

# HOTEL BOOKING DEMAND CLASSIFICATION

**FINAL PROJECT:**

Purwadhika Digital Technology School

**PRESENTED BY:**

BETA Group

**MEMBER:**

Azhar Muhammad Fikri Fuadi  
Chasanah Nur Ikayanti  
Muhammad Zulfiqar

# The Team



**Azhar Muhammad  
Fikri Fuadi**

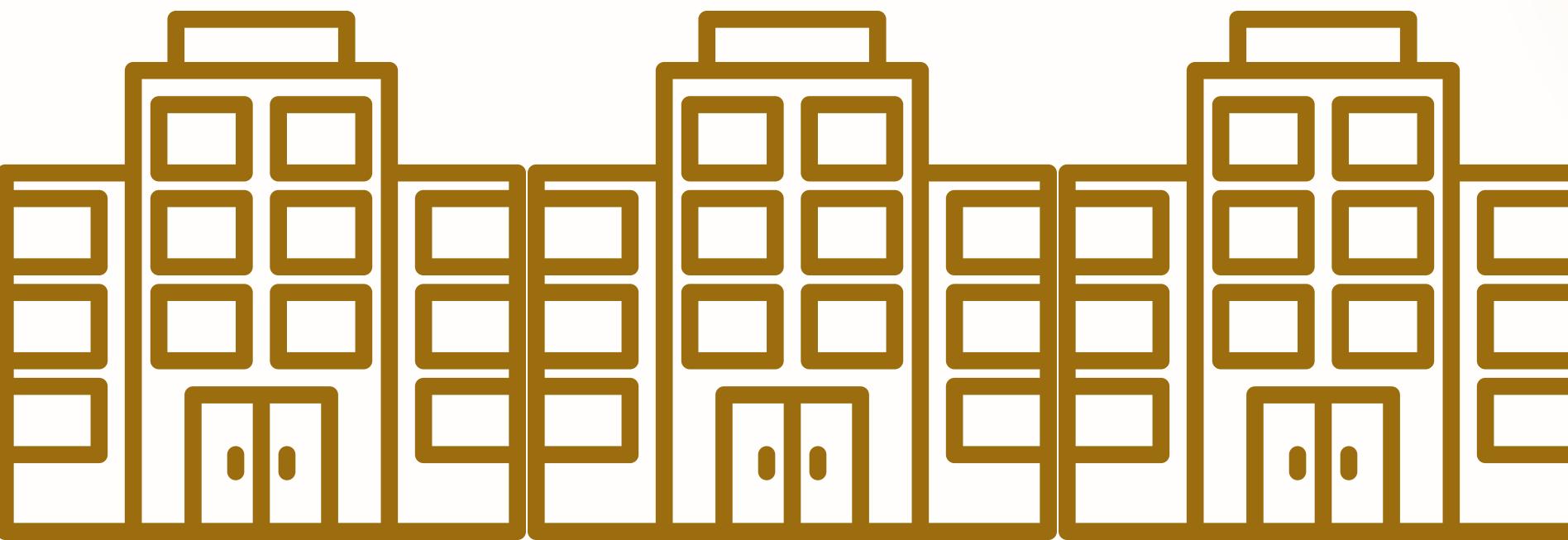


**Chasanah Nur  
Ikayanti**



**Muhammad  
Zulfiqar**

# TABLE OF CONTENTS



1	<b>Business Understanding</b>
2	<b>Data Understanding</b>
3	<b>Exploratory Data Analysis (EDA)</b>
4	<b>Preprocessing</b>
5	<b>Methodology (Analysis / Modeling)</b>
6	<b>Conclusion and Recommendation</b>

# STUDY CASE

DATA SOURCE:



Data in Brief  
Volume 22, February 2019, Pages 41-49



Data Article

## Hotel booking demand datasets

Nuno Antonio<sup>a b</sup>   , Ana de Almeida<sup>a c d</sup>, Luis Nunes<sup>a b d</sup>

### City Hotel di Lisbon

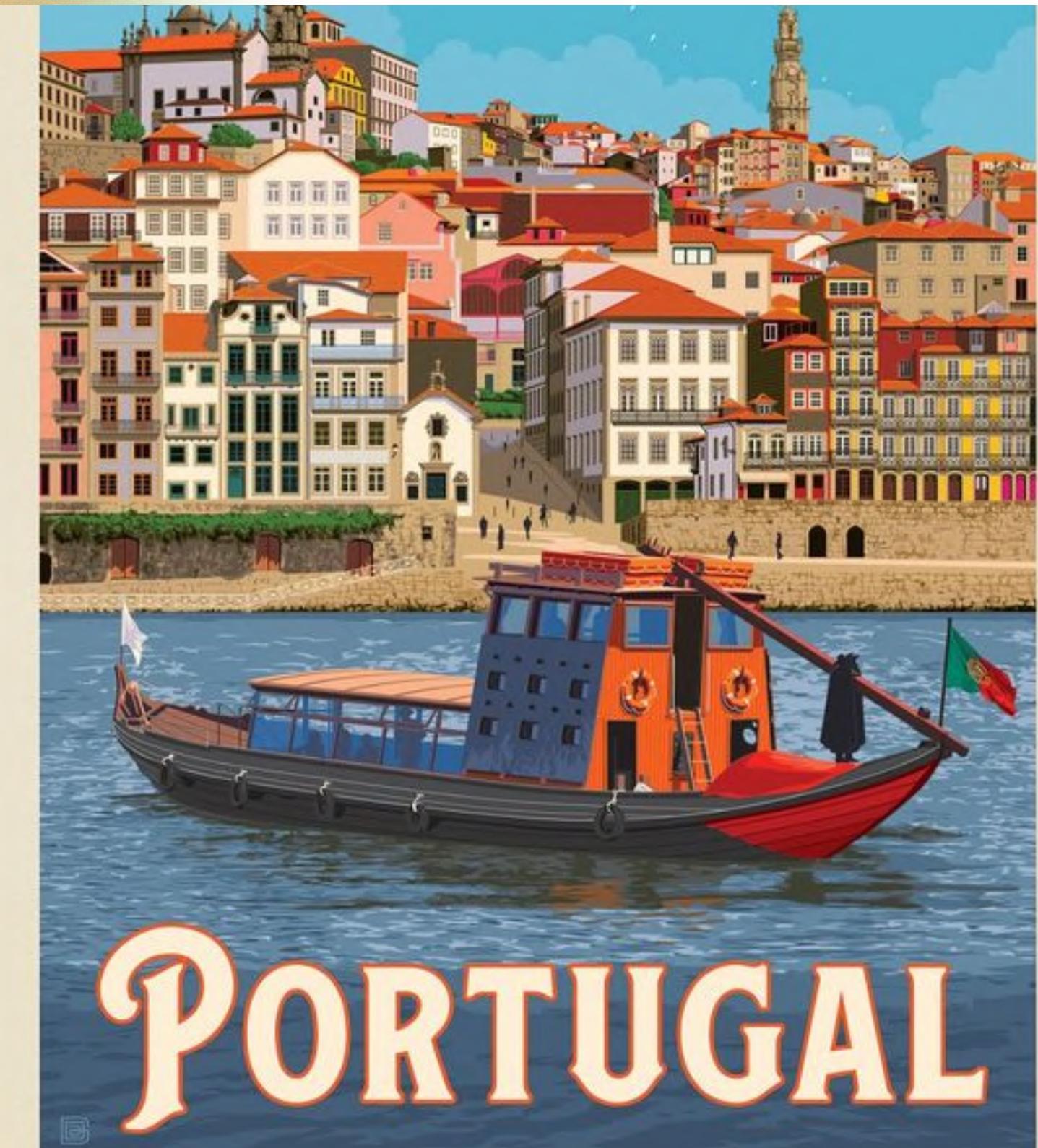
Memiliki 200 kamar:

- 40% tipe kamar standard
- 30% tipe kamar deluxe
- 20% tipe kamar premium
- 10% tipe kamar family

### Resort Hotel di Algarve

Memiliki 200 kamar:

- 40% tipe kamar standard
- 30% tipe kamar deluxe
- 20% tipe kamar premium
- 10% tipe kamar family



# STUDY CASE



Objek wisata populer di Lisbon



Kastel São Jorge



Marquis of Pombal Square



Elevator Santa Justa



Objek wisata populer di Algarve



Alun-alun Lama Albufeira



Marina Vilamoura



Marina Lagos

Lisbon

Algarve

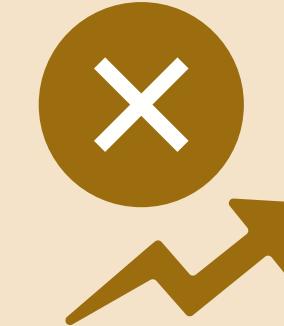
# BUSINESS UNDERSTANDING



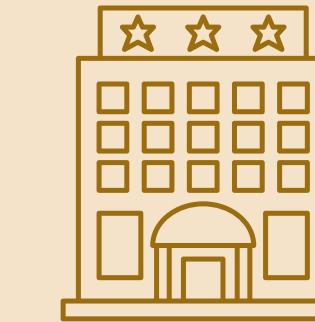
LisGarve City & Resort Hotel adalah sebuah penginapan yang berada di Portugal. Saat ini mengalami penurunan pendapatan karena banyak mengalami pembatalan pemesanan yang menyebabkan kerugian bagi pihak hotel. Mereka ingin mengetahui tindakan yang harus diambil oleh pihak manajemen hotel untuk mengurangi peluang terjadinya pembatalan pemesanan dengan mengidentifikasi karakteristik yang dimiliki oleh hotel.



**Revenue Hotel  
Menurun**



**Tingkat Cancellation  
Tinggi**



**Karakteristik Hotel**

- Jenis Hotel apa? (**What**)
- Kapan tingkat pembatalan tinggi? (**When**)
- Negara asal customer? (**Where**)
- Bagaimana pelayanan hotel? (**How**)

# PROBLEM

Hotel mengalami penurunan pendapatan akibat pembatalan pemesanan. Sehingga butuh informasi calon konsumen yang akan melakukan pembatalan

# GOALS

Meningkatkan revenue hotel dengan cara mengetahui karakteristik konsumen melakukan cancel dan memprediksi pemesan yang akan membatalkan pemesanan hotel untuk mengetahui pelanggan yang melakukan pembatalan pemesanan

# HOW?

- Meprediksi peluang terjadinya pembatalan pemesanan
- Mengidentifikasi faktor-faktor customers yang melakukan pembatalan
- Menerapkan strategi khusus untuk customer yang teridentifikasi melakukan cancel
- Meningkatkan retensi pelanggan

# ANALYTIC APPROACH

Analisis eksploratif data untuk memahami tren dan pola perilaku pelanggan. Selanjutnya, membangun model klasifikasi menggunakan algoritma machine learning. Melibatkan analisis interpretasi model untuk mengidentifikasi fitur-fitur kunci yang mempengaruhi keputusan pelanggan

## ASSUMPTION

Menurut Terminology Glossary Industri Perhotelan, standar harga breakfast di hotel biasanya berkisar antara 10% hingga 25% dari harga kamar per malamnya.

Perbandingan biaya Retensi Pelanggan dengan biaya kehilangan apabila customer melakukan cancel:

- Harga rata-rata kamar hotel : 91.85 Euro
- Total Biaya Marketing (Double Booking, Gaji Karyawan, Budget Campaign) :  $11\% \times 91.85 \text{ Euro} = 10 \text{ Euro}$

## METRIC EVALUATION

### Positif Class Interested (Recall, Presisi, F1)

membuat model evaluation yang tertarik kepada kelas Positif yaitu (Recall, Presisi, F1) namun dilihat secara sederhana diatas recall memiliki dampak yang cukup besar karena biaya yang ditanggung pihak hotel akan lebih besar. Namun akan dicek lebih mendalam ketiganya untuk ditemukan metric yang terbaik



# DATA UNDERSTANDING

## NUMERICAL FEATURES

Attribute	Data Type	Description
lead_time	int	Jarak hari antara tanggal booking hotel yang telah dimasukkan ke dalam Hotel Property Management System (PMS) hingga tanggal yang dijadwalkan untuk check-in.
arrival_date_year	int	Tahun dari tanggal yang dijadwalkan untuk check-in.
arrival_date_week_number	int	Urutan minggu dalam setahun dari tanggal yang dijadwalkan untuk check-in.
arrival_date_day_of_month	int	Tanggal yang dijadwalkan untuk check-in.
stays_in_wekend_nights	int	Jumlah malam di akhir pekan (Sabtu atau Minggu) yang dipesan oleh pelanggan untuk menginap di hotel.
stay_in_week_nights	int	Jumlah malam dalam seminggu (Senin – Jumat) yang dipesan oleh pelanggan untuk menginap di hotel.
adults	int	Jumlah orang dewasa.
children	float	Jumlah anak-anak.
babies	int	Jumlah bayi.
previous_cancellations	int	Jumlah booking hotel sebelumnya yang dibatalkan oleh pemesan sebelum booking hotel terkini.
preious_bookings_not_canceled	int	Jumlah booking hotel sebelumnya yang tidak dibatalkan oleh pemesan sebelum booking hotel terkini.
booking_changes	int	Jumlah perubahan yang dilakukan pada booking hotel.
day_in_waiting_list	int	Jumlah hari untuk setiap booking hotel yang masuk ke dalam waiting list sebelum dikonfirmasi kepada pemesan.
adr	float	Rata-rata tarif harian atau Average Daily Rate yang ditentukan untuk membagi jumlah semua transaksi dengan jumlah malam yang dipesan.
required_car_parking_spaces	int	Jumlah ruang parkir mobil yang dibutuhkan oleh pemesan.
total_of_special_requests	int	Jumlah permintaan khusus yang dibuat oleh pemesan.



# DATA UNDERSTANDING

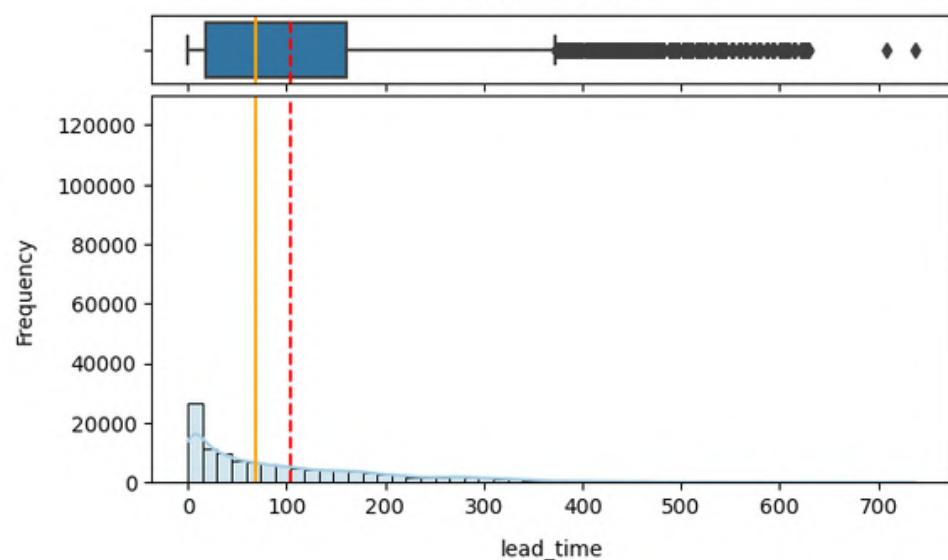
## CATEGORICAL FEATURES



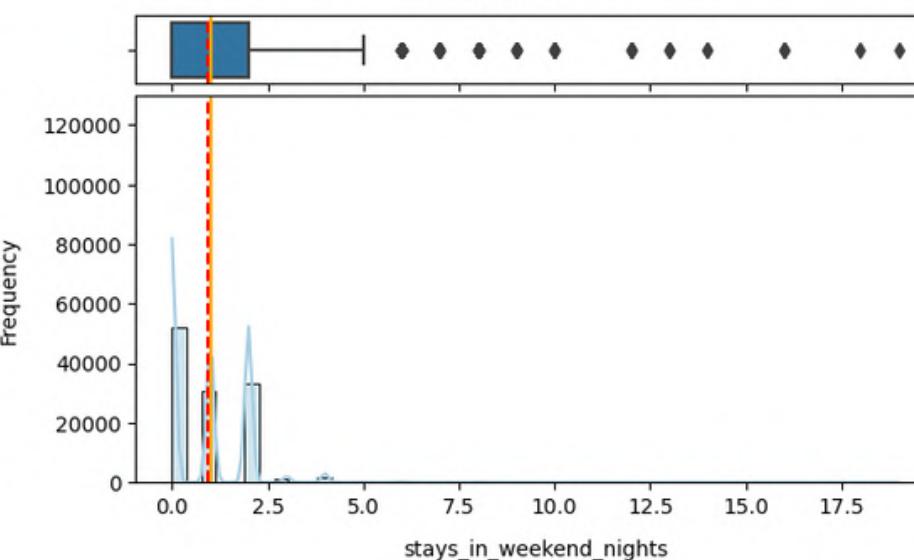
Attribute	Data Type	Description
is_canceled	int	Nilai yang menunjukkan apakah booking hotel dibatalkan atau tidak dibatalkan
hotel	str	Tipe hotel yang disediakan.
arrival_date_month	str	Bulan dari tanggal yang dijadwalkan untuk check-in.
meal	str	Jenis makanan yang dipesan.
country	str	Negara asal pemesan.
market_segment	str	Pemilihan untuk segmentasi pasar.
distribution_channel	str	Saluran distribusi untuk setiap booking hotel.
is_repeated_guest	int	Nilai yang menunjukkan apakah nama pemesan adalah pemesan berulang atau bukan
reserved_room_type	str	Kode untuk tipe kamar yang dipesan.
assigned_room_type	str	Kode untuk tipe kamar yang ditetapkan untuk pemesanan.
deposit_type	str	Indikasi apakah pemesan melakukan deposit untuk menjamin booking hotel.
agent	float	ID dari travel agency yang melakukan booking hotel.
company	float	ID dari perusahaan yang melakukan booking hotel atau yang bertanggung jawab untuk membayar booking-nya.
customer_type	str	Jenis booking hotel.
reservation_status	str	Status terakhir booking hotel.
reservation_status_date	str	Tanggal saat status terakhir ditetapkan.

# EDA - NUMERICAL VARIABLE

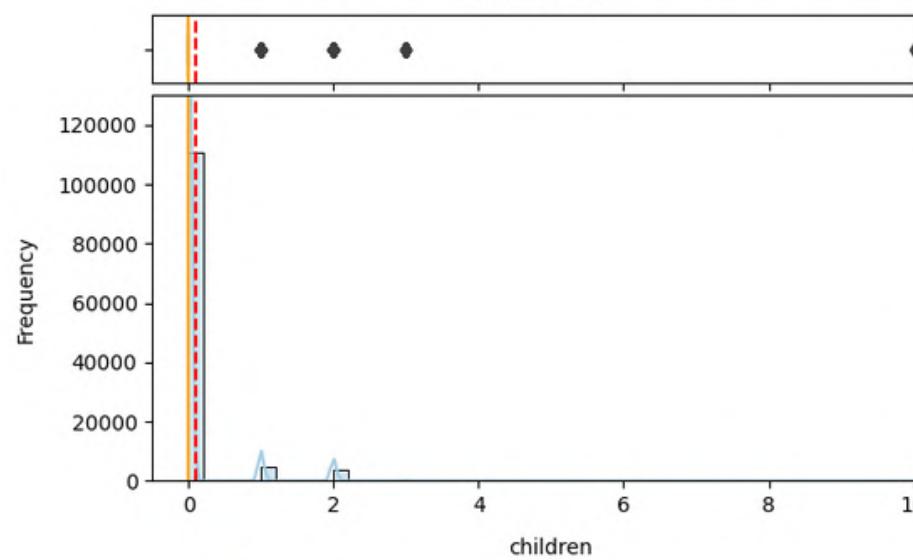
**Distribution of lead\_time**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 69.0



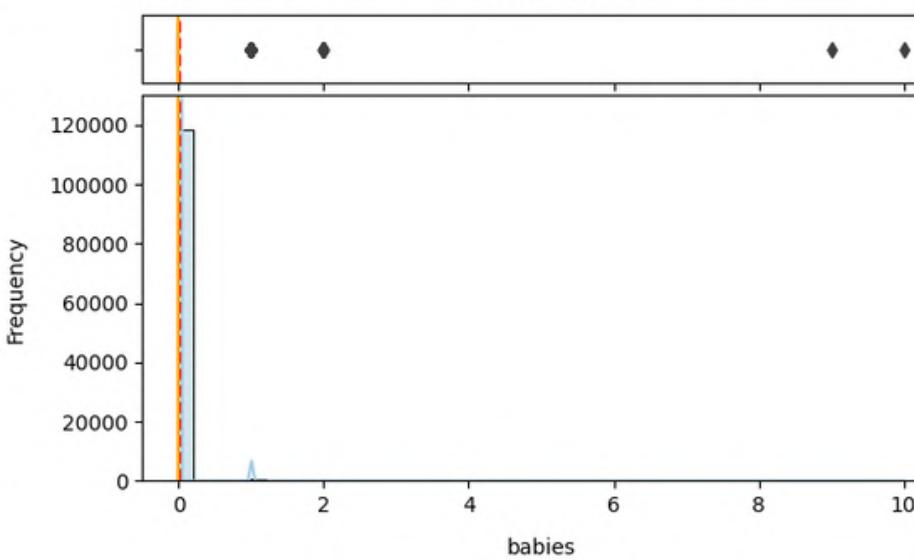
**Distribution of stays\_in\_weekend\_nights**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 1.0



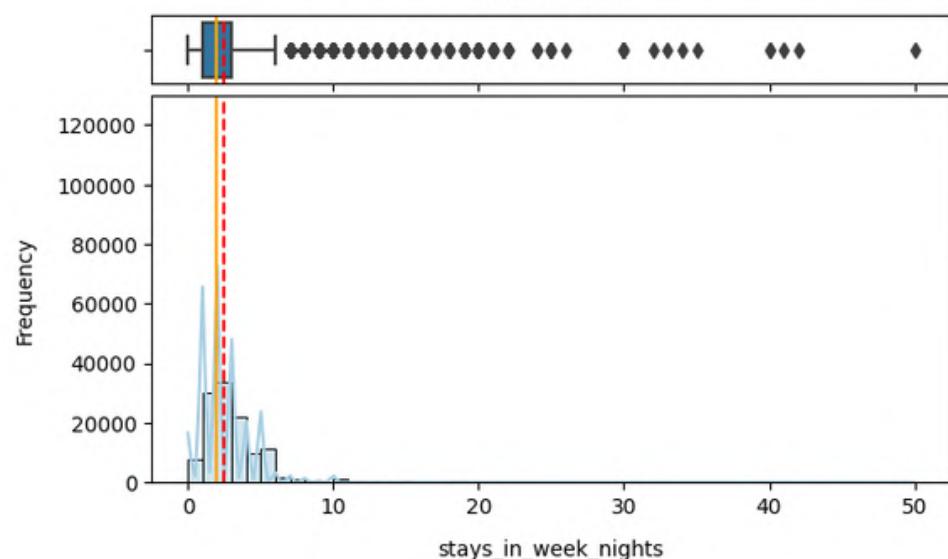
**Distribution of children**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 0.0



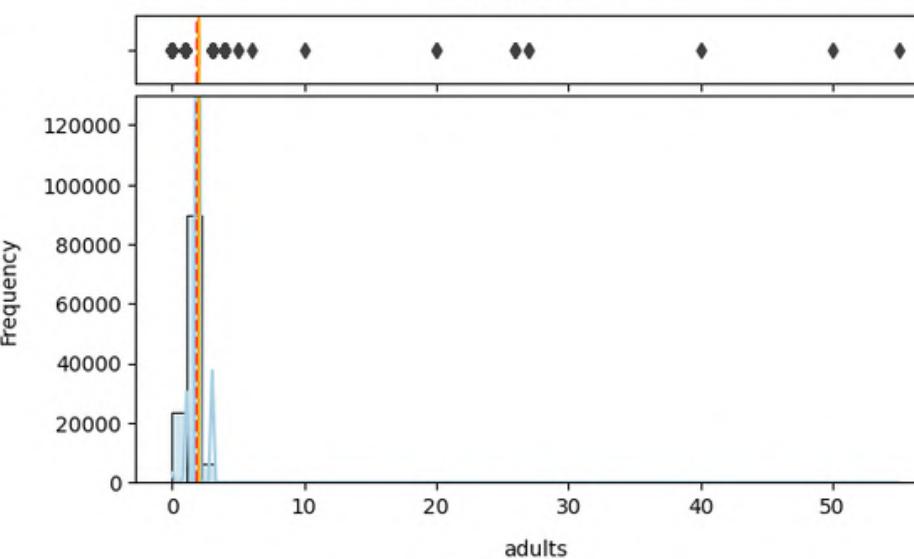
**Distribution of babies**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 0.0



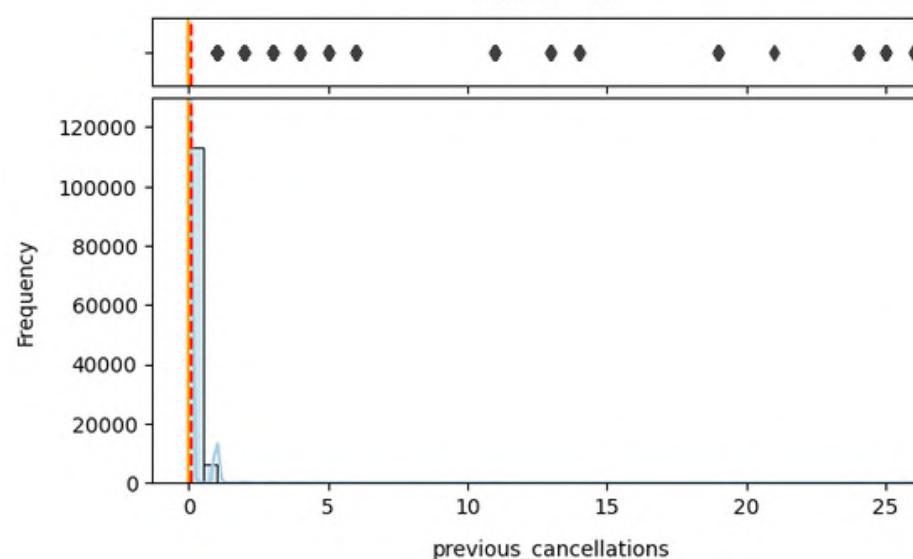
**Distribution of stays\_in\_week\_nights**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 2.0



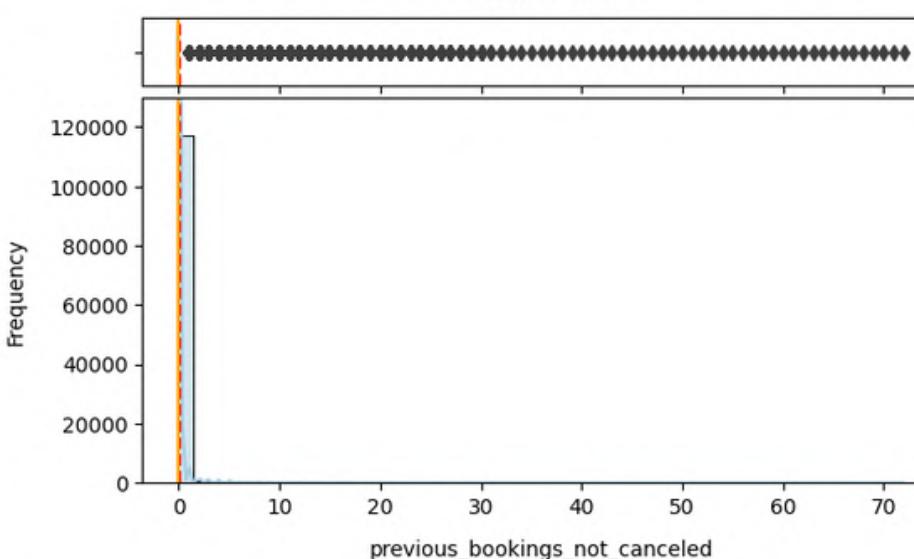
**Distribution of adults**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 2.0



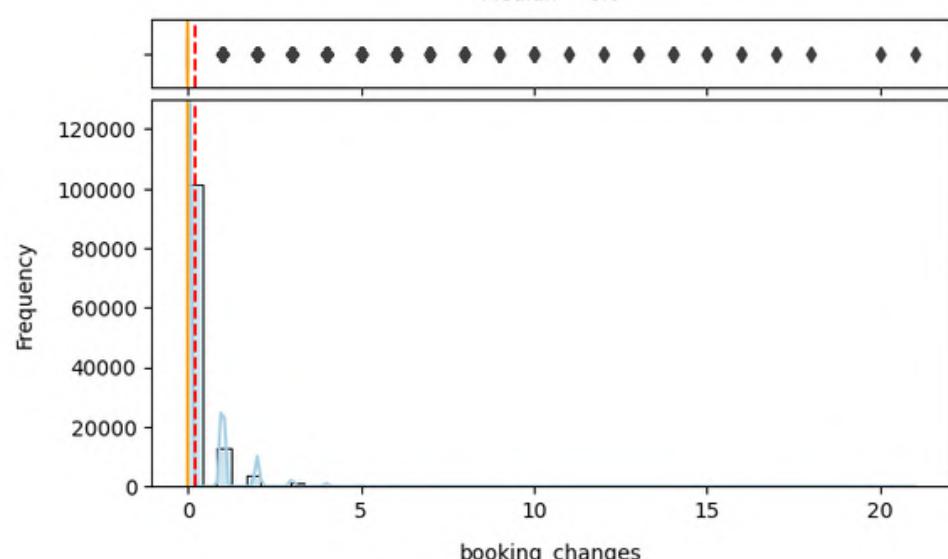
**Distribution of previous\_cancellations**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 0.0



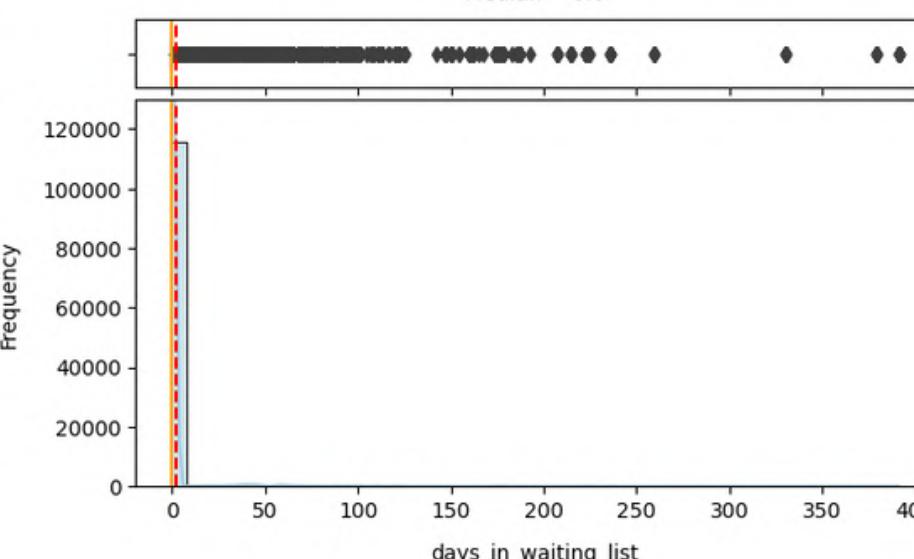
**Distribution of previous\_bookings\_notCanceled**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 0.0



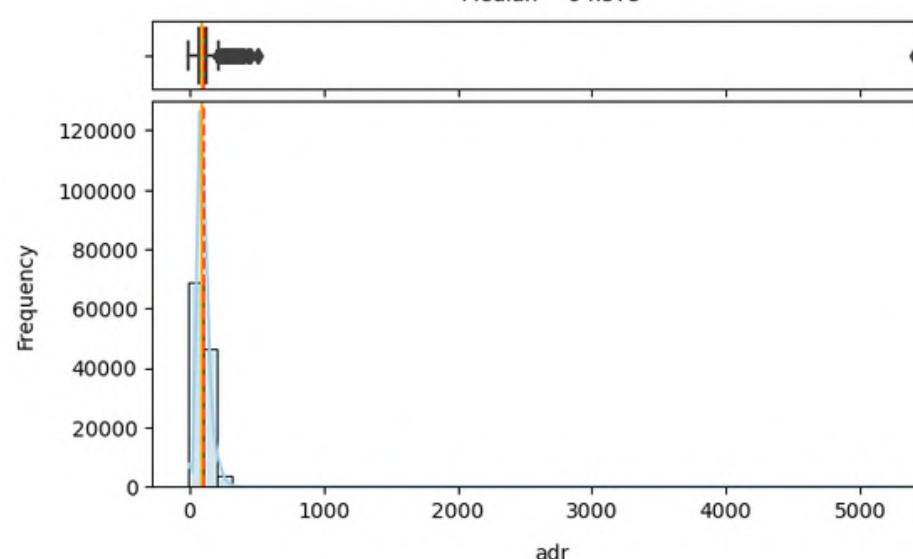
**Distribution of booking\_changes**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 0.0



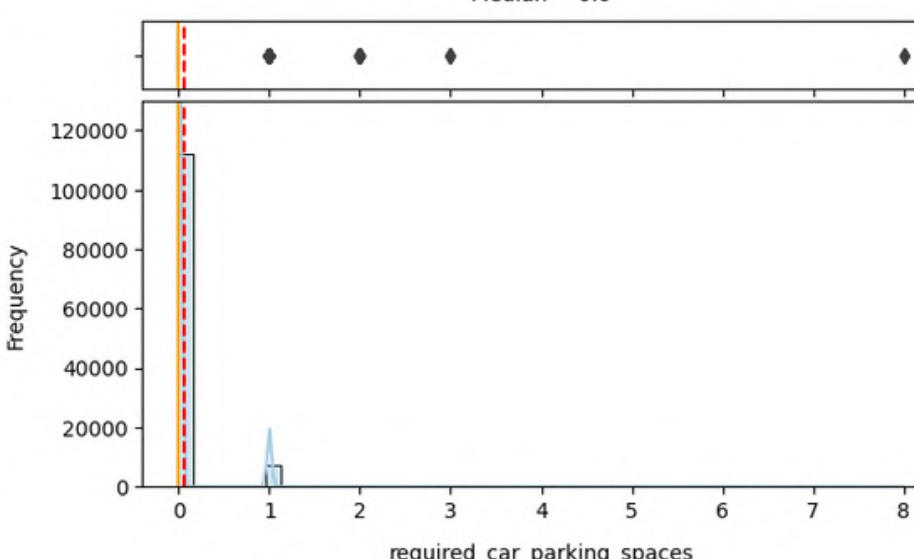
**Distribution of days\_in\_waiting\_list**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 0.0



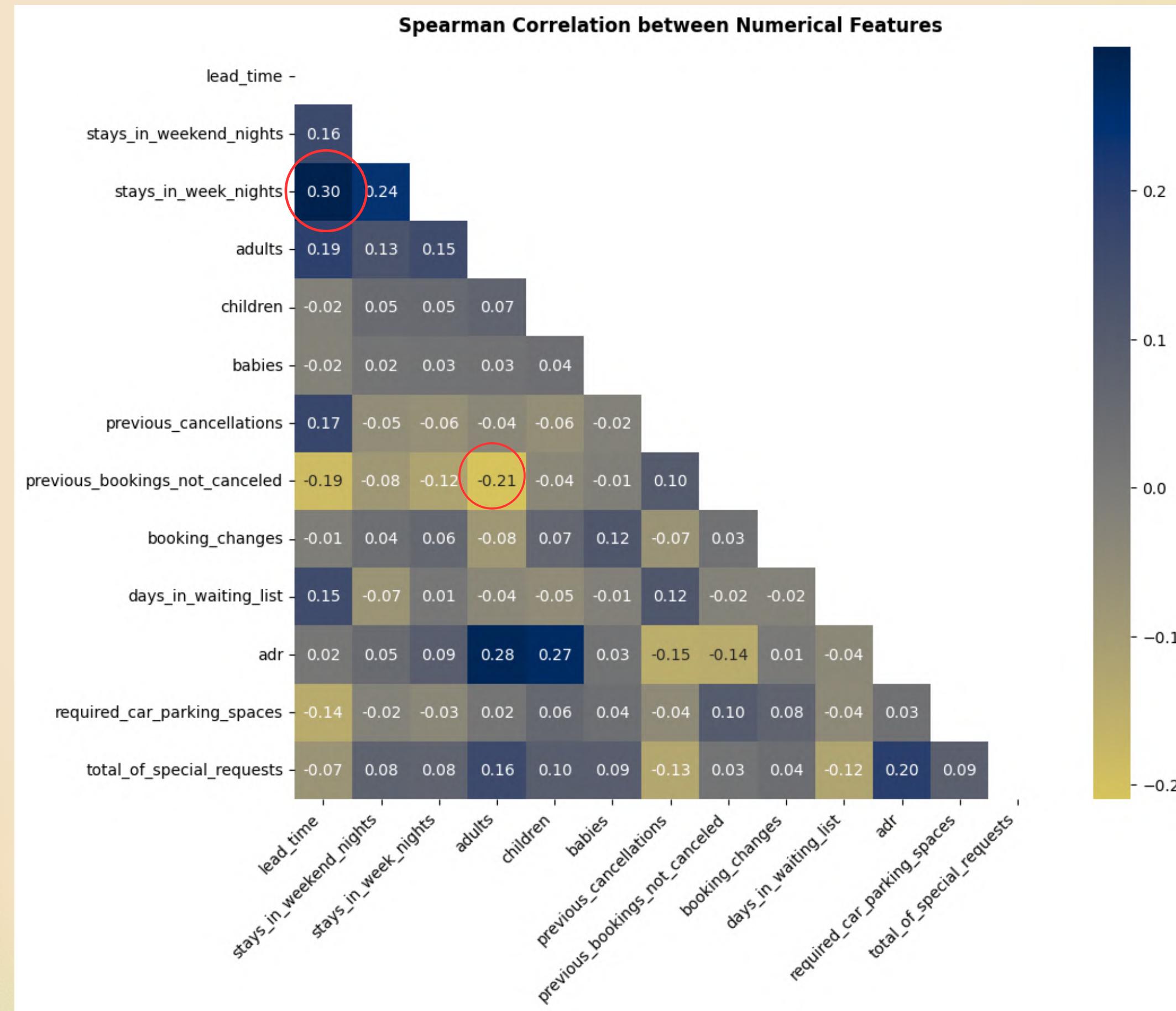
**Distribution of adr**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 94.575



**Distribution of required\_car\_parking\_spaces**  
pvalue = 0.00000, data TIDAK terdistribusi normal  
Median = 0.0

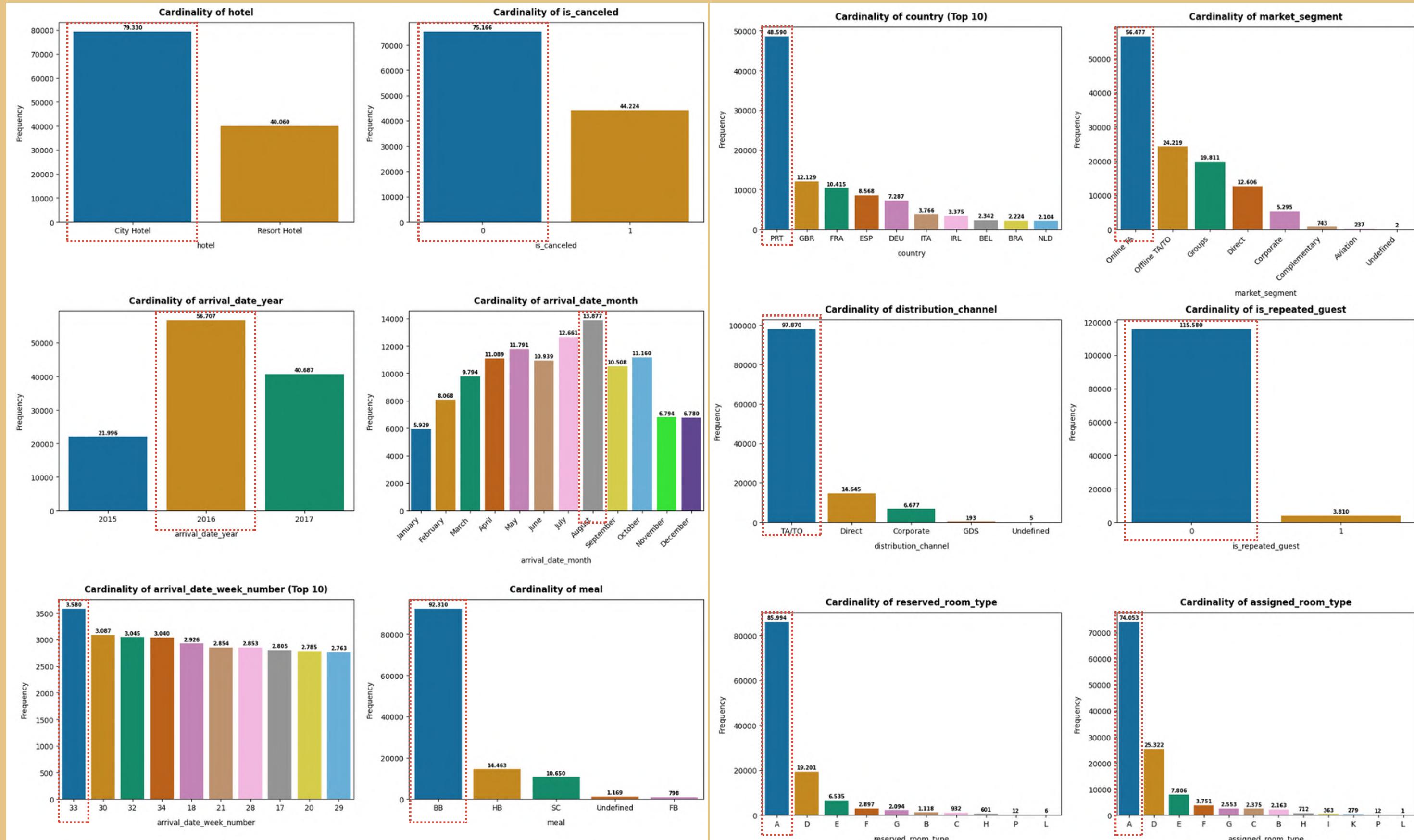


# EDA - NUMERICAL VARIABLE

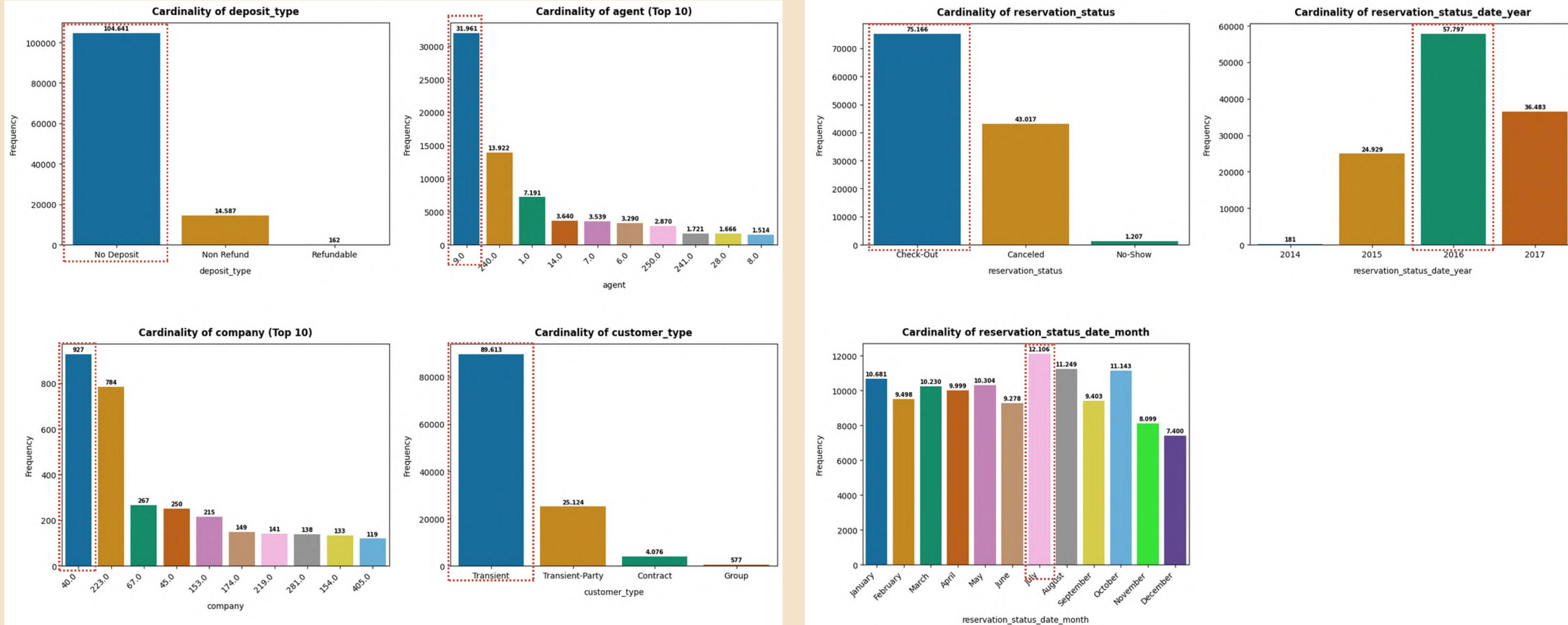




# EDA - CATEGORICAL VARIABLE

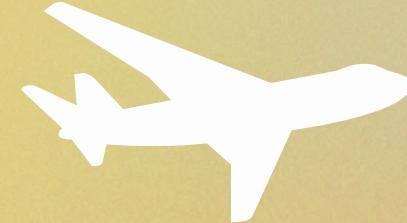


# EDA - CATEGORICAL VARIABLE





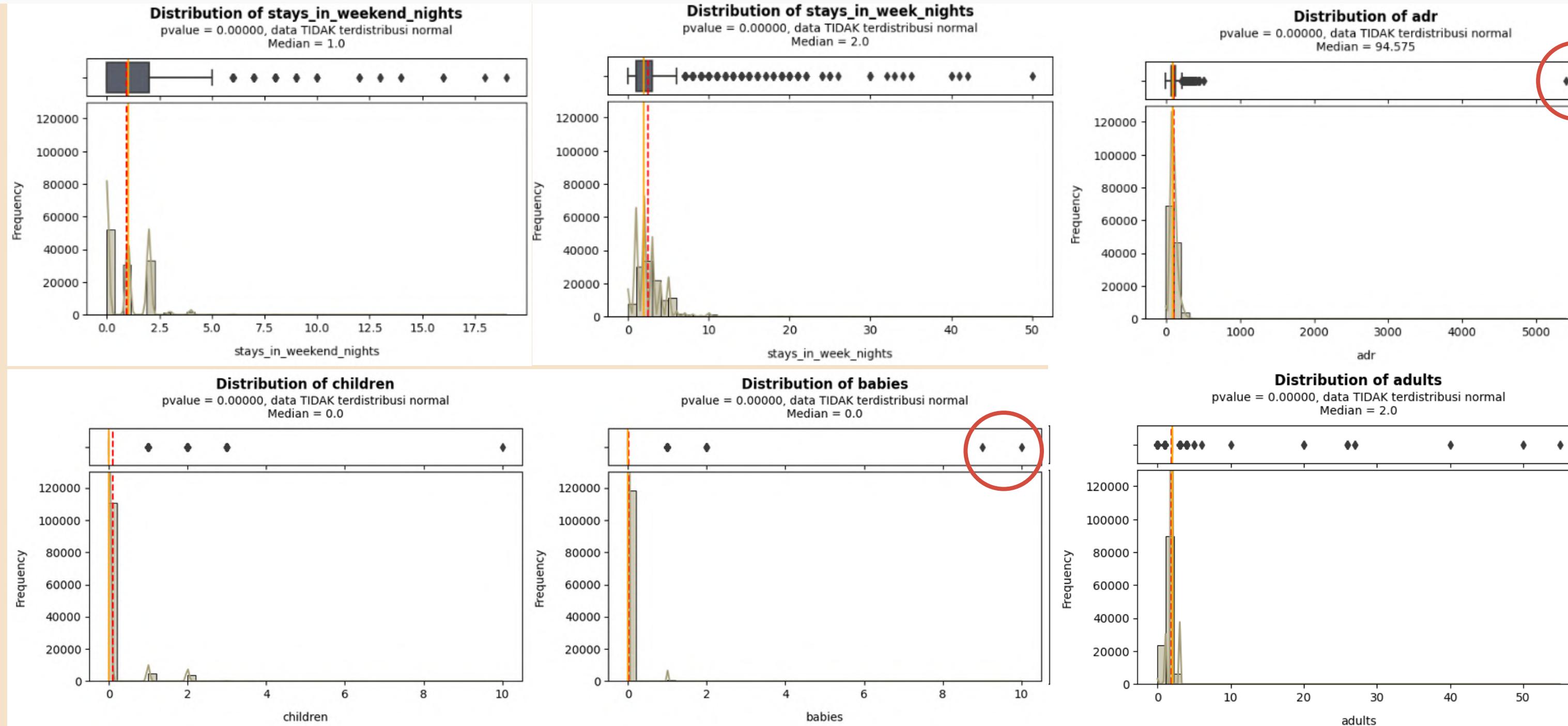
# PREPROCESSING



 <b>DATA DUPLICATED</b>		 <b>MISSING VALUE</b>	
Data sebelumnya: 119,390 Data sekarang: 87,396		<ul style="list-style-type: none"><li>• Agent (16340 Data) --&gt; Unknown</li><li>• Company (112593 Data) --&gt; Unknown</li><li>• Children (4 data) --&gt; 0</li><li>• Country (488 Data) --&gt; Unknown</li></ul>	
 <b>DATA OUTLIER &amp; ANOMALIES</b>		 <b>FEATURE ENGINEERING</b>	
Terdapat data outlier, namun tidak dilakukan penghapusan data. Handling data anomali dengan domain knowledge		<ul style="list-style-type: none"><li>• Analysis</li><li>• Modeling</li></ul>	



# HANDLING ANOMALIES



- **Weekend & Week Night** bernilai 0, hal ini tidak relevan karena minimal pemesanan 1 hari
- **Adr** memiliki nilai 5400 Euro, handlingnya diganti dengan 143.1. Pedekatan mengisi nilai dengan nilai berdasarkan data dengan karakteristik yang sama
- **Adr** memiliki nilai 0, namun tidak termasuk market segment complementary
- **Children**, Babies, dan Adults bernilai 0, hal tidak relevan karena minimal jumlah pemesanan bernilai 1
- **Babies** bernilai 9 dan 10, hal ini tidak relevan karena jumlah bayi dan jumlah orang dewasa (2) dan type kamar City Hotel



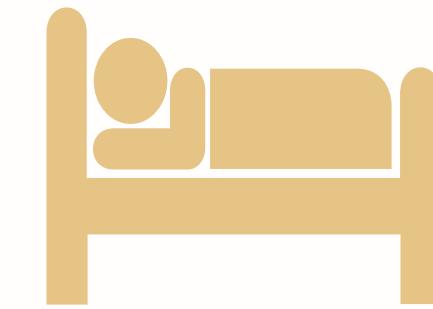
# DATA ANALYSIS



# Google

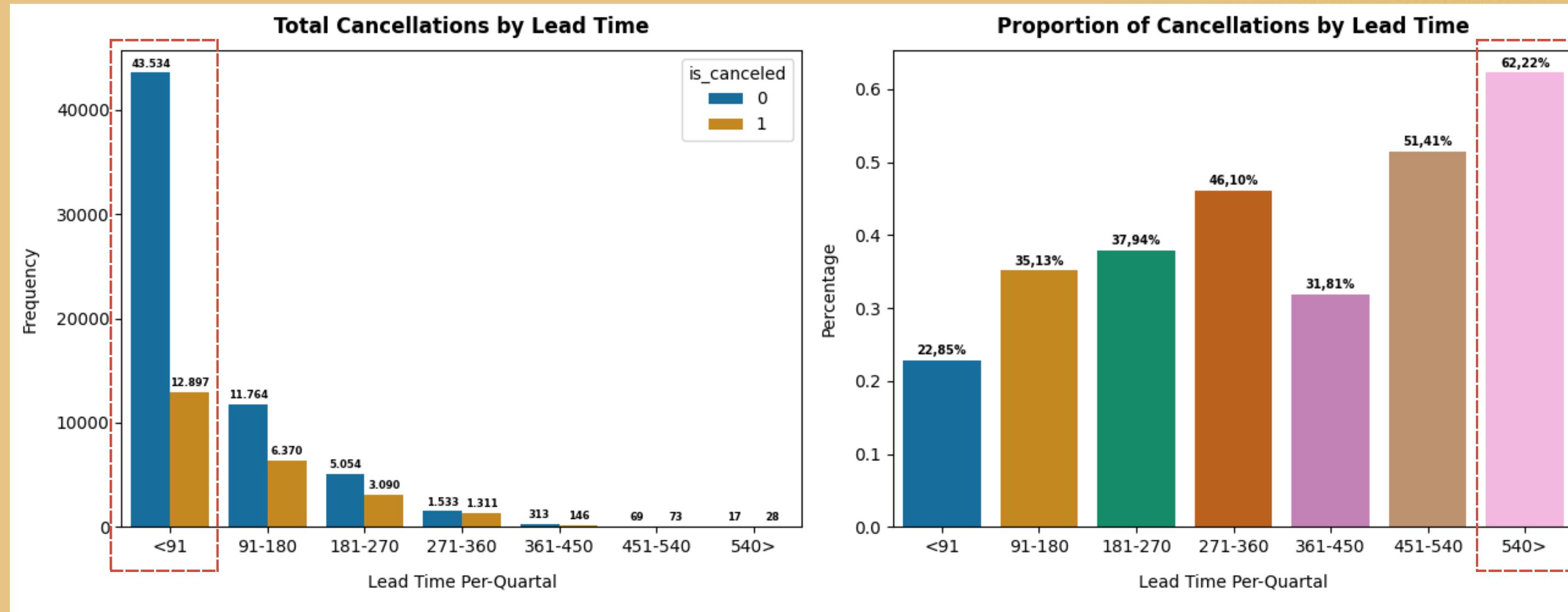


Pengaruh Waktu Terhadap Pembatalan Booking Hotel





# Karakteristik jarak hari antara tanggal pemesanan hotel hingga tanggal yang dijadwalkan untuk check-in hotel?

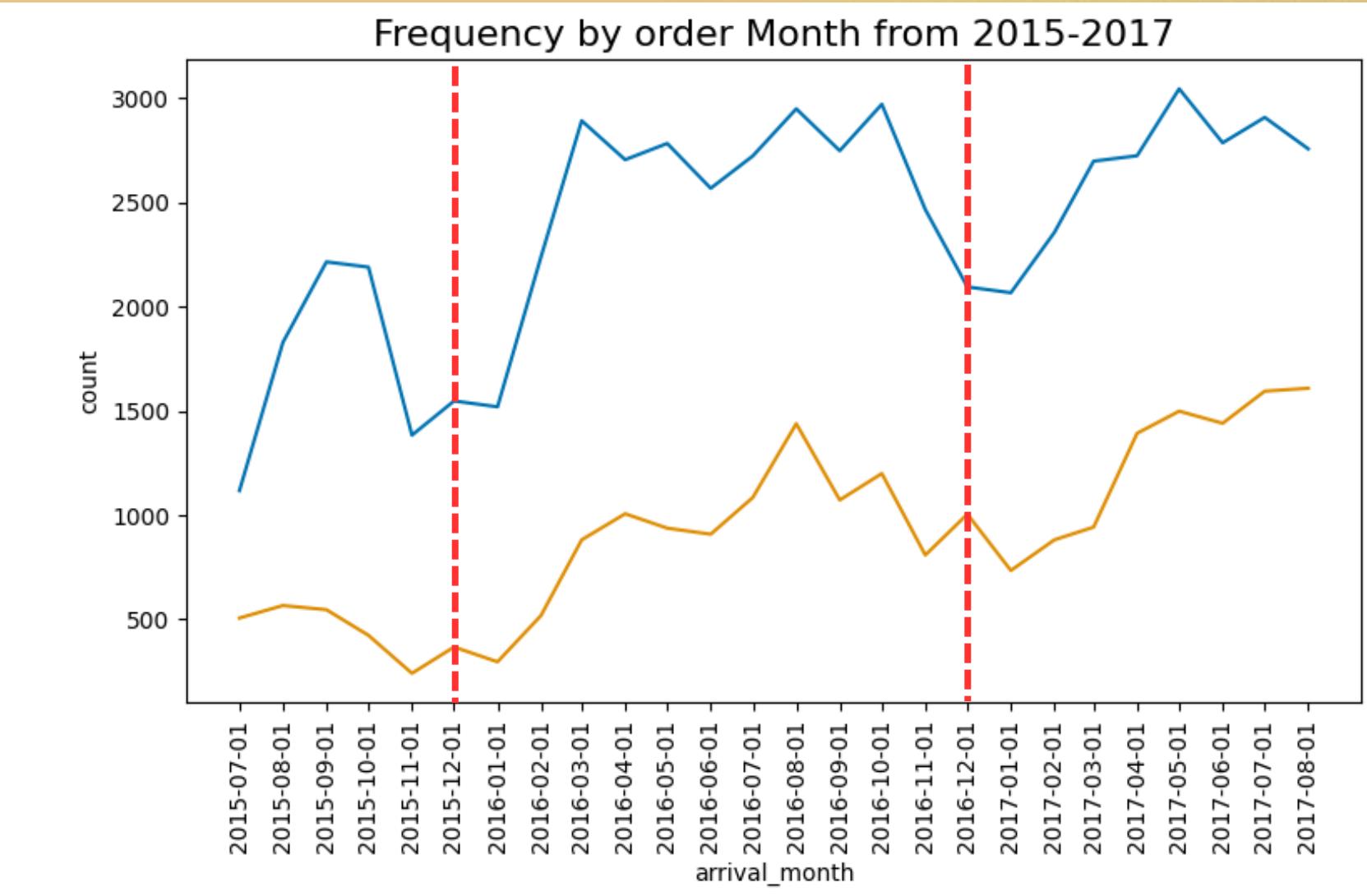
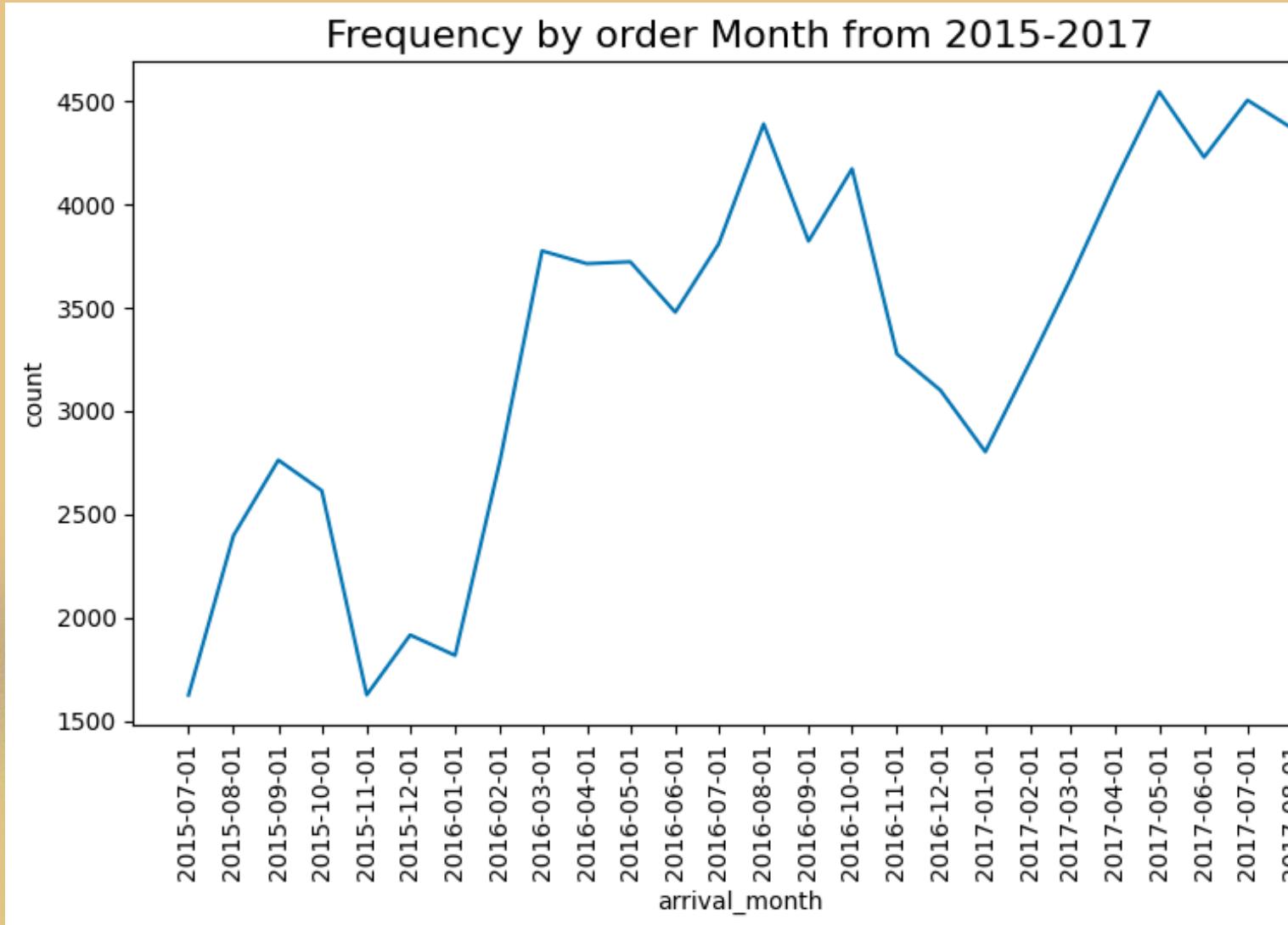


## Leadtime

- Pemesan yang memesan hotel yang < 90 hari (leadtime) memiliki jumlah cancel paling banyak dengan 12.966 transaksi
- Proporsi pembatalan pemesanan yang jedanya semakin lama jumlah proporsinya cenderung semakin tinggi
- **Semakin lamanya leadtime, semakin tinggi peluang terjadinya pembatalan booking hotel**



# Pengaruh Banyaknya pemesan hotel di bulan tertentu terhadap persentase pembatalan

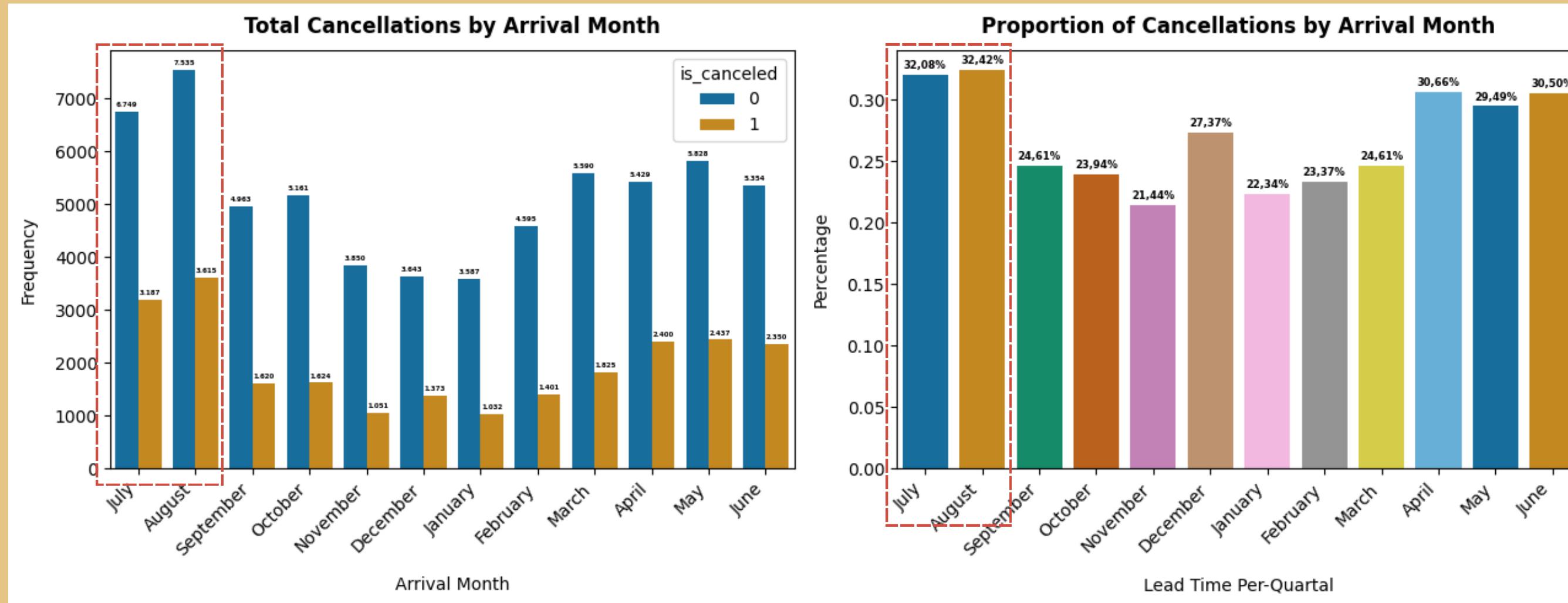


## Frequency

- High Frequency : **Bulan Juli - Agustus & Agustus - September**
- Low Frequency : **Bulan November - Januari**



# Pengaruh Banyaknya pemesan hotel di bulan tertentu terhadap persentase pembatalan



## High Season

seasonal mempengaruhi keputusan pelanggan pada season tertentu untuk melakukan transaksi

High Demand : **Summer Seasons dan Festival Anggur** (Juli-September)

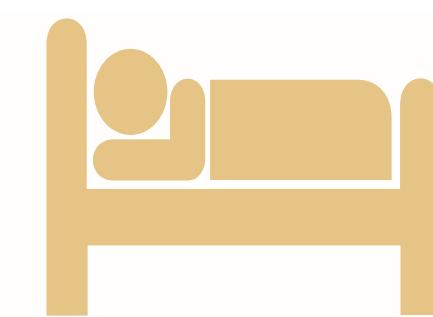
Low Demand : **Winter Seasons** (November-Januari)

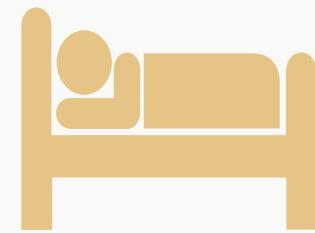


# Google

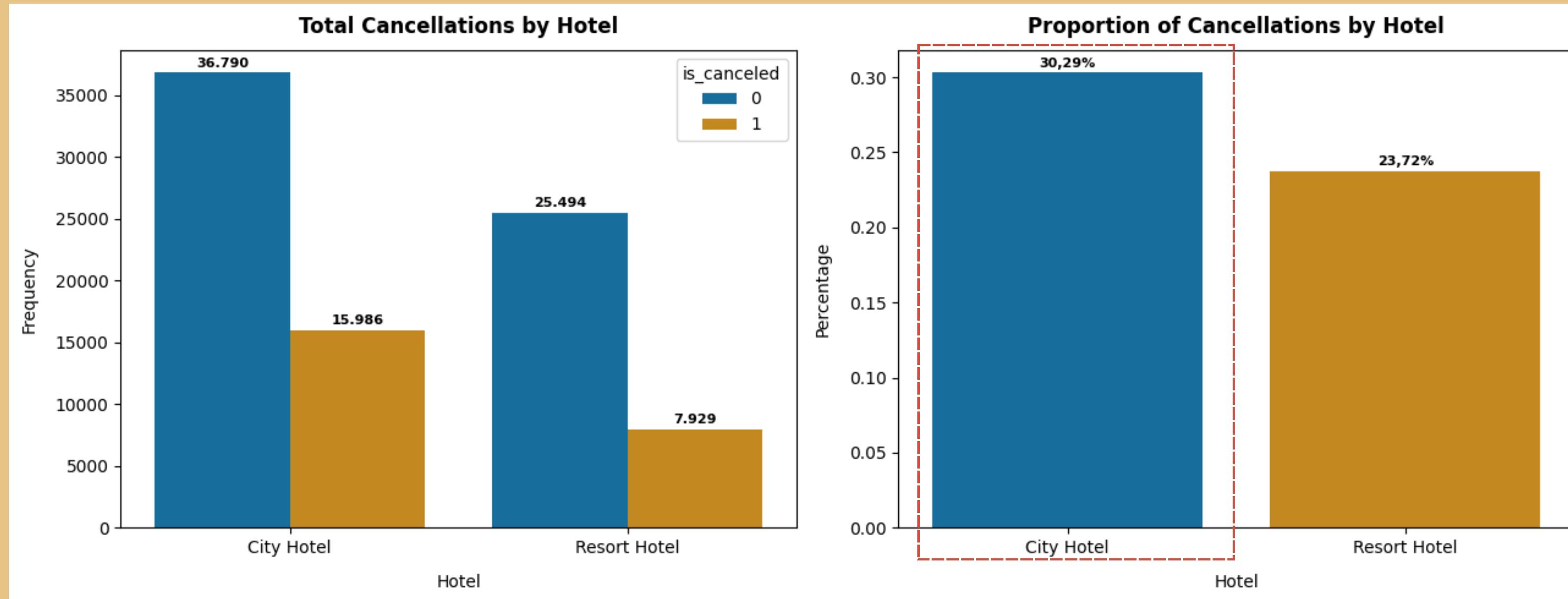


Pengaruh Fasilitas Terhadap Pembatalan Booking Hotel



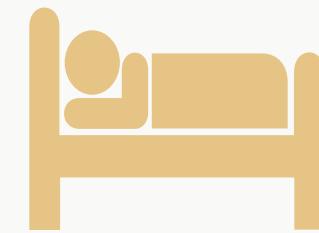


# Jenis Hotel yang melakukan pembatalan pemesanan

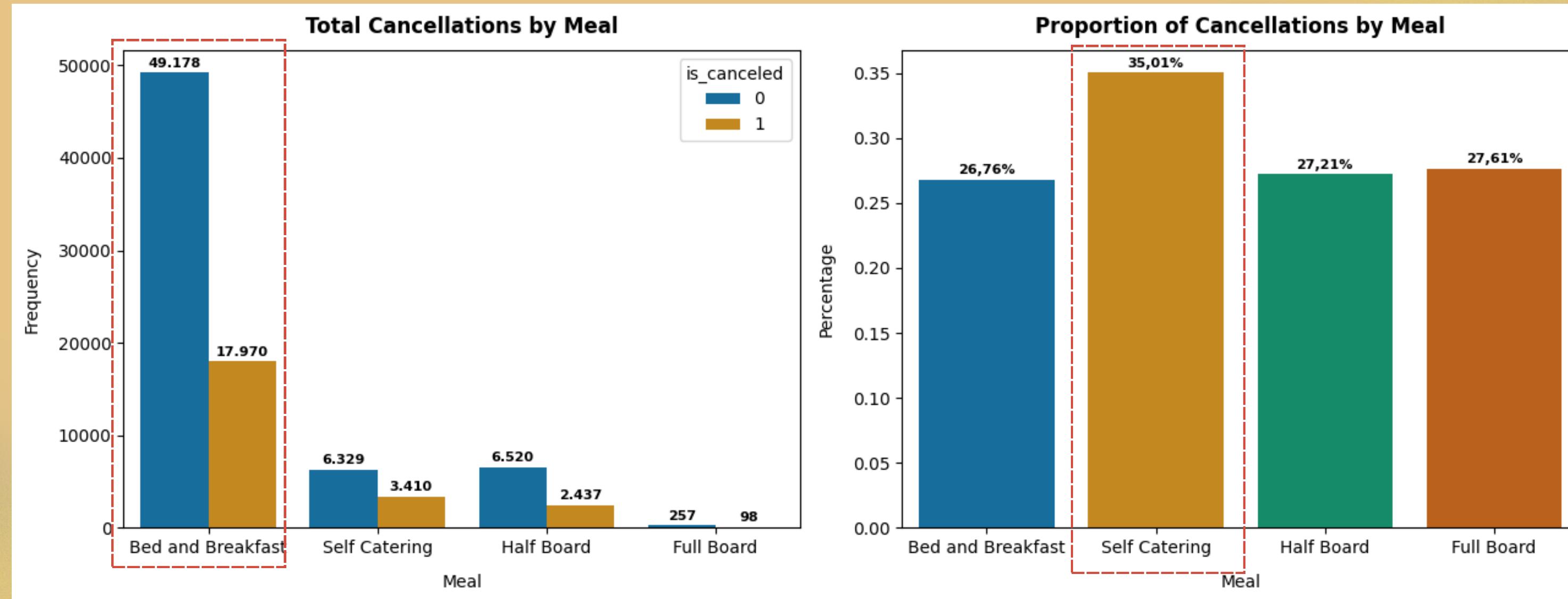


## Insight :

- Jumlah pemesan hotel **paling banyak melakukan cancel** berdasarkan jenis hotel berasal dari **City Hotel** dengan 16,047 dengan dua kali lipat dibandingkan Resort Hotel.
- **30%** pemesan city hotel membatalkan pemesanan booking hotel, yang berarti peminat pemesan hotel **city hotel** memang paling banyak dari segi banyaknya pemesanan ataupun banyaknya yang melakukan pembatalan dari proporsi

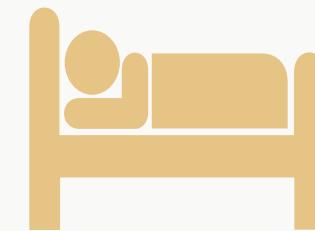


# Jenis Paket Breakfast Terbanyak yang Mengakibatkan Pembatalan Pemesanan Hotel

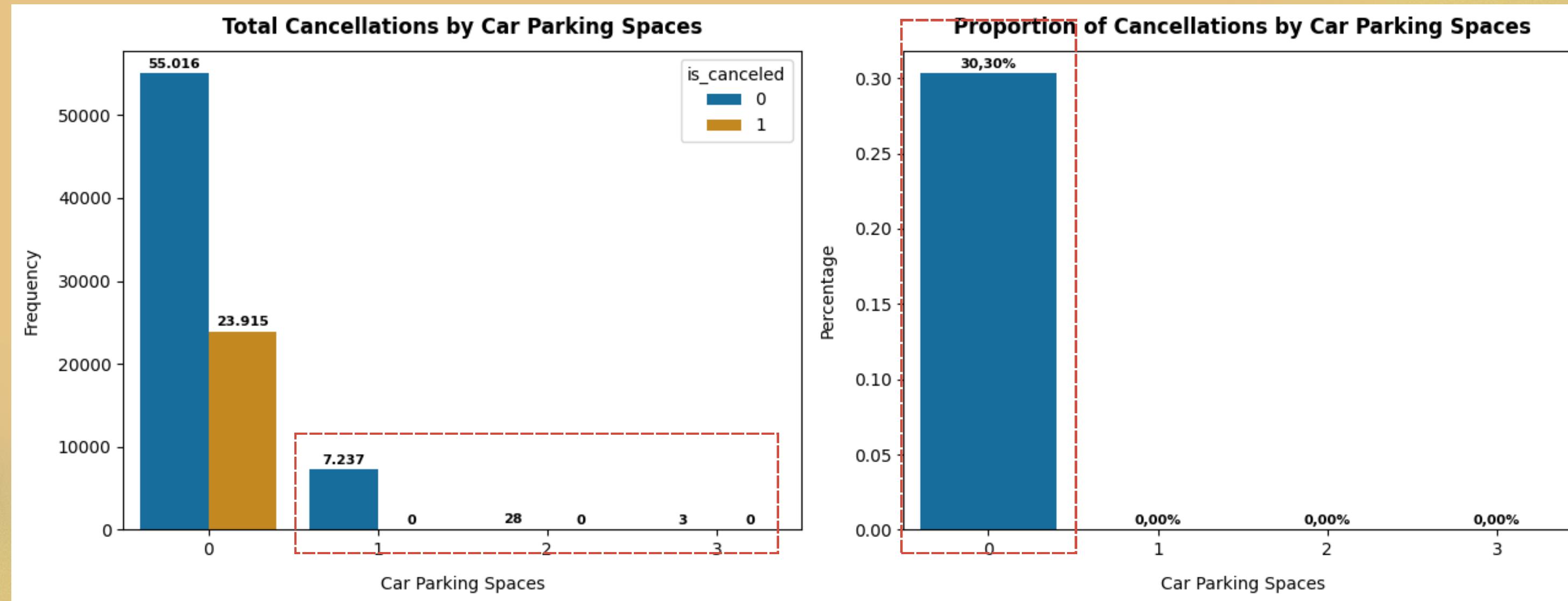


## Insight :

- Proporsi cancel terbesar ada pada **Self Catering** dengan persentase **34%** kemungkinan hal itu terjadi karena rata-rata harga untuk **Self Catering** dan **Bed and breakfast** hanya berbeda sedikit sekitar **€5** para konsumen sudah mendapatkan menu sarapan

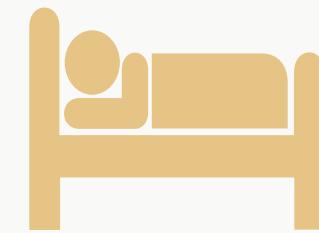


# Pengaruh Banyaknya Lahan Parkir yang disediakan Hotel Terhadap Pembatalan Pemesanan Hotel

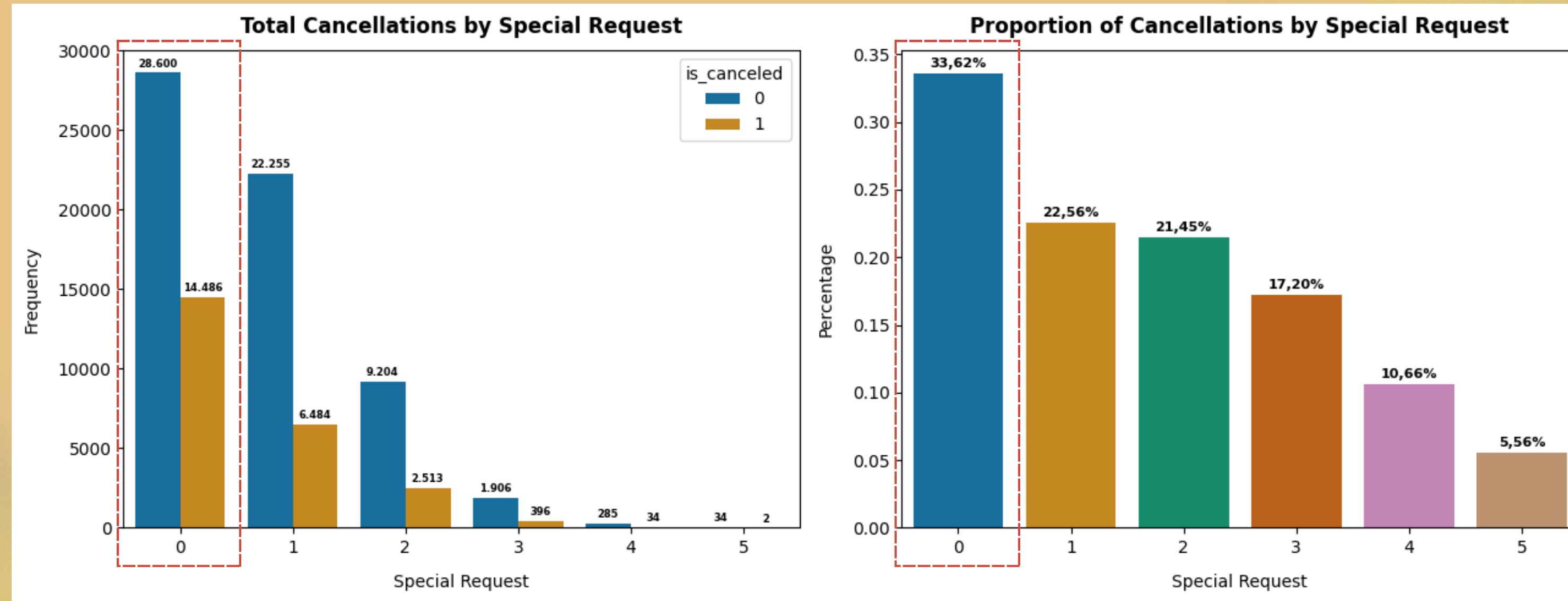


## Insight :

- mayoritas pengunjung tidak membutuhkan car space, hal ini mungkin dikarenakan pengunjung tidak membawa kendaraan pribadi. Terdapat **24023 orang** yang tidak memerlukan **car space** namun melakukan cancellation
- **Car space parking** bukan menjadi faktor dilakukannya pembatalan pemesanan

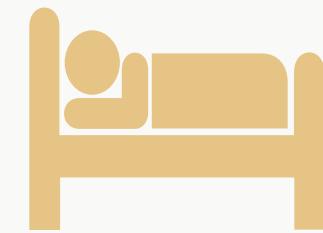


# Banyaknya Permintaan Konsumen Terhadap Fasilitas Hotel Mempengaruhi Pembatalan Pemesanan Hotel

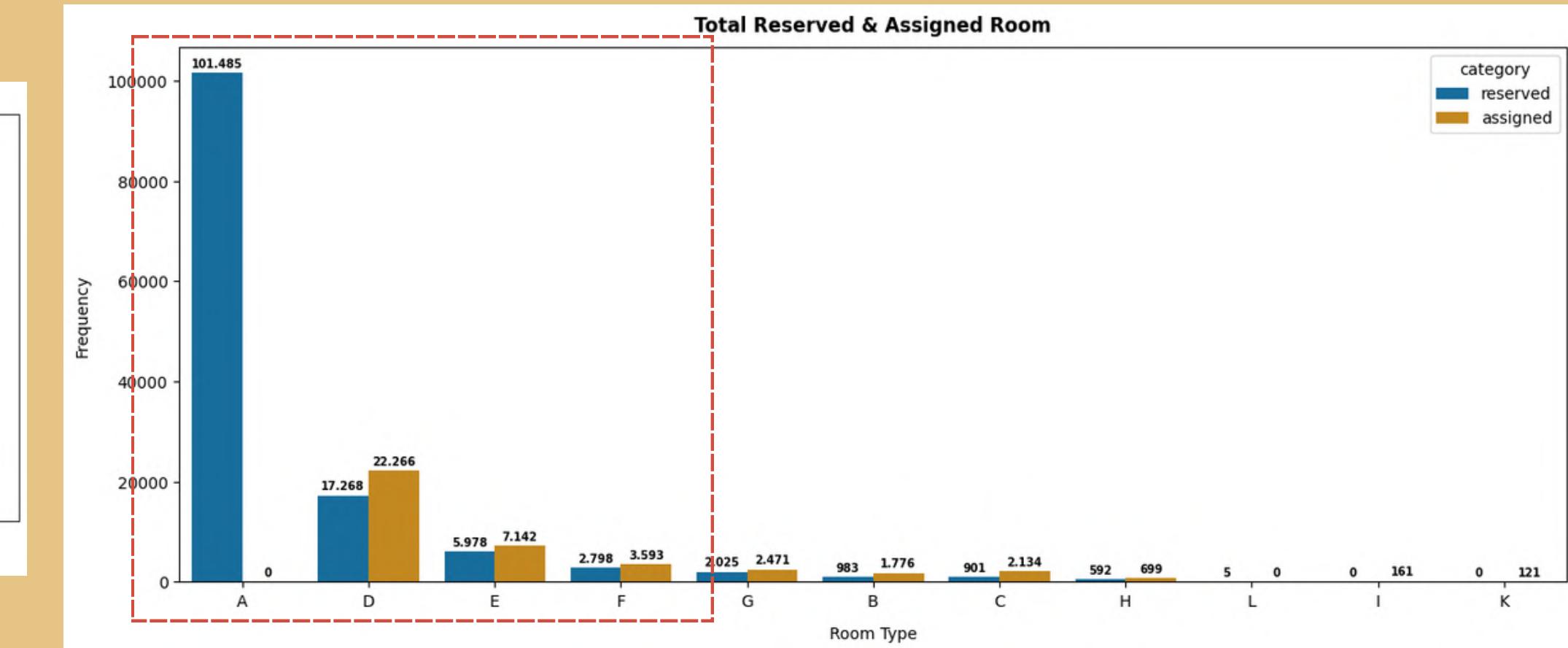
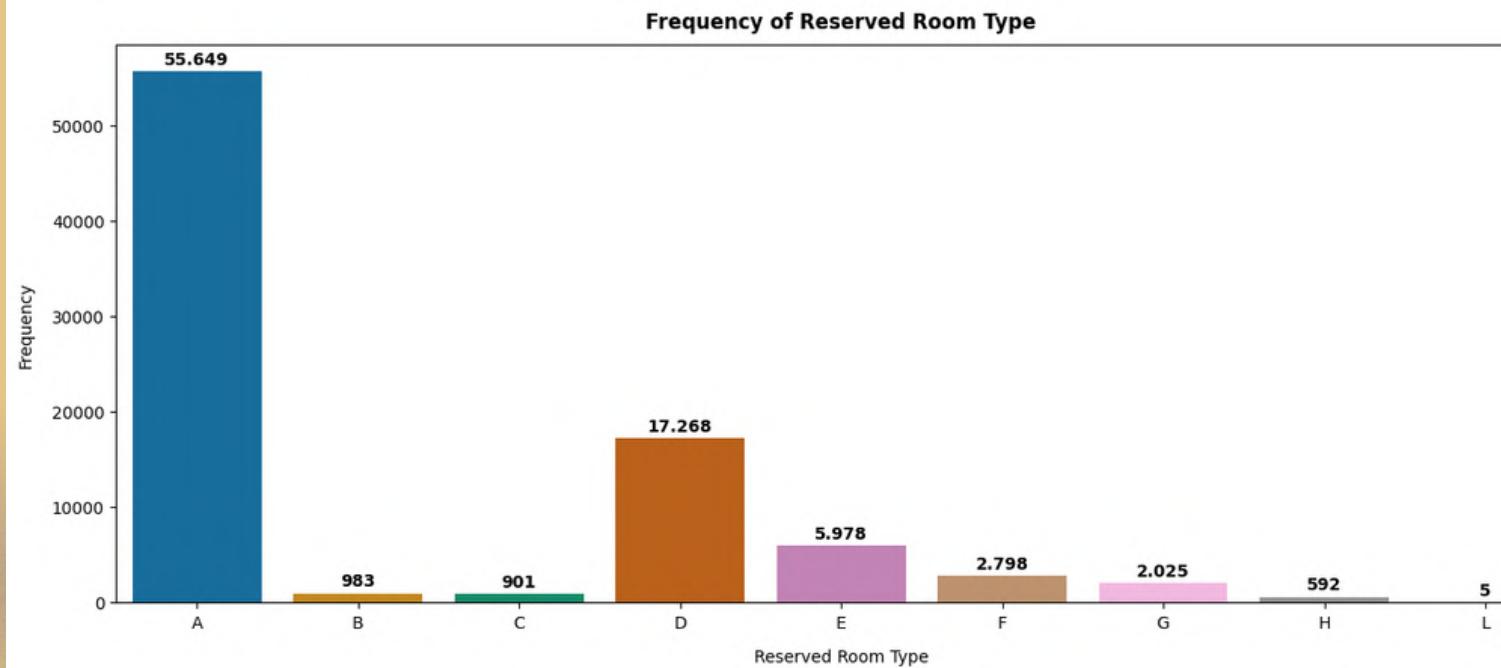


## Insight :

- Semakin banyaknya jumlah spesial request, jumlah pembatalan hotel semakin kecil. Namun hal ini dikarenakan jumlah transaksi dengan jumlah **spesial request > 3** lebih sedikit dari transaksi tanpa spesial request.



# Pengaruh perbedaan pemesana type kamar dan type yang disediakan hotel



## Insight :

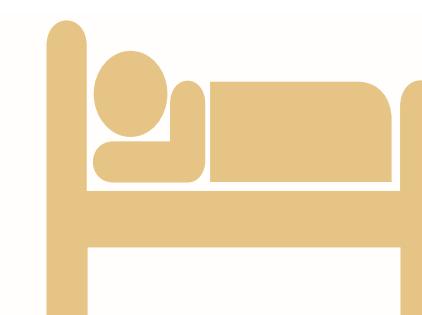
- Minat pelanggan secara garis besar terhadap Room Type A, D dan E
- Banyak pelanggan mendapatkan kamar tidak sesuai permintaan ketika booking. Hal ini dapat menyebabkan pelanggan melakukan cancellation atau menurunkan kepuasan pelanggan
- Sehingga kita merekomendasikan terhadap pihak hotel untuk melakukan layanan seperti berikut ini:
- Pihak hotel harus menanyakan kesediaan customer atas pergantian kamar, memberikan complementary terhadap customer, memberikan free upgrade, melakukan sistem ketersediaan kamar yang akurat



# Google

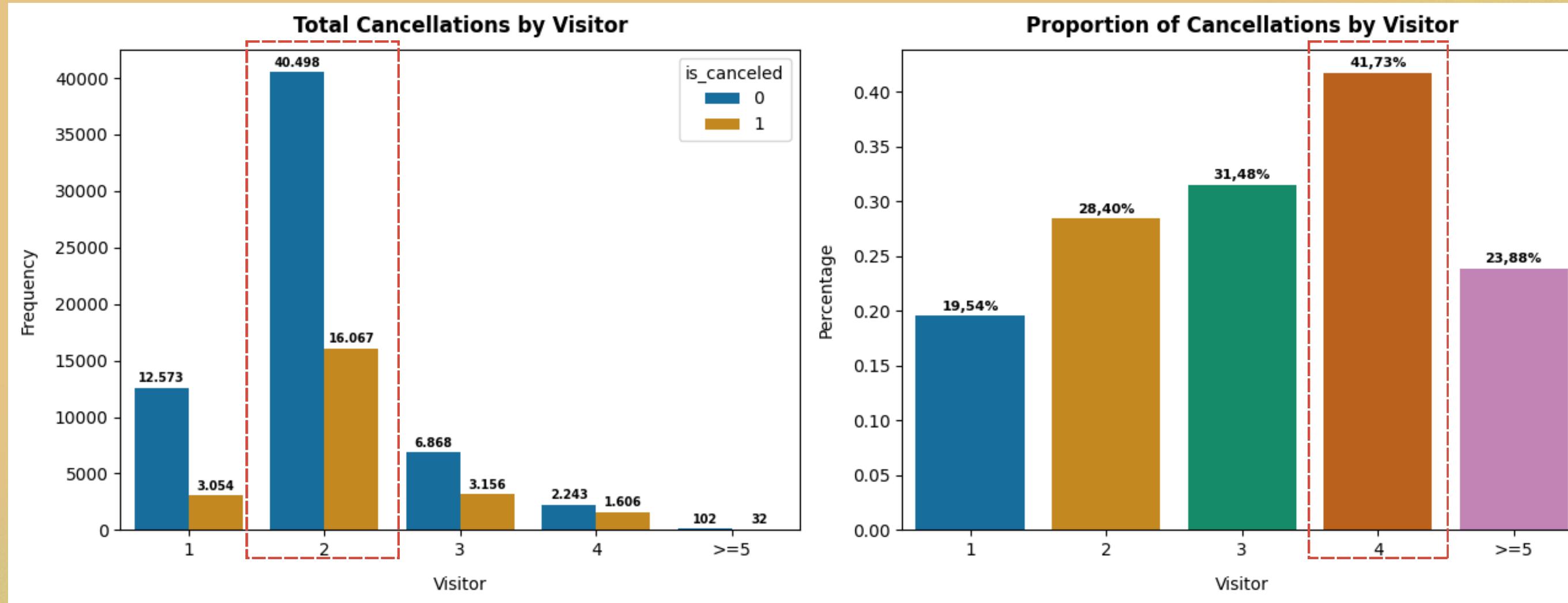


Pengaruh Karakteristik Pelanggan Terhadap Pembatalan Booking Hotel





# Pengaruh jumlah pengunjung dalam satu transaksi terhadap pembatalan pemesanan hotel

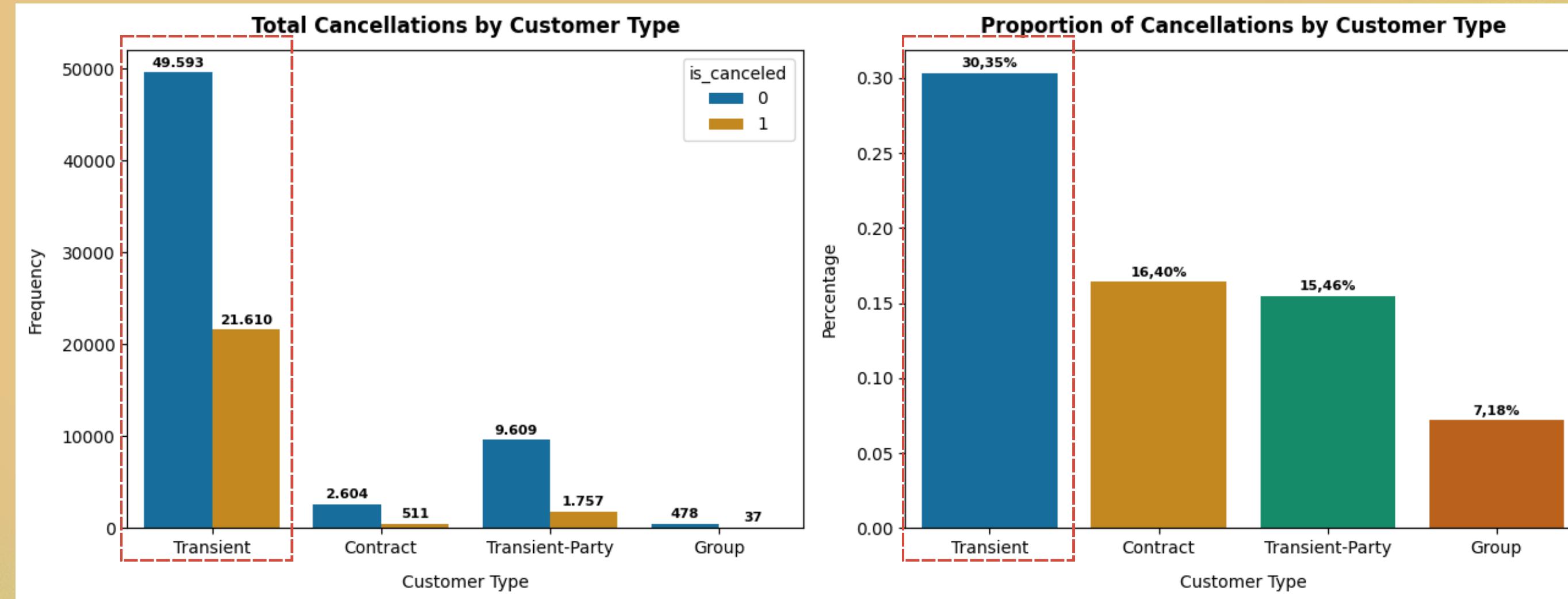


## Insight :

- Jumlah visitor dalam satu transaksi dengan **jumlah 2 orang cenderung tinggi peluang terjadinya cancel**
- Untuk data **diatas 5** hanya sedikit data, jadi data tersebut dianggap anomali
- Proporsi cancel visitor cenderung naik semakin banyaknya pengunjung pemesan hotel. Hal ini mungkin ada faktor lain seperti ketersediaan kamar dan jenis kamar yang dipesan



# pengaruh tipe pelanggan terhadap pembatalan pemesanan hotel

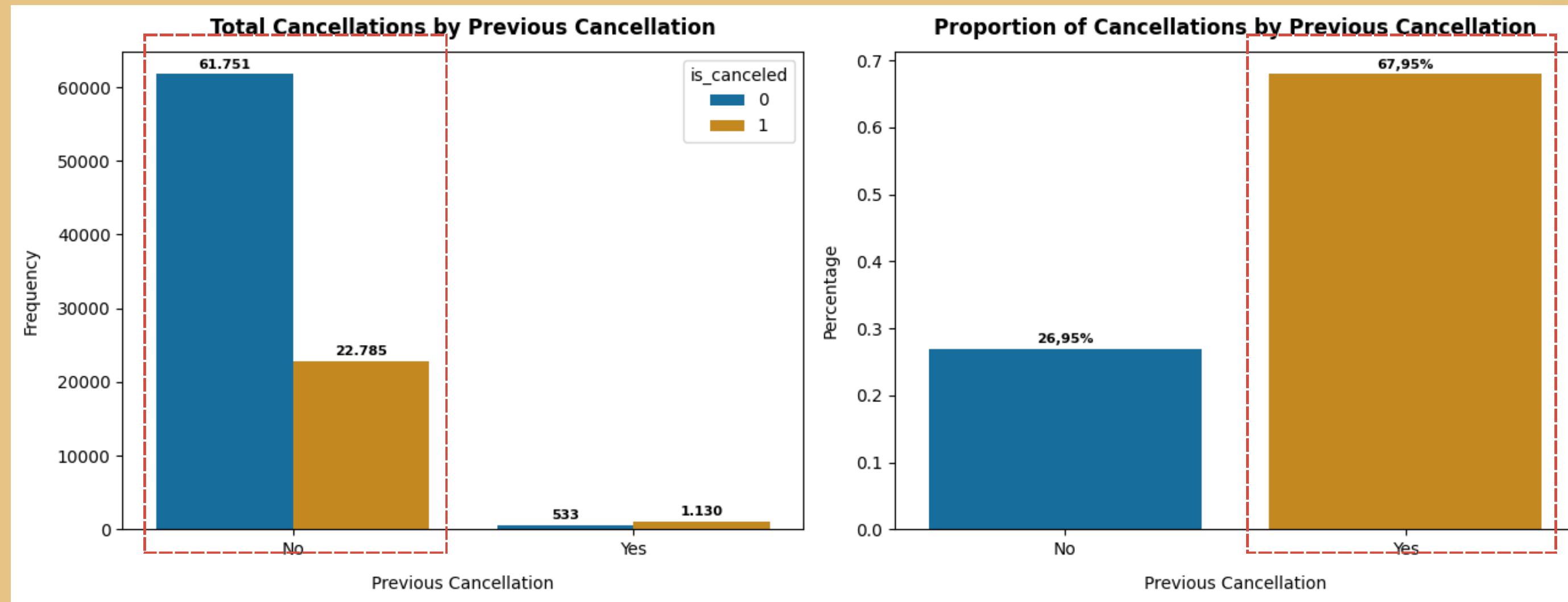


## Insight :

- Jumlah dan proporsi konsumen transient melakukan pembatalan paling banyak dan jika dilihat transient merupakan tipe pelanggan yang memesan secara individual dan tidak terikat kontrak jangka panjang seperti wisatawan yang ingin berlibur atau perjalanan bisnis dalam waktu singkat
- Kami merekomendasikan untuk tim marketing membuat paket promo seperti membuat iklan yang ditargetkan kepada akun sosial media pribadi berupa promo liburan atau promo akhir pekan



# pengaruh banyaknya reservasi hotel yang dibatalkan sebelumnya terhadap pembatalan pemesanan hotel terkini

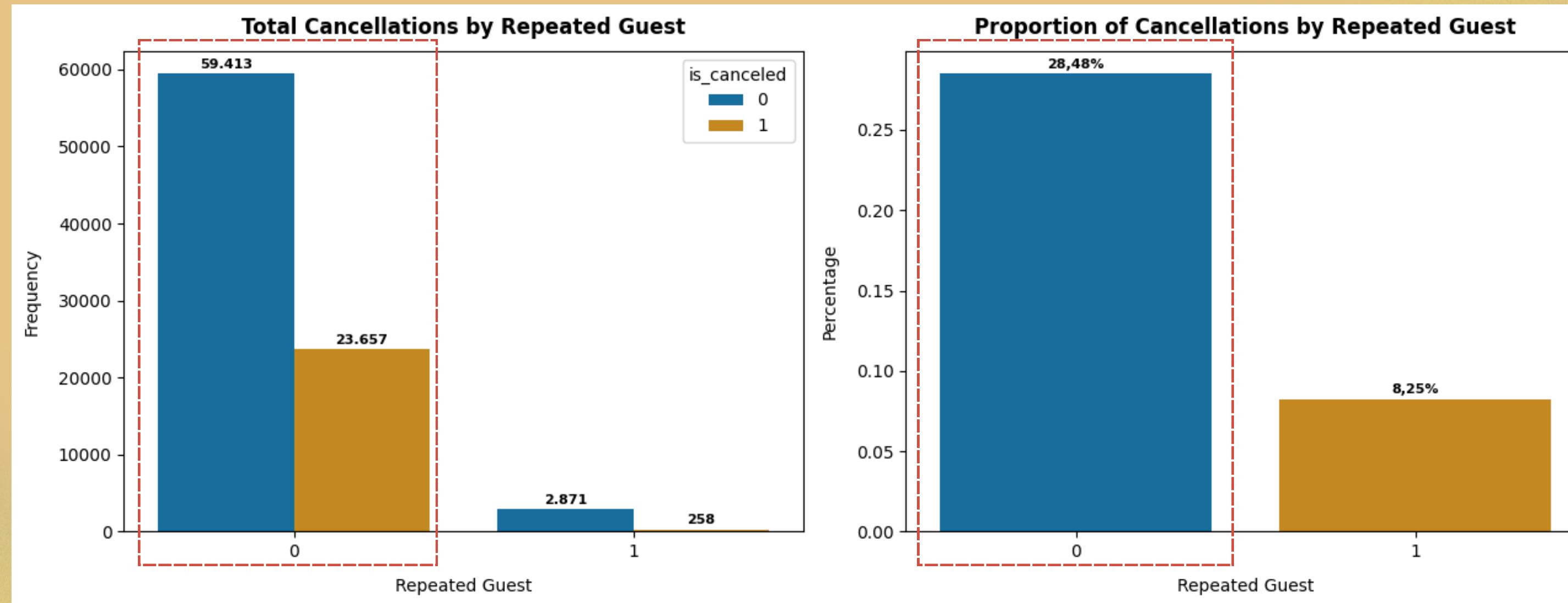


## Insight :

- Berdasarkan Faktor **previous cancellation** diketahui bahwa customer yang sebelumnya melakukan pembatalan proporsinya lebih besar melakukan pembatalan ditransaksi berikutnya
- Rekomendasi terhadap pihak hotel dapat memberikan treatment khusus terhadap konsumen yang sebelumnya pernah melakukan cancellation seperti : Penawaran reschedule apabila memang calon customer tersebut memiliki ketidaktepatan/tidak sesuaian tanggal dan Penawaran apakah customer tersebut memiliki special request yang dapat dipenuhi oleh pihak hotel



# pengaruh pemesan hotel repeated guest terhadap pembatalan pemesanan hotel



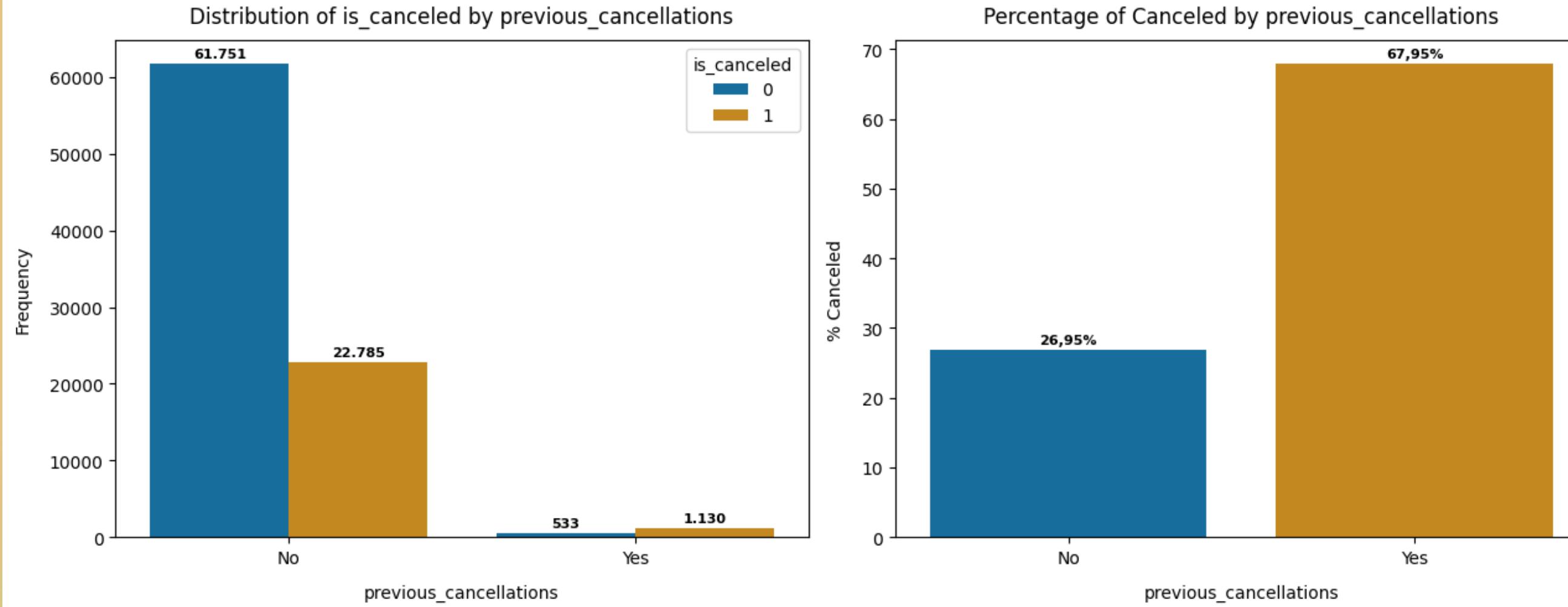
## Insight :

- Customer yang belum pernah melakukan repeat booking lebih besar dari customer yang pernah melakukan booking secara jumlah maupun proporsi
- Hal ini mungkin disebabkan karena customer yang pernah melakukan booking telah mengetahui fasilitas dan layanan yang didapatkan oleh pihak hotel dibandingkan customer yang belum pernah melakukan repeat booking.
- saran kami terhadap pihak hotel yaitu memberikan potongan pada pelanggan baru agar tertarik untuk melakukan transaksi kembali dan memberikan pendekatan terhadap alasan cancellation pengunjung



# pengaruh pemesan hotel yang sebelumnya melakukan reservasi terhadap cancellation

## Relationship between Previous Cancellation and Hotel Cancellation

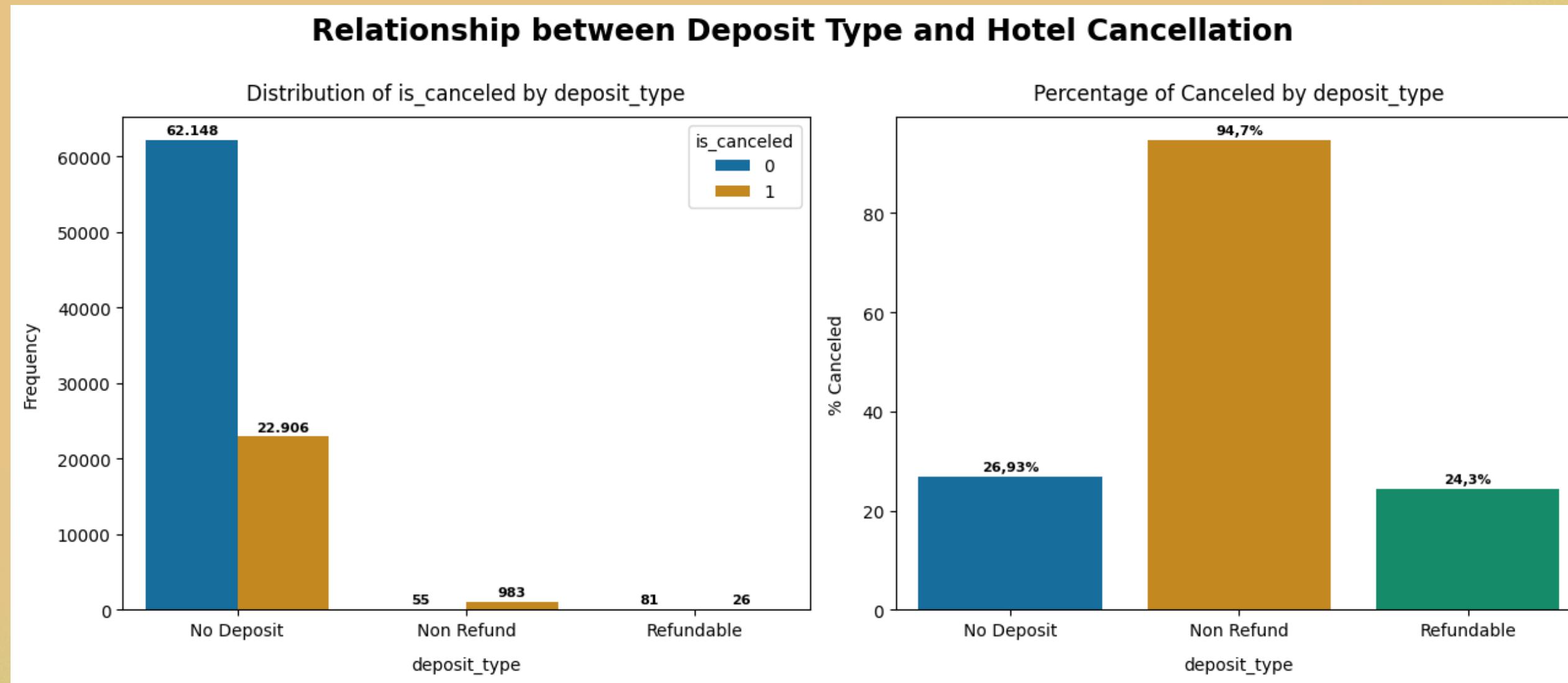


## Insight :

- pelanggan yang sebelumnya melakukan pembatalan memiliki persentase pembatalan yang lebih tinggi pada transaksi berikutnya.
- kami merekomendasikan kepada pihak hotel untuk memberikan perlakuan khusus kepada konsumen yang sebelumnya pernah melakukan pembatalan. Beberapa rekomendasi tersebut melibatkan: Menawarkan opsi untuk *reschedule* dan Memberikan penawaran terkait permintaan khusus yang dapat dipenuhi oleh pihak hotel.



# pengaruh tipe deposit yang dipilih oleh pelanggan terhadap pembatalan pemesanan hotel



## Insight :

- persentase tertinggi yang berpeluang melakukan cancel ada pada pelanggan yang memilih kategori Non Refund
- pelanggan cenderung memilih pembatalan daripada kehilangan uang deposit atau biaya hotel yang tidak dapat dikembalikan
- kami merekomendasikan pihak hotel untuk menerapkan kebijakan konfirmasi kedatangan kepada pelanggan satu hari sebelum tanggal check-in
- membatasi opsi *No Deposit* dengan mengarahkan pelanggan pada pilihan deposit yang lebih fleksibel atau kebijakan *reschedule*



# PREPROCESSING



## Encoding : One Hot Encoding

Data ini tidak memiliki urutan dan memiliki nilai unique yang sedikit.

- **hotel**
- **arrival\_date\_month**
- **meal**
- **market\_segment**
- **distribution\_channel**
- **reserved\_room\_type**
- **assigned\_room\_type**
- **deposit\_type**
- **customer\_type**



## Define Target & Features

- Target : **is\_canceled**
- Features : Numerical & Categorical



## Splitting

- 80 % : Data Train set
- 20% : Data test set

**Scaler :** RobustScaler

**Resampler :**

- SMOTENC
- SMOTE
- RandomUnderSampler
- RandomOverSampler
- NearMiss

**Analytic Approach :**

- Logistic Regression
- K-Nearest Neighbors
- Decision Tree
- Random Forest
- Gradient Boosting
- AdaBoost
- XGBoost
- LightGBM



# BENCHMARKING MODEL



without sampling

Model	Mean Recall	Recall Score
Random Forest	0.541	0.545
XGBoost	0.536	0.539
K-Nearest Neighbors	0.515	0.521
LightGBM	0.510	0.512
AdaBoost	0.468	0.478
Gradient Boosting	0.463	0.461
Decision Tree	0.436	0.441
Logistic regression	0.421	0.416

with sampling

Model - NearMiss	Mean Recall	Recall Score
Random Forest	0.784	0.784
XGBoost	0.784	0.783
LightGBM	0.776	0.781
Gradient Boosting	0.754	0.752
Decision Tree	0.753	0.751
AdaBoost	0.742	0.697
K-Nearest Neighbors	0.643	0.645
Logistic regression	0.631	0.629

Berdasarkan hasil diatas dapat dilihat dari 8 model yang telah diuji, didapatkan 2 model yang direkomendasikan berdasarkan Metric Evaluation yang terkecil untuk memprediksi data test set yaitu:

- **Random Forest Classifier**
- **XGBoost Classifier**

Metric Evaluation **Recall** untuk menghindari **FN (False Negatif)** sebagai prioritas utama.



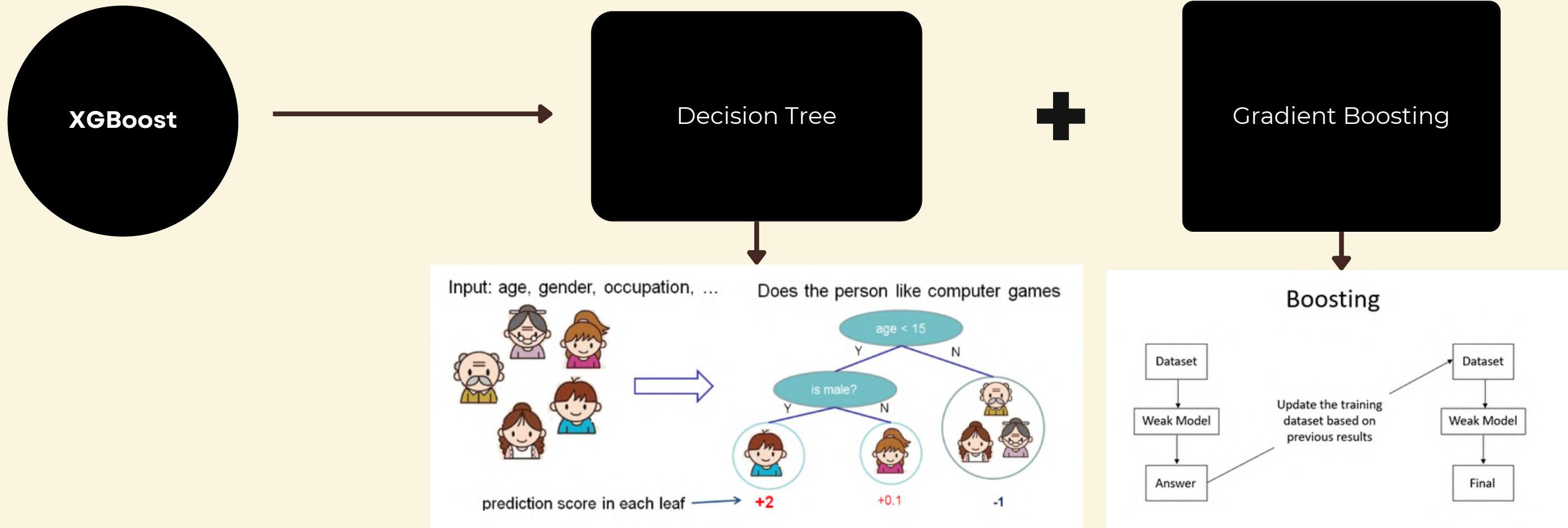
# HYPERPARAMETER TUNING



Model	Mean Recall	Mean Precision	Mean F1	subsample	scale_pos_weight	n_estimators	min_child_weight	max_depth	max_delta_step	max_bin	gamma	eta	colsample_bytree
XGBoost Benchmark (Recall)	0.784	-	-	-	-	-	-	-	-	-	-	-	-
XGBoost Tuned (Recall)	0.988	-	-	0.9	30	434	3	5	3	300	0	0.05	0.7
XGBoost Benchmark (Precision)	-	0.436	-	-	-	-	-	-	-	-	-	-	-
XGBoost Tuned (Precision)	-	0.588	-	0.7	10	887	3	10	9	300	0.1	0.2	0.9
XGBoost Benchmark (F1 Score)	-	-	0.560	-	-	-	-	-	-	-	-	-	-
XGBoost Tuned (F1 Score)	-	-	0.65	0.9	10	934	7	10	8	275	0.1	0.1	0.8

	Model	Predict Score
0	XGBoost Benchmark (Recall)	0.783
1	XGBoost Tuned (Recall)	0.990
2	XGBoost Benchmark (Precision)	0.435
3	XGBoost Tuned (Precision)	0.587
4	XGBoost Benchmark (F1 Score)	0.559
5	XGBoost Tuned (F1 Score)	0.651

# XGBOOST



## Fungsi Tujuan:

- Model XGBoost mengoptimalkan fungsi tujuan yang mencakup bagian kerugian (loss function) dan komponen regularisasi (penalization)
- Meminimalkan kesalahan prediksi dan menghindari overfitting

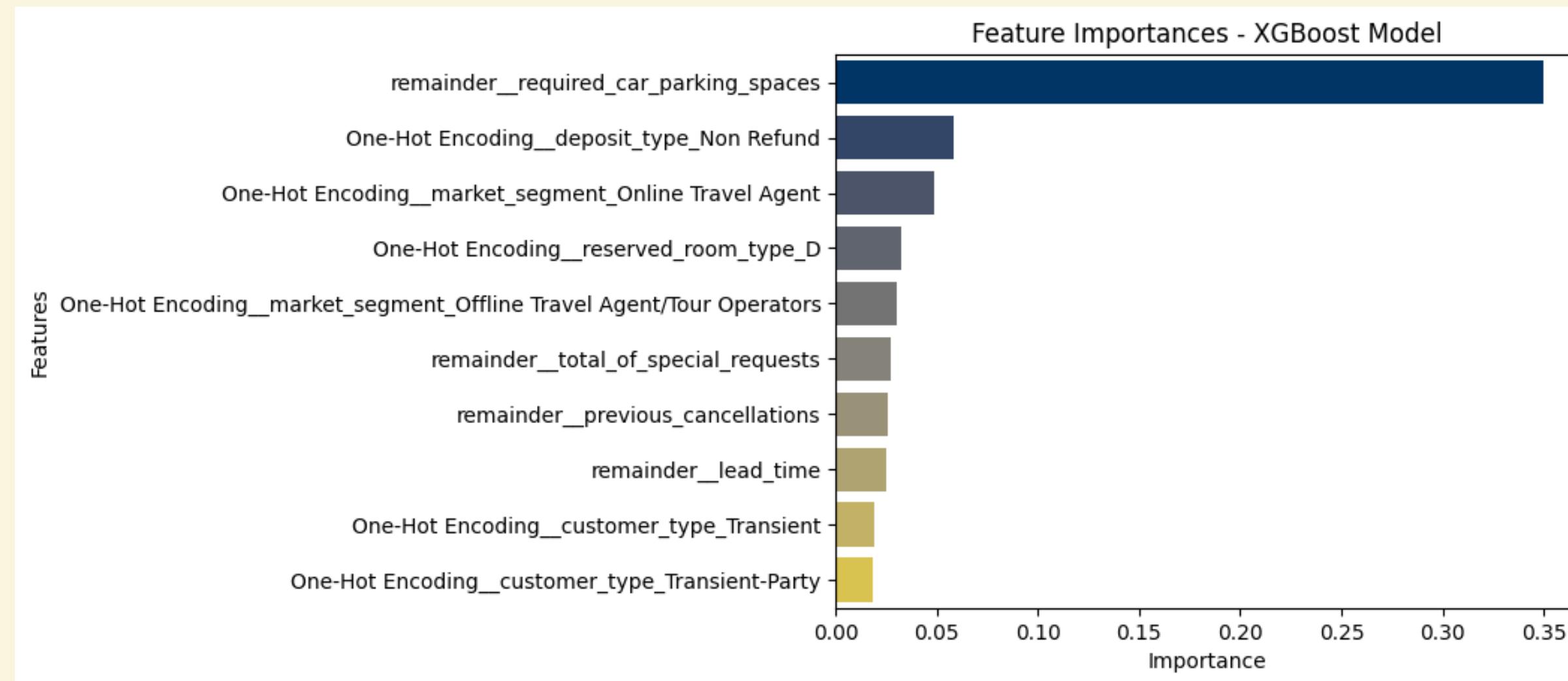
## Based modeling : Decision Tree & Gradient Boosting

Cara kerja :

- Menggabungkan beberapa model prediksi sederhana yang disebut "Weak Learner" untuk membentuk model yang lebih kompleks dan efektif



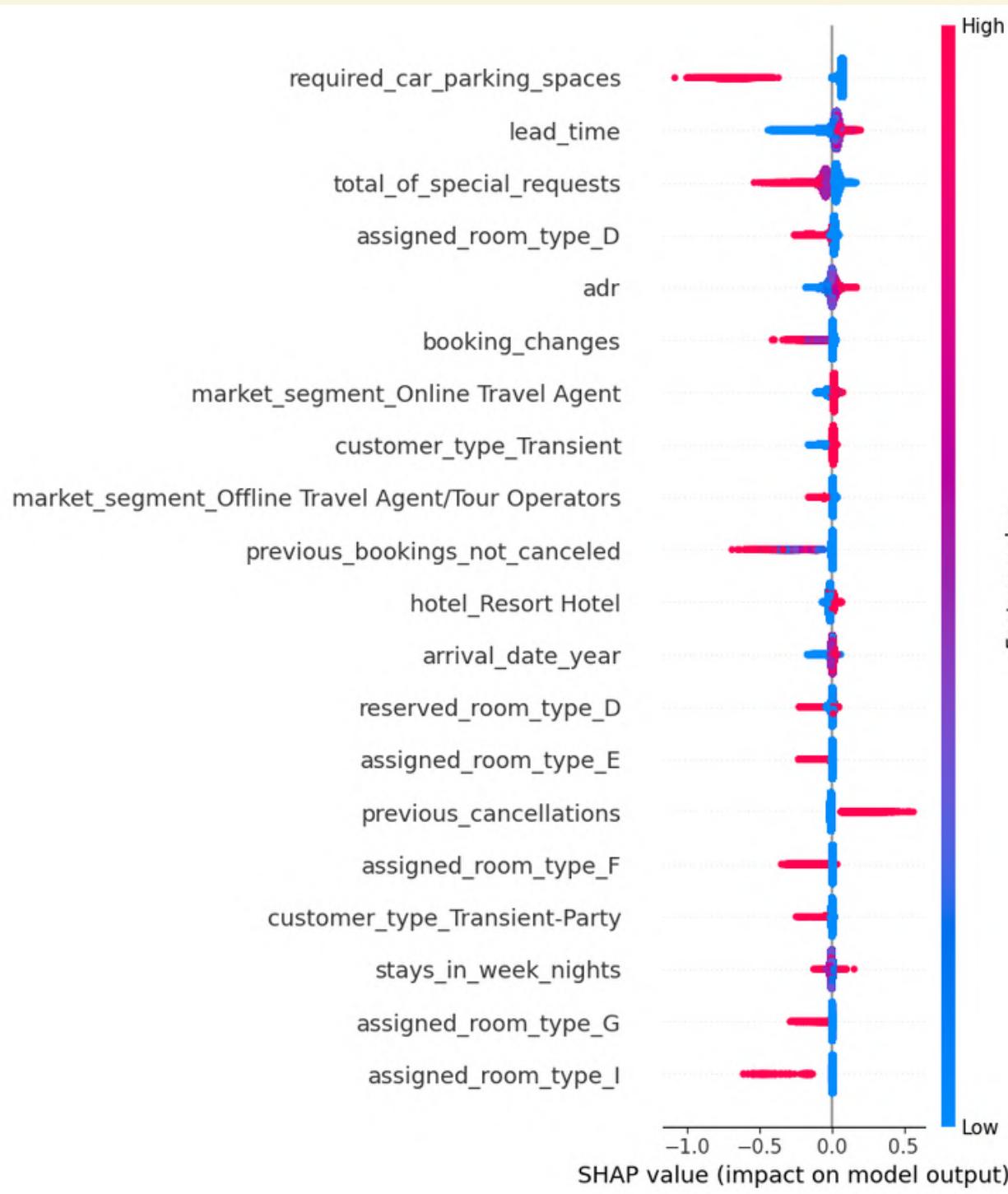
# FEATURE IMPORTANT



beberapa fitur/kolom yang paling penting pada **feature importance** mayoritas cukup sesuai dengan sebagian EDA yang kami analisis

- **required\_car\_parking\_spaces** : tidak terlalu sesuai dengan analisis. pada Analisis target melakukan cancel pada saat nilainya 0
- **deposit\_type** : target cenderung akan cancel jika deposit typenya non refund
- **Market Segment** : target cenderung akan melakukan cancel pada market segment Online Travel Agent

# SHAP



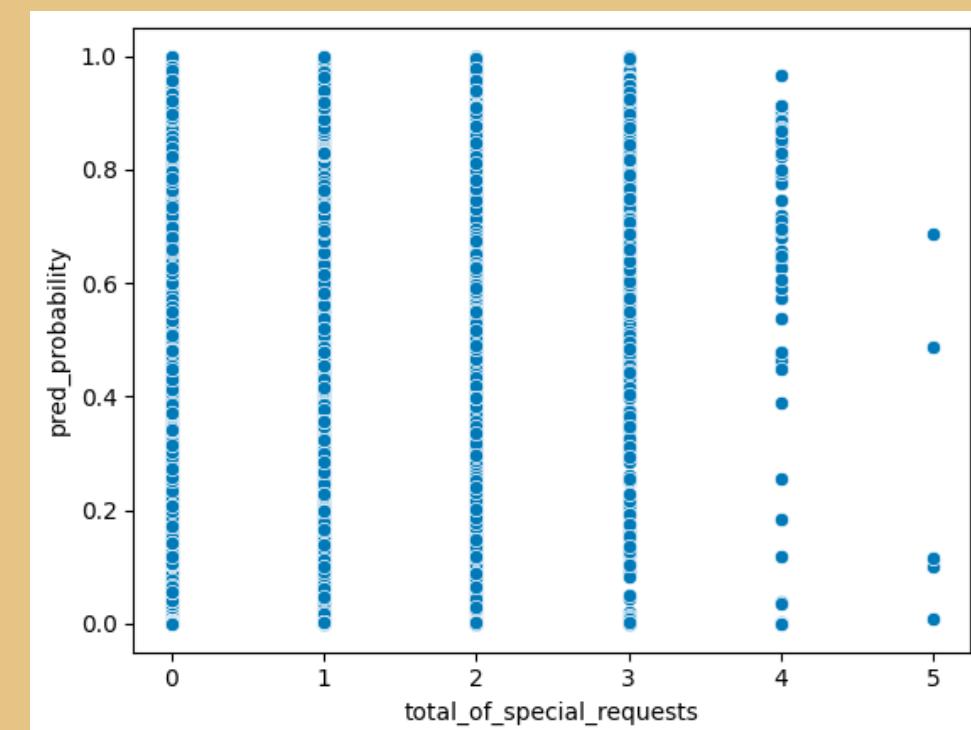
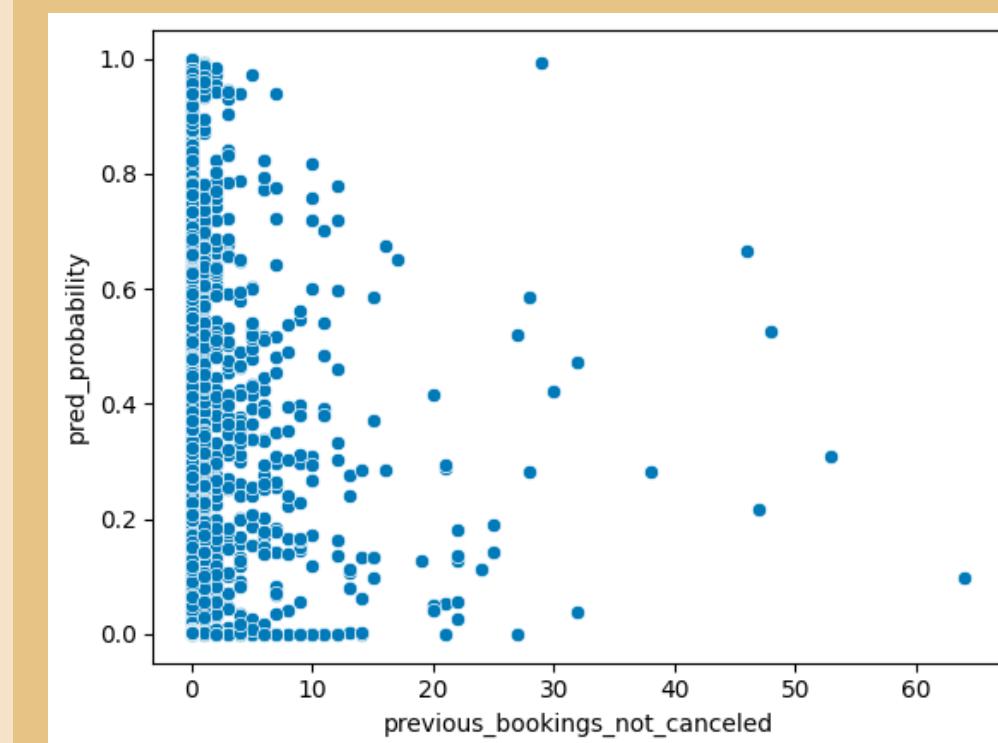
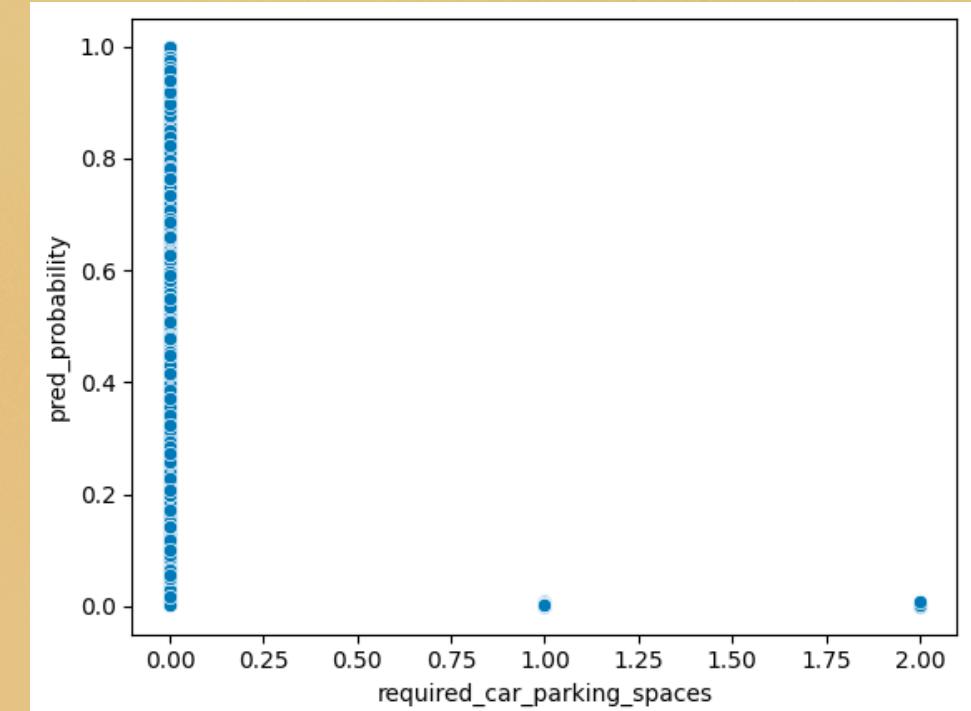
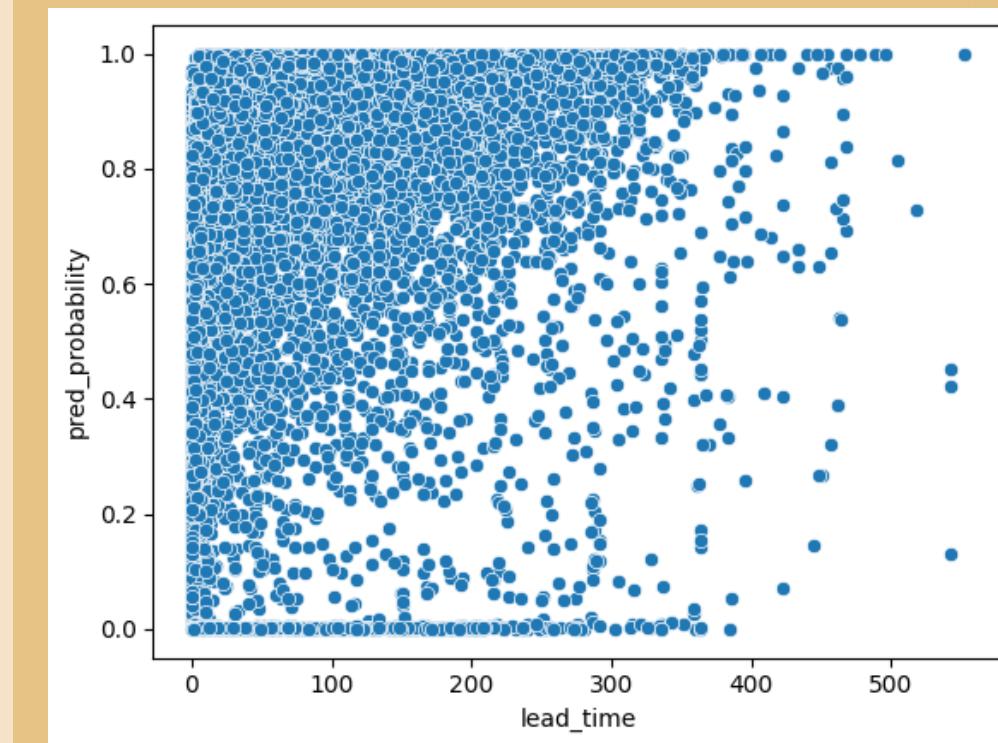
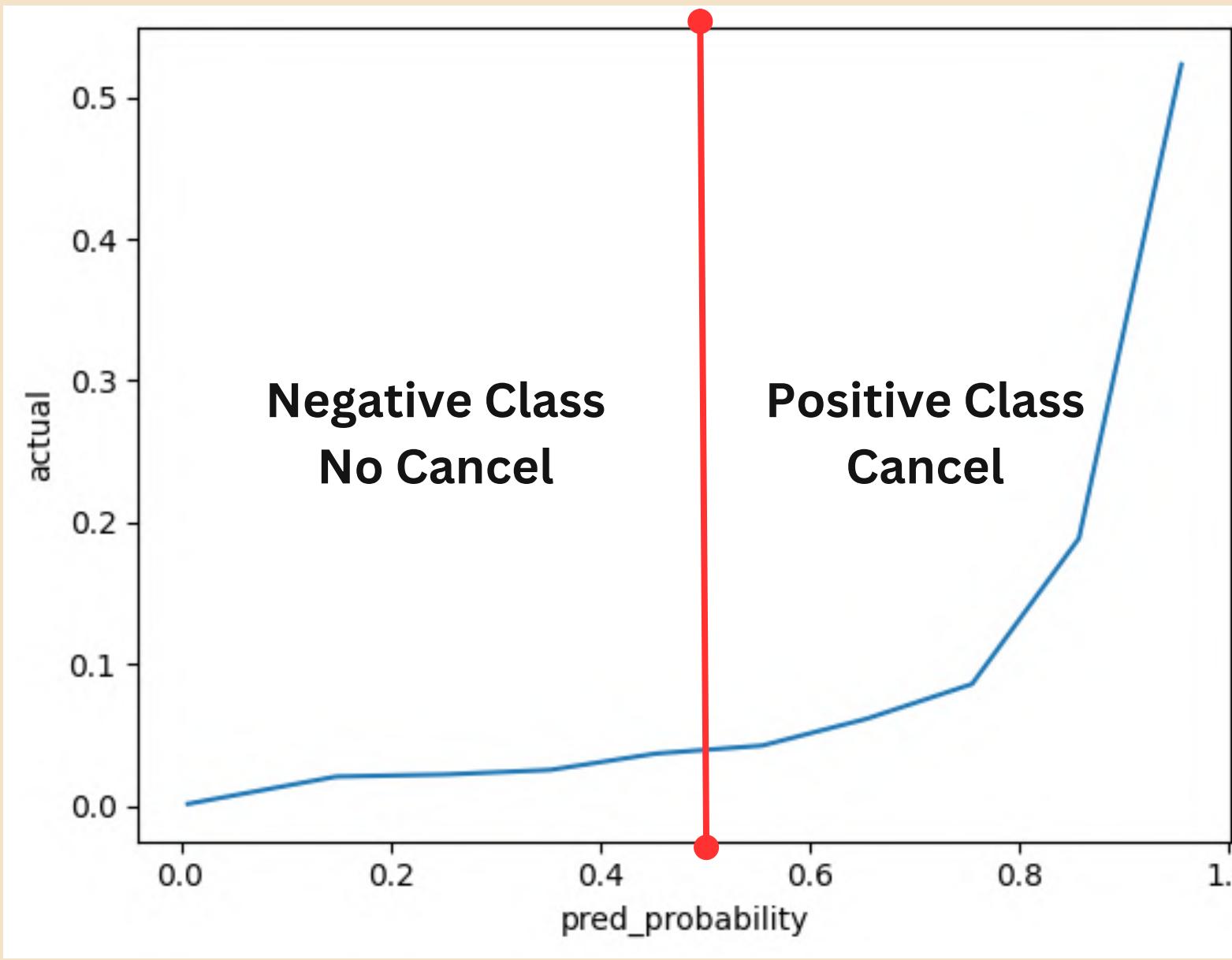
Berdasarkan **SHAP** value, kita melihat beberapa fitur yang penting diantaranya:

- **required\_car\_parking\_spaces**: Customer yang meminta lahan parkir mobil semakin banyak akan semakin kecil kemungkinannya melakukan pembatalan hotel
- **lead\_time**: customer yang melakukan booking dengan jeda waktu yang lama akan semakin besar potensi melakukan pembatalan booking hotel
- **total\_of\_special\_request**: semakin sedikit permintaan spesial dari customer terhadap penyedia hotel semakin tinggi kemungkinan customer tersebut melakukan pembatalan booking.

Interpretasi SHAP Value diatas memiliki hasil yang cukup sesuai dengan hasil analisa kami pada Data Analis



# DATA ACTUAL VS DATA PREDICTION



# MODEL LIMITATION

Feature	Data Type	Limitasi
<code>lead_time</code>	Integer	Jarak hari antara tanggal <i>booking</i> hotel hingga tanggal yang dijadwalkan untuk <i>check-in</i> di rentang 0-709
<code>arrival_date_year</code>	Integer	Tahun dijadwalkan <i>check-in</i> di rentang 2015-2017
<code>stays_in_weekend_nights</code>	Integer	Jumlah malam di akhir pekan (Sabtu atau Minggu) yang di- <i>booking</i> di rentang 0-19
<code>stays_in_week_nights</code>	Integer	Jumlah malam dalam seminggu (Senin - Jum'at) yang di- <i>booking</i> di rentang 0-50
<code>adults</code>	Integer	Jumlah orang dewasa di rentang 0-4
<code>children</code>	Float	Jumlah anak-anak di rentang 0-10
<code>babies</code>	Integer	Jumlah bayi di rentang 0-2
<code>is_repeated_guest</code>	Integer	Nilai yang berasal dari pemesan yang berulang atau tidak di rentang 0-1
<code>previous_cancellations</code>	Integer	Jumlah Pemesanan yang dibatalkan sebelumnya di rentang 0-26
<code>previous_bookings_not_canceled</code>	Integer	Jumlah <i>booking</i> hotel sebelumnya yang tidak dibatalkan di rentang 0-72
<code>booking_changes</code>	Integer	Jumlah perubahan yang dilakukan pada <i>booking</i> hotel di rentang 0-18
<code>days_in_waiting_list</code>	Integer	Jumlah hari untuk setiap <i>booking</i> hotel yang masuk ke dalam <i>waiting list</i> di rentang 0-391
<code>adr</code>	Float	Rata-rata tarif harian atau <i>Average Daily Rate</i> di rentang 0-510
<code>required_car_parking_spaces</code>	Integer	Jumlah ruang parkir mobil yang dibutuhkan pemesan di rentang 0-3
<code>total_of_special_requests</code>	Integer	Jumlah permintaan khusus yang dibuat oleh pemesan di rentang 0-5

# COST ANALYSIS

Sebelum Model dan sesudah menggunakan model:

**€ 532,457**

Tanpa Model

Metrics	Score Model	Keuntungan
Recall	99	1.000.455 €
Precision	58	961.833 €
F1 Score	65	989.371 €

Berdasarkan Evaluation Cost Metrics yang dipilih adalah **Recall**. Setelah meninjau Model dapat memprediksi data dengan cukup tepat dan memberikan keuntungan yang lebih besar.

**€ 467,998**

Income

**187%**

Total Keuntungan yang didapatkan



# CONCLUSION



## Analisis

1. Bagaimana karakteristik jarak hari dan waktu tertentu mempengaruhi kebiasaan pembatalan pemesanan hotel?
  - Pembatalan pemesanan hotel rentan dilakukan ketika antara **waktu pemesanan dengan waktu kedatangan jaraknya terlalu jauh**
  - Waktu yang tingkat cancellationnya tinggi terjadi pada high season seperti liburan **summer season** atau kisaran bulan **Juli-September** (Festival Anggur)
2. Bagaimana karakteristik fasilitas dan pelayanan hotel mempengaruhi pelanggan dalam membatalkan pesanan hotel?
  - Jenis hotel **City hotel** memiliki proporsi lebih tinggi dilakukannya cancel
  - Customer hotel yang memesan paket **self catering** memiliki peluang lebih tinggi untuk melakukan pembatalan
  - **Car space parking** bukan menjadi faktor utama yang menyebabkan pengunjung melakukan pembatalan
  - Customer dengan **jumlah permintaan khusus** proporsi pembatalan hotel cenderung berkurang
3. Bagaimana karakteristik pelanggan yang melakukan pembatalan pemesanan hotel?
  - Semakin banyaknya jumlah **visitor**, proporsi tingkat pembatalan hotel cenderung semakin naik
  - Customer dengan kategori **transient** berpeluang lebih besar melakukan pembatalan pemesanan
  - Pelanggan yang **sebelumnya melakukan pembatalan** memiliki persentase pembatalan yang lebih tinggi
  - Pelanggan dengan persentase tertinggi yang berpeluang melakukan cancel ada pada pelanggan yang memilih kategori **Non Refund**,



# CONCLUSION



## Modeling

- Model prediktif yang mampu mengidentifikasi potensi customer melakukan cancel didapat menggunakan algoritma XGBoost.
- Parameter ini digunakan untuk imbalance data dengan `Scale\_post\_weight`: 30.
- Hyperparameter tuning dilakukan untuk mendapatkan nilai metriks evaluasi yang maksimal. dengan mengontrol kompleksitas model secara langsung dengan `min\_child\_weight` : 3 dan `gamma` : 0
- Beberapa Fitur yang paling berpengaruh terhadap target (Is\_canceled) diantaranya `required\_car\_parking\_spaces`, `deposit\_type`, dan `market\_segment`.
- Model XGBoost yang dipilih pada metrics evaluation ini adalah **RECALL** karena memiliki score sebesar **99%** dan memberikan keuntungan yang cukup besar dari segi Bisnis jika dilihat melalui Cost Analysis.
- Tanpa ML, Tim Marketing memberikan penawaran (Double Booking) atau promo bundling lain secara acak. Dengan Machine learning kita bisa memprioritaskan kepada Pelanggan Potensial untuk menghemat budget. Kita juga dapat membuat campaign dengan target audience yang lebih terarah dan terukur dengan adanya machine learning.

Selain itu juga **RECALL** memberikan penanganan nilai error yang cenderung lebih aman untuk dataset ini. Score 99 artinya bahwa model mampu mengidentifikasi dan menangkap data yang memiliki potensi pembatalan booking hotel dengan baik sesuai dengan tujuan penelitian yang dilakukan.



# RECOMENDATION



## Rekomendasi agar performa model lebih baik :

- Untuk meningkatkan hasil analisis, pihak hotel memperbaiki sistem website/dashboard hotel yang digunakan oleh user dengan penggunaan Menu Dropdown atau pilihan yang dibatasi untuk menghindari human error
- Menetapkan aturan input yang jelas dan mengkomunikasikannya kepada pengguna, seperti format tanggal yang diharapkan, tipe data yang diperbolehkan, dan sebagainya.
- Menambahkan kolom atau informasi tentang identitas pemesan atau unique ID pemesan

## Rekomendasi berdasarkan Data Analisis :

- Membatasi waktu pemesanan ke waktu check in maksimal tidak lebih dari setahun
- Menerapkan sistem reschedule/refund dan memberikan kebijakan proporsional penalty
- Meningkatkan pemesanan booking pada winter dan mengurangi cancellation pada summer yang merupakan high season
- Memberikan penawaran menarik berupa promosi free sarapan untuk pemesan self catering
- Menawarkan kebutuhan special request kepada pemesan yang tidak melakukan special request terlebih dahulu
- Membuat paket promo kepada type transient pada waktu akhir pekan
- Menerapkan sistem deposit dan reschedule sebagai pertimbangan customer melakukan cancellation
- Menerapkan sistem double booking pada high season atau pada bulan rentan tinggi terjadinya cancel seperti pada bulan Agustus
- Membuat campaign yang ditargetkan berupa promo liburan atau promo akhir pekan dengan target audience type customer transient

# **TERIMA KASIH**