

PHÂN TÍCH VÀ DỰ ĐOÁN XU HƯỚNG NGÀNH KHÁCH SẠN





NỘI DUNG

01

02

03

04

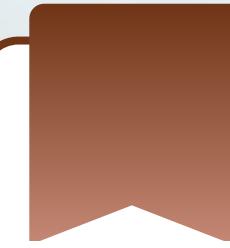
Giới thiệu

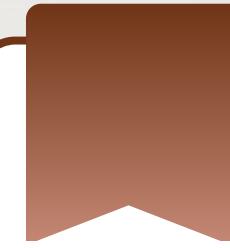
Xử lý dữ liệu

Phân tích

Đề xuất

1.GIỚI THIỆU

 Thành viên
nhóm

 Bộ dữ liệu

1. GIỚI THIỆU

THÀNH VIÊN NHÓM



HỒNG BẢO

Trưởng nhóm



DIỆU MY

Thành viên



DUY PHONG

Thành viên



HỒNG NHI

Thành viên

1. GIỚI THIỆU

Bộ dữ liệu **Hotel Booking Dataset** chứa thông tin về đặt phòng khách sạn từ tháng 7/2015 cho đến tháng 8/2018 bao gồm: Thông tin chi tiết về đặt phòng khách sạn của khách hàng và các thuộc tính của khách sạn
(nguồn: <https://www.kaggle.com/datasets/saadharoon27/hotel-booking-dataset/data>)

hotel: The type of hotel, either "City Hotel" or "Resort Hotel."

is_canceled: Binary value indicating whether the booking was cancelled (1) or not (0).

lead_time: Number of days between booking and arrival.

arrival_date_year: Year of arrival date.

arrival_date_month: Month of arrival date.

arrival_date_week_number: Week number of arrival date.

arrival_date_day_of_month: Day of the month of arrival date.

stays_in_weekend_nights: Number of weekend nights (Saturday or Sunday) the guest stays.

stays_in_week_nights: Number of weekday nights (Monday to Friday) the guest stays.

adults: Number of adults.

children: Number of children.

babies: Number of babies.

meal: Type of meal booked.

country: Country of origin.

market_segment: Market segment designation.

distribution_channel: Booking distribution channel.

is_repeated_guest: Binary value indicating whether the guest is a repeated guest (1) or not (0). **credit_card:** Guest's credit card details. (Not Real)

previous_cancellations: Number of previous booking cancellations.

previous_bookings_not_canceled: Number of previous bookings not cancelled.

reserved_room_type: Code of room type reserved.

assigned_room_type: Code of room type assigned at check-in.

booking_changes: Number of changes/amendments made to the booking.

deposit_type: Type of deposit made.

agent: ID of the travel agency.

company: ID of the company.

days_in_waiting_list: Number of days on the waiting list before booking.

customer_type: Type of booking.

adr: Average daily rate.

required_car_parking_spaces: Number of car parking spaces required.

total_of_special_requests: Number of special requests made.

reservation_status: Reservation last status.

reservation_status_date: Date of the last status.

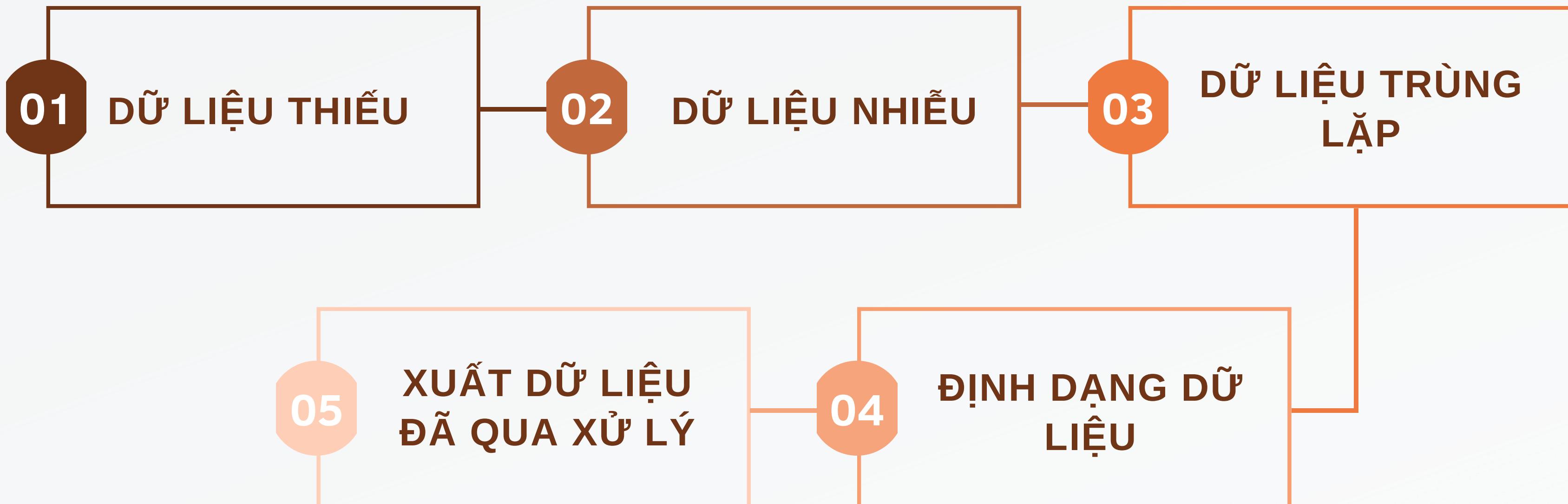
name: Guest's name. (Not Real)

email: Guest's email address.(Not Real)

phone-number: Guest's phone number. (Not Real)

2. XỬ LÝ DỮ LIỆU

 python™ Sử dụng Python để xử lý dữ liệu



2. XỬ LÝ DỮ LIỆU

DỮ LIỆU THIẾU

01

Cột children thiếu
4 giá trị

02

Cột company thiếu
112,593 giá trị

03

Cột agent thiếu
16,340 giá trị

04

Cột country thiếu
488 giá trị

Thay bằng 0

Giả định cho đặt phòng đó không có trẻ em

Thay bằng 0

Giả định cho việc không biết đến từ công ty nào

Thay bằng 0

Giả định cho việc không biết đến từ đơn vị OTA nào

Thay bằng 'Unknown'

Giả định việc không biết đến từ quốc gia nào

2. XỬ LÝ DỮ LIỆU

DỮ LIỆU NHIỀU

Dữ liệu ngoại lai

Có 1 giá trị âm trong cột adr
(Average Daily Rate)



Loại bỏ dòng giữ liệu chứa
giá trị âm

Dữ liệu nhất quán

180 booking trong bộ dữ liệu
mà không có khách
223 booking trong bộ dữ liệu
không có người lớn mà có trẻ
em



Loại bỏ 403 dòng dữ liệu
bất hợp lý

Dữ liệu không quan trọng

Tất cả các giá trị của các cột:
name, email, phone-number
và credit_card là thông tin giả
định



Loại bỏ 4 cột thông tin
giả định

2. XỬ LÝ DỮ LIỆU

DỮ LIỆU TRÙNG LẶP

- Không có dữ liệu trùng lặp

ĐỊNH DẠNG DỮ LIỆU

- Đa số các cột có định dạng phù hợp

- Cột reservation_status_date

object  datetime64[ns]

- Cột children, agent và company

float64  int

2. XỬ LÝ DỮ LIỆU

XUẤT DỮ LIỆU ĐÃ QUA XỬ LÝ

```
RangeIndex: 118986 entries, 0 to 118985
Data columns (total 32 columns):
 #   Column           Non-Null Count  Dtype  
 ---  -- 
 0   hotel            118986 non-null   object 
 1   is_canceled      118986 non-null   int64  
 2   lead_time         118986 non-null   int64  
 3   arrival_date_year 118986 non-null   int64  
 4   arrival_date_month 118986 non-null   object 
 5   arrival_date_week_number 118986 non-null   int64  
 6   arrival_date_day_of_month 118986 non-null   int64  
 7   stays_in_weekend_nights 118986 non-null   int64  
 8   stays_in_week_nights   118986 non-null   int64  
 9   adults            118986 non-null   int64  
 10  children          118986 non-null   int64  
 11  babies             118986 non-null   int64  
 12  meal               118986 non-null   object 
 13  country            118986 non-null   object 
 14  market_segment      118986 non-null   object 
 15  distribution_channel 118986 non-null   object 
 16  is_repeated_guest    118986 non-null   int64  
 17  previous_cancellations 118986 non-null   int64  
 18  previous_bookings_not_canceled 118986 non-null   int64  
 19  reserved_room_type    118986 non-null   object 
 20  assigned_room_type    118986 non-null   object 
 21  booking_changes       118986 non-null   int64  
 22  deposit_type          118986 non-null   object 
 23  agent                118986 non-null   int64  
 24  company               118986 non-null   int64  
 25  days_in_waiting_list 118986 non-null   int64  
 26  customer_type         118986 non-null   object 
 27  adr                  118986 non-null   float64 
 28  required_car_parking_spaces 118986 non-null   int64  
 29  total_of_special_requests 118986 non-null   int64  
 30  reservation_status     118986 non-null   object 
 31  reservation_status_date 118986 non-null   datetime64[ns]
dtypes: datetime64[ns](1), float64(1), int64(19), object(11)
```

3. PHÂN TÍCH

Phân tích
xu hướng
của ngành
khách sạn



01

Sự biến động theo mùa

02

Phân khúc thị trường và kênh phân phối

03

Phân khúc khách hàng

04

Thời gian lưu trú

05

Tỷ lệ hủy đặt phòng

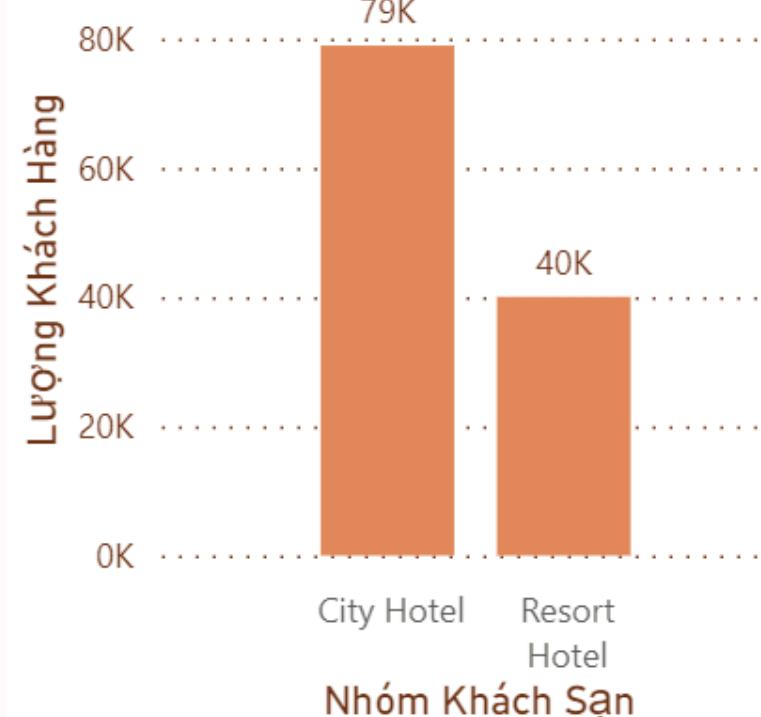
3. PHÂN TÍCH

OVERVIEW

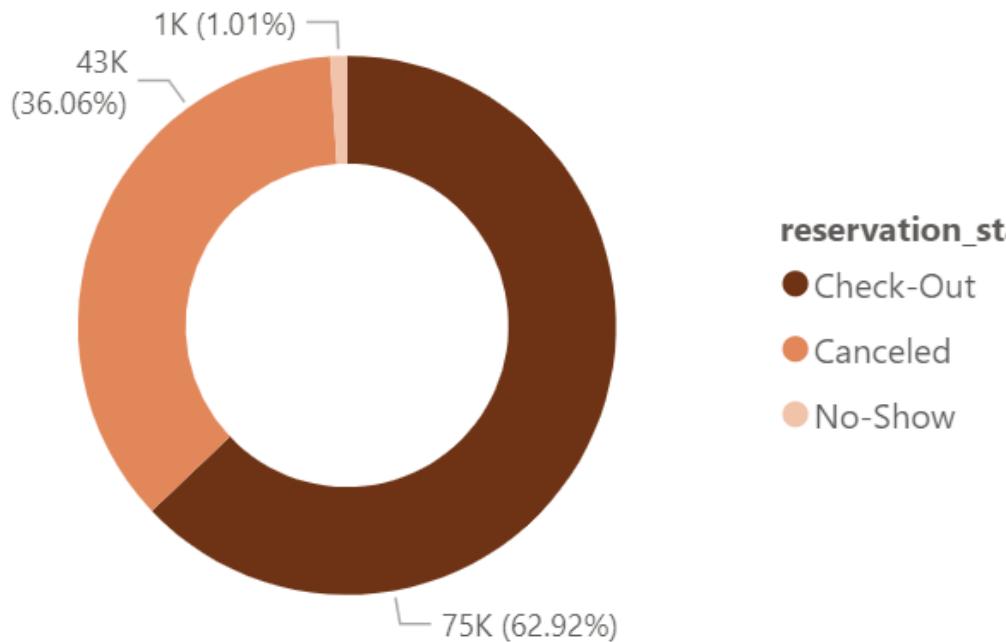
Tổng khách hàng
118,986

Tổng thành phố
178

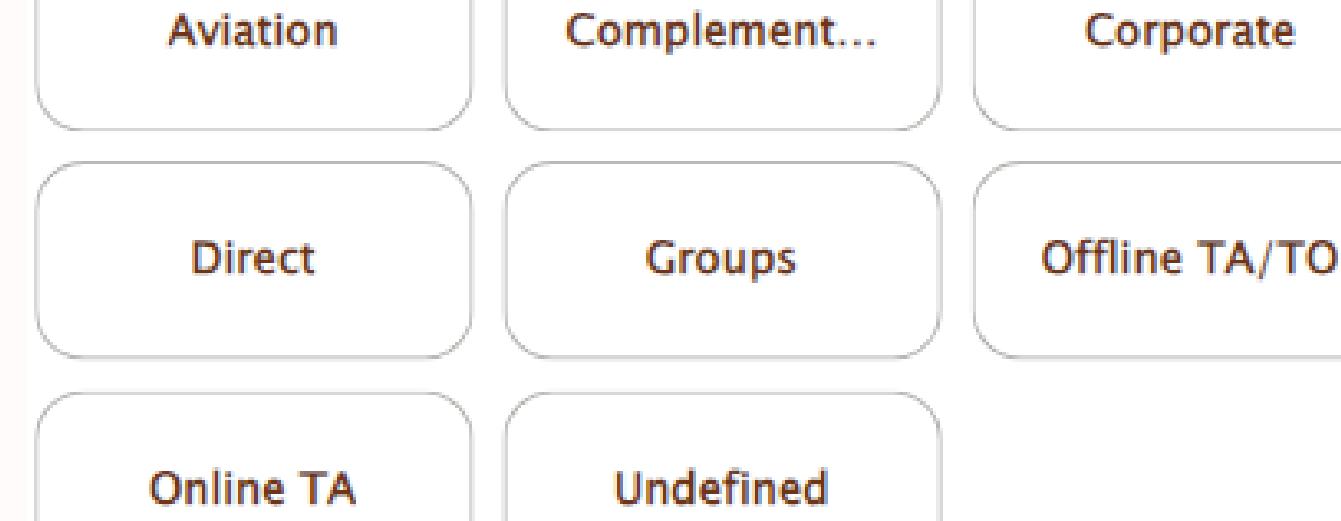
Lượng Khách Hàng theo Nhóm Khách Sạn



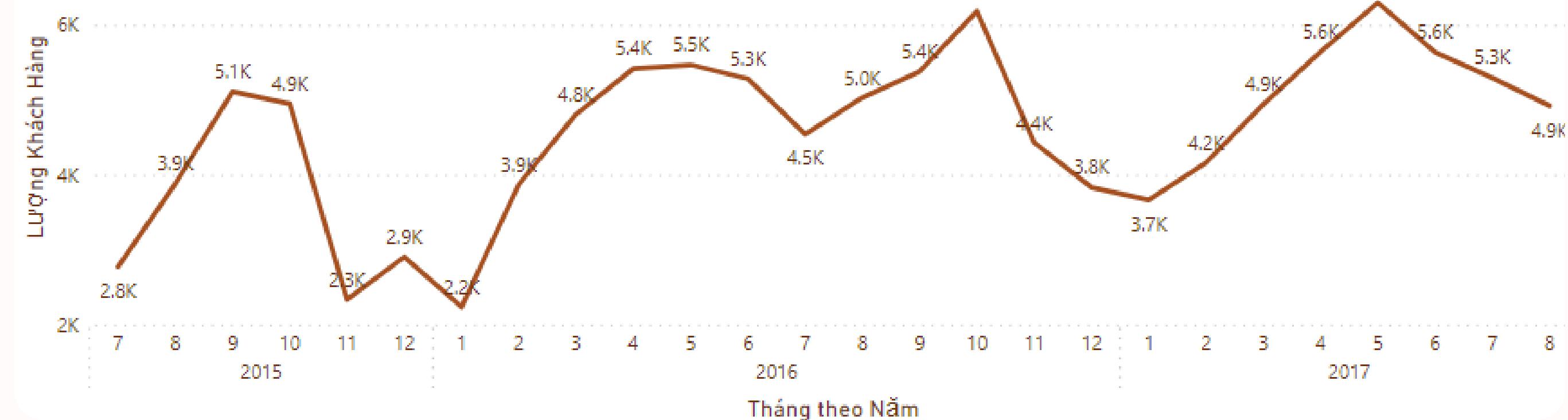
Tổng lượng khách hàng theo tình trạng đặt chỗ



Phân Khúc Thị Trường



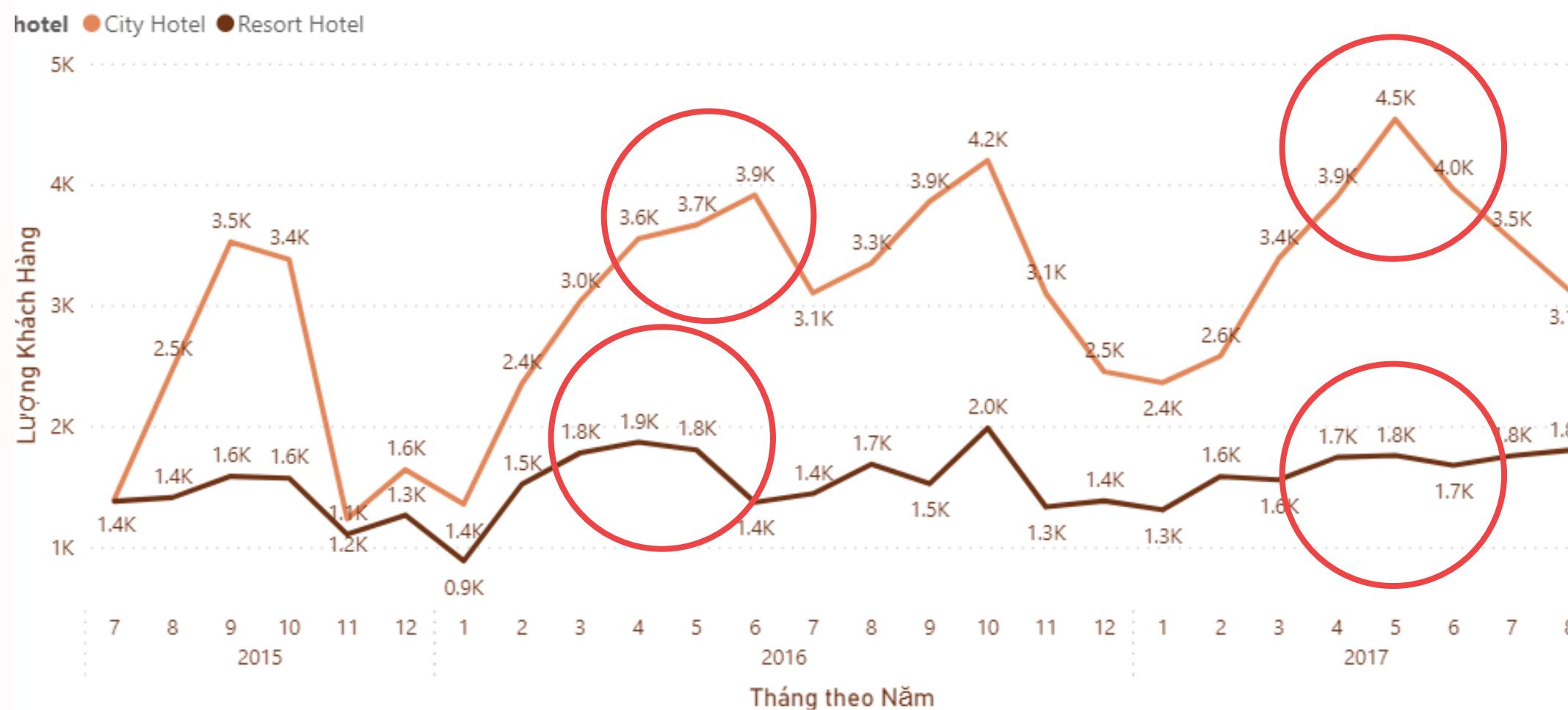
Tổng lượng khách hàng theo Tháng trong Năm



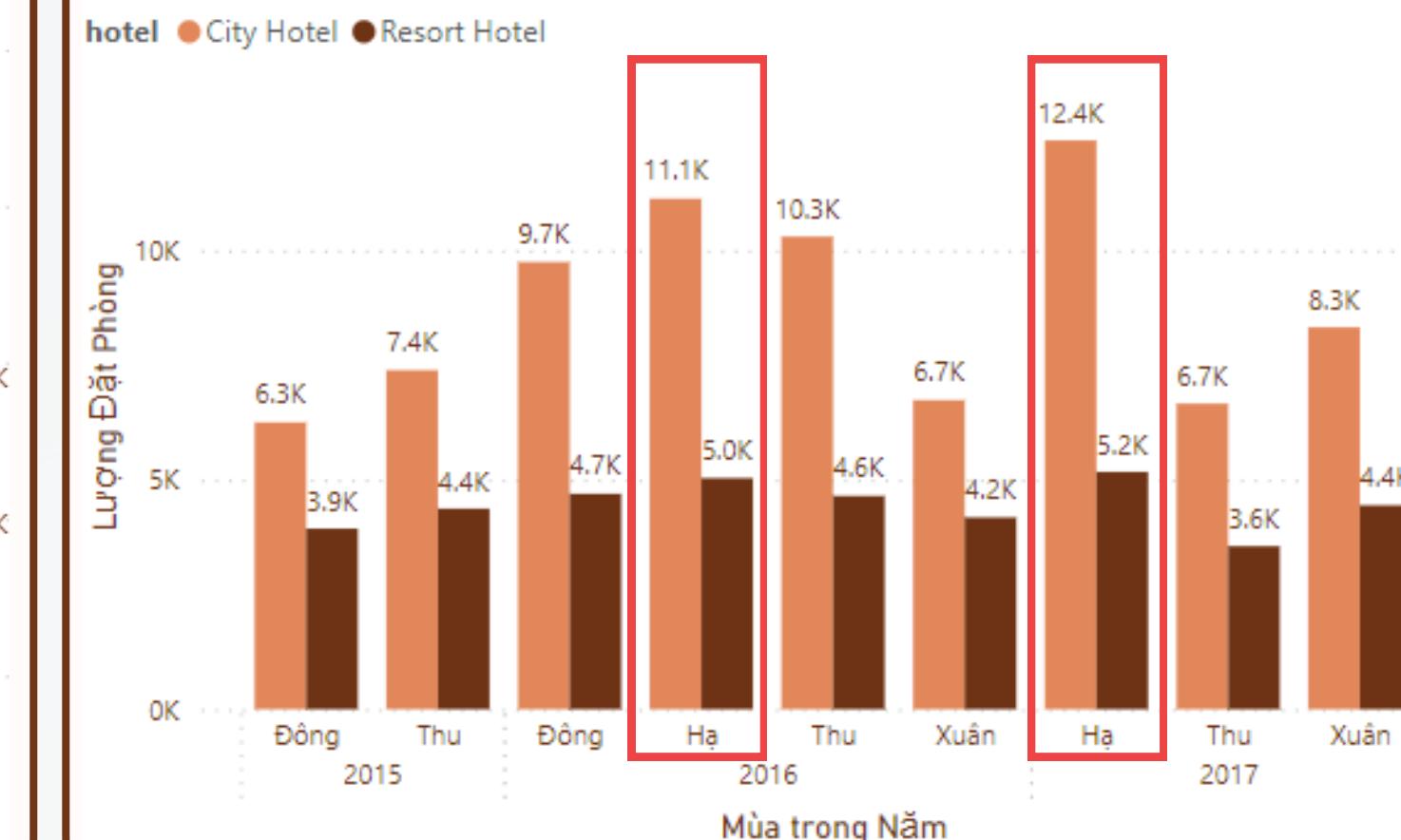
3. PHÂN TÍCH

Sự biến động theo mùa

Tổng lượng khách hàng theo Tháng trong Năm



Tổng lượng khách hàng theo Mùa trong Năm

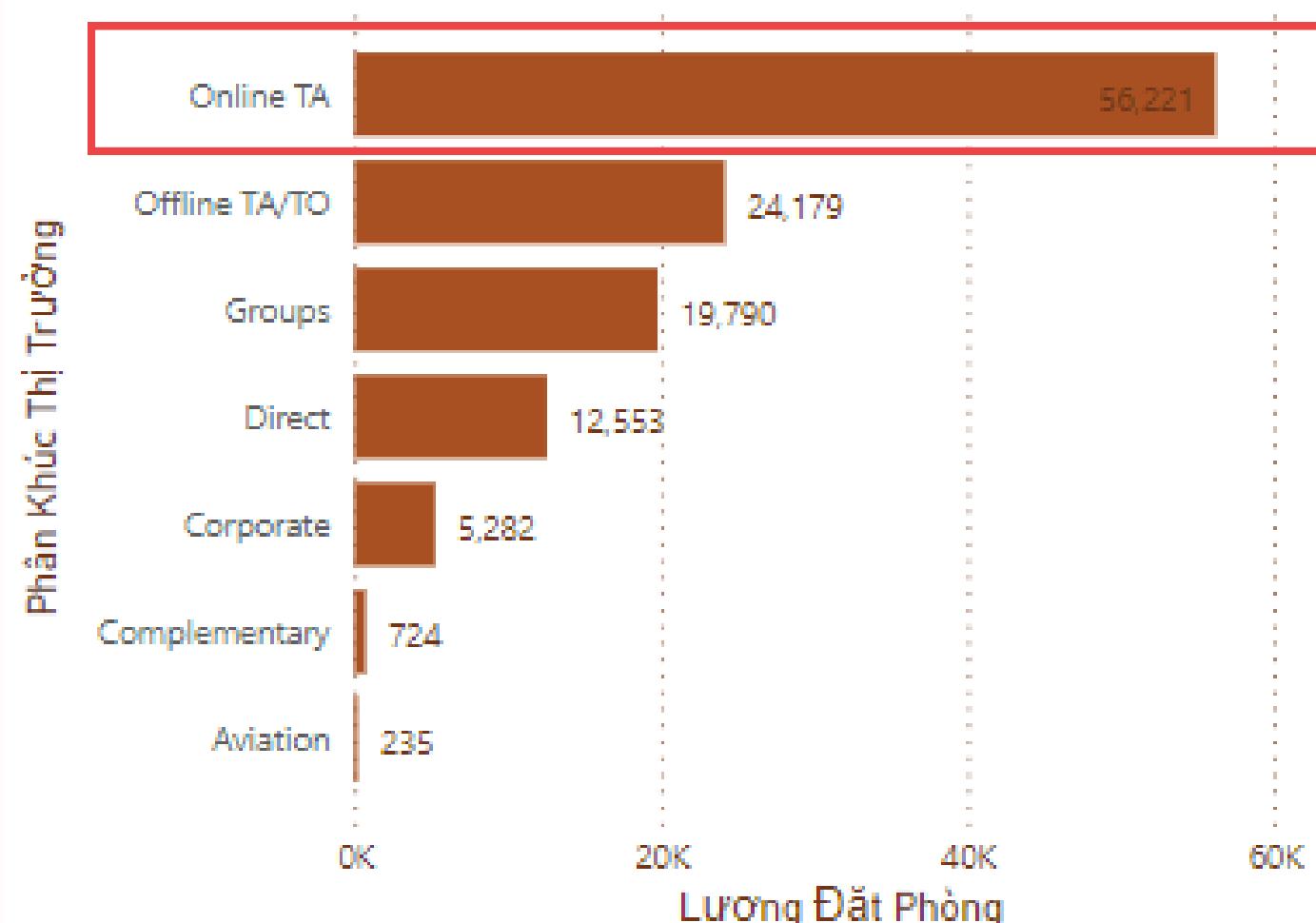


- Phản ánh các **mùa cao điểm** và **thấp điểm** trong ngành khách sạn
- Mùa cao điểm khoảng từ tháng 4 đến tháng 6. Nhu cầu du lịch tăng cao vào mùa hè
- City Hotel có số lượng đặt phòng cao hơn so với Resort Hotel quanh năm

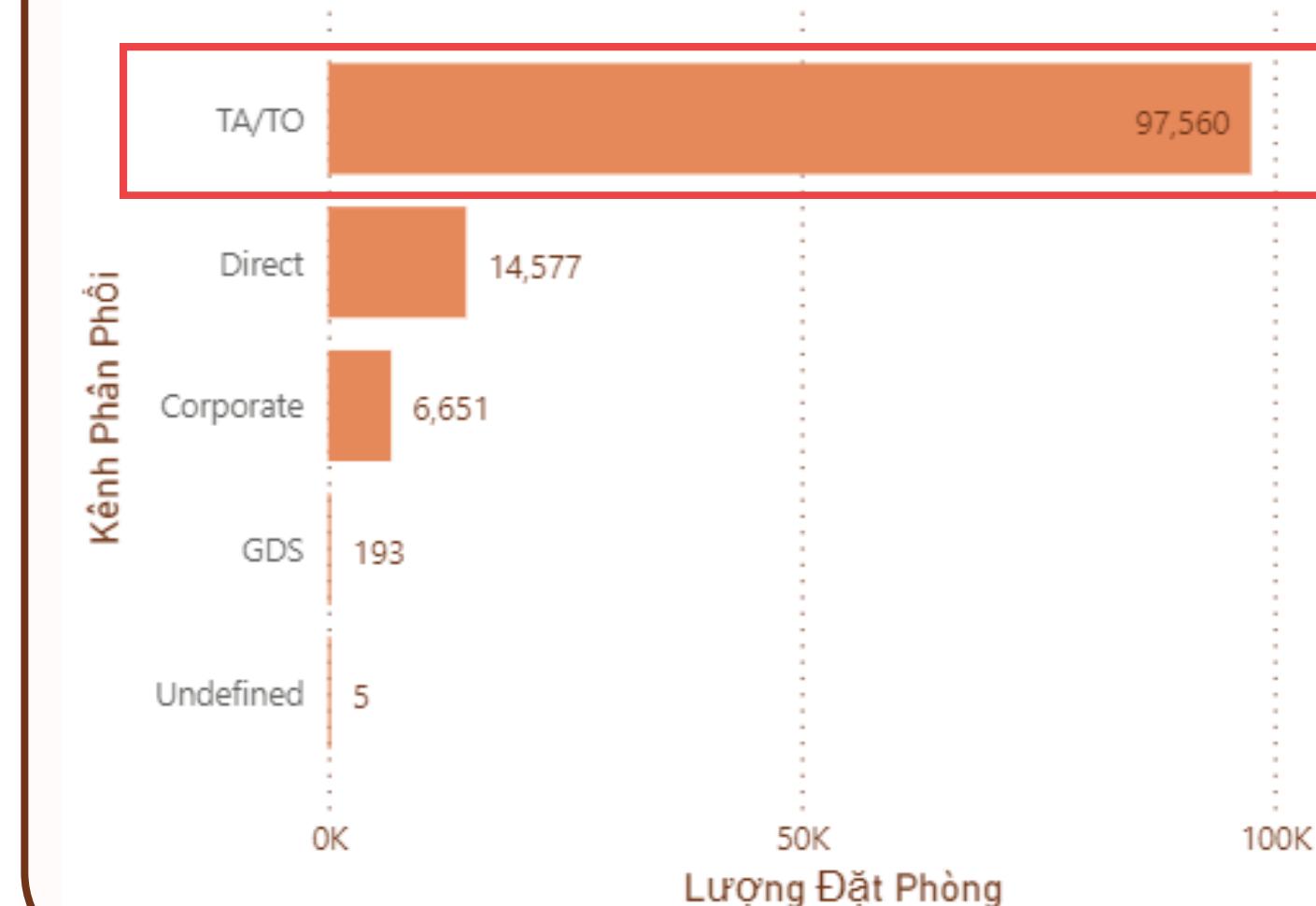
3. PHÂN TÍCH

Phân khúc thị trường và kênh phân phối

Số Lượng Đặt Phòng theo Phân Khúc Thị Trường



Số Lượng Đặt Phòng theo Kênh Phân Phối



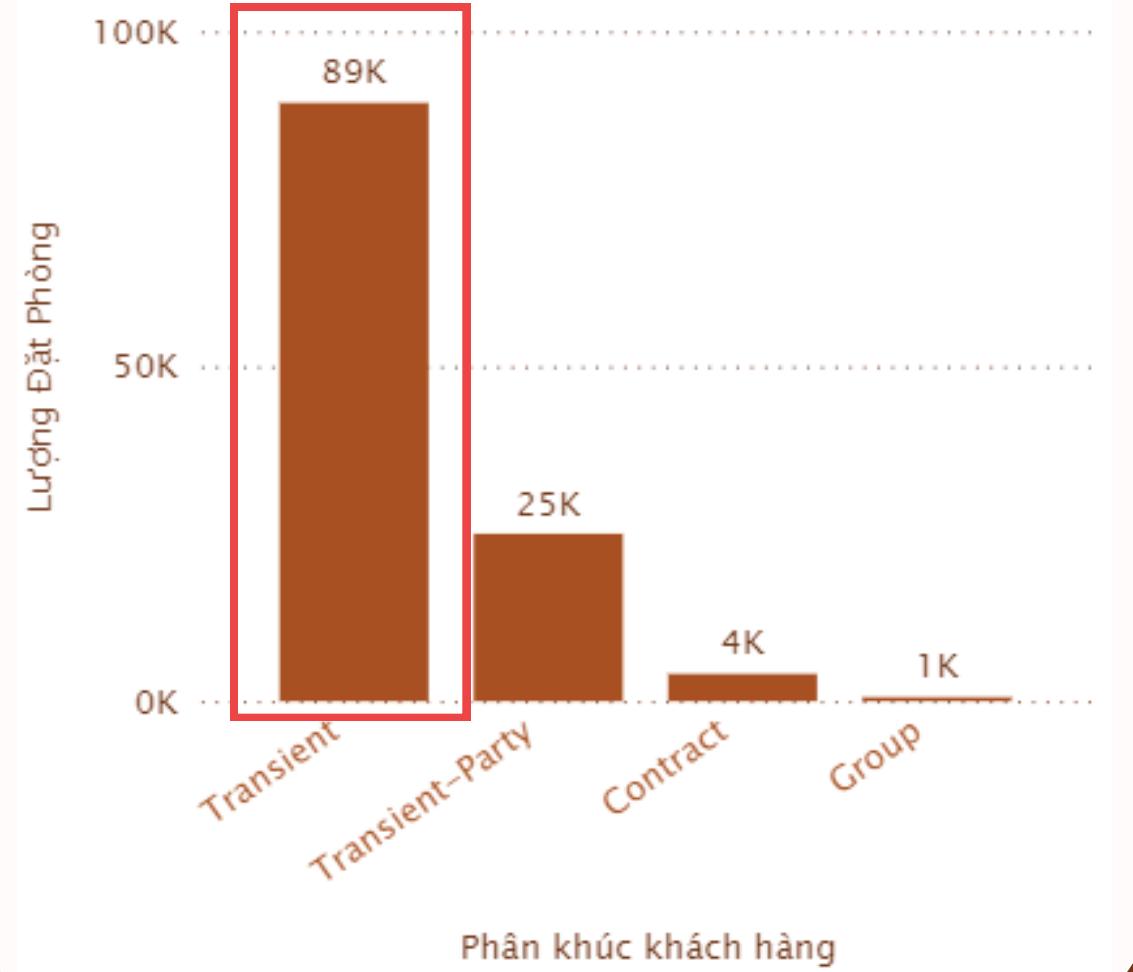
Phân khúc thị trường **Online TA** chiếm ưu thế lớn

Kênh **TA/TO** chiếm ưu thế lớn trong việc phân phối và bán phòng

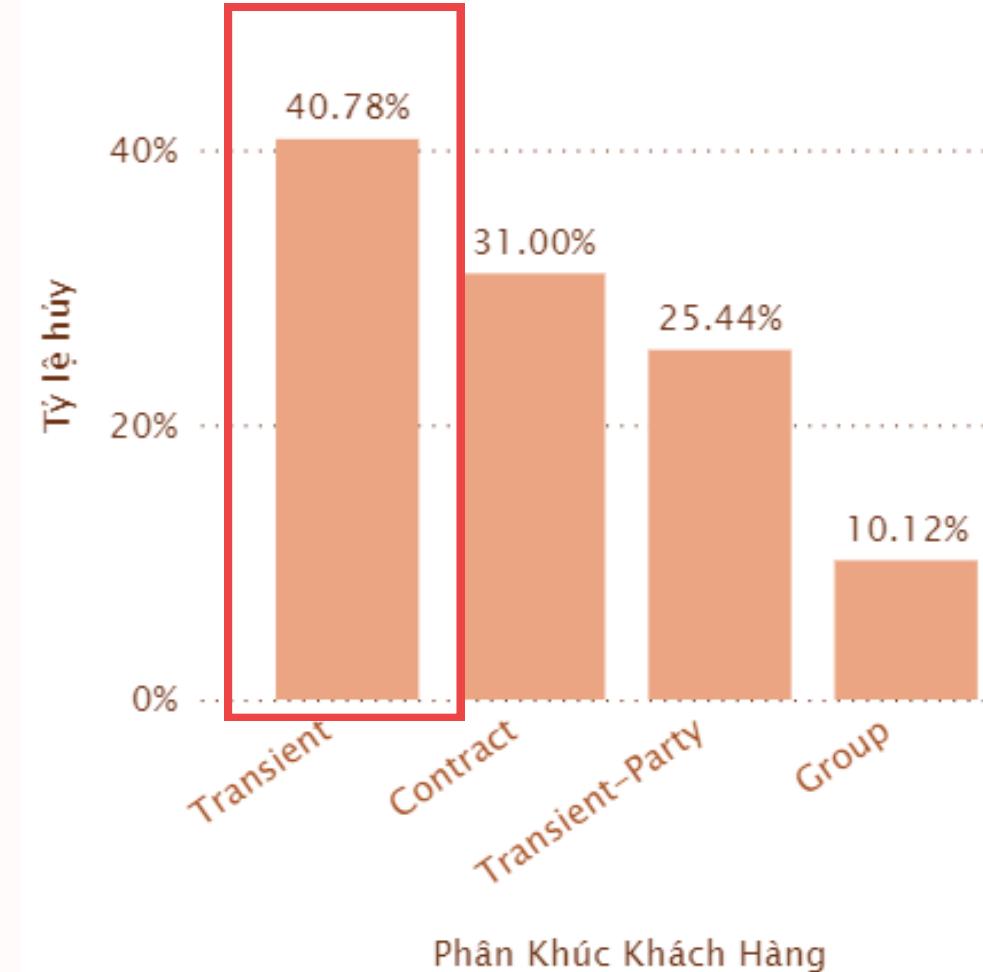
3. PHÂN TÍCH

Phân khúc khách hàng

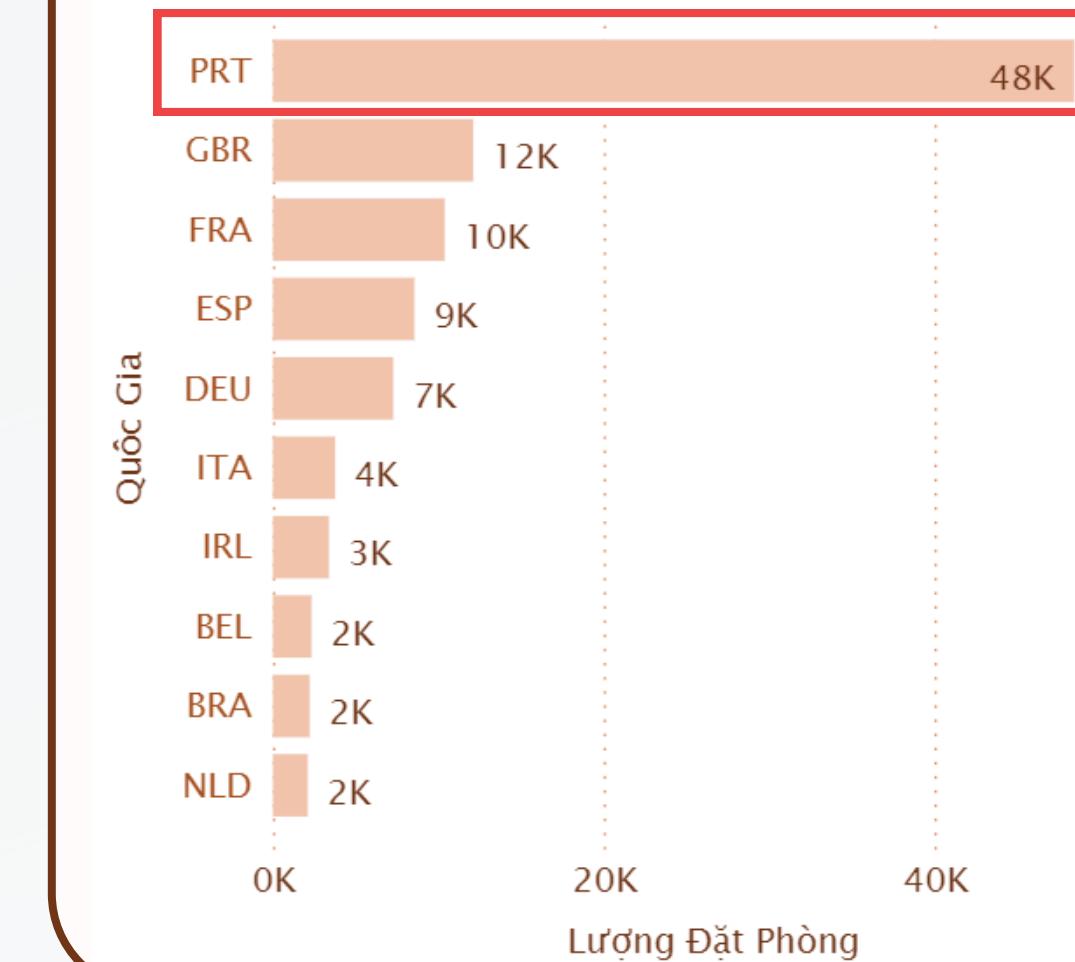
Số Lượng Đặt Phòng theo Phân Khúc Khách Hàng



Tỷ Lệ Hủy Phòng theo Phân Khúc Khách Hàng



Top 10 Quốc gia có lượng đặt phòng cao nhất

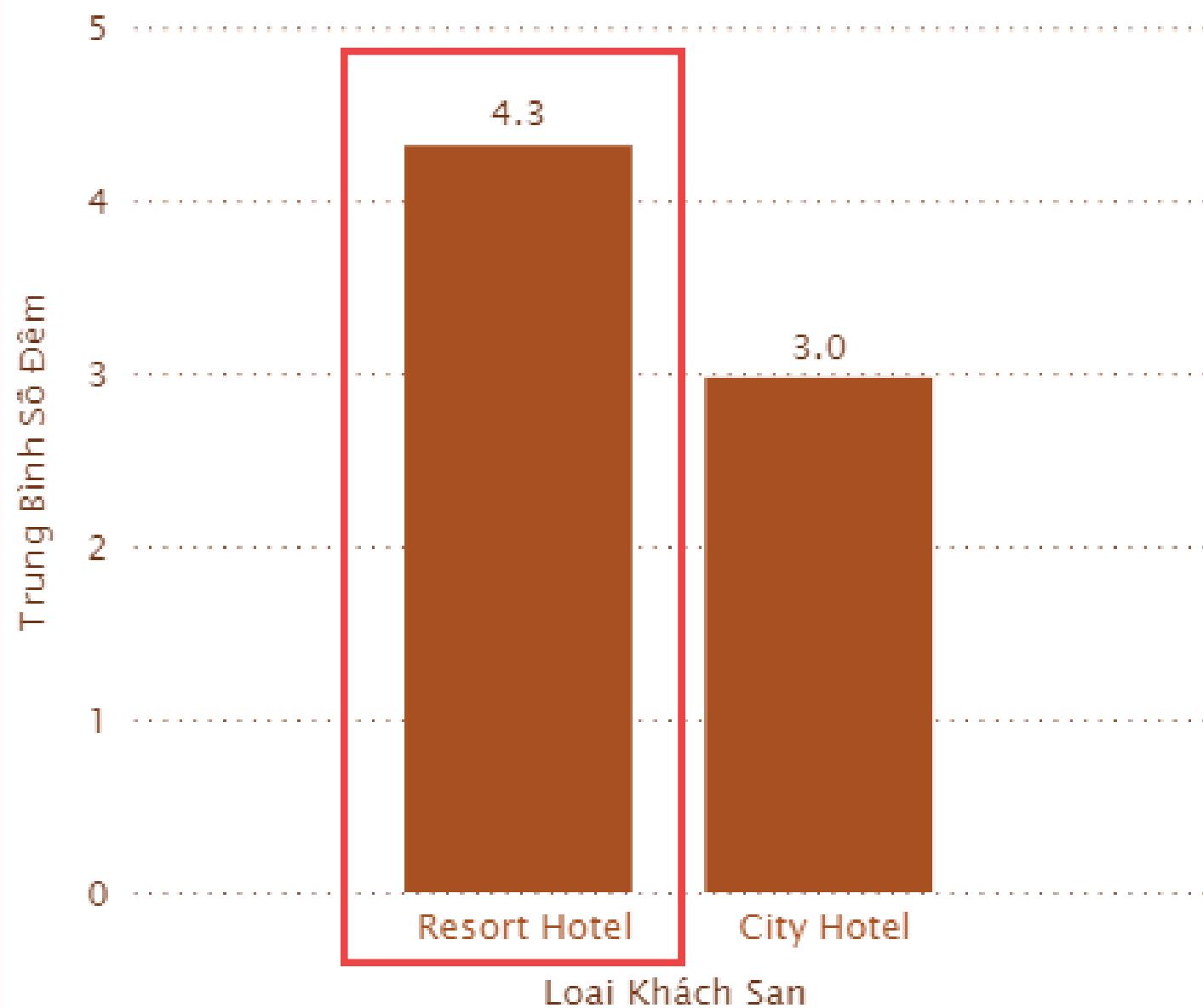


- Transient là loại khách hàng phổ biến nhất với **89,337 đặt phòng**, chiếm **75% tổng số lượng đặt phòng**. Tuy nhiên **Transient** cũng là nhóm khách hàng có **tỷ lệ hủy phòng cao nhất (40,78%)**.
- Khách hàng từ quốc gia Bồ Đào Nha chiếm lượng khách hàng vượt trội.

3. PHÂN TÍCH

Thời Gian Lưu Trú

Thời Gian Lưu Trú Trung Bình Theo Loại Khách Sạn



- Khách du lịch thường **dành nhiều thời gian hơn** tại các **khách sạn nghỉ dưỡng (Resort Hotel)**
- Phản ánh **mục đích chính của chuyến đi**
- Nghỉ dưỡng tại **khách sạn nghỉ dưỡng** thường liên quan đến **kỳ nghỉ dài hơn và thư giãn**
- Khách sạn thành phố thường được sử dụng cho **chuyến đi ngắn ngày hoặc công việc.**

3. PHÂN TÍCH

Tỷ Lệ Hủy Phòng

Tỷ lệ hủy phòng **37.08%**

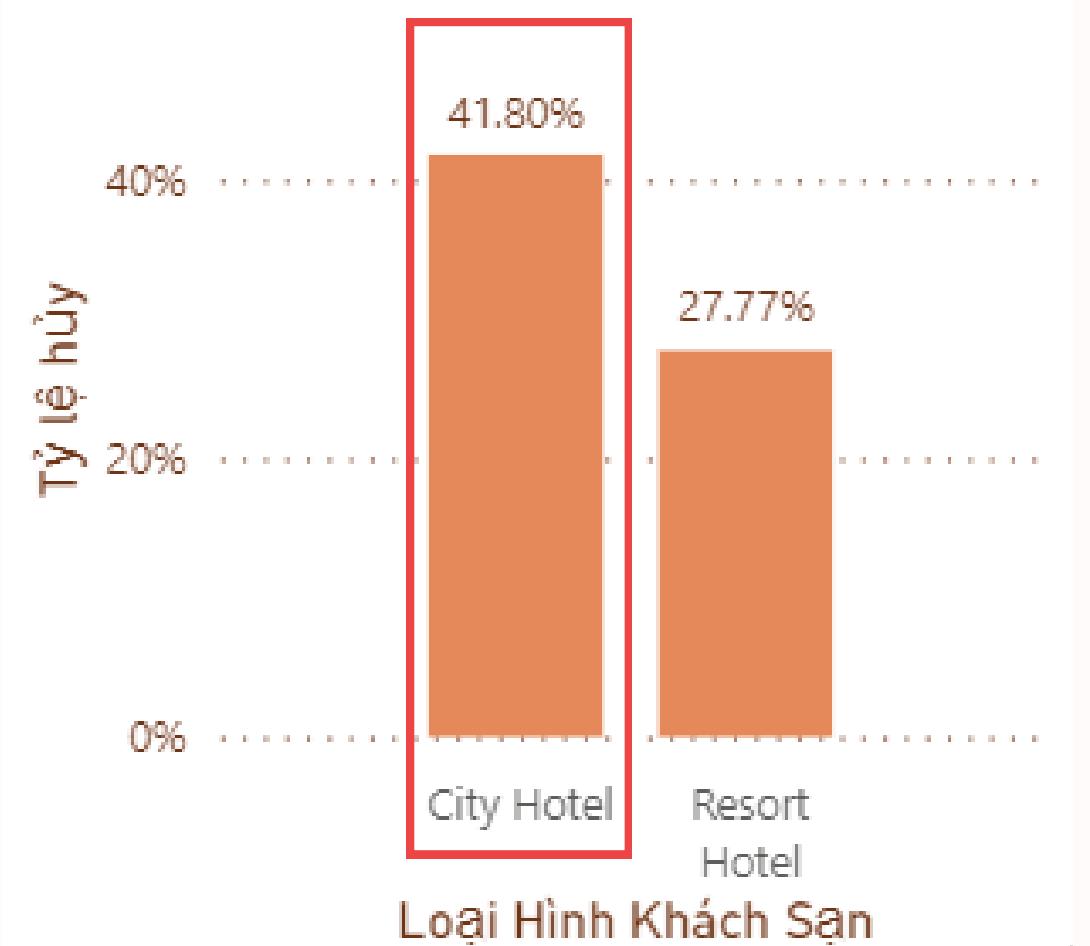
44,115

trường hợp
hủy phòng

74,871

trường hợp không
hủy phòng

Tỷ Lệ Hủy Phòng theo Loại Hình Khách Sạn



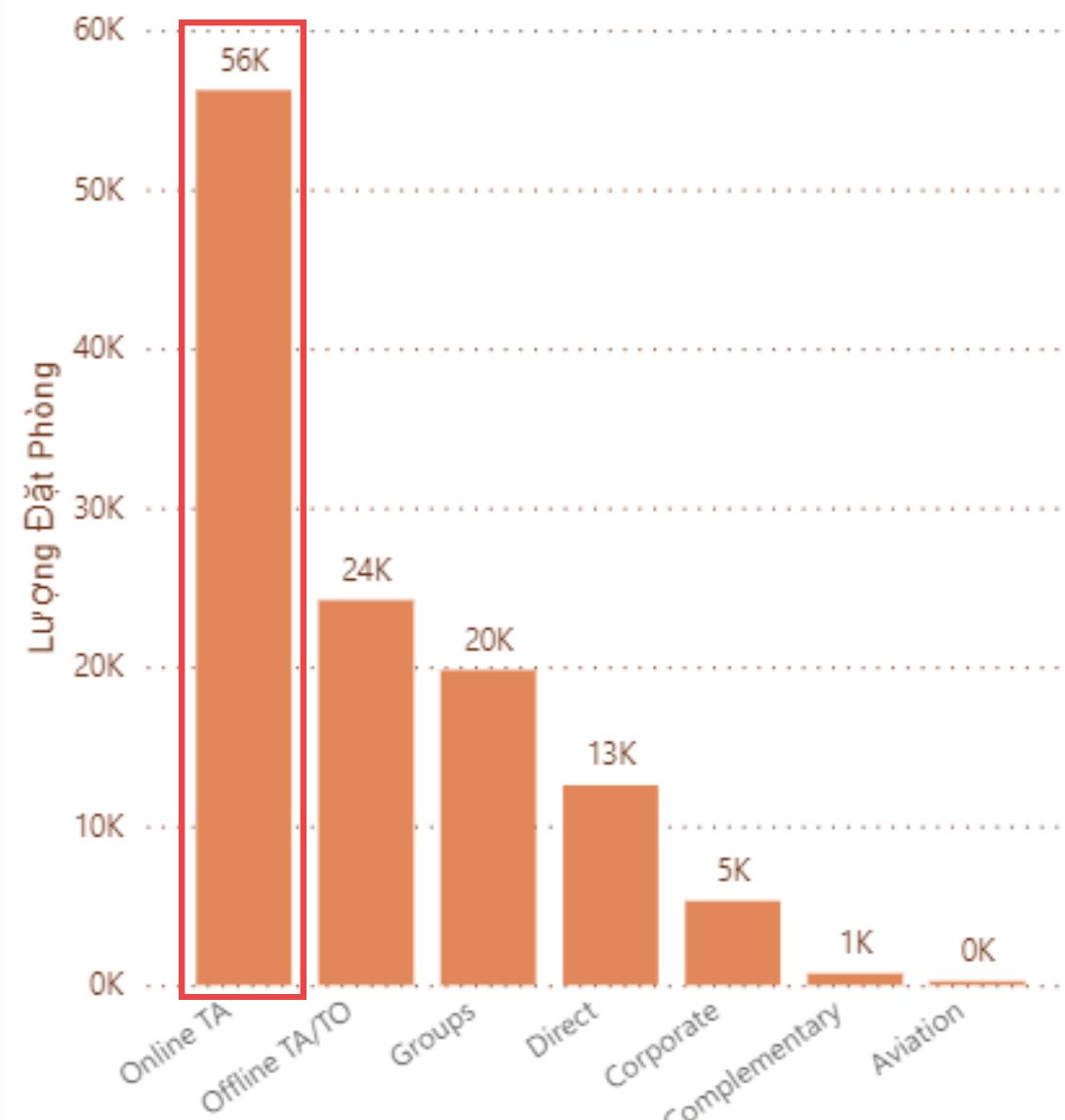
- Chuyển đi đến thành phố thường linh hoạt hơn và có thể thay đổi hoặc hủy dễ dàng hơn so với kỳ nghỉ dài ngày tại khách sạn nghỉ dưỡng.

3. PHÂN TÍCH

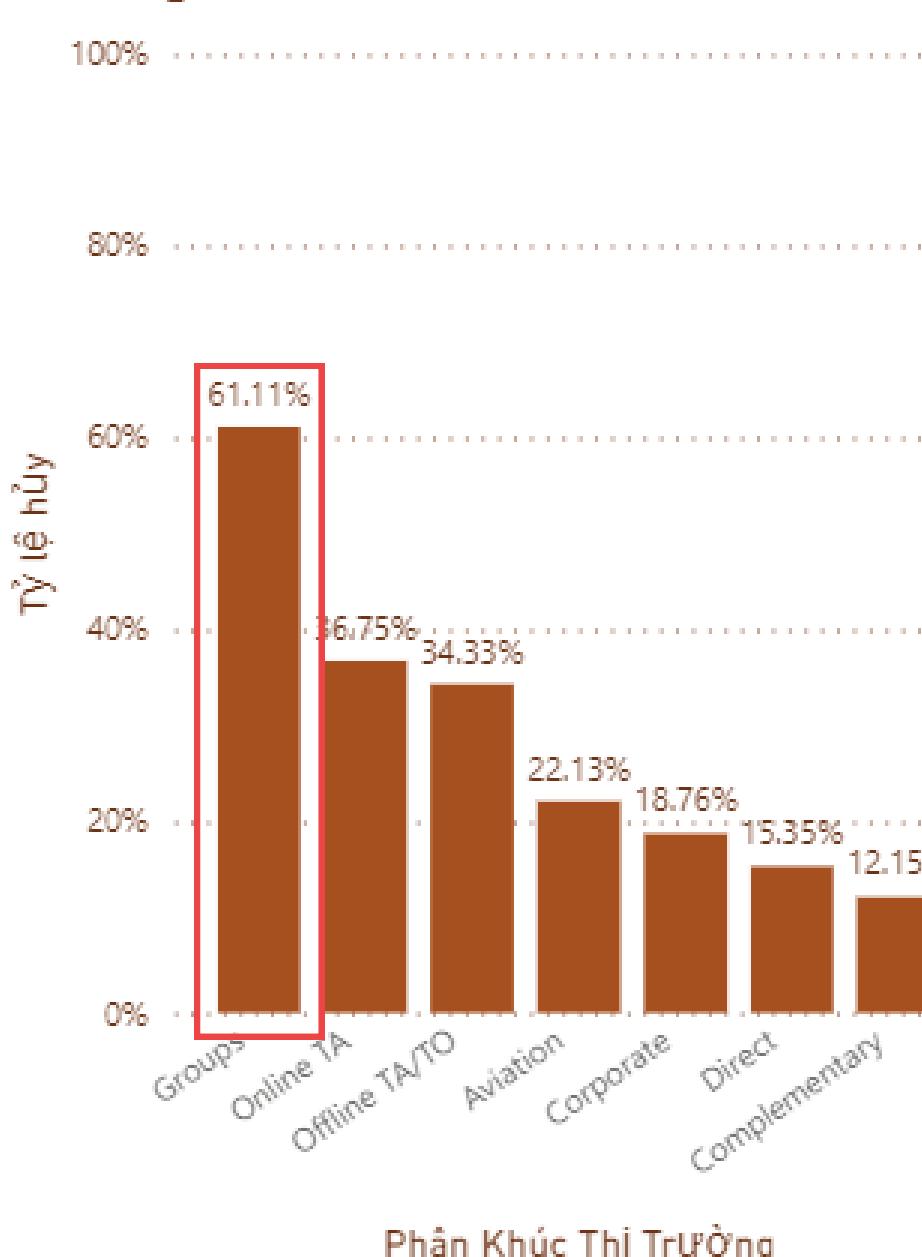
Tỷ Lệ Hủy Phòng

Một số yếu tố có thể ảnh hưởng đến tỷ lệ huỷ đặt phòng

Số Lượng Đặt Phòng theo Phân Khúc Thị Trường



Tỷ Lệ Hủy Phòng theo Phân Khúc Thị Trường



Phân Khúc Thị Trường

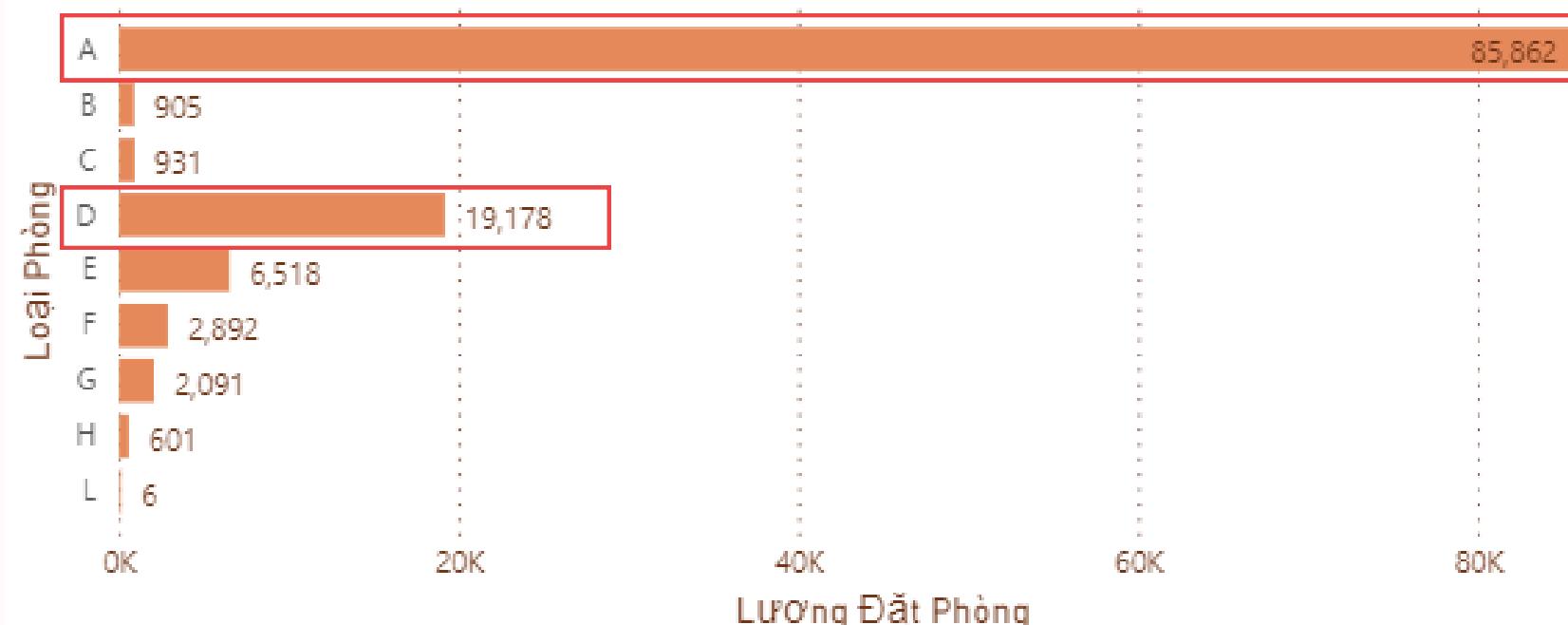
- Các nhóm đặt phòng qua **các đại lý du lịch (Offline và Online)** và đặt phòng nhóm (**Groups**) có tỷ lệ hủy cao.
- **Phản ánh sự không chắc chắn hoặc thay đổi kế hoạch** dễ dàng hơn khi đặt qua các kênh TA

3. PHÂN TÍCH

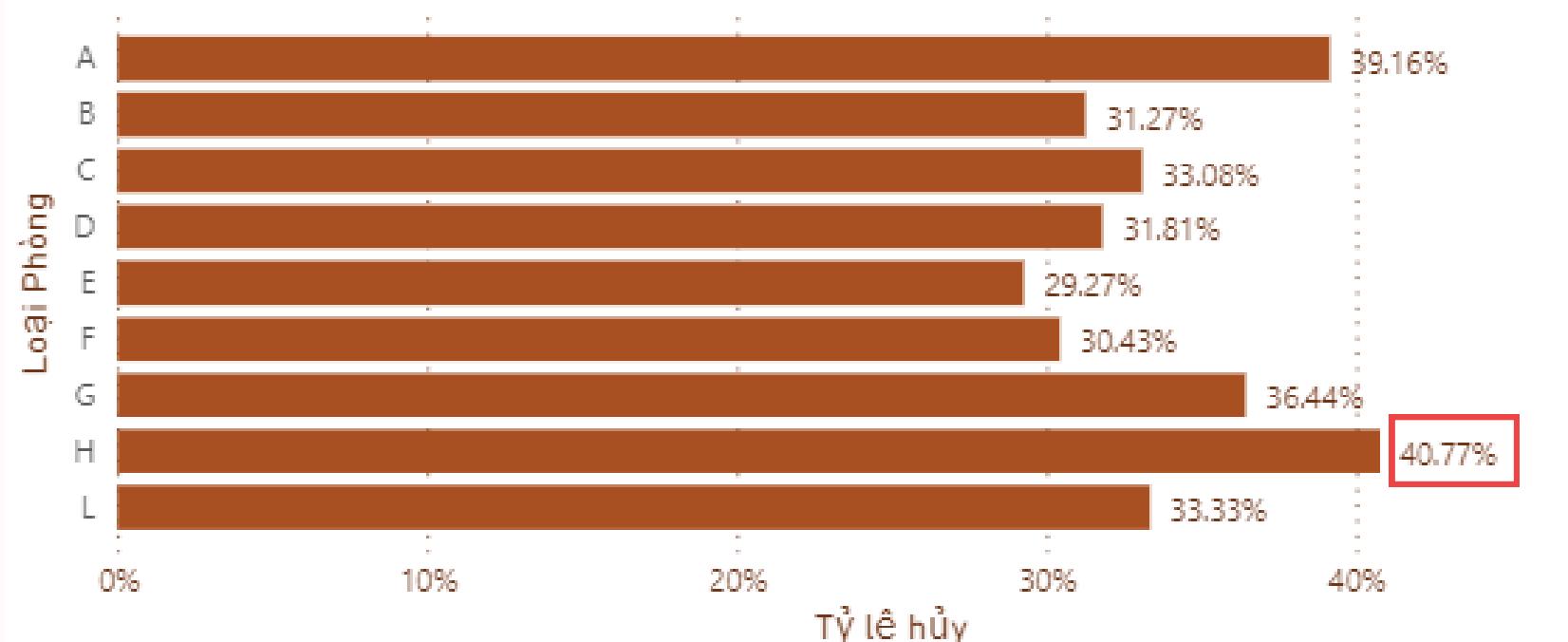
Tỷ Lệ Hủy Phòng

Một số yếu tố có thể ảnh hưởng đến tỷ lệ huỷ đặt phòng

Số Lượng Đặt Phòng Theo Loại Phòng



Tỷ Lệ Hủy Phòng Theo Loại Phòng



Loại Phòng Đặt Trước

- Loại phòng A và D được đặt nhiều nhưng tỷ lệ hủy phòng cao nhất thuộc nhóm khách hàng đặt phòng loại H

- Do nhu cầu cao và tính linh hoạt trong việc thay đổi hoặc hủy bỏ các loại phòng phổ biến

3. PHÂN TÍCH

Tỷ Lệ Hủy Phòng

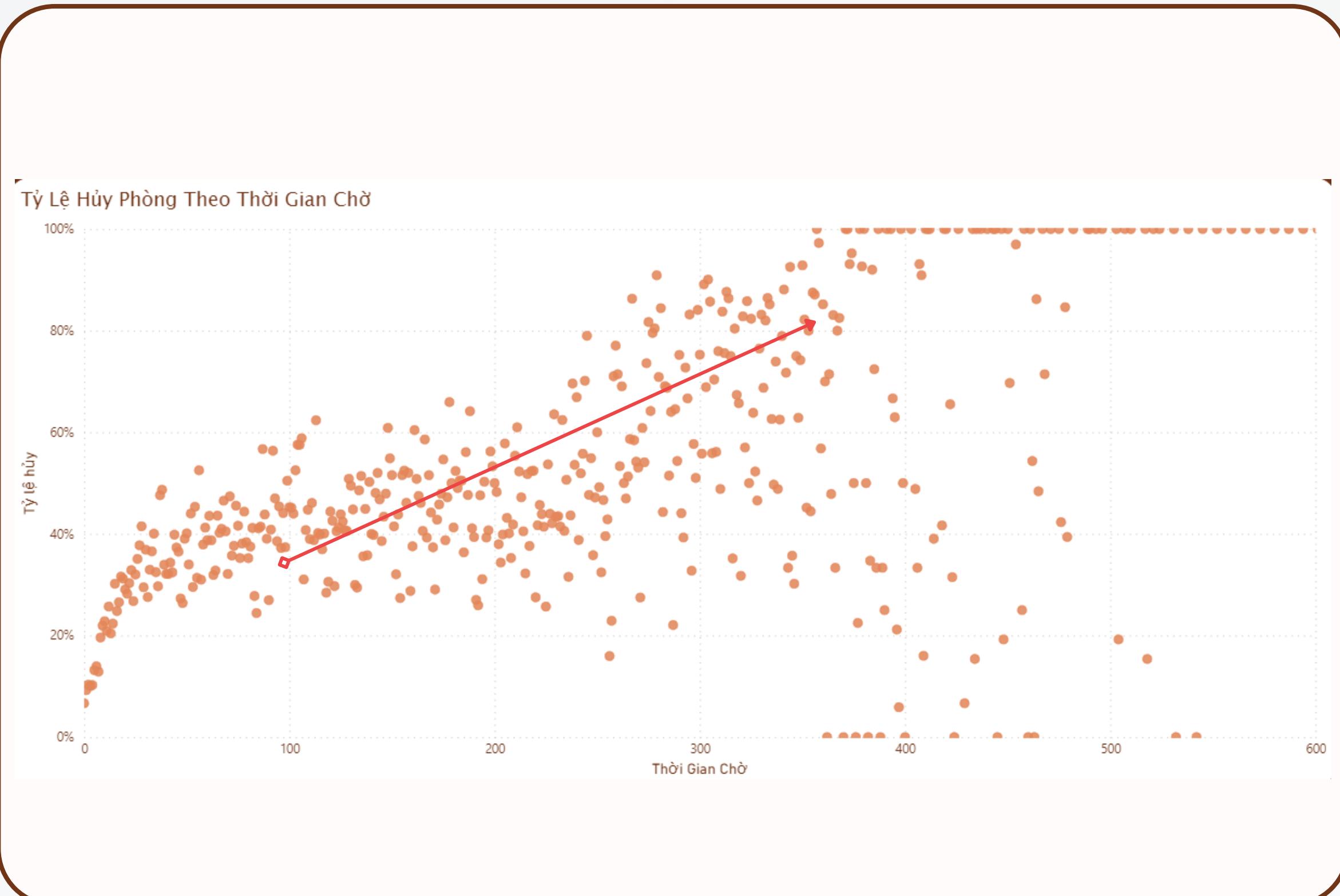
Một số yếu tố có thể ảnh hưởng đến tỷ lệ huỷ đặt phòng

Thời gian chờ

Thời gian chờ (thời gian dẫn đầu đặt phòng) là một yếu tố quan trọng ảnh hưởng đến tỷ lệ hủy đặt phòng

Khách hàng có xu hướng duy trì các đặt phòng được thực hiện gần với ngày đến

Thời gian chờ càng lâu thì tỷ lệ hủy càng cao



3. PHÂN TÍCH

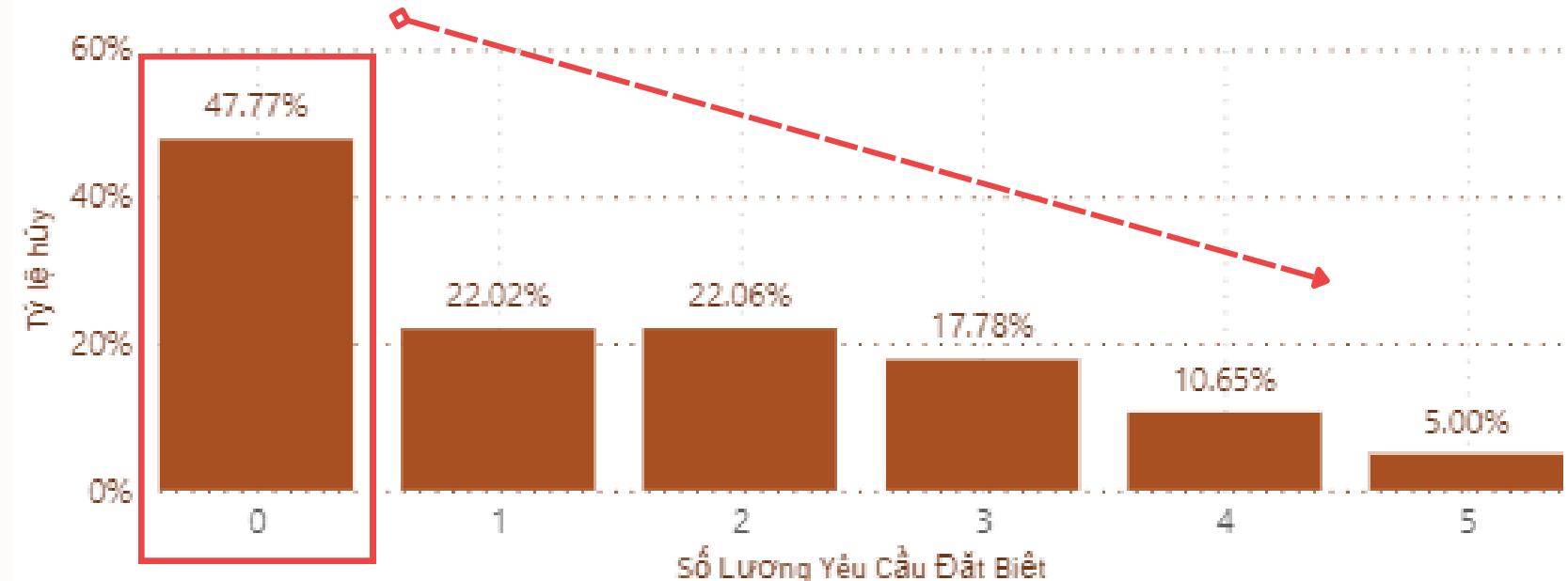
Tỷ Lệ Hủy Phòng

Một số yếu tố có thể ảnh hưởng đến tỷ lệ huỷ đặt phòng

Trung Bình Yêu Cầu Đặt Biệt Mỗi Lượt Đặt Phòng Theo Loại Khách Sạn



Tỷ Lệ Hủy Phòng Theo Số Yêu Cầu Đặt Biệt Của Mỗi Lượt Đặt Phòng



Yêu Cầu Đặc Biệt Khi Đặt Phòng

- **Resort Hotel** có số lượng yêu cầu đặc biệt trung bình cao hơn **City Hotel**
- Có sự giảm đáng kể về tỷ lệ hủy đặt phòng khi số lượng yêu cầu đặc biệt tăng lên
- Cho thấy rằng khi khách hàng có nhiều yêu cầu đặc biệt hơn, họ có xu hướng duy trì lịch trình

3. PHÂN TÍCH

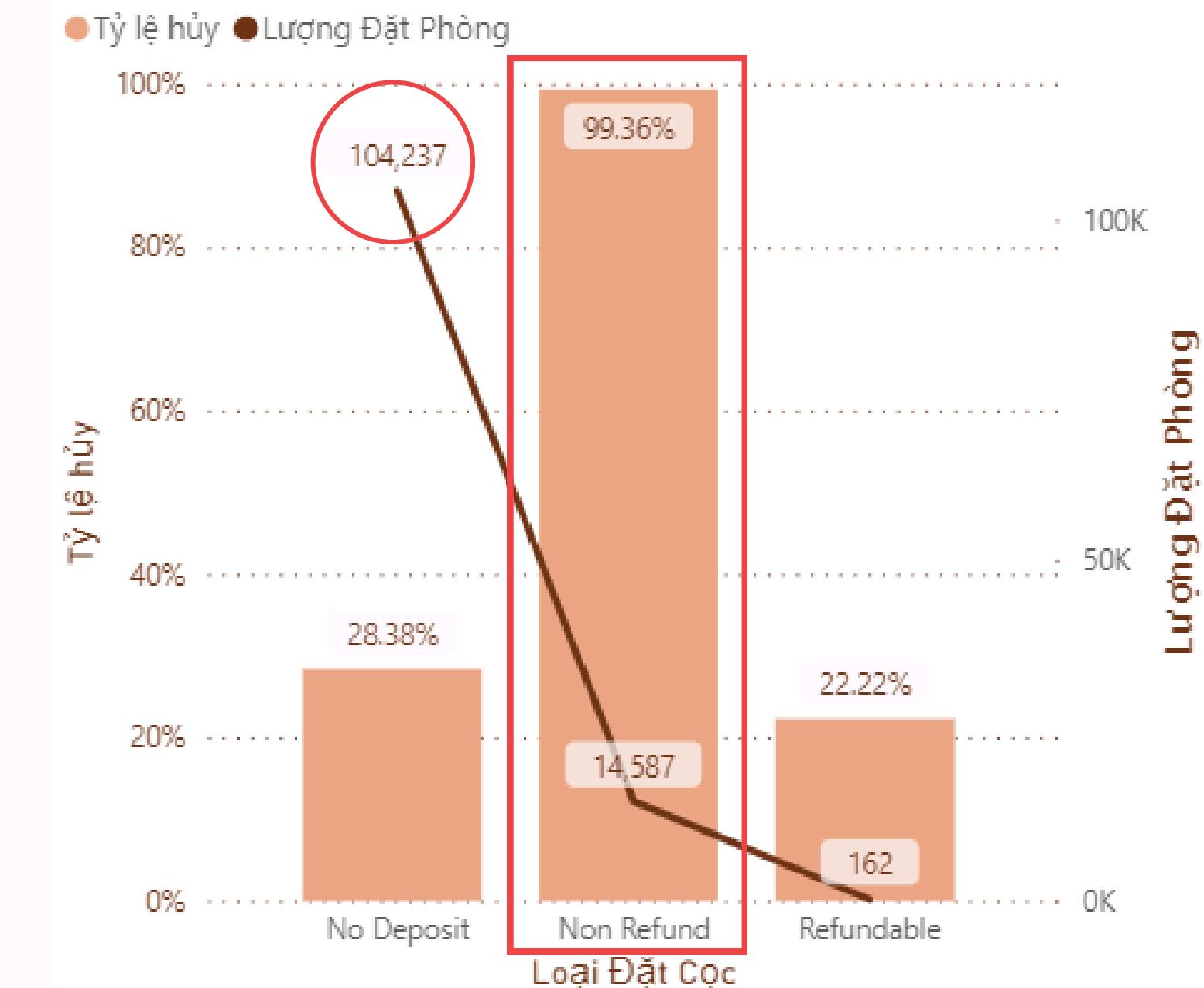
Tỷ Lệ Hủy Phòng

Một số yếu tố có thể ảnh hưởng đến tỷ lệ huỷ đặt phòng

Loại Đặt Cọc Khi Đặt Phòng Khách Sạn

- Loại Không Hoàn Tiền (Non Refund) có tỷ lệ hủy rất cao, gần 100%
- Loại Không Đặt Cọc (No Deposit) chiếm số lượng đặt phòng lớn nhất và có tỷ lệ hủy đặt phòng ở mức vừa phải
- Loại Hoàn Tiền (Refundable) có tỷ lệ hủy thấp nhất, nhưng chỉ chiếm một phần nhỏ trong tổng số đặt phòng

Tỷ Lệ Hủy Và Lượng Đặt Phòng Theo Loại Đặt Cọc Khi Đặt Phòng Khách Sạn



TÓM TẮT

Sự biến động theo mùa	<ul style="list-style-type: none">Tình hình đặt phòng có sự biến động theo mùa, với đỉnh cao vào tháng hè. City Hotel có số lượng đặt phòng cao hơn so với Resort Hotel quanh năm
Phân khúc thị trường và kênh phân phối	<ul style="list-style-type: none">Kênh đặt phòng qua Travel Agents/Tour Operators (TA/TO) chiếm ưu thế lớn. Số lượng đặt phòng cao hơn qua Online TA.
Loại khách hàng	<ul style="list-style-type: none">Transient là loại khách hàng phổ biến nhất. Contract và Group chiếm một số lượng nhỏ
Thời gian lưu trú	<ul style="list-style-type: none">City Hotel: 3 đêm. Resort Hotel: 4.3 đêm. Khách hàng thường dành nhiều thời gian hơn tại các khách sạn nghỉ dưỡng so với khách sạn thành phố
Tỷ lệ hủy đặt phòng	<ul style="list-style-type: none">Tổng tỷ lệ hủy là 37.08%. City Hotel có tỷ lệ hủy cao hơn (41.8%) so với Resort Hotel (27.8%)Kênh Online TA chiếm ưu thế về đặt phòng, nhưng có tỷ lệ hủy cao (36.7%). Groups có tỷ lệ hủy cao nhất (61.1%).Loại phòng A là phổ biến nhất, nhưng có tỷ lệ hủy cao (39.2%). Loại phòng H có tỷ lệ hủy cao nhất (40.8%).Tỷ lệ hủy giảm khi đặt phòng gần thời gian đến (24.0%), nhưng tăng lên với thời gian dẫn đầu dài (>180 ngày - 62.2%).Resort Hotel có số lượng yêu cầu đặc biệt cao hơn. Tỷ lệ hủy giảm khi số lượng yêu cầu đặc biệt tăng lên.Chính sách Không Đặt Cọc có tỷ lệ hủy ổn định và tổng số đặt phòng cao, thu hút khách hàng muốn giữ chỗ mà không cần thanh toán trước, bên cạnh đó tỷ lệ hủy tăng với các chính sách yêu cầu đặt cọc.

4. ĐỀ XUẤT



4. ĐỀ XUẤT



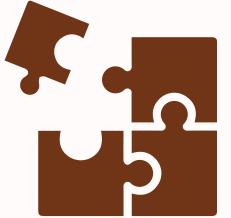
Nên tiếp thị qua các kênh OTA để tiếp cận được lượng khách hàng lớn



Thời gian lưu trú của các khách hàng ở khách sạn nghỉ dưỡng cao hơn ở khách sạn thành phố, nên thiết kế và thêm một số **dịch vụ** đem lại trải nghiệm tốt cho khách du lịch tại khách sạn nghỉ dưỡng, từ đó tăng doanh thu và tỷ lệ quay lại



Mùa cao điểm của du lịch là **vào tháng 4-6** (vào mùa hè) và tỷ lệ huỷ đặt phòng với thời gian đặt trước 30 ngày luôn thấp vậy nên **chiến dịch marketing các promotion** nên chạy vào tháng 2-3 sẽ dễ tăng tỷ lệ **chuyển đổi**



Với **tỷ lệ hủy phòng chiếm 37%**, có thể áp dụng chiến lược **yêu cầu khách khi đặt phòng** thì thanh toán trước, hoặc **đặt cọc** một phần áp dụng kèm chính sách hoàn hủy linh hoạt để khách hàng an tâm thanh toán

**THANK YOU
FOR LISTENING**