Sangat tidak wajar, kemungkinan ada error input atau disalahgunakan untuk "booking sementara".



**No Deposit = 28.3% pembatalan**  
→ Tingkat pembatalan tinggi karena pelanggan tidak merasa berkomitmen.

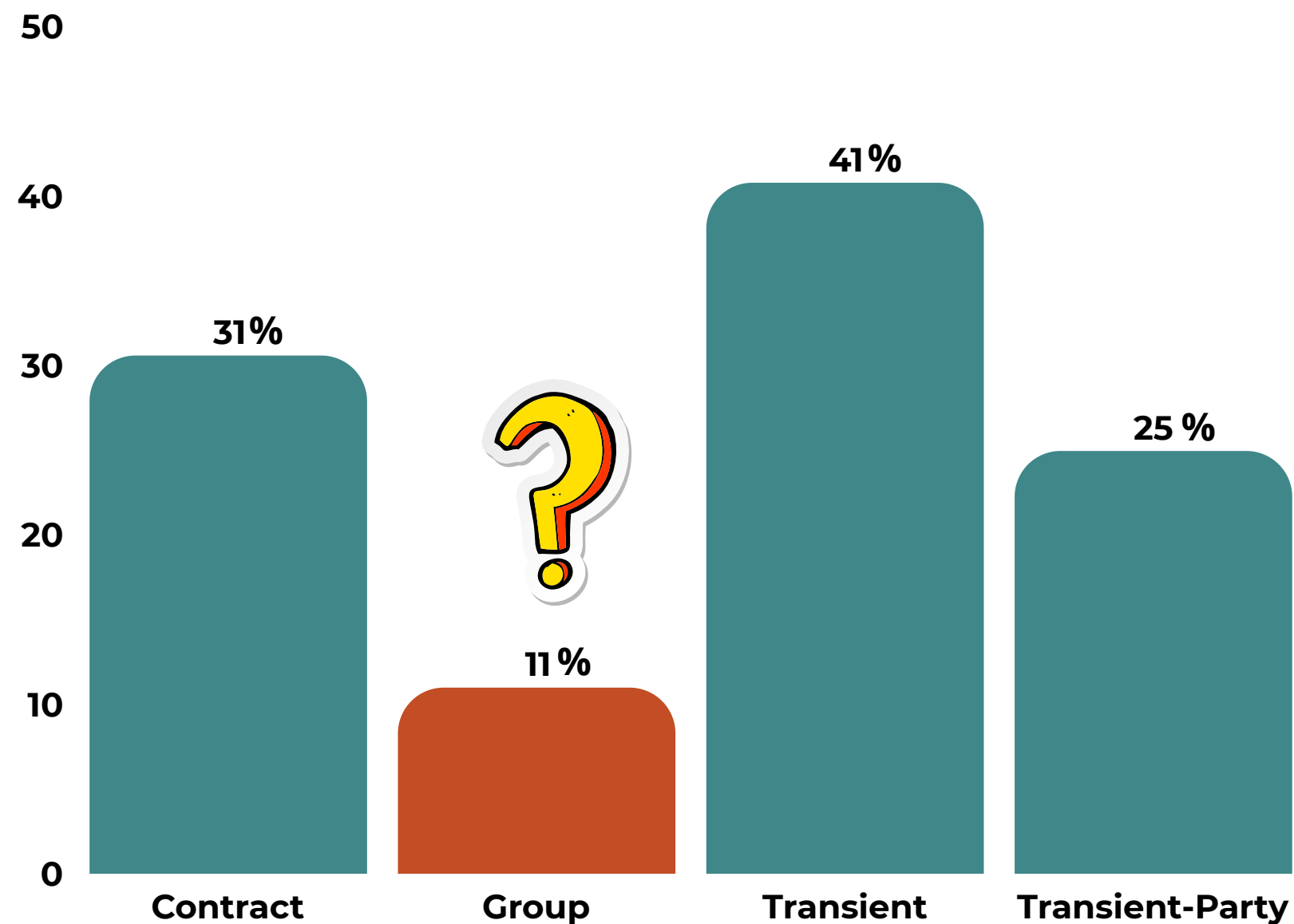


**Refundable Deposit = 21.8% pembatalan**  
→ Pilihan paling stabil, menunjukkan komitmen pelanggan tanpa menghilangkan fleksibilitas.

Ya, ada pengaruh yang jelas dan terukur antara tipe deposit dan kemungkinan pembatalan. Ini adalah salah satu fitur penting yang perlu dipertimbangkan baik dalam analisis maupun strategi bisnis ke depan.

# Tipe Pelanggan: Siapa Paling Rentan Membatalkan?

Tingkat Pembatalan berdasarkan Tipe Customer



**Transient** = tingkat pembatalan tertinggi (40.78%) → Individu tanpa paket atau kontrak, cenderung fleksibel dan spontan.

**Group** = paling stabil (10.98%) → Booking besar dengan perencanaan dan komitmen kolektif yang kuat.

**Transient-Party** = pembatalan menengah (24.94%) → Efek komitmen kelompok kecil menurunkan niat pembatalan.

**Contract** = tetap tinggi (30.58%) meskipun korporat → Kemungkinan besar karena fleksibilitas jadwal atau revisi kerja sama.

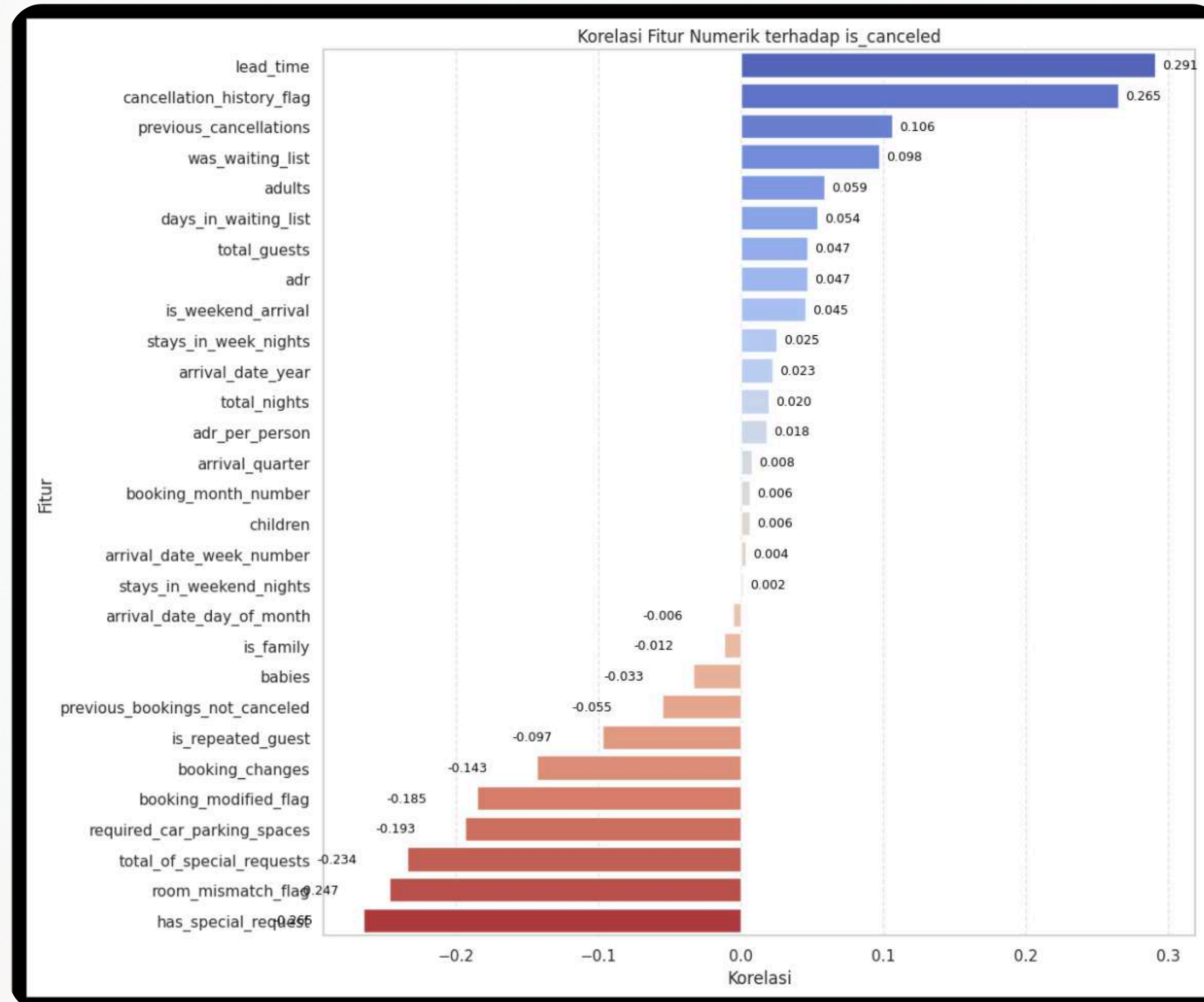


# **PEMBUATAN DAN EVALUASI MODEL**





# Korelasi Fitur Numerik terhadap is\_canceled





# MODEL ML YANG DICOBA

## Random Forest

Gabungan banyak  
pohon keputusan



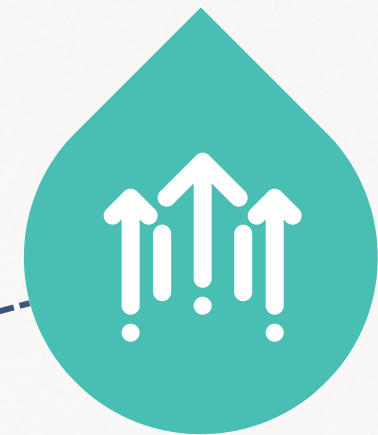
## Decision Tree

Membuat keputusan  
berbasis pohon aturan



## XGBoost

*Model boosting  
yang powerful  
dan akurat .*





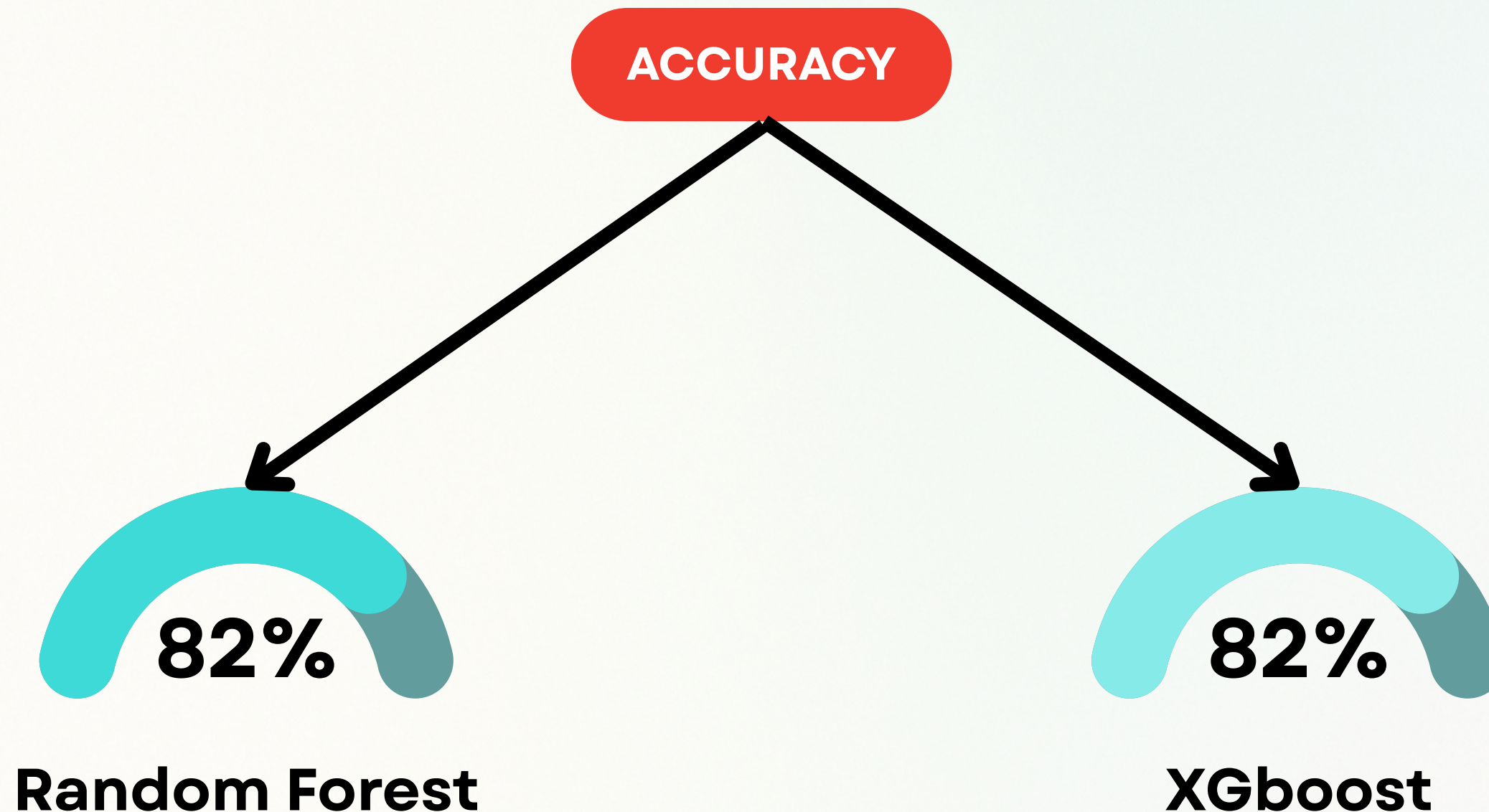
# PERBANDINGAN MODEL DECISION TREE, RANDOM FOREST, DAN XGBOOST

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	72.7%	72.2%	73.7%	72.1%
Random Forest	78.3%	77.4%	79.1%	77.6%
XGBoost	80.3%	84.2%	74.4%	76.1%

- XGBoost unggul dalam akurasi (80.3%) dan precision (84.2%), cocok untuk meminimalkan false positive pembatalan.
- Random Forest memiliki keseimbangan terbaik antara precision (77.4%) dan recall (79.1%), cocok untuk model general-purpose dengan kestabilan tinggi.



# PERBANDINGAN MODEL DECISION TREE, RANDOM FOREST, DAN XGBOOST

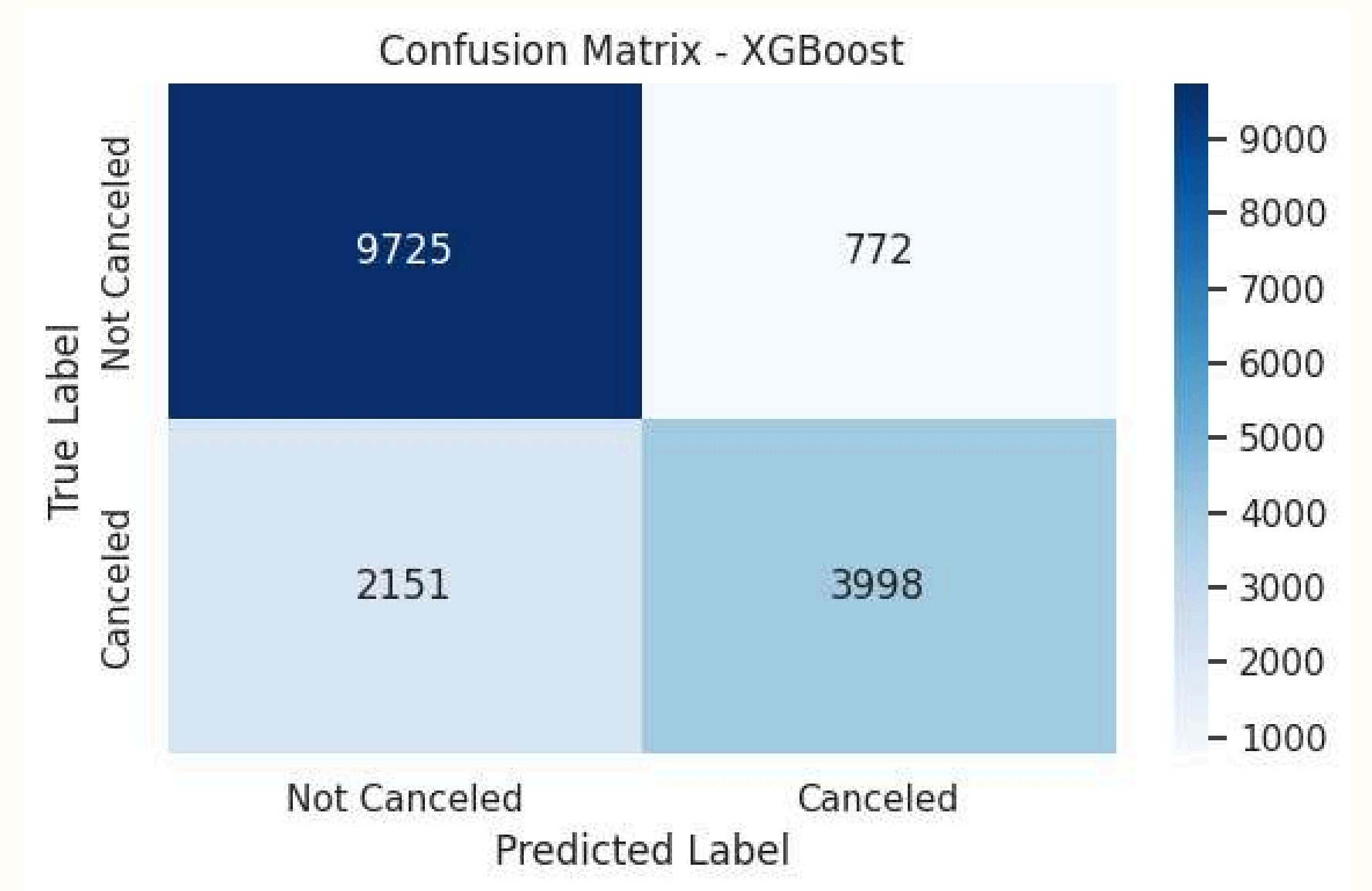
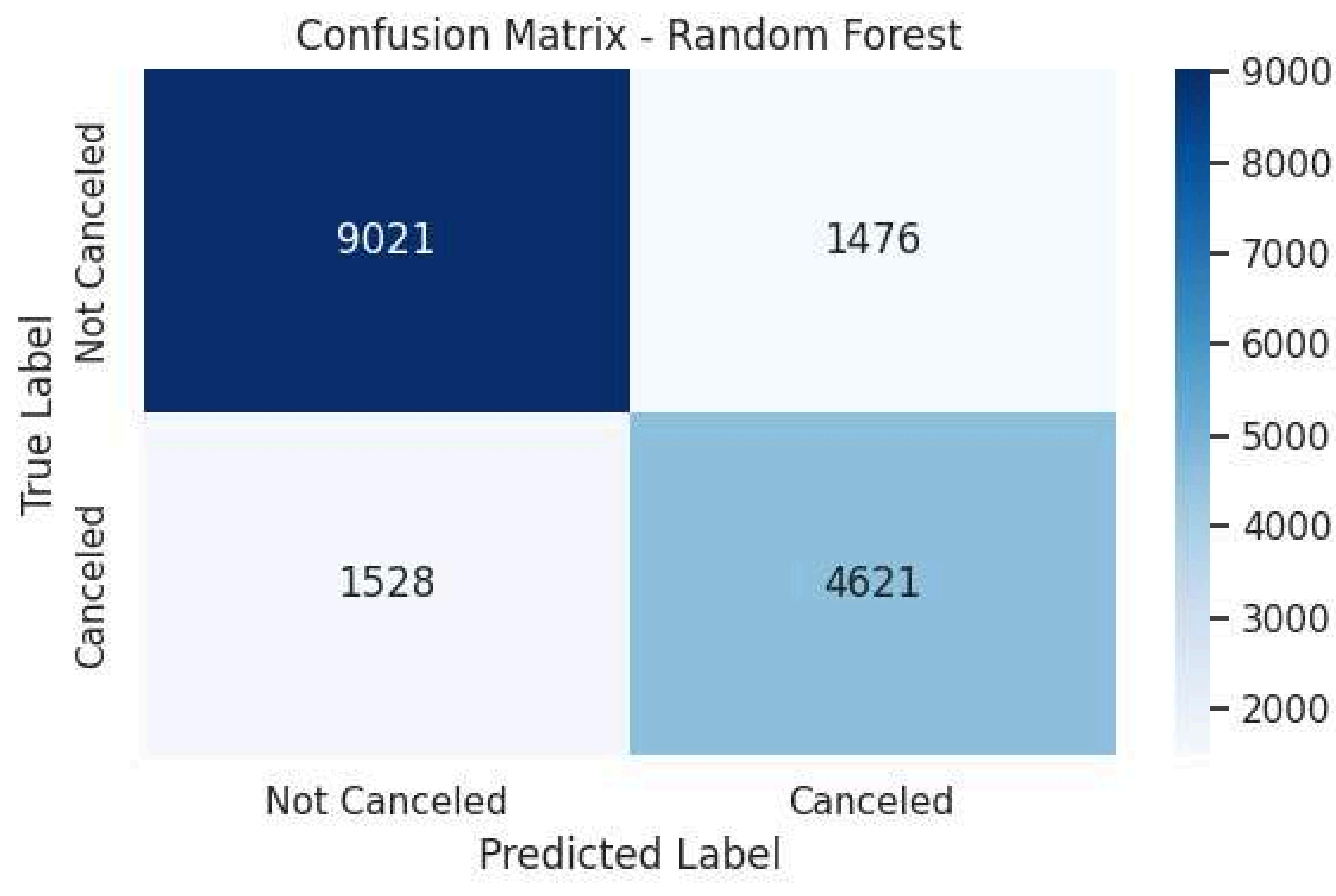


- Gunakan Random Forest jika tujuan utama Anda adalah mendeteksi sebanyak mungkin pembatalan (recall tinggi untuk Canceled), agar strategi pencegahan dapat dilakukan lebih luas.
- Pilih XGBoost jika Anda ingin lebih yakin terhadap setiap prediksi pembatalan (precision tinggi untuk Canceled), sehingga tindakan yang diambil lebih tepat sasaran dan efisien.



# PERBANDINGAN KINERJA MODEL

Metric	Random Forest	XGBoost
Accuracy	82%	82%
Precision	86% (No), 76% (Yes)	82% (No), 84% (Yes)
Recall	86% (No), 75% (Yes)	93% (No), 65% (Yes)
F1-Score	86% (No), 75% (Yes)	87% (No), 73% (Yes)
Macro Avg F1	81%	80%
Support Total	16,646	16,646







# **ANALISIS STATISTIK DAN BUSINESS INSIGHT**



# Data Talks: Fakta Menarik dari Analisis Statistik Booking

1

2

Judul

Insight Utama

Metrik Pendukung

 Lead Time	Semakin lama waktu booking → makin tinggi pembatalan	Mean: 104 vs 47 hari
 Harga & Lama Inap	Harga tinggi & short stay = rawan batal	ADR > 150 = risiko tinggi
 Special Request	Permintaan khusus = sinyal komitmen	1.1 (tidak batal) vs 0.3
 Tipe Deposit	"Not Refund" justru paling banyak batal	99.3% vs 28.3% vs 21.8%
 Tipe Customer	Transient paling sering batal	Transient: 40.7% vs Group: 10.9%

# Insight Booking Hotel: Strategi Mengurangi Pembatalan

## Total Booking

 83.293

On-Time Check-In Rate



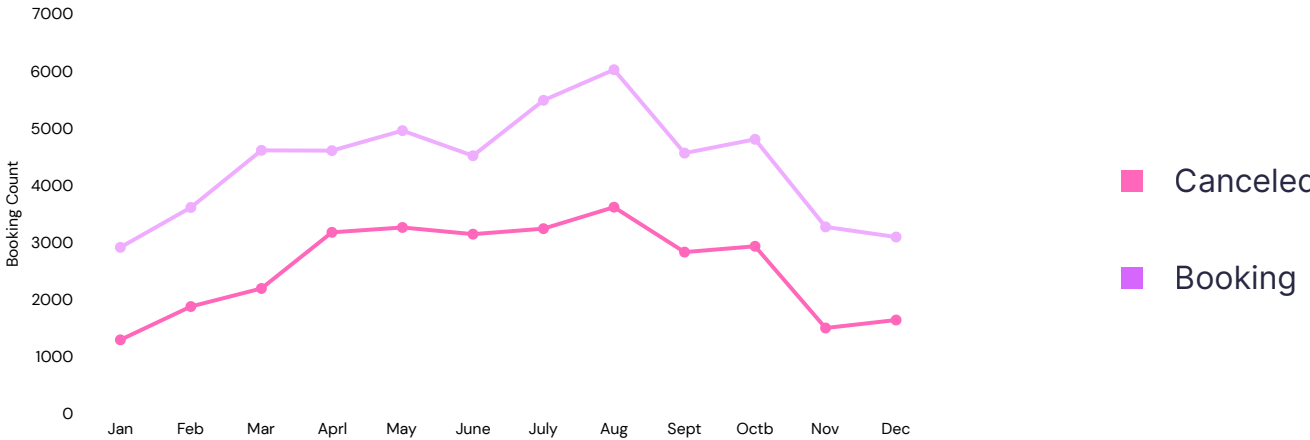
## Total Canceled

 30.777

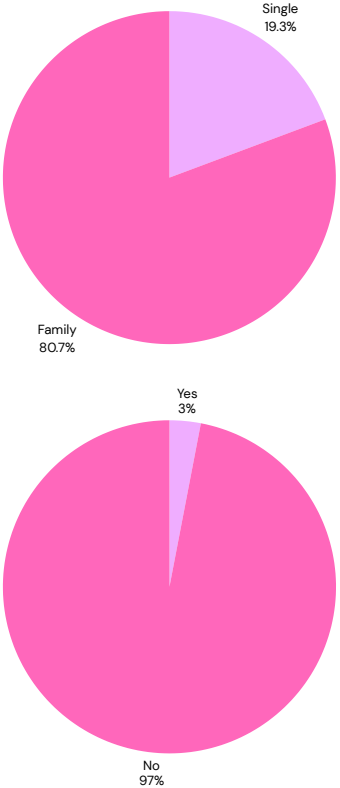
On-Time Delivery Rate



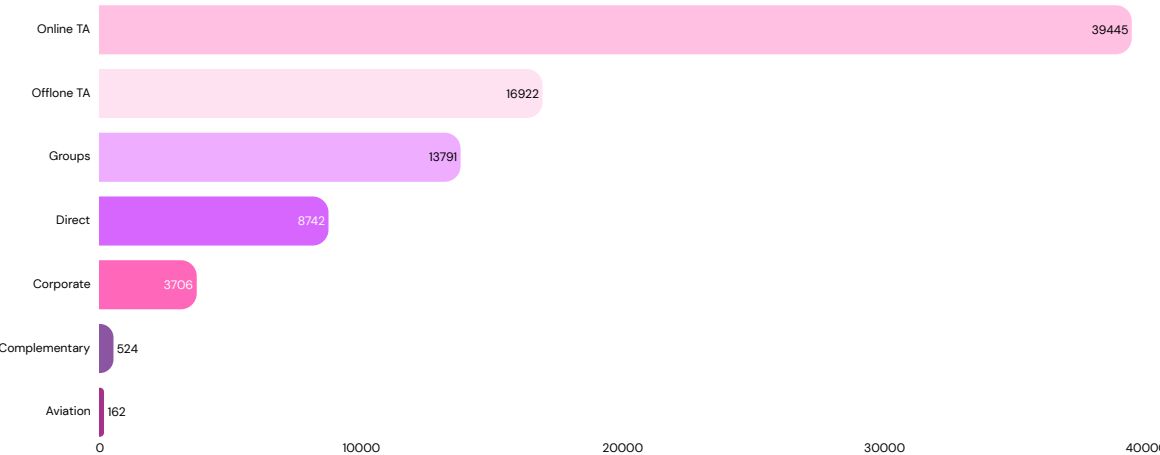
## Tren Booking And Canceled



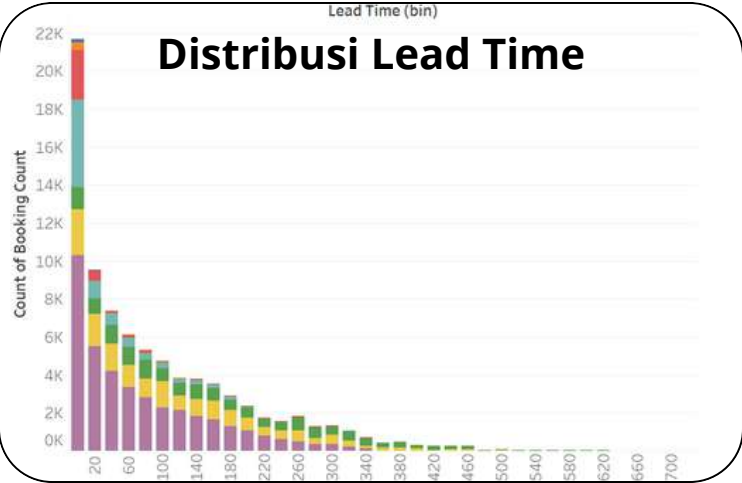
## Distribusi Tamu Keluarga & Tamu Berulang



## Segmentasi Market



## Distribusi Lead Time



### Market Segment

- Aviation
- Complementary
- Corporate
- Direct
- Groups
- Offline TA/TO
- Online TA
- Undefined



# PREDIKSI PEMBATALAN BOOKING: SOLUSI CERDAS UNTUK OPTIMASI OPERASIONAL HOTEL

## 1. 🎯 Use Case Bisnis

Model prediksi membantu menentukan apakah reservasi akan dibatalkan sejak awal, menggunakan data seperti jenis kamar, channel pemesanan, lead time, dan lainnya.

## 2. 🚨 Mitigasi Risiko Pembatalan

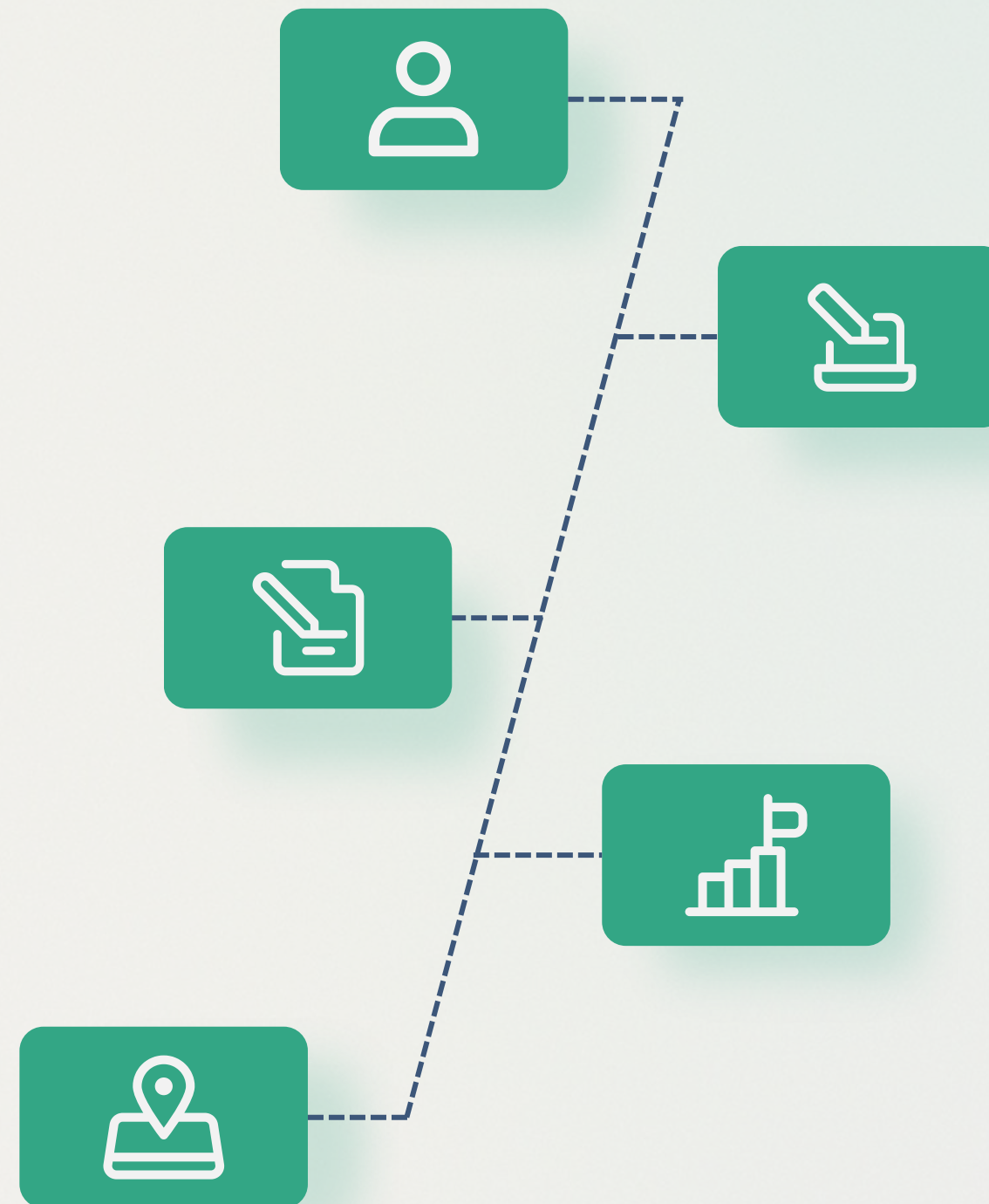
Booking dengan probabilitas pembatalan >70% dapat ditindaklanjuti melalui:

- Email reminder
- Diskon konfirmasi ulang
- Syarat pembayaran di muka / deposit

## 3. 🏨 Optimasi Kapasitas & Sumber Daya

Dengan prediksi akurat, manajemen dapat:

- Menyusun ulang alokasi kamar untuk meminimalkan potensi kamar kosong
- Meningkatkan efisiensi tim housekeeping & operasional



## 4. 📊 Evaluasi Model – Konteks Bisnis

- Random Forest unggul dalam recall kategori Canceled → baik untuk menghindari kerugian karena no-show
- XGBoost memiliki precision tinggi → ideal jika bisnis menghindari false alarm

## 5. Siap Digunakan dalam Operasional

- Akurasi model 82% → cukup andal sebagai sistem pendukung keputusan
- Cocok untuk integrasi sistem CRM, reservasi, dan manajemen risiko



# DARI ANALISIS KE AKSI



# Thank you very much!

PRESENTED DIMAS HARDIANTO

## Contatc Us:



dimashardianto1994@gmail.com



+6285332612529