



# DATA SCIENCE UNTUK SOLUSI RESERVASI HOTEL

Data Science Challenge 2023





# MY PROFILE

Hi, Saya Maurino Audrian Putra, mahasiswa penuh semangat dari Jurusan Matematika di Universitas Islam Bandung. Dengan gembira, saya ingin berbagi proyek terbaru saya yang bertajuk 'Data Science untuk Solusi Reservasi Hotel.' Mari kita bersama-sama menjelajahi bagaimana Data Science dapat memberikan solusi inovatif dalam dunia perhotelan!"



Maurino Audrian Putra



audrianmaurino@gmail.com



# TABLE OF CONTENTS

01

Define Questions  
& Goals

02

Gathering Data

03

Assessing Data

04

Data Cleaning

05

Deep Analytics &  
EDA

06

Feature  
Engineering

07

Machine  
Learning

08

Evaluation

09

Rekomendasi  
Bisnis



# 01

DEFINE  
QUESTIONS &  
GOALS



# RESERVASI HOTEL

Hotel adalah bisnis akomodasi yang menyediakan kamar untuk tamu dengan fasilitas seperti tempat tidur, kamar mandi, dan layanan tambahan.

Teknologi juga memainkan peran penting dalam operasi hotel, termasuk reservasi online dan sistem manajemen properti.



Hotel





# 1. GOALS & QUESTION



Meningkatkan  
profitabilitas dan  
kepuasan  
pelanggan.

1. Bagaimana distribusi **lead time**?
2. Apakah ada **tren musiman** dalam reservasi sepanjang tahun?
3. **Jenis pelanggan** yang paling sering reservasi?
4. **Deposit** jenis apa yang sukses dalam reservasi?
5. Top 5 **negara** dengan tingkat keberhasilan reservasi tertinggi?
6. **Korelasi** antara lead time dan permintaan kamar?
7. Jenis **meal** favorit pelanggan?
8. **Market segment** dengan tingkat keberhasilan reservasi tertinggi?
9. Cara **reservasi** atau **penjualan** dilakukan?
10. Persentase pelanggan sebagai **Repeated Guest**?
11. Berapa banyak **pelanggan tidak mendapatkan tipe kamar yang sesuai**?
12. **RFM** Analysis



# 02

GATHERING  
DATA



## 2. GATHERING DATA

Dataset yang digunakan diperoleh dari link GDrive berikut:

<https://bit.ly/DatasetChallengeDS>

FileHomeInsertPage LayoutFormulasDataReviewViewHelpTell me what you want to do

Clipboard

Format Painter

Font

Alignment

Number

Styles

Conditional Formatting

Cell Styles

Insert

Delete

Format

Format

AutoSum

Fill

Sort & Filter

Find & Select

Add-ins

Add-ins

Editing

Add-ins

A1

fx

hotel

	A	B	C	D	E	F	G	H	I	J	
		lead_time	arrival_date_year	arrival_date_mn	arrival_date_wk	arrival_date_da	stays_in_week	stays_in_week_adults	children	babies	
1	hotel										
2	Fort Worth University Drive Courtyard Fort Worth, TX	46	2019 July	27	3	1	0	2	0		
3	Sacramento Airport Natomas Courtyard Sacramento, CA	98	2017 November	48	23	1	2	1	0		
4	Residence Inn Las Vegas Convention Center Las Vegas, NV	171	2019 July	30	28	0	1	2	0		
5	Courtyard by Marriott Rio de Janeiro Barra da Tijuca Barra da Tijuca, Brazil	2	2018 January	1	2	0	1	2	0		
6	Leipzig Marriott Hotel Leipzig, Germany	167	2019 May	20	14	2	1	2	0		
7	Marriott Puerto Vallarta Resort & Spa Puerto Vallarta, Mexico	69	2018 July	29	16	2	3	0			
8	Protea Hotel by Marriott Midrand Midrand, South Africa	30	2018 March	10	5	2	3	2	0		
9	Courtyard by Marriott Rio de Janeiro Barra da Tijuca Barra da Tijuca, Brazil	229	2018 July	28	12	1	4	2	0		
10	Hotel Maria Cristina, San Sebastian San Sebastian, Spain	175	2019 April	17	29	2	5	2	0		
11	Treasure Beach, Barbados Barbados	211	2019 May	21	24	0	4	2	0		
12	Renaissance Hamburg Hotel Hamburg, Germany	262	2018 October	41	7	2	3	2	0		
13	Renaissance Hamburg Hotel Hamburg, Germany	222	2017 September	37	6	4	6	2	0		
14	Waves, Barbados Barbados	4	2017 July	31	31	0	2	2	0		
15	Courtyard by Marriott Rio de Janeiro Barra da Tijuca Barra da Tijuca, Brazil	184	2019 August	32	7	1	2	2	0		
16	Courtyard by Marriott Aberdeen Airport Aberdeen, United Kingdom	257	2019 May	19	9	0	4	2	0		
17	Orlando International Drive/Convention Center Courtyard Orlando, FL	444	2018 December	49	1	0	3	1	0		
18	Turtle Beach, Barbados Barbados	274	2018 April	15	7	0	2	2	0		
19	Berlin Marriott Hotel Berlin, Germany	0	2019 March	10	5	1	0	1	0		
20	Sacramento Airport Natomas Courtyard Sacramento, CA	2	2017 November	45	4	0	2	1	0		
21	Des Moines West/Clive Courtyard Clive, IA	186	2019 July	28	15	4	6	3	0		
22	Indianapolis Airport Courtyard Indianapolis, IN	26	2018 April	15	4	1	0	1	0		
23	African Pride Melrose Arch, Autograph Collection Johannesburg, South Africa	56	2019 July	30	24	1	4	2	0		
24	The St. Regis Osaka Osaka, Japan	95	2019 August	32	12	2	3	2	0		
25	Protea Hotel Fire & Ice by Marriott Johannesburg Melrose Arch Johannesburg, South Africa	192	2017 December	50	10	0	1	2	0		
26	Renaissance Hamburg Hotel Hamburg, Germany	151	2018 January	4	19	0	3	1	0		
27	Treasure Beach, Barbados Barbados	290	2019 July	30	29	0	2	2	0		

Ready

Accessibility: Investintate





Data Assessment

# 03

## ASSESSING DATA



# 3. ASSESSING DATA

Proses yang bertujuan untuk **mengidentifikasi masalah** yang terdapat dalam data dan **memastikan data tersebut berkualitas**.

Tampilan 5 baris teratas dari dataset

	hotel	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week
0	Fort Worth University Drive Courtyard Fort Wor...	46	2019	July	27	3		1
1	Sacramento Airport Natomas Courtyard Sacrament...	98	2017	November	48	23		1
2	Residence Inn Las Vegas Convention Center Las ...	171	2019	July	30	28		0
3	Courtyard by Marriott Rio de Janeiro Barra da ...	2	2018	January	1	2		0
4	Leipzig Marriott Hotel Leipzig, Germany	167	2019	May	20	14		2

Berikut adalah penjelasan untuk setiap kolom pada dataset:

1. **hotel**: Nama hotel atau penginapan.
2. **lead\_time**: Jumlah hari antara tanggal pemesanan dan tanggal kedatangan.
3. **arrival\_date\_year**: Tahun kedatangan tamu.
4. **arrival\_date\_month**: Bulan kedatangan tamu.
5. **arrival\_date\_week\_number**: Nomor minggu dalam tahun untuk tanggal kedatangan.
6. **arrival\_date\_day\_of\_month**: Hari kedatangan tamu dalam bulan.
7. **stays\_in\_weekend\_nights**: Jumlah malam yang dihabiskan tamu di akhir pekan.
8. **stays\_in\_week\_nights**: Jumlah malam yang dihabiskan tamu pada hari kerja.
9. **adults**: Jumlah tamu dewasa.
10. **children**: Jumlah anak-anak yang menyertai tamu.
11. **babies**: Jumlah bayi yang menyertai tamu.
12. **meal**: Tipe paket makanan yang dipilih oleh tamu (misalnya, BB untuk Bed & Breakfast).
13. **country**: Negara asal tamu.
14. **market\_segment**: Segmen pasar melalui mana pemesanan dilakukan (misalnya, Online TA untuk Travel Agents online).
15. **distribution\_channel**: Saluran distribusi pemesanan (misalnya, TA/TO untuk Travel Agents/Tour Operators).
16. **is\_repeated\_guest**: Menunjukkan apakah tamu merupakan pengunjung berulang atau tidak (0 untuk tidak, 1 untuk ya).
17. **previous\_cancellations**: Jumlah pemesanan yang telah dibatalkan oleh tamu sebelumnya.
18. **previous\_bookings\_not\_canceled**: Jumlah pemesanan yang tidak dibatalkan oleh tamu sebelumnya.
19. **reserved\_room\_type**: Tipe kamar yang telah dipesan oleh tamu.
20. **assigned\_room\_type**: Tipe kamar yang sebenarnya diberikan kepada tamu saat check-in.
21. **booking\_changes**: Jumlah perubahan yang dilakukan pada pemesanan.
22. **deposit\_type**: Jenis deposit yang telah dibayarkan oleh tamu (misalnya, No Deposit untuk tanpa deposit).
23. **agent**: ID agen perjalanan yang membuat pemesanan.
24. **company**: ID perusahaan yang membuat pemesanan.
25. **days\_in\_waiting\_list**: Jumlah hari pemesanan berada dalam daftar tunggu sebelum dikonfirmasi.
26. **customer\_type**: Jenis tamu (misalnya, Transient untuk tamu individu).
27. **adr**: Average Daily Rate, rata-rata biaya kamar per malam.
28. **required\_car\_parking\_spaces**: Jumlah tempat parkir yang dibutuhkan oleh tamu.
29. **total\_of\_special\_requests**: Jumlah permintaan khusus oleh tamu.
30. **reservation\_status**: Status pemesanan terakhir (misalnya, Check-Out untuk tamu yang sudah check-out).
31. **reservation\_status\_date**: Tanggal terakhir diupdate status pemesanan.
32. **bookingid**: ID unik untuk setiap pemesanan.



## 3.1 Pengecekan Missing Value

```
# cek missing value
df.isna().sum()

hotel                0
lead_time            0
arrival_date_year    0
arrival_date_month   0
arrival_date_week_number 0
arrival_date_day_of_month 0
stays_in_weekend_nights 0
stays_in_week_nights 0
adults              0
children            1
babies              0
meal                0
country              142
market_segment       0
distribution_channel 0
is_repeated_guest    0
previous_cancellations 0
previous_bookings_not_canceled 0
reserved_room_type   0
assigned_room_type    0
booking_changes       0
deposit_type         0
agent                4925
company              33636
days_in_waiting_list 0
customer_type        0
adr                  0
required_car_parking_spaces 0
total_of_special_requests 0
reservation_status    0
reservation_status_date 0
bookingID            0
dtype: int64
```

Informasi di atas menunjukkan terdapat **1** missing value pada kolom **children**, **142** pada kolom **country**, **4925** pada kolom **agent**, dan **33636** (**94%** dari keseluruhan data) pada kolom **company**.

Untuk **kolom company** akan **dihapus** karena jika sebagian besar data pada suatu kolom hilang, informasi yang tersedia tidak dapat dianggap sebagai representasi yang baik dari kolom tersebut.

Terkait hal itu dan kolom-kolom yang terdapat missing value lainnya akan ditangani pada tahap data cleaning nanti.



## 3.2 Pengecekan Data yang Duplikat

```
# cek data duplikat  
print("Jumlah duplikasi: ", df.duplicated().sum())  
  
Jumlah duplikasi: 0
```

Berdasarkan informasi di atas, Data **tidak memiliki** baris yang **duplikat**.



## 3.3 Pengecekan Parameter Statistik

```
# parameter statistik  
df.describe()
```

	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	
count	35698.000000	35698.000000	35698.000000	35698.000000	35698.000000	35698.000000	35698.000000	35698.000000
mean	103.421760	2018.162782	27.229257	15.787607	0.925626	2.497115	1.859376	35698.000000
std	106.223079	0.708424	13.576197	8.782366	0.987463	1.876008	0.513811	35698.000000
min	0.000000	2017.000000	1.000000	1.000000	0.000000	0.000000	0.000000	35698.000000
25%	18.000000	2018.000000	16.000000	8.000000	0.000000	1.000000	2.000000	35698.000000
50%	69.000000	2018.000000	28.000000	16.000000	1.000000	2.000000	2.000000	35698.000000
75%	160.000000	2019.000000	38.000000	23.000000	2.000000	3.000000	2.000000	35698.000000
max	709.000000	2019.000000	53.000000	31.000000	16.000000	41.000000	27.000000	35698.000000

Hasil di atas menunjukkan bahwa **tidak terdapat keanehan** dalam ringkasan parameter statistik dari dataset.



# 04

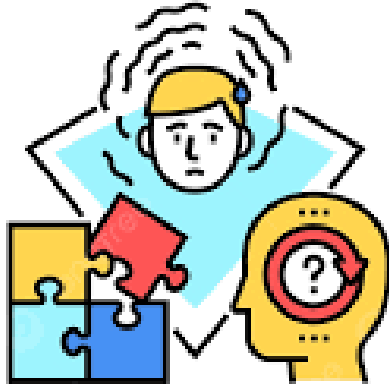
DATA  
CLEANING



## 4. DATA CLEANING

Proses yang bertujuan untuk **membersihkan** berbagai **masalah** yang telah teridentifikasi dalam proses assessing data

### PROBLEM



### MASALAH YANG TELAH TERIDENTIFIKASI:

- **Tipe data** dari kolom children, agent, dan company seharusnya adalah int
- **Missing value** kolom children, country, agent, company.
- Untuk **kolom company** akan **dihapus** karena jika sebagian besar data pada suatu kolom hilang,



## 4.1 Menangani Missing Value (1)

### 4.1.1 Kolom children

```
# cek kolom children yang mengandung missing value  
df[df.children.isna()]
```

ie	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	children
2	2017	August	32	3	1	0	2	NaN

Missing value pada kolom children **tidak dihapus**, tetapi **diimputasi dengan nilai median** untuk menjaga informasi penting.

```
# isi missing value dengan median  
df['children'] = df.children.fillna(df.children.median())
```

#### **NOTE:**

Menggunakan median untuk mengisi missing value dianggap lebih stabil dan kurang dipengaruhi oleh outliers dibandingkan dengan mean.





## 4.1 Menangani Missing Value (2)

### 4.1.2 Kolom country

```
# cek kolom country yang mengandung missing value  
df[df.country.isna()]
```

	hotel	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_w
18	Sacramento Airport Natomas Courtyard Sacrament...	2	2017	November	45	4	0	
155	Residence Inn Las Vegas Convention Center Las ...	0	2018	October	44	24	1	
317	Frankfurt Marriott Hotel Frankfurt, Germany	0	2018	March	13	22	0	
1119	W London – Leicester Square London, United Kin...	0	2017	November	48	26	0	
1949	W London – Leicester Square London, United Kin...	2	2019	January	4	25	0	
...	...	...	...	...	...	...	...	...

karena **kolom country** mengandung hanya **142 (0.4%) missing value** dari keseluruhan data dan jika dihapus tidak akan menimbulkan efek yang signifikan terhadap analisis data.

maka **missing value** pada kolom **country di-drop**



## 4.1 Menangani Missing Value (3)

### 4.1.4 Kolom company

```
# cek kolom company yang mengandung missing value  
df[df.company.isna()]
```

	hotel	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_w
0	Fort Worth University Drive Courtyard Fort Wor...	46	2019	July	27	3	1	
1	Sacramento Airport Natomas Courtyard Sacrament...	98	2017	November	48	23	1	
2	Residence Inn Las Vegas Convention Center Las ...	171	2019	July	30	28	0	
3	Courtyard by Marriott Rio de Janeiro Barra da ...	2	2018	January	1	2	0	
4	Leipzig Marriott Hotel Leipzig, Germany	167	2019	May	20	14	2	
...	...	...	...	...	...	...	...	...

Kolom **company** di-drop karena mengandung lebih dari 50%, yaitu 33636 missing value atau **94%** dari keseluruhan data.

Jika lebih dari setengah data dalam suatu kolom hilang, **informasi** yang dapat diambil dari kolom tersebut menjadi sangat **tidak pasti**.



## 4.2 Menangani Tipe Data yang Salah (1)

**Kolom company di-drop** karena mengandung lebih dari 50%, yaitu 33636 missing value atau **94%** dari keseluruhan data.

Jika lebih dari setengah data dalam suatu kolom hilang, **informasi** yang dapat diambil dari kolom tersebut menjadi sangat **tidak pasti**.

```
# cek tipe data
df.dtypes

hotel                object
lead_time            int64
arrival_date_year     int64
arrival_date_month   object
arrival_date_week_number int64
arrival_date_day_of_month int64
stays_in_weekend_nights int64
stays_in_week_nights int64
adults               int64
children             float64
babies               int64
meal                 object
country              object
market_segment       object
distribution_channel object
is_repeated_guest    int64
previous_cancellations int64
previous_bookings_not_canceled int64
reserved_room_type   object
assigned_room_type   object
booking_changes       int64
deposit_type         object
agent                float64
days_in_waiting_list int64
customer_type         object
adr                  float64
required_car_parking_spaces int64
total_of_special_requests int64
reservation_status    object
reservation_status_date datetime64[ns]
bookingID             int64
dtype: object
```



## 4.2 Menangani Tipe Data yang Salah (2)

- Mengganti tipe data kolom '**children**' dari **float** ke **integer**
- Mengganti tipe data kolom '**agent**' dari **float** ke **integer**
- Mengubah tipe data kolom '**arrival\_data\_month**' nama bulan menjadi urutan bulan

# cek tipe data df.dtypes	
hotel	object
lead_time	int64
arrival_date_year	int64
arrival_date_month	object
arrival_date_week_number	int64
arrival_date_day_of_month	int64
stays_in_weekend_nights	int64
stays_in_week_nights	int64
adults	int64
children	float64
babies	int64
meal	object
country	object
market_segment	object
distribution_channel	object
is_repeated_guest	int64
previous_cancellations	int64
previous_bookings_not_canceled	int64
reserved_room_type	object
assigned_room_type	object
booking_changes	int64
deposit_type	object
agent	float64
days_in_waiting_list	int64
customer_type	object
adr	float64
required_car_parking_spaces	int64
total_of_special_requests	int64
reservation_status	object
reservation_status_date	datetime64[ns]
bookingID	int64
dtype:	object



# 05

DEEP  
ANALYTICS &  
EDA



## 5. DEEP ANALYTICS & EXPLORATORY DATA ANALYSIS (EDA)

### Tujuan:

Keduanya bertujuan untuk menggali wawasan dari data.

### Teknik Analisis:

Deep Analytics: Menggunakan teknik analisis canggih seperti model statistik kompleks.

EDA: Terlibat dalam visualisasi dan statistik deskriptif.

### Langkah Awal:

EDA sering langkah awal sebelum deep analytics.

### Pengambilan Keputusan:

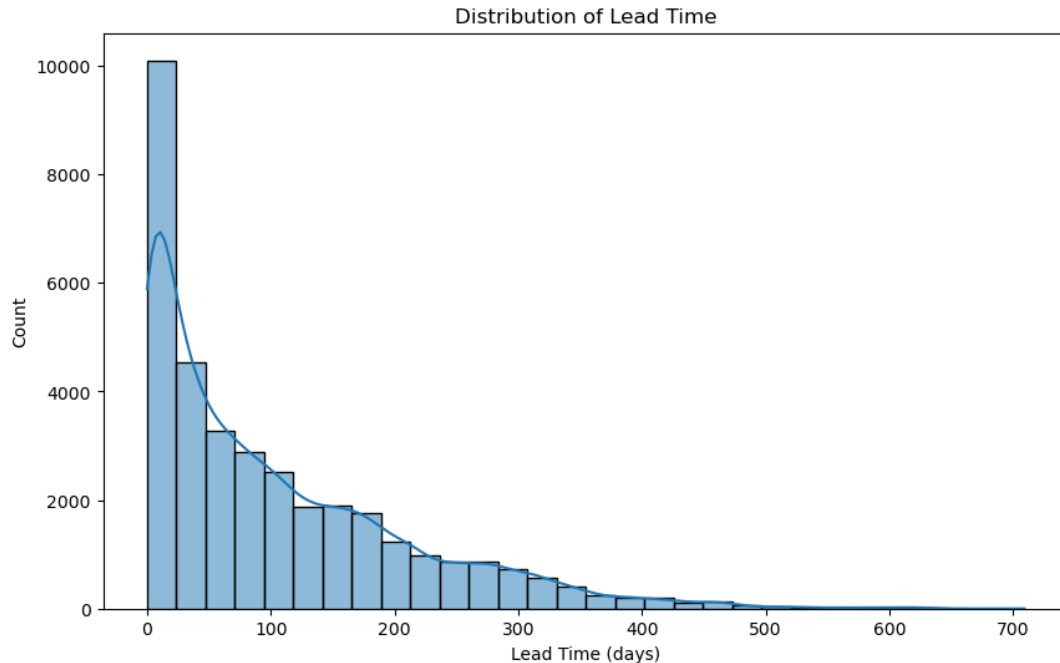
Keduanya mendukung pengambilan keputusan, dengan deep analytics memberikan wawasan mendalam.





# 5.1 Deep Analytics & EDA

Bagaimana Distribusi lead time (waktu antara pemesanan dan kedatangan)?



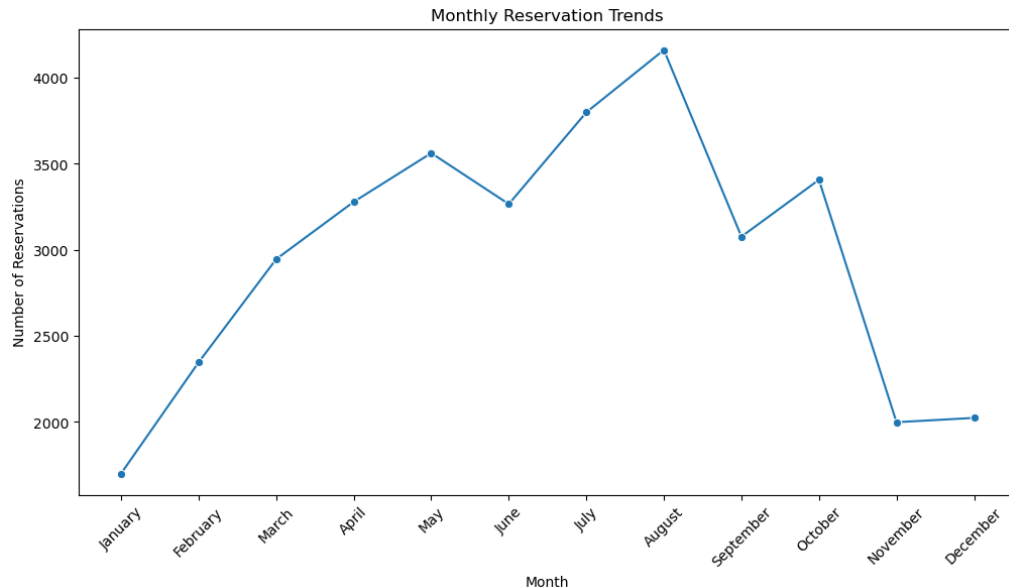
Distribusi Lead Time berjenis **Positive Skew** (right-skewed dsitribution).

Ini dapat diartikan bahwa **sebagian besar reservasi memiliki lead time yang relatif pendek**, tetapi ada beberapa reservasi dengan lead time yang sangat panjang.



## 5.2 Deep Analytics & EDA

Apakah ada pola tren atau musiman dalam jumlah reservasi sepanjang tahun?



Berdasarkan line plot di samping, bulan **Agustus** memiliki jumlah reservasi **terbanyak**.

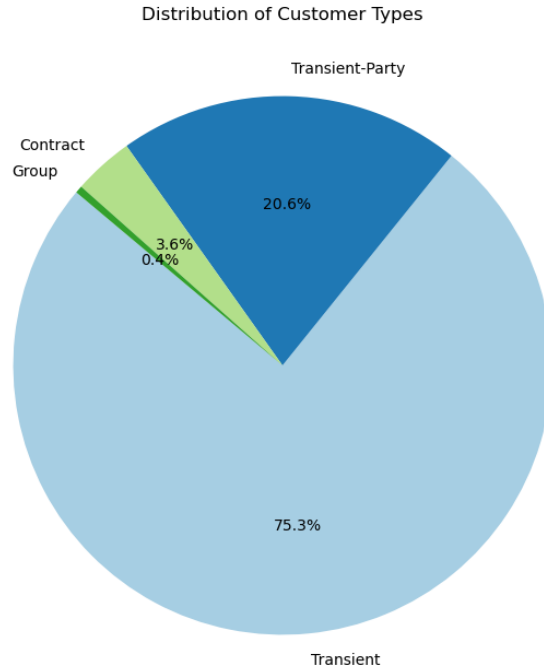
Ini mungkin dikarenakan bulan Agustus memiliki beberapa acara dan kejadian yang dapat menjadi faktor peningkatan jumlah reservasi hotel di Eropa, seperti liburan Musim Panas, festival dan acara, cuaca baik, dan musim liburan.





## 5.3 Deep Analytics & EDA

Apa jenis pelanggan yang paling sering melakukan reservasi?



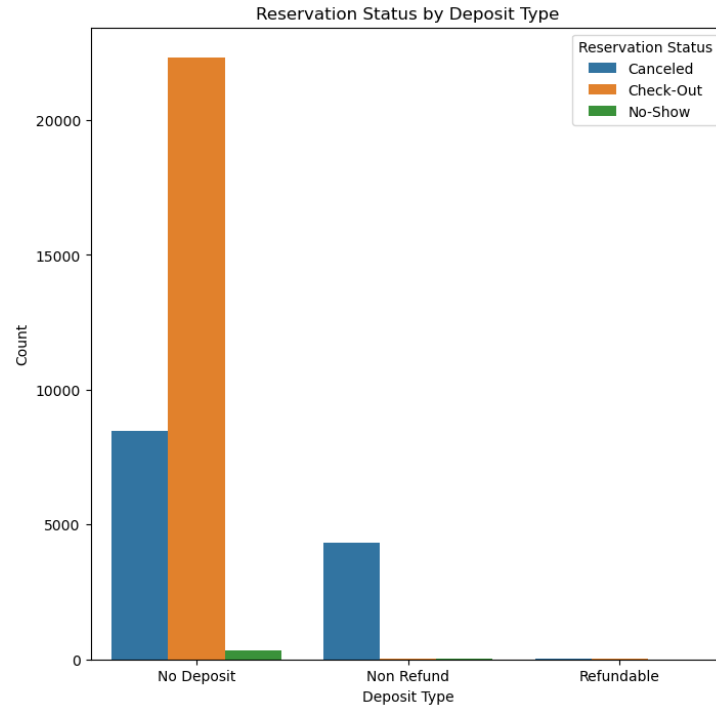
Berdasarkan perolehan di samping, **Transient** adalah jenis pelanggan yang paling sering melakukan reservasi

Pelanggan jenis "Transient" adalah pelanggan reguler atau individu yang melakukan reservasi tanpa adanya kontrak atau perjanjian khusus.



## 5.4 Deep Analytics & EDA

Jenis Deposit apa yang menghasilkan keberhasilan reservasi?



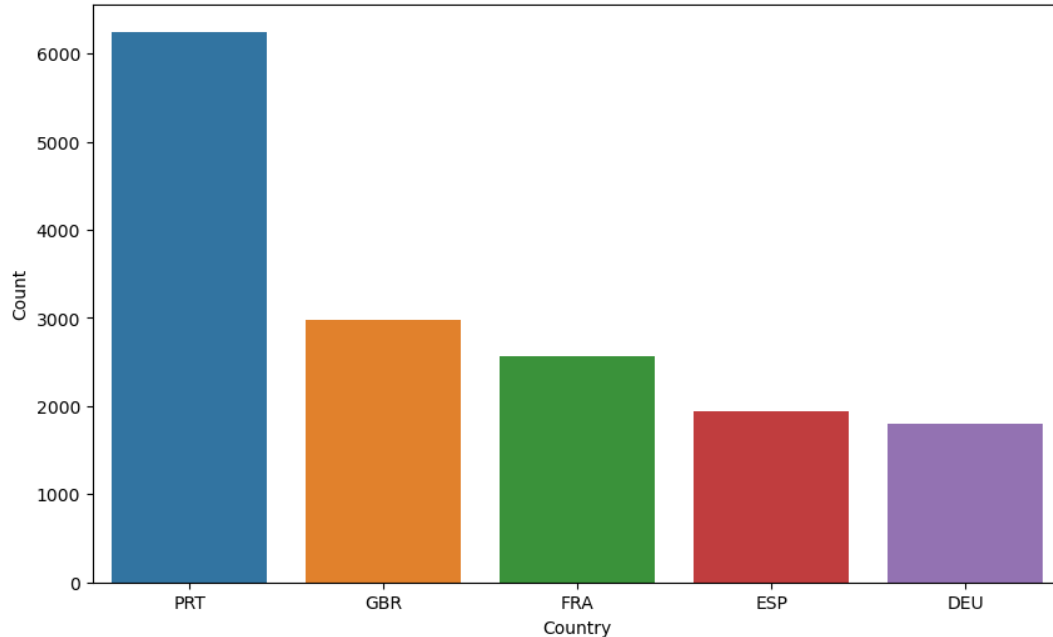
Berdasarkan informasi di samping, Jenis deposit menggunakan **No Deposit** memiliki tingkat **keberhasilan reservasi (Check-Out)** paling tinggi.



## 5.5 Deep Analytics & EDA

Top 5 Country yang tingkat keberhasilan reservasinya tertinggi?

Top 5 Countries with Highest Check-Outs



Berdasarkan standar ISO 3166-1 alpha-3, Top 5 Countries yang dimaksud adalah:

- **PRT:** Portugal
- **GBR:** United Kingdom (Inggris)
- **FRA:** France (Prancis)
- **ESP:** Spain (Spanyol)
- **DEU:** Germany (Jerman)



## 5.6 Deep Analytics & EDA

Apakah terdapat korelasi antara lead time dan jumlah permintaan kamar?

```
correlation_lead_time_booking_changes = df['lead_time'].corr(df['booking_changes'])  
print(f'Korelasi antara lead time dan booking changes: {correlation_lead_time_booking_changes}')
```

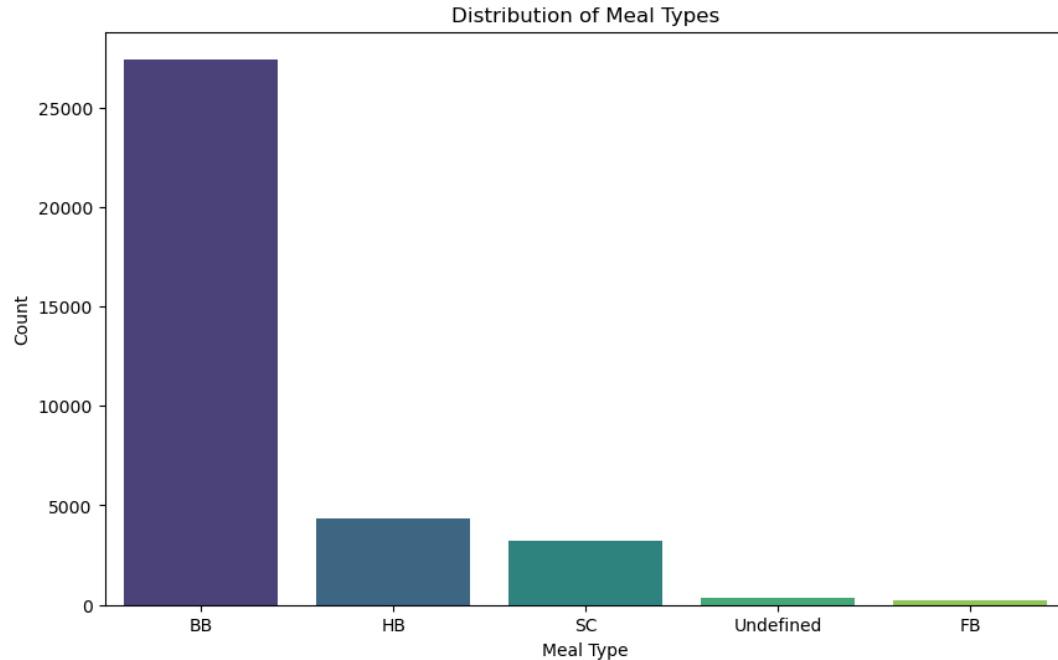
Korelasi antara lead time dan booking changes: 0.0016468119509200215

**Tidak ada korelasi** antara lead time dan jumlah permintaan kamar, karena nilai **korelasinya sangat rendah**



## 5.7 Deep Analytics & EDA

Jenis Meal Apa yang menjadi Favorit Pelanggan?



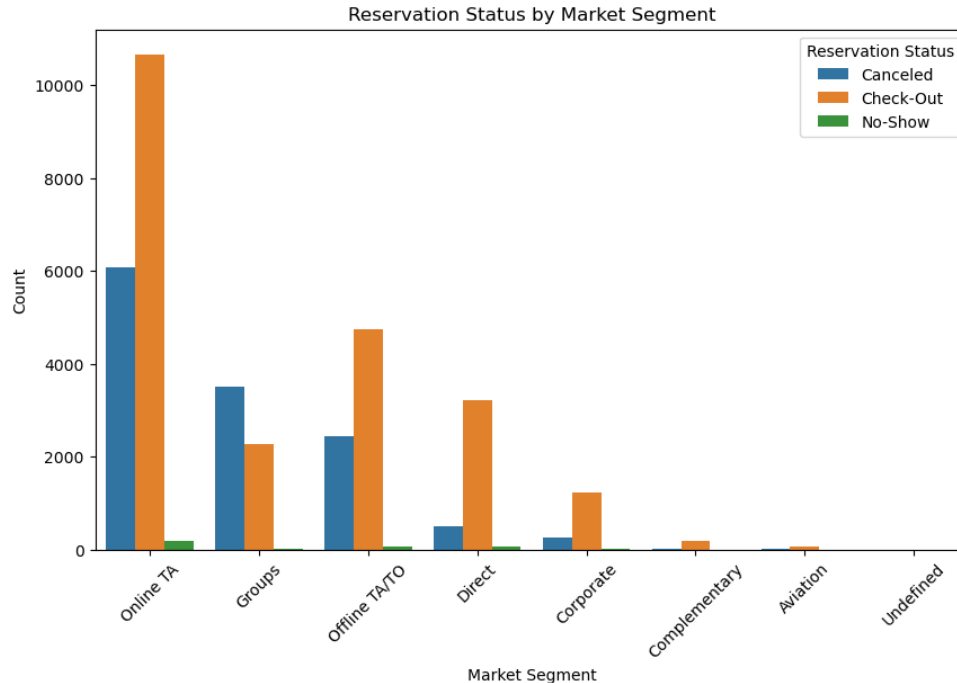
Berdasarkan informasi yang diperoleh, meal **BB (Bed & Breakfast)** lah menjadi favorit pelanggan

**BB (Bed & Breakfast)** adalah tamu yang hanya mendapatkan sarapan sebagai bagian dari paket. Makanan untuk waktu yang lain tidak termasuk



## 5.8 Deep Analytics & EDA

Market segment mana yang memiliki tingkat keberhasilan reservasi tertinggi?



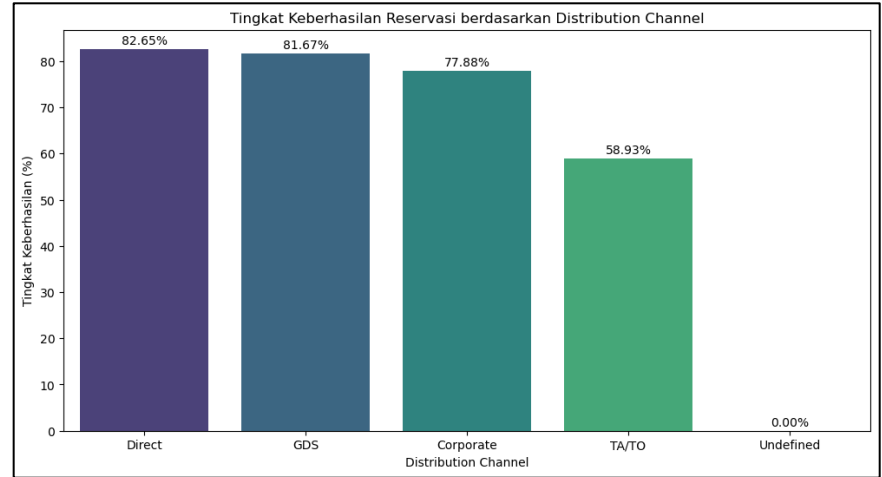
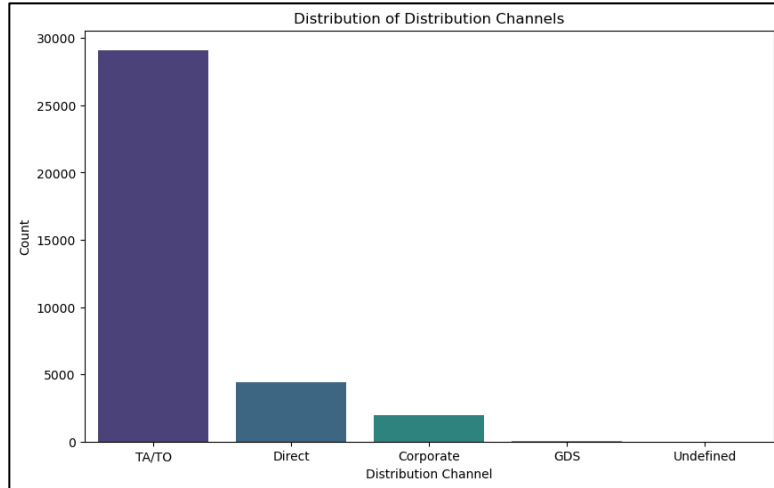
Berdasarkan informasi yang diperoleh, segment **Online TA** yang memiliki tingkat keberhasilan reservasi tertinggi.

Serta pada market segment **Groups** lebih banyak pelanggan meng-*cancel* daripada men-*check-out*



## 5.9 Deep Analytics & EDA

Bagaimana reservasi atau penjualan dilakukan?



Cara bagaimana reservasi atau penjualan dilakukan **terbanyak** adalah menggunakan **TA/TO** (Travel Agents/ Tour Operators)

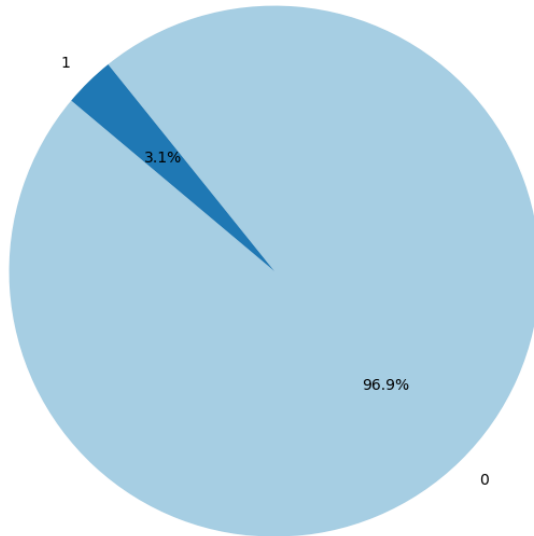
Namun, secara Tingkat keberhasilan pelanggan yang Check-Out, Reservasi secara **Direct** lah yang memiliki **tingkat keberhasilan tertinggi**, yaitu **82.65%**.



## 5.10 Deep Analytics & EDA

Berapa Persentase Pelanggan yang termasuk Repeated Guest?

Distribution of Repeated Guests



Berdasarkan hasil analisis, hanya **3.1%** yang merupakan **Repeated Guest** (berlabel 1).

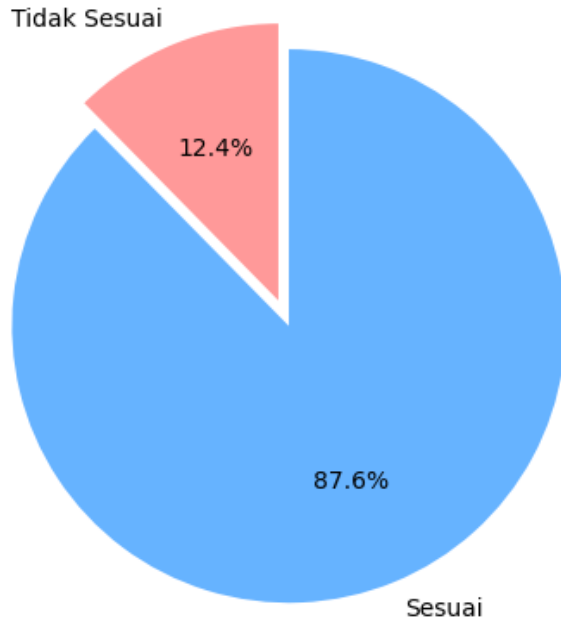




## 5.11 Deep Analytics & EDA

Berapa banyak pelanggan yang tidak mendapatkan tipe kamar yang sesuai?

Pemilihan Tipe Kamar Sesuai atau Tidak Sesuai

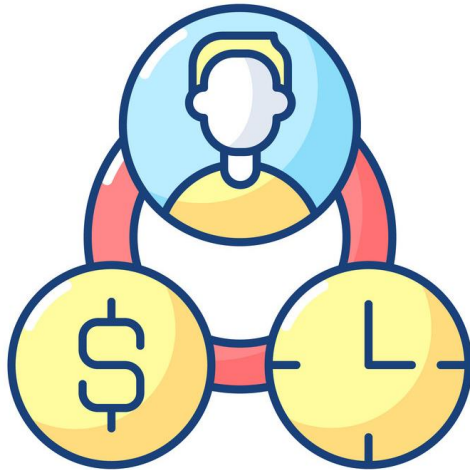


Berdasarkan Analisis yang dilakukan, ada sebanyak **12.4%** dari keseluruhan customer **tidak mendapat kamar yang sesuai**.



## 5.12 Deep Analytics & EDA (1)

### Definisi RFM Analysis



**RFM Analysis** adalah metode untuk mengelompokkan pelanggan berdasarkan tiga dimensi kunci

- **Recency** (kebaruan pembelian),
- **Frequency** (frekuensi pembelian),
- **Monetary** (nilai uang yang dihabiskan).

Dalam hal ini, 'reservation\_status\_date' digunakan untuk recency, jumlah reservasi untuk frequency, dan 'adr' (Average Daily Rate) untuk monetary.



## 5.12 Deep Analytics & EDA (2)

### Hasil RFM Analysis (1)

customer_type	recency	frequency	monetary
0	0	1283	113704.55
1	7	158	13702.32
2	0	26778	2883100.51
3	4	7337	633642.64



## 5.12 Deep Analytics & EDA (3)

### Hasil RFM Analysis (2)

#### Recency (R):

- 'Contract' dan 'Transient' memiliki recency rendah (0), menunjukkan pembelian baru-baru ini.
- 'Group' recency sedang (7), juga baru-baru ini melakukan pembelian.
- 'Transient-Party' recency sedang (4), relatif baru-baru ini melakukan pembelian.

#### Frequency (F):

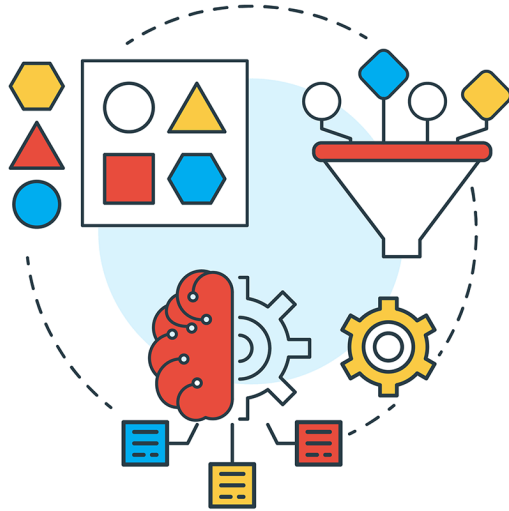
- 'Transient' memiliki frekuensi tinggi (26.778 reservasi).
- 'Contract' frekuensi tinggi (1.283 reservasi).
- 'Transient-Party' frekuensi menengah (7.337 reservasi).
- 'Group' frekuensi rendah (158 reservasi).

#### Monetary (M):

- 'Transient' nilai monetary tinggi.
- 'Contract' nilai monetary tinggi.
- 'Transient-Party' dan 'Group' nilai monetary lebih rendah.

#### Kesimpulan:

- 'Transient' dan 'Contract' mungkin segmen berharga dengan nilai R, F, dan M tinggi.
- 'Group' mungkin segmen lebih kecil, dengan nilai R dan F yang lebih rendah, tetapi tetap penting.



# 06

## FEATURE ENGINEERING



# 6 Feature Engineering



## 6.1 Total Guests

Fitur baru yang menunjukkan total jumlah tamu.

***Total Guests***

*= adults + children + babies*



# 6 Feature Engineering



## 6.2 Stay Duration

Fitur baru yang menghitung durasi tinggal tamu dengan menjumlahkan **stays\_in\_weekend\_nights** dan **stays\_in\_week\_nights**.



# 6 Feature Engineering

## 6.3 Categorical Encoding

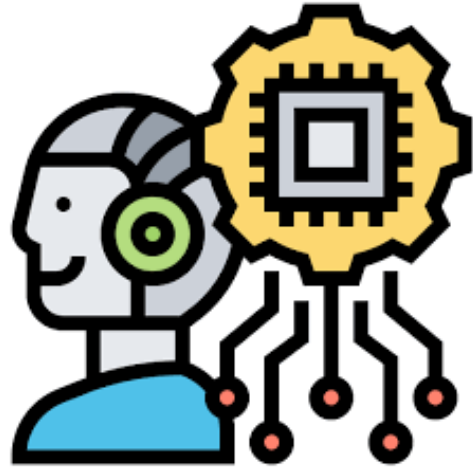
Mengonversi variabel-variabel kategorikal ke numerical representations.

```
from sklearn.preprocessing import LabelEncoder

# Categorical columns
categorical_columns = ['hotel', 'arrival_date_month',
                       'meal', 'country', 'market_segment',
                       'distribution_channel', 'reserved_room_type',
                       'assigned_room_type', 'deposit_type',
                       'customer_type']

# Apply Label Encoding to categorical columns
label_encoder = LabelEncoder()
for column in categorical_columns:
    df[column] = label_encoder.fit_transform(df[column])
```





# 07

Machine  
Learning Model



## 7.1 Feature Selection

```
# Pilih kolom yang akan digunakan sebagai fitur
features = ['lead_time', 'arrival_date_year',
            'stays_in_weekend_nights', 'stays_in_week_nights',
            'adults', 'children', 'babies',
            'is_repeated_guest', 'previous_cancellations',
            'previous_bookings_not_canceled', 'booking_changes',
            'adr', 'required_car_parking_spaces',
            'total_of_special_requests', 'room_type_match']

# Pilih kolom target
target = 'reservation_status'

# Buat dataframe fitur
X = df[features]

# Buat dataframe target
y = df[target]
```

[illegible]



## 7.3 Bangun Model Random Forest

```
from sklearn.ensemble import RandomForestClassifier

# Inisialisasi model
rf_model = RandomForestClassifier(random_state=42)

# Melatih model
rf_model.fit(X_train, y_train)
```

▼ RandomForestClassifier

RandomForestClassifier(random\_state=42)



# 08

Evaluation



## 8. Evaluation



ACCURACY

81.2%



F1-SCORE

81.2%



# 09

Rekomendasi  
Bisnis



# Rekomendasi Bisnis (1)



## Fokus Pemasaran di Bulan Agustus

Optimalkan promosi dan penawaran khusus pada bulan Agustus, bulan dengan jumlah reservasi tertinggi.



## Perkuat Pelanggan Tipe Transient:

Tingkatkan strategi pemasaran untuk menarik dan mempertahankan pelanggan tipe Transient, yang paling sering melakukan reservasi.





## Rekomendasi Bisnis (2)



### **Fleksibilitas Kebijakan Deposit**

Pertimbangkan kebijakan deposit yang lebih fleksibel atau model tanpa deposit untuk meningkatkan tingkat keberhasilan reservasi.



### **Targetkan Pasar Negara Utama**

Fokuskan pemasaran tambahan pada negara-negara dengan tingkat keberhasilan reservasi tertinggi.

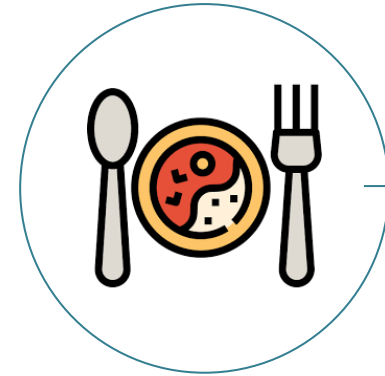


## Rekomendasi Bisnis (3)



### **Optimalkan Channel Distribusi**

Pertahankan hubungan dengan Travel Agents/Tour Operators dan pantau kinerja saluran distribusi untuk hasil terbaik.



### **Promosikan Meal Plan BB**

Tingkatkan promosi untuk meal plan Bed & Breakfast yang menjadi favorit pelanggan.



## Rekomendasi Bisnis (4)



### Peningkatan Layanan Repeated Guest

Fokus pada peningkatan layanan untuk meningkatkan retensi pelanggan Repeated Guest.



### Analisis RFM untuk Targeted Marketing

Manfaatkan analisis RFM untuk pemasaran yang lebih terarah, terutama pada pelanggan tipe 'Transient' dan 'Contract'.



## Rekomendasi Bisnis (5)



### **Analisis Cancellation untuk Market Segment Groups**

Selidiki penyebab pembatalan reservasi di market segment Groups dan atasi masalah yang mungkin muncul.



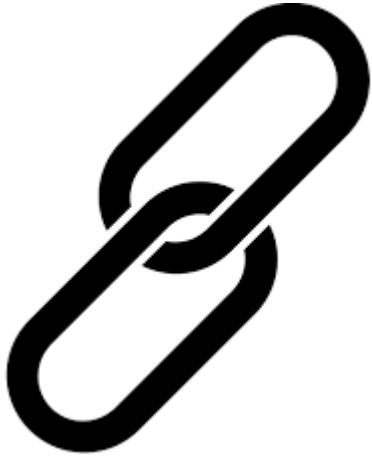
Kesuksesan

### **Peningkatan Ketersediaan Kamar Sesuai Permintaan**

Tingkatkan ketersediaan kamar sesuai permintaan pelanggan untuk mengurangi persentase pelanggan yang tidak mendapatkan tipe kamar yang diinginkan.



# Link Notebook



<https://colab.research.google.com/drive/1okxSZ8K141DnmgwNDRkyEG3RtCnP3AdF?usp=sharing>

**HAVE A GREAT DAY !**

