



PANDAS 08

Hotel Cancellation Booking Prediction

Anggota Kelompok:

1. Abdussalam Darmaatmaja
2. Dimas Jabbar
3. Bijak Ika Handhika
4. Rahmatian Jayanty Sholichah
5. Trully Ananda
6. Nathanael



PANDAS HOTEL

“Sangat penting bagi suatu perusahaan untuk selalu menganalisa performa bisnisnya. Pada kesempatan kali ini, kita akan lebih mendalami bisnis dalam bidang perhotelan. Fokus yang kita tuju adalah untuk mengetahui bagaimana perilaku pelanggan kita dalam melakukan pemesanan hotel, dan hubungannya terhadap tingkat pembatalan pemesanan hotel. Hasil dari insight yang kita temukan akan kita sajikan dalam bentuk data visualisasi agar lebih mudah dipahami dan bersifat lebih persuasif.”

Background

Rata-rata cancellation rate dari suatu bisnis hotel dari 2015 -2017 sebesar 37%



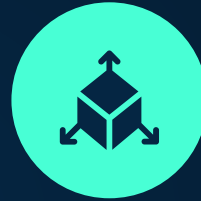
Problem

Cancellation rate yang cukup besar menyebabkan kerugian revenue



Goals

Membuat strategi untuk menanggulangi pembatalan booking hotel.



Business Metric

Persentase Cancellation rate (persentase pelanggan yang membatalkan pemesanan)



Objective

Mengembangkan model yang mampu memprediksi customer yang berpotensi melakukan pembatalan.

Project Flow

Raw Data



Data Preparation



Result



Exploratory
Data Analysis



Machine Learning
Modeling



DATASET

Ada 97.890 baris data yang terdiri dari profil pelanggan dan detail pemesanan.



Customer Identity

- Name
- Email
- Phone Number
- Credit Card



Feature

- Lead Time
- Arrival Date
- Stay Duration
- Deposit Type
- Agent and Company
- Customer Type
- Market Segment
- Reservation Status
- Reservation Status Date
- Waiting List
- Country
- Distribution Channel
- Service Package
- Booking History
- Room Type
- Booking Changes
- Average Daily Rate
- Children



Target

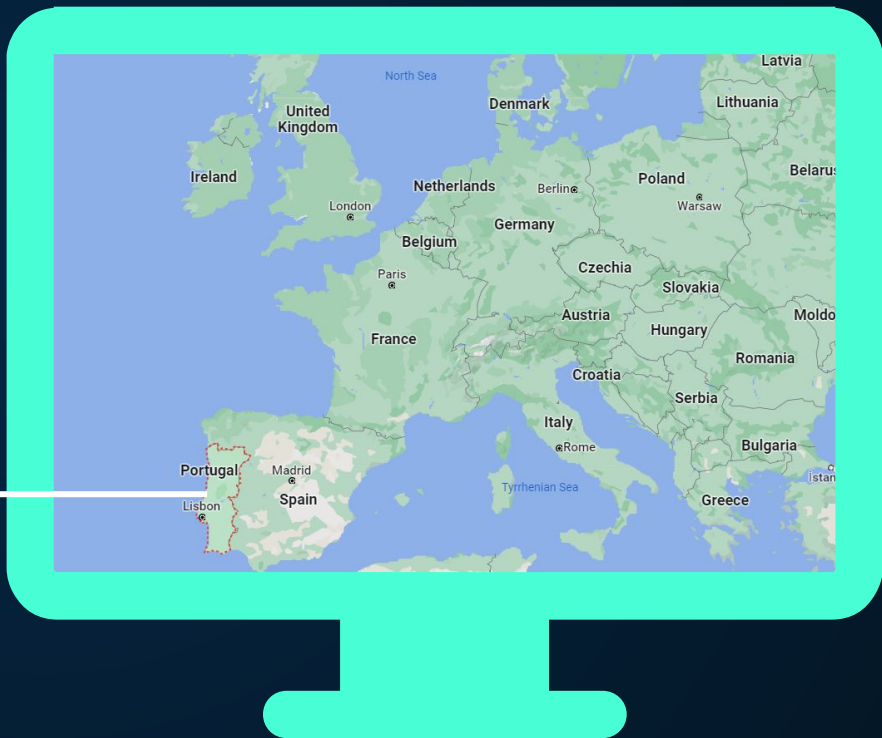
- Canceled
- Not-canceled

INSIGHT

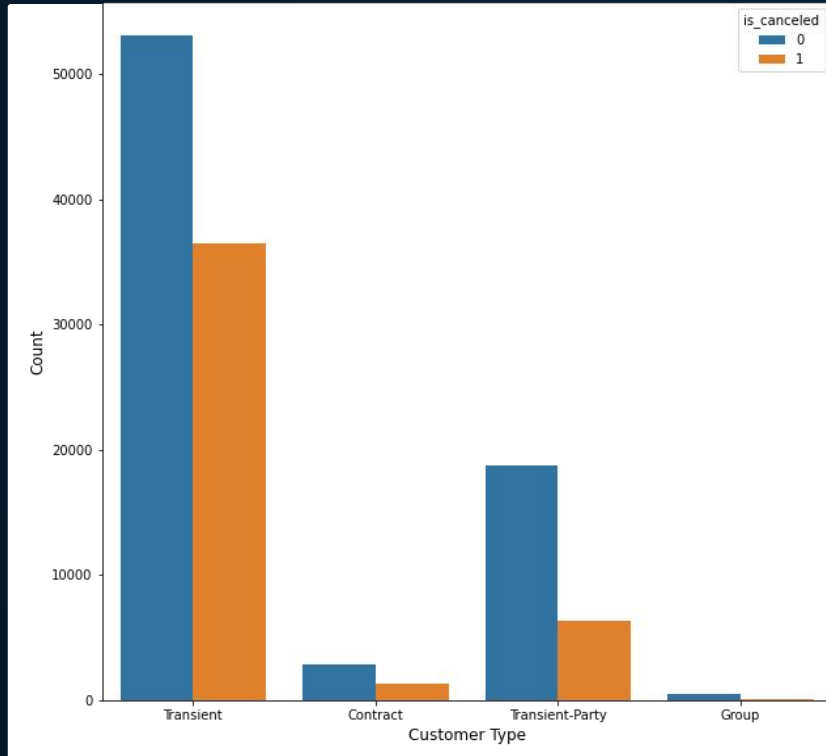


TOP 5 ASAL NEGARA CUSTOMER

Country	Banyaknya Booking
Portugal	48590
UK	12129
France	10415
Spain	8568
Germany	7287

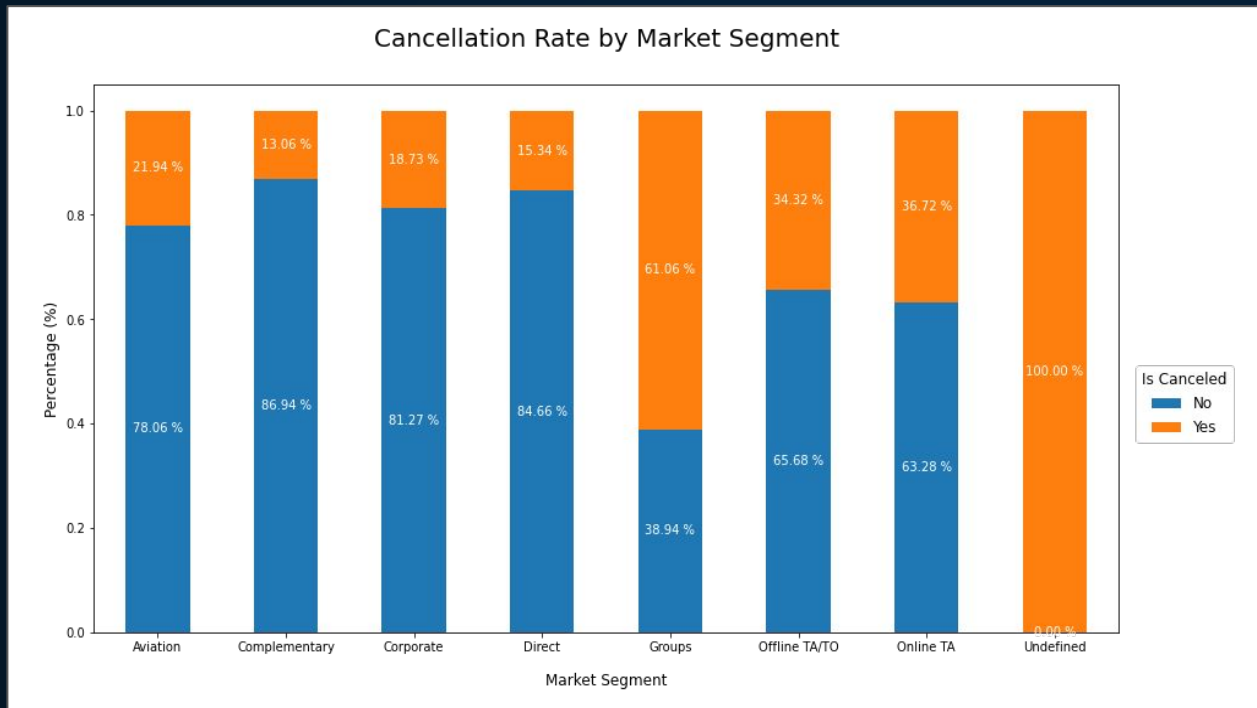


TIPE CUSTOMER & STATUS PEMBATALAN



Customer dengan tipe “Transient” mempunyai jumlah booked terbanyak jika dibandingkan dengan tipe customer lain, namun tipe customer ini juga mempunyai tingkat pembatalan yang paling tinggi juga.

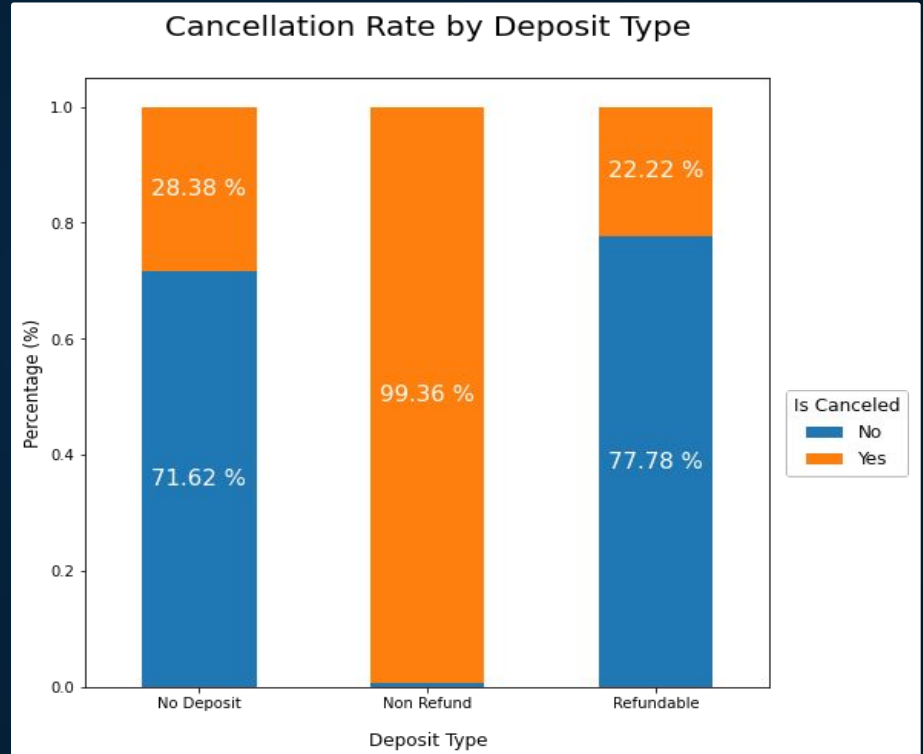
SEGMENT MARKET & STATUS PEMBATALAN



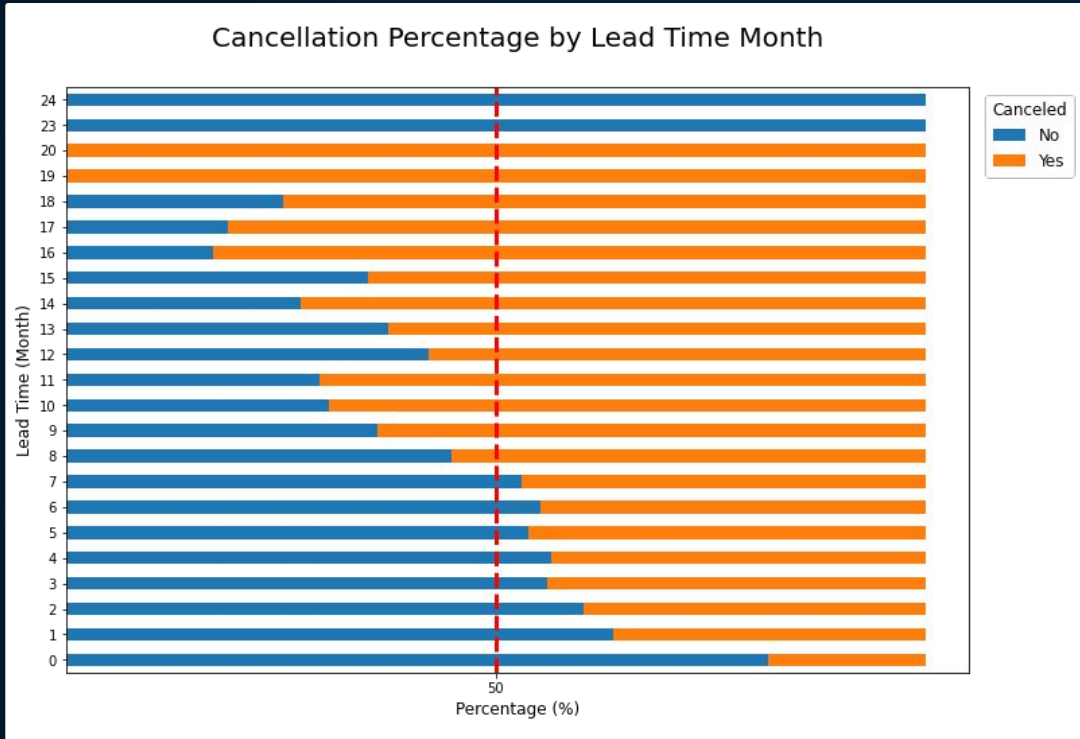
Cancellation tertinggi diperoleh pada segmen market “undefined” dan “Groups”

CANCELLATION RATE BERDASARKAN TIPE DEPOSIT

Customers dengan tipe deposit “Non-refund” mempunyai cancellation rate terbesar jika dibandingkan dengan tipe deposit lainnya.



CANCELLATION RATE BERDASARKAN LEAD TIME (MONTH)



Rasio cancellation rate $> 50\%$ untuk lead time (waktu tunggu antara booking dan check in) yang melebihi 8 bulan.



CLASSIFICATION MODEL

DATA CLEANSING & PRE-PROCESSING

Terdapat 4 data null pada Children,
jadi kami berasumsi bahwa
pelanggan tidak memiliki anak maka
diberikan "0"



Terdapat 468 Data Country
adalah null. kami membuat
kategori baru dengan "others"



Terdapat 112.593 Data Company dan 6.340
Data Agent adalah null. Jadi kami
menghapus data tersebut karena jumlah
tersebut lebih dari 20% total keseluruhan
baris data.



DATA CLEANSING & PRE-PROCESSING

Outlier Handling

Kami melakukan handle outliers dengan metode **Z-score** untuk memastikan semua features stabil untuk pemodelan.

Categoric data type pre-processing

Kolom dengan tipe **categorical** yang bersifat **nominal** akan di-handle dengan **one hot encoding**, sementara yang bersifat **ordinal** akan di-handle dengan **ordinal encoder**.

Numeric data type pre-processing

Kolom dengan tipe **numerik** akan di handle dengan **standard scaler** dan menggunakan **power transformation** dengan metode 'yeo-johnson'.

MODELLING

Dilakukan modelling dengan 3 algoritma

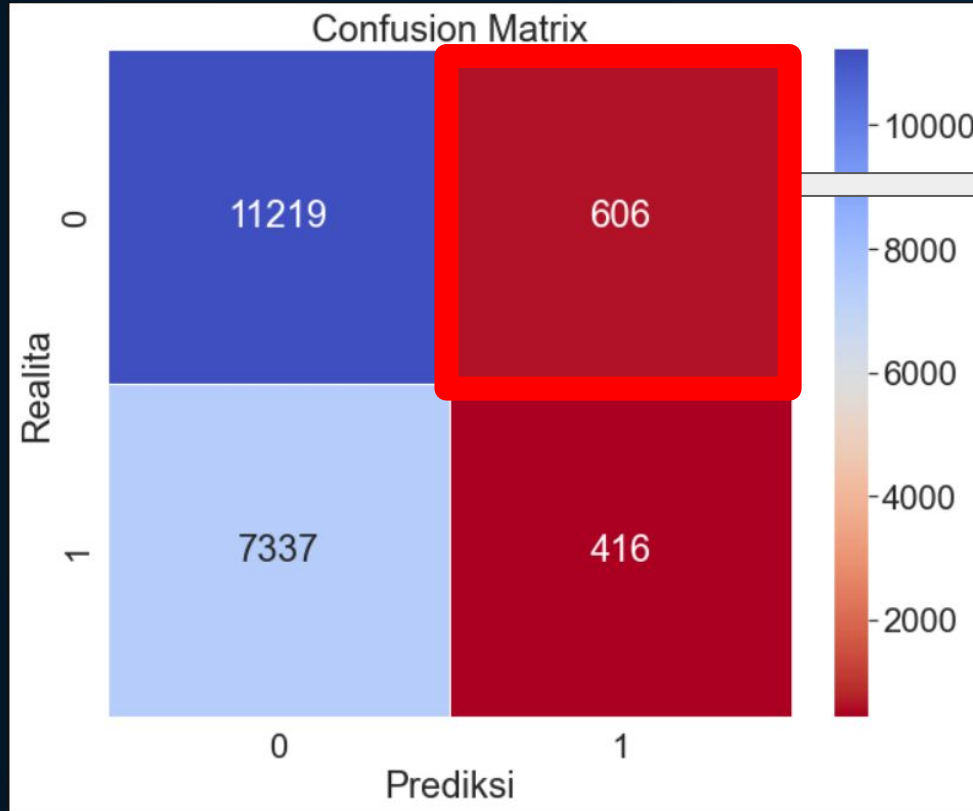


Model	Recall	Precision	F1 Score	AUC_ROC
KNN	0.48	0.62	0.54	0.71
Decision Tree	0.72	0.93	0.81	0.96
Random Forest	0.76	0.89	0.82	0.95

Fokus dari kelompok kami adalah meningkatkan skor **Precision** (Menekan false positive).

Model Decision Tree adalah model yang memberikan hasil terbaik setelah dilakukan pengujian pada 3 model di atas

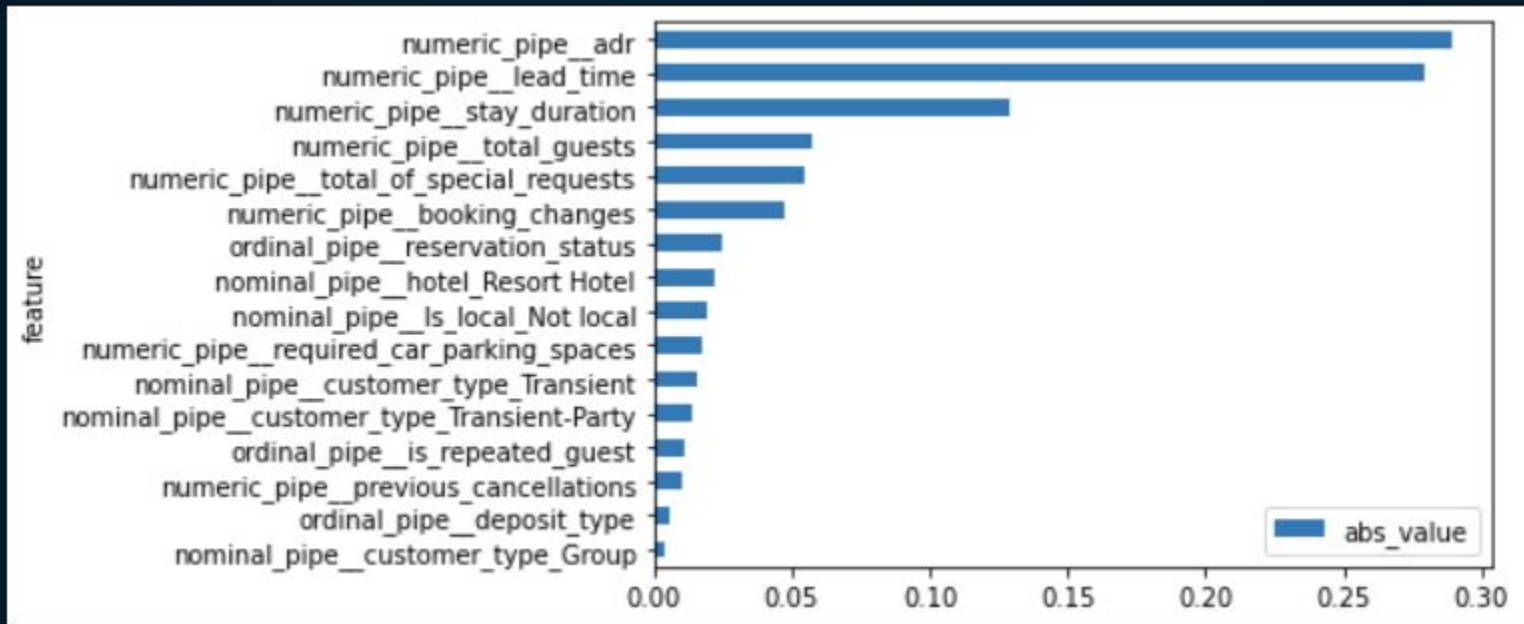
CONFUSION MATRIX



- False Positive harus ditekan untuk menghindari overbooking pada hotel, dimana Jika False Positive besar maka hotel dapat mengalami overbooking.
- Hal ini dikarenakan model memprediksi pelanggan cancel, padahal sebenarnya tidak melakukan cancel booking.



FEATURE IMPORTANCE



CONCLUSION

- Data berpusat di Portugal (kewarganegaraan pelanggan dan juga lokasi hotel).
- Metode **Decision Tree** dipilih untuk model. Ini memiliki kinerja terbaik secara keseluruhan. Dengan pembelajaran machine learning kami dapat memprediksi tingkat pembatalan di hotel dengan Precision sebesar 93%.
- Fitur yang paling berpengaruh dalam cancelation rate pelanggan adalah fitur numerik: lead time, adr (Average Daily Rate), stay duration.
- Tiga fitur ini berkorelasi positif dengan cancelation rate.



BUSINESS RECOMMENDATION



New Hotel Booking Policy

Kebijakan reservasi booking hotel

Kebijakan untuk mem-booking hotel maximal H-7 bulan sebelum check in.

Kebijakan pembatalan booking

Kebijakan mengenai pemberlakuan refund pada pembatalan booking.

Kebijakan penjadwalan ulang booking hotel

Alih-alih membatalkan pemesanan, kita dapat mengusulkan penjadwalan ulang. Ini dapat dilakukan kapan saja sebelum tanggal kedatangan.

SIMULASI KEBIJAKAN HOTEL BOOKING



New Hotel Booking Policy

Room Type	Sebelum	Sesudah	
	Cancellation policy	Cancellation policy	
	Full refund	Full refund	75% refund
Single room	D-day	D-1	D-day
Twin room	D-day	D-2	D-day
King room	D-day	D-7	D-1
Grand room	D-day	D-14	D-2

BUSINESS RECOMMENDATION



Overbooking strategy

Kebijakan penjualan kamar

Menambah jumlah kamar yang akan terjual berdasarkan tingkat pembatalan di bulan berikutnya (prediksi tingkat pembatalan per bulan).

Tindakan kontijensi

Kontijensi: Jika ada pelanggan yang diprediksi akan membatalkan meskipun dia tidak membatalkan.

Bermitra dengan perusahaan hotel lain dengan bintang yang sama dengan lokasi terdekat, untuk merelokasi pelanggan yang tidak mendapatkan kamar karena overbooked.

SIMULASI OVERBOOKING STRATEGY

Asumsi

Terdapat 200 kamar hotel
Prediksi tingkat pembatalan (40%)

Prediksi pembatalan

Prediksi Pembatalan : 40% dari 200 = 80 Kamar
Total Overbook Kamar : 200 kamar + 80 Kamar

Prediksi final pembatalan booking

kita peroleh dari model: Precision 93%
7% dari 80 Kamar = 6 Kamar diasumsikan sebagai kesalahan margin
Prediksi final untuk pembatalan booking : 74 Kamar
Total final overbook kamar: 200 kamar + 74 kamar = 274 kamar





Thank
you