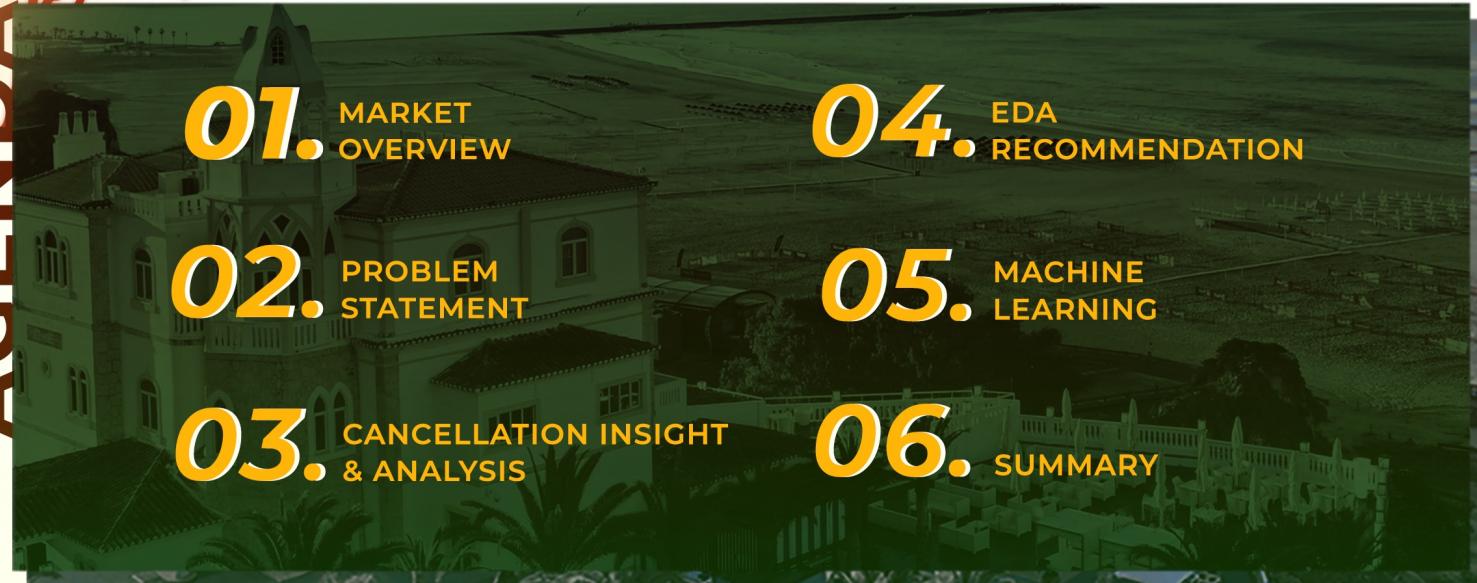


Consectetur adipisci velit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

# PREDICT HOTEL BOOKING *Cancellation* IN PORTUGAL

# AGENDA

# Agenda



AGENDA  
Agenda

## MỘT SỐ THÔNG TIN VỀ DỰ ÁN

- Dự án này là việc sử dụng máy học phân loại (machine learning classification) để cố gắng dự đoán việc đặt phòng sẽ bị huỷ bằng machine learning dựa trên dữ liệu lịch sử.
- Dữ liệu cho dự án này là từ Hotel Demand Booking<sup>[1]</sup>. Được lấy từ năm 2015 đến năm 2017 bao gồm khách sạn ở Vùng Algarve và Lisbon ở Bồ Đào Nha.

MARKET OVERVIEW

MARKET OVERVIEW

MARKET OVERVIEW

MARKET OVERVIEW

MARKET OVERVIEW

# MARKET OVERVIEW

# MARKET OVERVIEW



MARKET OVERVIEW

MARKET OVERVIEW

MARKET OVERVIEW

MARKET OVERVIEW

MARKET OVERVIEW

# NỀN CÔNG NGHIỆP DU LỊCH Ở BỒ ĐÀO NHA

Ở **Giải thưởng Du lịch Thế giới (WTA) lần thứ 29 năm 2022**, Bồ Đào Nha nhận được tổng cộng **12 giải thưởng**. Điều đó đưa nơi này lên **đầu danh sách những điểm đến du lịch tốt nhất thế giới**.

1. *Best island destination – Madeira*
2. *Best coastal city destination – Lisbon*
3. *Best city destination – Porto*
4. *Best airline company to Africa – TAP Air Portugal*
5. *Best airline company to South America – TAP Air Portugal*
6. *Best classic hotel – Olissipo Lapa Palace Hotel, Lisbon*
7. *Best Adventure tourism attraction – Paiva Walkways, Arouca*
8. *Best conservation company – Parques de Sintra – Monte da Lua*
9. *Best golf and villa resort – Dunas Douradas Beach Club, Almancil*
10. *Best luxury business hotel – Pestana Palace Hotel, Lisbon*
11. *Best responsible tourism project – Dark Sky Alqueva*
12. *Best boutique hotel operator – Amazing Evolution.*

# NỀN CÔNG NGHIỆP KHÁCH SẠN Ở BỒ ĐÀO NHA

Ngành công nghiệp khách sạn **phát triển nhanh nhất** trong ngành công nghiệp du lịch. Đặc biệt là với sự gia tăng của OTA (Online Travel Agent) khổng lồ khiến việc đặt phòng khách sạn trở nên dễ dàng hơn bao giờ hết.

Năm 2023 Lisbon trở thành **thành phố hấp dẫn thứ 2** sau Amsterdam về đầu tư khách sạn <sup>[1]</sup>

Doanh thu năm 2017 <sup>[1]</sup>

**3,6**  
triệu USD

Tổng số khách năm 2017 <sup>[2]</sup>

**~24**  
triệu người

Và một số giải thưởng từ WTA dành cho hotel đáng nói như *Best Five Star Hotel, Leading Design Hotel, Best Gold Hotel, Best Spa Hotel,...*

Với xu hướng phát triển nhanh của nền công nghiệp khách sạn **mang lại nhiều lợi ích** tuy nhiên đi kèm cũng là **một số thách thức**.

OBLEM STATEMENT

PROBLEM STATEMENT

PROBLEM STATEMENT

PROBLEM STATEMENT

PROBLEM STATE



**PROBLEM  
STATEMENT**

**PROBLEM  
STATEMENT**

OBLEM STATEMENT

PROBLEM STATEMENT

PROBLEM STATEMENT

PROBLEM STATEMENT

PROBLEM STAT

## VIỆC HỦY ĐẶT PHÒNG ẢNH HƯỞNG NHƯ THẾ NÀO ĐỐI VỚI KHÁCH SẠN?

Tỷ lệ huỷ phòng ở khách sạn



- Doanh thu thấp hơn do không bán được phòng
- Việc hủy phòng thường xuyên làm giảm thứ hạng khách sạn trên kết quả tìm kiếm OTA
- Hủy bỏ tạo thêm công việc & cảng thẳng không cần thiết
- RevPAR thấp hơn (Doanh thu trên mỗi phòng có sẵn) khi bán rẻ hơn vào phút cuối (Nếu có)

## VẬY LÀM SAO ĐỂ GIẢI QUYẾT TRÌNH TRẠNG HỦY PHÒNG ?

# VIỆC HỦY ĐẶT PHÒNG ẢNH HƯỞNG NHƯ THẾ NÀO ĐỐI VỚI KHÁCH SẠN?



Customer Journey Mapping for Coaches and Consultants

## CÂU HỎI ĐẶT RA:

1. Các nhóm đối tượng khách hàng ảnh hưởng tới tỷ lệ huỷ phòng như thế nào?
2. Nếu khách hàng đặt phòng A mà nhận phòng B thì có huỷ phòng không?
3. Khách hàng cũ quay lại liệu có huỷ phòng không?
4. Khách hàng hay huỷ phòng ở kiểu khách sạn nào?
5. Cặp đôi, bạn bè, gia đình hay người độc thân sẽ là nhóm đối tượng huỷ nhiều nhất?
6. Khách đến từ đâu thì hay huỷ phòng nhất?
7. Thường khách hàng đặt qua đâu thì sẽ huỷ phòng nhiều?
8. Nếu khách hàng đặt phòng từ sớm hoặc sát ngày thì ảnh hưởng như nào tới tỉ lệ huỷ
9. Khách sạn chốt booking càng nhanh thì khách hàng có huỷ phòng không?
10. Kiểu bữa ăn nào sẽ bị huỷ nhiều nhất
11. Khách yêu cầu có bãi đỗ xe thì có huỷ phòng hay không?
12. Khách hàng có những yêu cầu đặt biệt thì ảnh hưởng như thế nào đến việc huỷ phòng

## GIỚI HẠN DỰ ÁN

- Dự án hủy đặt phòng khách sạn này **chỉ áp dụng cho các đặt phòng khách sạn ở Vùng Lisbon và Vùng Algarve**, cả hai địa điểm đều nằm ở **Bồ Đào Nha**.
- Dự đoán hủy bỏ với ứng dụng web này bên ngoài cả hai khu vực có thể không có kết quả chính xác do hạn chế về vị trí, sự khác biệt về việc huỷ

DATA ANALYSIS

DATA ANALYSIS

DATA ANALYSIS

DATA ANALYSIS

DATA ANALYSIS

DATA ANALYSIS



# DATA ANALYSIS

# DATA ANALYSIS



DATA ANALYSIS

DATA ANALYSIS

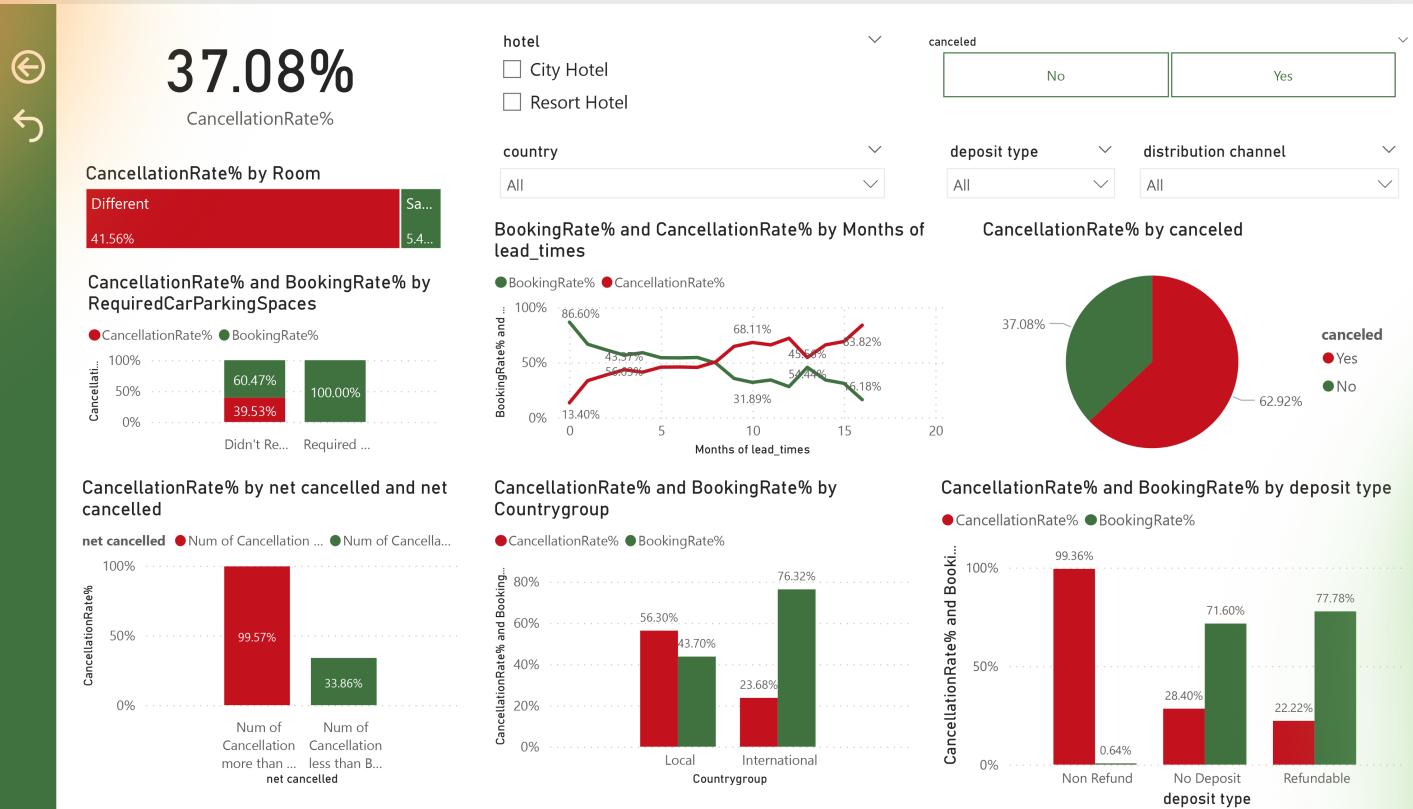
DATA ANALYSIS

DATA ANALYSIS

DATA ANALYSIS

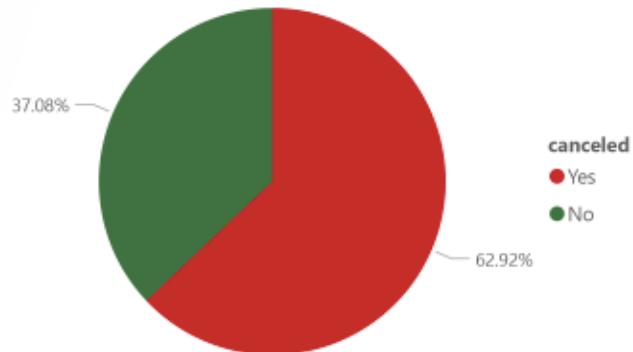
DATA ANALYSIS

# Overall



# TỶ LỆ HUỶ PHÒNG Ở BỒ ĐÀO NHA

## BOOKING CANCELLATION REPORT



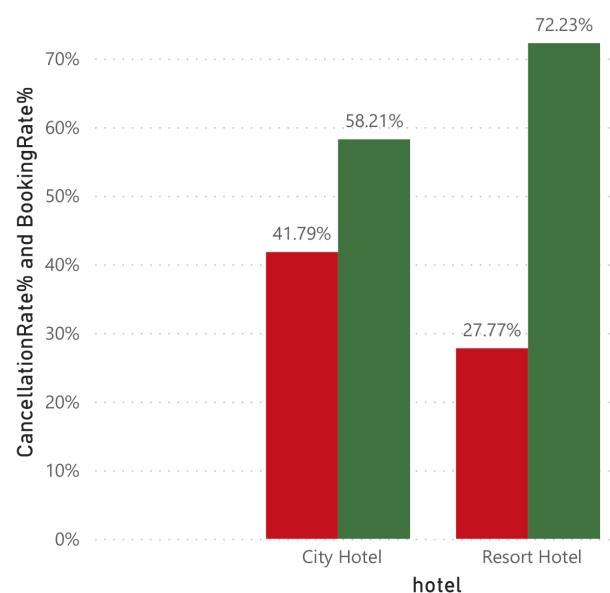
Biểu đồ thể hiện tỷ lệ huỷ phòng khách sạn

Tỷ lệ huỷ phòng khách sạn ở Bồ Đào Nha là **37,08%** gần bằng tỷ lệ huỷ phòng khách sạn trung bình của châu Âu năm 2017 là **41,03%**

# Liệu City Hotel tỷ lệ huỷ nhiều hơn Resort Hotel không?

CancellationRate% and BookingRate% by hotel

● CancellationRate% ● BookingRate%

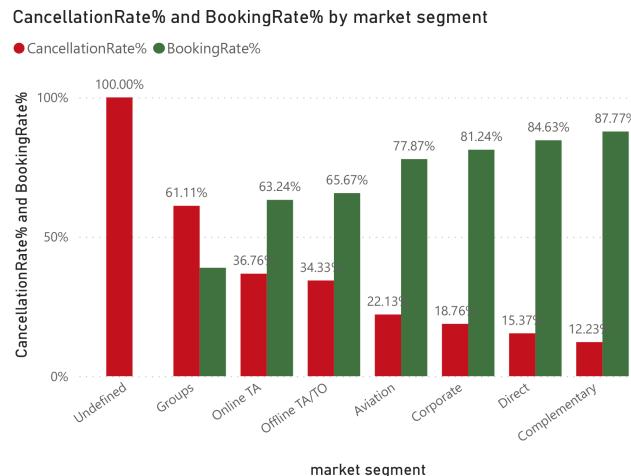


- **41,79%** tỷ lệ huỷ phòng ở **City Hotel** và
- **27,77%** tỷ lệ huỷ phòng ở **Resort Hotel**

Tỷ lệ huỷ phòng của city hotel nhiều trong khi tỷ lệ đặt phòng khách sạn thấp

Biểu đồ thể hiện tỷ lệ huỷ phòng khách sạn  
theo kiểu khách sạn

# Nhóm khách hàng nào ảnh hưởng tới việc huỷ phòng nhiều nhất?



Biểu đồ thể hiện tỷ lệ huỷ phòng khách sạn ở các nhóm khách hàng

## Chú thích:

**Groups:** Nhóm khách hàng bao gồm những khách nghỉ dưỡng kết hợp thời gian lưu trú tại khách sạn của họ với một sự kiện hoặc hình thức hoạt động hoặc kinh doanh khác diễn ra tại khách sạn hoặc gần khách sạn

**TA/TO( Travel Agent/Travel Operation:** Các đại lý du lịch

**Activation:** Khách doanh nhân

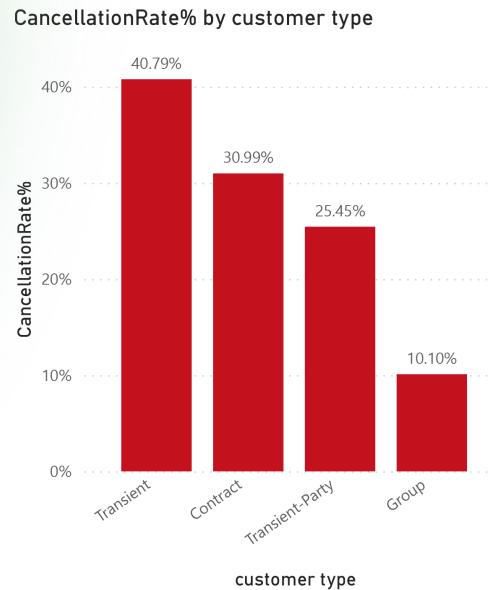
**Corporate:** Khách từ các tổ chức

**Direct:** Khách hàng đặt trực tiếp

**Complementary:** Khách hàng bổ sung

- Nhóm khách hàng không xác định có tỷ lệ huỷ **100%**
- Nhóm khách hàng **Groups** có tỷ lệ huỷ **61,11%**
- Nhóm khách hàng **Online Travel Agent, Offline Travel Agent/ Tour Opperation** có tỷ lệ huỷ từ **34,33% - 36,76%**
- Nhóm khách hàng còn lại có tỉ lệ huỷ từ **12,23% - 22,13%**

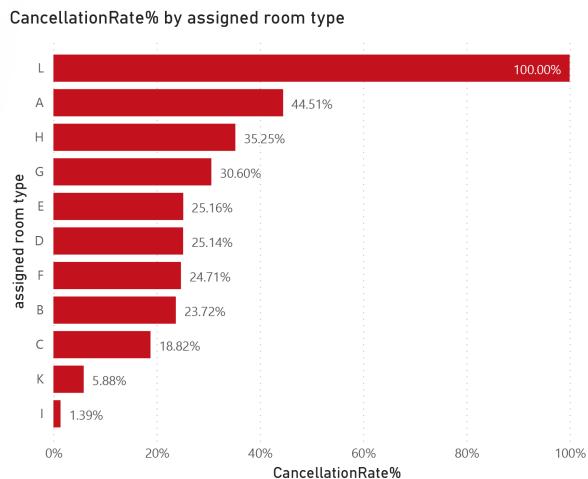
## Kiểu khách hàng nào ảnh hưởng tới việc huỷ phòng nhiều nhất?



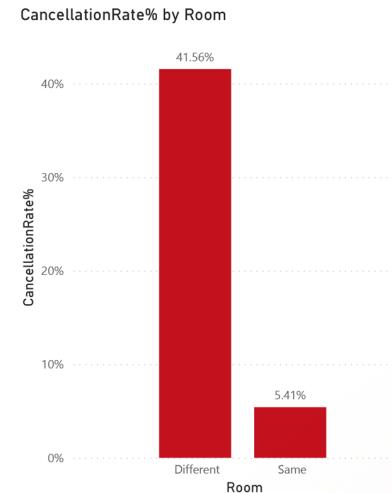
Biểu đồ thể hiện tỷ lệ huỷ phòng khách sạn ở các  
kiểu khách hàng

- **Kiểu khách hàng Transient** (Khách hàng ngắn hạn) có tỷ lệ huỷ ~ **40,8%**
- **Kiểu khách hàng Contract** (Khách hàng hợp đồng) có tỷ lệ huỷ ~ **31%**
- **Kiểu khách hàng Transient - Party** (Khách hàng tổ chức tiệc ở khách sạn thời gian ngắn) có tỷ lệ huỷ ~ **25,5%**
- **Kiểu khách hàng Group** có tỷ lệ huỷ ~ **10,1%**

# Khách hàng nhận phòng khác với phòng mình đặt thì có huỷ phòng hay không?



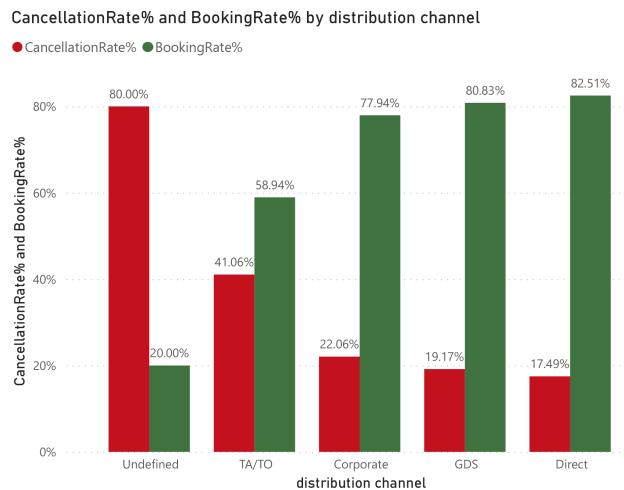
Biểu đồ thể hiện tỷ lệ huỷ phòng ở kiểu phòng  
được nhận



Biểu đồ thể hiện tỷ lệ huỷ phòng ở kiểu phòng  
nhận giống hay khác phòng đặt

Từ hai biểu đồ trên cho thấy nếu **khách hàng nhận phòng khác với phòng mình đặt** thì tỷ lệ huỷ là **41,56%**. Phòng chắc chắn sẽ bị huỷ là L, tiếp đó là A với 41,56%, H với 32,25%, G với 30,6%.

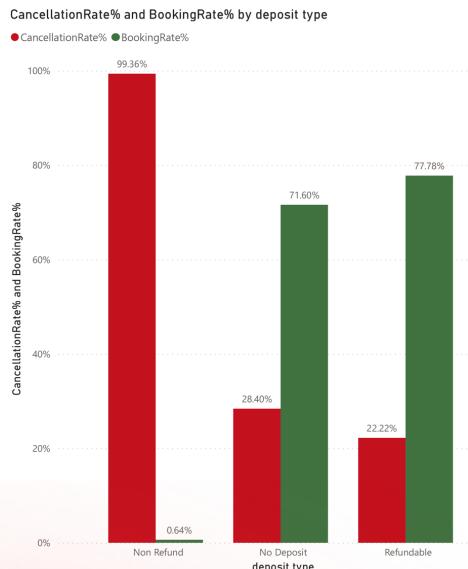
## Kênh phân phối ảnh hưởng như thế nào tới việc huỷ phòng?



Biểu đồ thể hiện tỷ lệ huỷ phòng ở các kênh phân phối

Dựa trên biểu đồ, dường như **khách hàng sẽ huỷ phòng phần lớn ở những kênh không có thông tin rõ ràng** với tỷ lệ huỷ là **80%**. **Đại lý du lịch** có tỷ lệ huỷ là **41,06%**, tiếp đó là các **tổ chức** với tỷ lệ huỷ là **22,06%**.

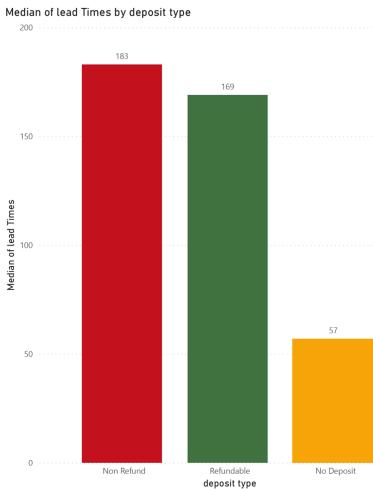
# Khách hàng không đặt cọc thì có huỷ phòng hay không?



Biểu đồ thể hiện tỷ lệ huỷ phòng ở mỗi kiểu đặt cọc

- Tỷ lệ khách hàng huỷ phòng ở nhóm đặt cọc ở **Non – Refund** lên tới **99,36%**.
- Tỷ lệ khách hàng huỷ phòng ở nhóm đặt cọc ở **No – Deposit** là **28,40%**.
- Tỷ lệ khách hàng huỷ phòng ở nhóm đặt cọc ở **Refundable** là **22,22%**.

## Khách hàng đặt cọc quá sớm thì có huỷ phòng hay không?

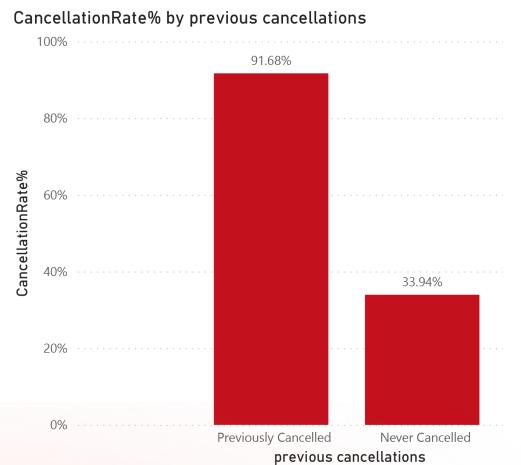


Biểu đồ thể hiện thời gian trung bình ở các kiểu đặt cọc

- Thời gian đặt cọc ở **Non – Refund** trung bình vào khoảng **183 ngày**
- Thời gian đặt cọc ở **Refundable** trung bình rơi vào khoảng **169 ngày**
- Thời gian đặt cọc ở **No – Deposit** trung bình rơi vào khoảng **57 ngày**

Dựa trên phân tích trước, nhận thấy rằng đặt phòng không hoàn tiền có tỷ lệ hủy cao nhất so với loại đặt phòng khác và dựa trên phân tích về thời gian chờ và hủy, ta thấy rằng thời gian chờ dài hơn có tỷ lệ hủy cao hơn so với đặt trước ngắn hơn

## Khách hàng huỷ phòng trước đó thì lần kế tiếp đặt họ có huỷ tiếp hay không

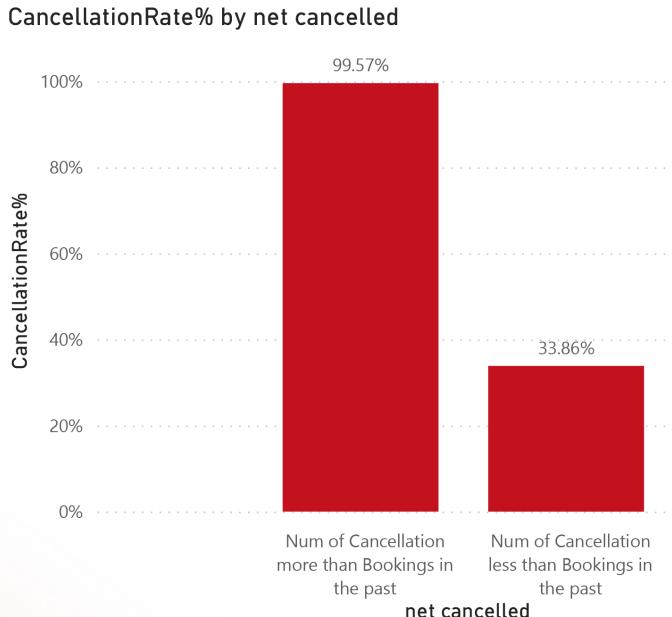


Biểu đồ thể hiện tỷ lệ huỷ phòng đối với lần đặt phòng trước đó

- Tỷ lệ huỷ phòng của khách hàng đã từng huỷ phòng trước đó là **91,68%**.
- Đối với khách hàng chưa bao giờ huỷ thì lần đặt tiếp theo sẽ huỷ với tỉ lệ **33,94%**.

Điều này cho thấy rằng đối với **những người đã từng huỷ phòng trước đó** sẽ có xu hướng huỷ phòng tiếp.

# Khách hàng đã từng huỷ phòng trước đó nhiều hơn việc đặt thành công thì lần đặt phòng tiếp sẽ huỷ?



Biểu đồ thể hiện tỷ lệ huỷ phòng đối với việc  
đặt phòng bị huỷ nhiều hơn hay ít hơn đặt  
phòng thành công

Dựa trên biểu đồ, ta thấy rằng:

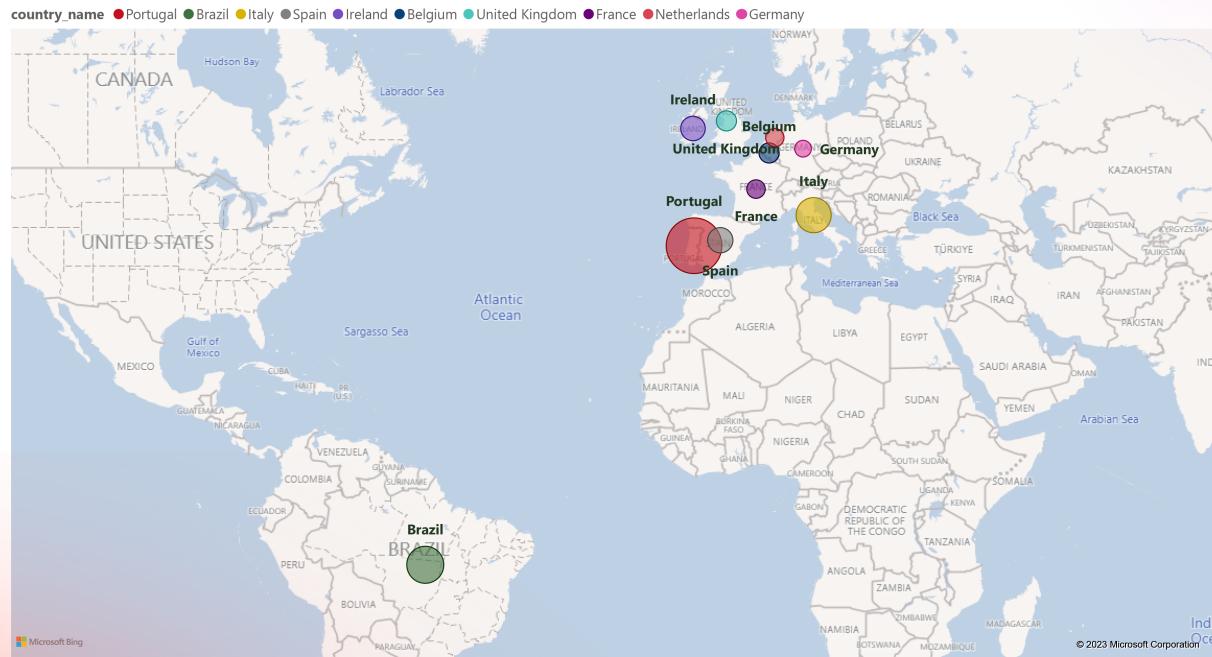
- Nếu người A có 10 lần đặt, 8 lần huỷ, 2 lần đặt thành công thì: **99,57%** họ sẽ huỷ phòng trong đợt đặt phòng tiếp theo
- Nếu người A có 10 lần đặt, 2 lần huỷ, 8 lần đặt thành công thì: **33,86%** họ sẽ huỷ phòng trong đợt đặt phòng tiếp theo

Nếu khách hàng đã từng huỷ phòng  
nhiều hơn việc đặt thành công trước  
đó thì chắc chắn sẽ huỷ tiếp

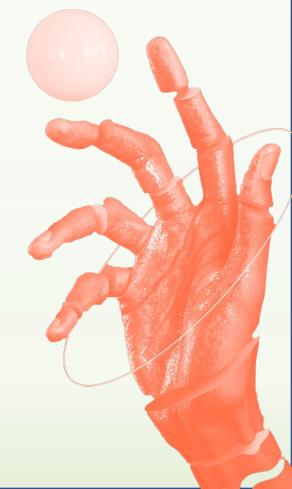
# Nguồn gốc khách hàng có ảnh hưởng tới việc huỷ phòng hay không?

Country name

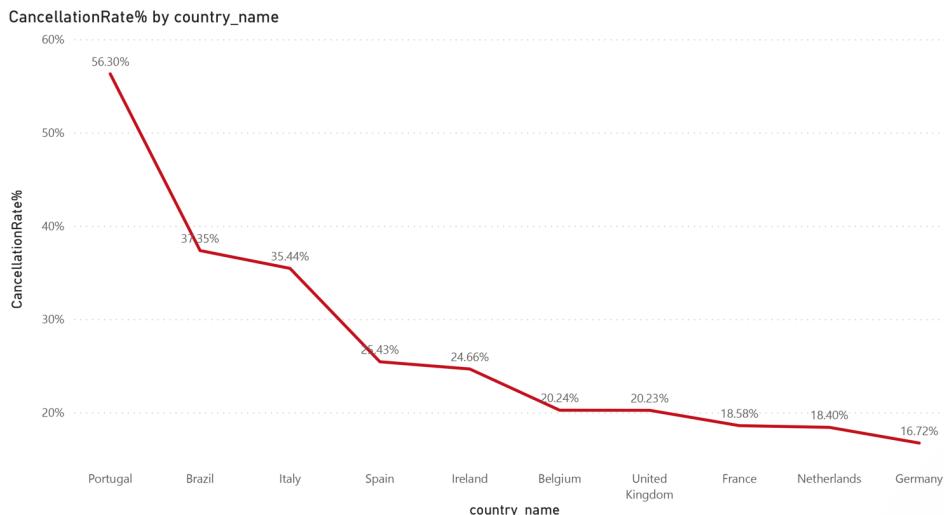
CancellationRate% by country\_name and country\_name



Biểu đồ thể hiện tỷ lệ huỷ phòng ở các quốc gia khách hàng sống



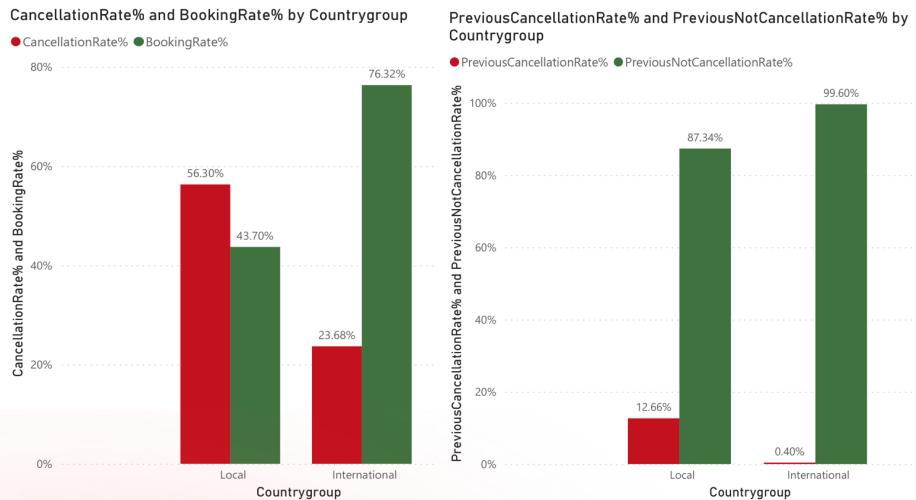
# Nguồn gốc của khách hàng có ảnh hưởng tới việc đặt phòng khách sạn hay không?



Biểu đồ thể hiện tỷ lệ huỷ phòng ở các quốc gia mà khách hàng sống

Tỷ lệ huỷ phòng cao nhất là ở **Portugal** với tỷ lệ **56,3%**. Sau đó là **Brazil** với tỷ lệ **37,35%** và **Italy** có tỷ lệ **35,44%**

# Nhóm khách hàng địa phương huỷ phòng nhiều hơn nhóm khách hàng quốc tế?



Biểu đồ thể hiện tỷ lệ huỷ phòng ở khách hàng địa phương và quốc tế

Tỷ lệ huỷ phòng ở địa phương cao với tỷ lệ **56,3%** trong khi việc đặt phòng cũng gần bằng với tỷ lệ **43,7%**.

Dựa trên phân tích trước đó, ta thấy rằng việc đặt phòng đã bị huỷ trước đó có tỷ lệ huỷ là **91,68 %**

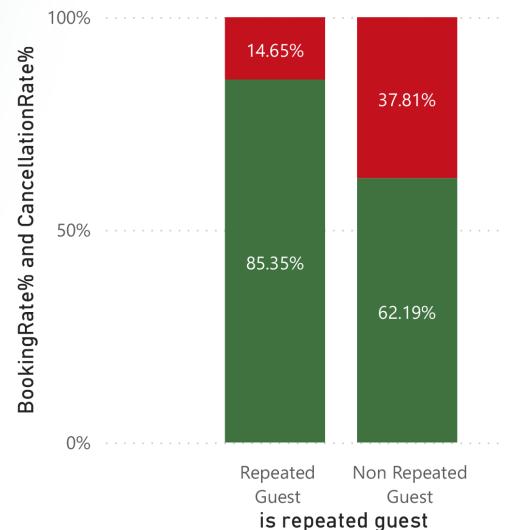
- **99,60%** đối với việc đặt phòng từ khách quốc tế đặt phòng chưa bao giờ bị huỷ bỏ trước đây
- Trong khi đối với đặt phòng tại địa phương, **12.66%** số lượt đặt phòng đã được huỷ trước đây

Số lượng huỷ trước cao hơn chắc chắn là một trong những yếu tố khiến đặt phòng nội địa có tỷ lệ huỷ cao hơn so với đặt phòng quốc tế

# Khách hàng đã từng đặt phòng trước đó liệu lần đặt phòng tiếp theo có huỷ phòng hay không

BookingRate% and CancellationRate% by is repeated guest

● BookingRate% ● CancellationRate%

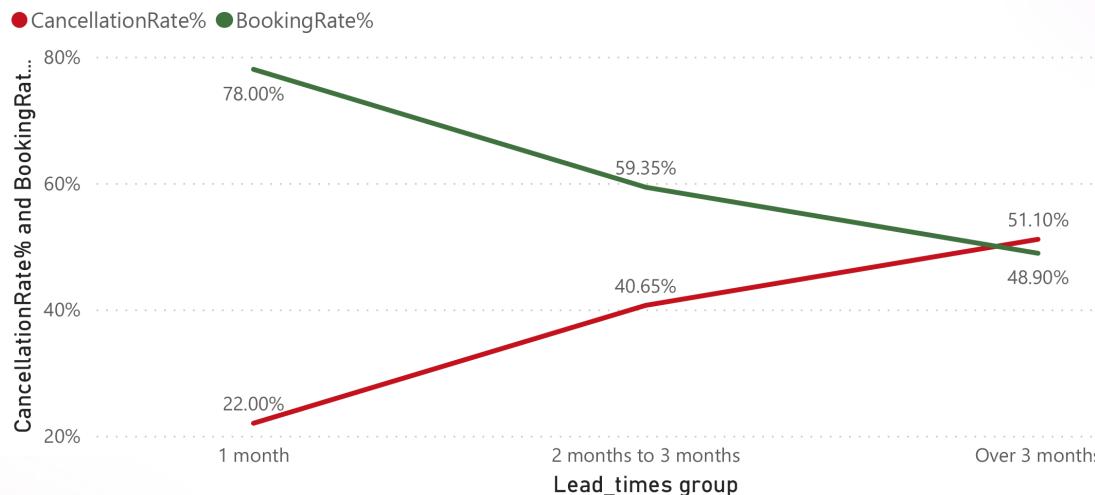


Biểu đồ thể hiện tỷ lệ huỷ phòng ở  
khách hàng lặp lại

- Khách hàng đã đặt phòng trước đó quay lại thì sẽ có tỉ lệ huỷ phòng thấp **14,65%** đối với lần đặt tiếp theo
- Khách hàng mới đặt lần đầu thì sẽ có tỉ lệ huỷ phòng **37,81%** cao hơn so với khách quen

# Khách hàng đặt phòng sớm thì có huỷ phòng hay không?

CancellationRate% and BookingRate% by Lead\_times group

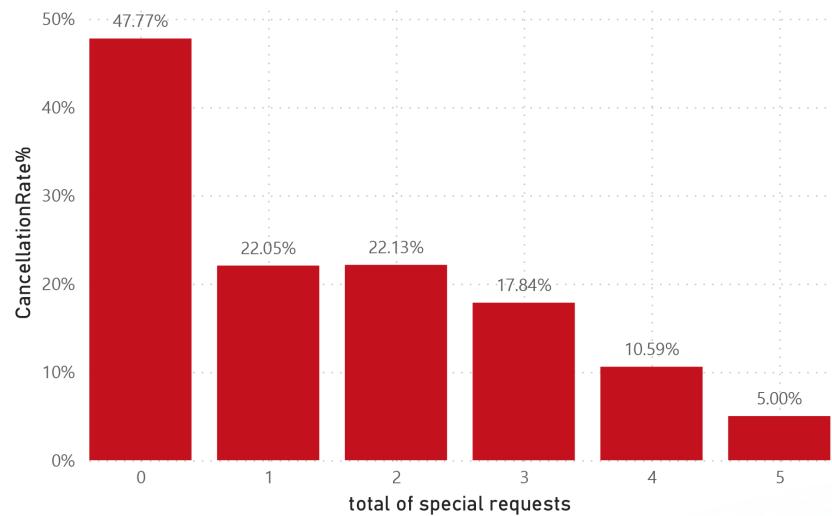


Biểu đồ thể hiện tỷ lệ huỷ phòng ở các mốc thời gian đặt phòng so với ngày đến

Tỷ lệ huỷ phòng **tỉ lệ thuận** với thời gian đặt phòng so với lúc đến. Từ 2 tới 3 tháng tỷ lệ huỷ phòng là **40,65%** và nếu trên 3 tháng thì tỉ lệ huỷ tăng lên **51,10%**. Vậy **thì nếu khách hàng đặt phòng càng sớm thì tỉ lệ huỷ càng cao**

# Khách hàng đưa ra yêu cầu đặc biệt nhiều thì có huỷ phòng hay không?

CancellationRate% by total of special requests



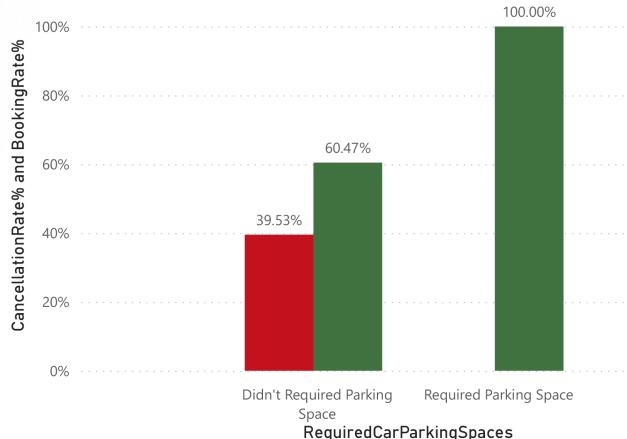
- Tỷ lệ huỷ phòng dường *nhu giảm dần nếu số lượng yêu cầu đặc biệt của khách hàng tăng lên.*
- Ở mức không có yêu cầu, tỷ lệ huỷ là **47,77%** và ở mức 1 – 5 yêu cầu tỷ lệ huỷ dao động từ **5%** (mức 5) tới **22,05%** (mức 1)

Biểu đồ thể hiện tỷ lệ huỷ phòng đối với số yêu cầu đặc biệt của khách hàng

# Khách hàng đưa ra yêu cầu có bãi đỗ xe thì có huỷ phòng hay không?

CancellationRate% and BookingRate% by RequiredCarParkingSpaces

● CancellationRate% ● BookingRate%



- Tỷ lệ huỷ phòng dường *nhu không có*, tỷ lệ **0%** *nếu nhu có* yêu cầu có bãi đậu xe ô tô .
- Ở mức không có yêu cầu, tỷ lệ huỷ là **39,53%**

Biểu đồ thể hiện tỷ lệ huỷ phòng đối với số yêu cầu có bãi đậu xe ô  
tô của khách hàng

EDA RECOMMENDATION

EDA RECOMMENDATION

EDA RECOMMENDATION

EDA RECOMMENDATION



EDA RECOMMENDATION  
EDA RECOMMENDATION

EDA RECOMMENDATION

EDA RECOMMENDATION

EDA RECOMMENDATION

EDA RECOMMENDATION

# EDA Recommendation

## Giới hạn đặt cọc không hoàn lại cho nhóm khách hàng Groups

Từ phân tích trên ta có thể thấy tỷ lệ huỷ của nhóm khách hàng Groups cao. Nên ta sẽ giới hạn việc đặt cọc của nhóm khách hàng Groups thành **chỉ được cọc không hoàn lại (Non – Refund)**. Tránh việc gây giảm doanh thu

## Giới hạn thời gian có thể đặt phòng

Ta thấy khách hàng sẽ huỷ phòng nhiều hơn nếu đặt trước hơn 3 tháng. Tuy nhiên ta không thể cài chế độ chỉ cho book phòng trước 3 tháng vì có thể ảnh hưởng tới khả năng nhận diện, tương tác và có thể là rủi ro về doanh thu của khách sạn. Vì vậy ta có thể cài đặt hệ thống đặt phòng **chỉ nhận đặt cọc phòng không hoàn lại dưới 3 tháng**.

# EDA Recommendation

## Tăng số lượng ở nhóm khách hàng đặt phòng trực tiếp

Từ phân tích trên ta có thể thấy tỷ lệ huỷ của nhóm khách hàng đặt trực tiếp ~ 15,5% và khách hàng bổ sung thấp ~ 12%. Vì vậy ta thực hiện các chiến dịch để **nâng cao độ nhận diện của khách sạn với các nhóm đặt trực tiếp**. Có thể bằng các giải pháp sau:

- Tạo các chương trình giảm giá, ưu đãi coupon khi đi lần sau, khách hàng thân thiết khi đặt trực tiếp qua kênh khách sạn
- Tối ưu hoá, cải thiện UI/UX trang web
- Tích hợp Web Booking Engine, biến website thương hiệu của bạn thành nguồn đặt phòng trực tiếp tiềm năng
- Cho phép khách đặt phòng trực tiếp qua trang Facebook của thương hiệu
- Phát triển sự hiện diện của bạn trên các công cụ tìm kiếm siêu dữ liệu
- Tăng danh tiếng trực tuyến cho khách sạn
- Xử lý cẩn thận việc từ bỏ đặt phòng

# EDA *Recommendation*

## **Thiết lập kiểu cọc cho nhóm khách hàng đã huỷ đặt phòng vào những lần trước**

Đặt chỗ trước đó đã bị hủy có tỷ lệ hủy là 92%. Nhìn vào mô hình này chúng ta thấy rằng đặt phòng đã bị hủy trước đó rất có thể sẽ bị hủy bỏ một lần nữa. Để bảo vệ khách sạn khỏi mất doanh thu do loại hủy bỏ này, khách sạn cần thiết lập thanh toán đặt phòng trong **đặt cọc tạm ứng/không hoàn lại**, cho đặt phòng đã bị hủy bỏ trước. Nếu khách hàng có đặt chỗ trước đó huỷ nhiều hơn thành công thì thiết lập thanh toán cọc tạm ứng. Ngoài ra hãy **đánh dấu đỏ** những khách hàng này trên trang web để dễ dàng nhận biết.

# EDA Recommendation

## Tăng số lượng khách hàng ở địa phương và quốc tế

Khách hàng địa phương tỷ lệ huỷ cạnh tranh với tỷ lệ đặt phòng thành công trong khi khách hàng quốc tế thì có tỷ lệ huỷ thấp ~ 24%. Vì vậy chúng ta cần **giảm tỷ lệ huỷ phòng của khách địa phương xuống và nâng cao tỷ lệ đặt phòng của khách quốc tế**. Như phân tích ở trên, đối với những khách hàng quay lại vào lần sau sẽ có khả năng huỷ thấp hơn. Tận dụng ưu thế ở địa phương thì dễ quay lại khách sạn. Ta có cách giải pháp sau:

- Nâng cao chất lượng dịch vụ ở khách sạn
- Tăng cường quảng bá trên các nền tảng online hoặc offline để tăng nhận diện thương hiệu, tận dụng ưu thế Bồ Đào Nha là nơi du lịch lý tưởng ở Châu Âu
- Chạy các chương trình ưu đãi, khách hàng thân thiết
- Cho phép tổ chức sự kiện,....
- Sử dụng influencers/ KOLs để quảng bá

# EDA Recommendation

## Tăng số lượng khách hàng yêu cầu bãi đỗ xe và yêu cầu đặc biệt

Ta nhận thấy khách hàng có yêu cầu đặc biệt và có bãi giữ xe tỷ lệ huỷ cực thấp, 22% đến 5% đối với khách hàng có yêu cầu đặc biệt và 0% đối với khách yêu cầu có bãi đỗ xe ( 7407 lượt ~ 10% tổng lượt đặt phòng thành công). Vì vậy ta cần **nâng cao sự thu hút đối với các khách hàng có những yêu cầu này:**

- Highlight có bãi đỗ xe ở các trang đặt phòng
- Tạo chất lượng dịch vụ tốt để khách hàng có thể dễ dàng yêu cầu

MACHINE LEARNING

MACHINE LEARNING

MACHINE LEARNING

MACHINE LEARNING

MACHINE LEARNI

# MACHINE LEARNING

# MACHINE LEARNING



MACHINE LEARNING

MACHINE LEARNING

MACHINE LEARNING

MACHINE LEARNING

MACHINE LEARNING

# MACHINE Learning

## VỀ DỰ ÁN

- 1. Bài toán:** Tạo ra mô hình cho phép người quản lý khách sạn/người quản lý doanh thu thực hiện các hành động đối với các đặt phòng được xác định là "có khả năng bị hủy",
- 2. Mục đích:**
  - Góp phần quản lý doanh thu của khách sạn.
  - Cho phép các nhà quản lý khách sạn giảm thiểu tổn thất doanh thu do hủy đặt phòng và giảm thiểu rủi ro liên quan đến đặt trước quá nhiều (chi phí phân bổ lại, tiền mờ hoặc bồi thường dịch vụ và đặc biệt quan trọng hiện nay là chi phí branding thương hiệu).
  - Cho phép các nhà quản lý khách sạn thực hiện các chính sách hủy đặt phòng một cách dễ dàng hơn mà không làm tăng sự không chắc chắn. Điều này có khả năng chuyển thành nhiều doanh số hơn, tạo ra nhiều các đặt phòng thành công hơn.

Trước khi tiến hành build model, ta tạo thêm 2 features mới dựa trên việc phân tích trên:

1. Tạo net\_cancelled với 2 giá trị là 1 và 0. Nếu việc huỷ đặt phòng ở những lần trước lớn hơn việc đặt thành công thì giá trị 1 và ngược lại là 0.
2. Tạo room với 2 giá trị là 1 và 0. Nếu nhận cùng phòng mình đặt là 0 và đặt phòng A nhưng nhận phòng B thì là 1.

Tiến hành drop các cột không dùng đến dựa trên cor ở trang sau:

```
['hotel', 'adr', 'arrival_date_year', 'arrival_date_month',
 'arrival_date_day_of_month', 'arrival_date_week_number', 'adults', 'children', 'babies',
 'days_in_waiting_list', 'stays_in_weekend_nights', 'reserved_room_type', 'country_name', 'country',
 'reservation_status', 'meal', 'reservation_status_date'])
```

Và rồi làm các thủ tục Onehot Encoder, Scale, Split Data thành tệp Train và Test như bình thường

## MARKET OVERVIEW

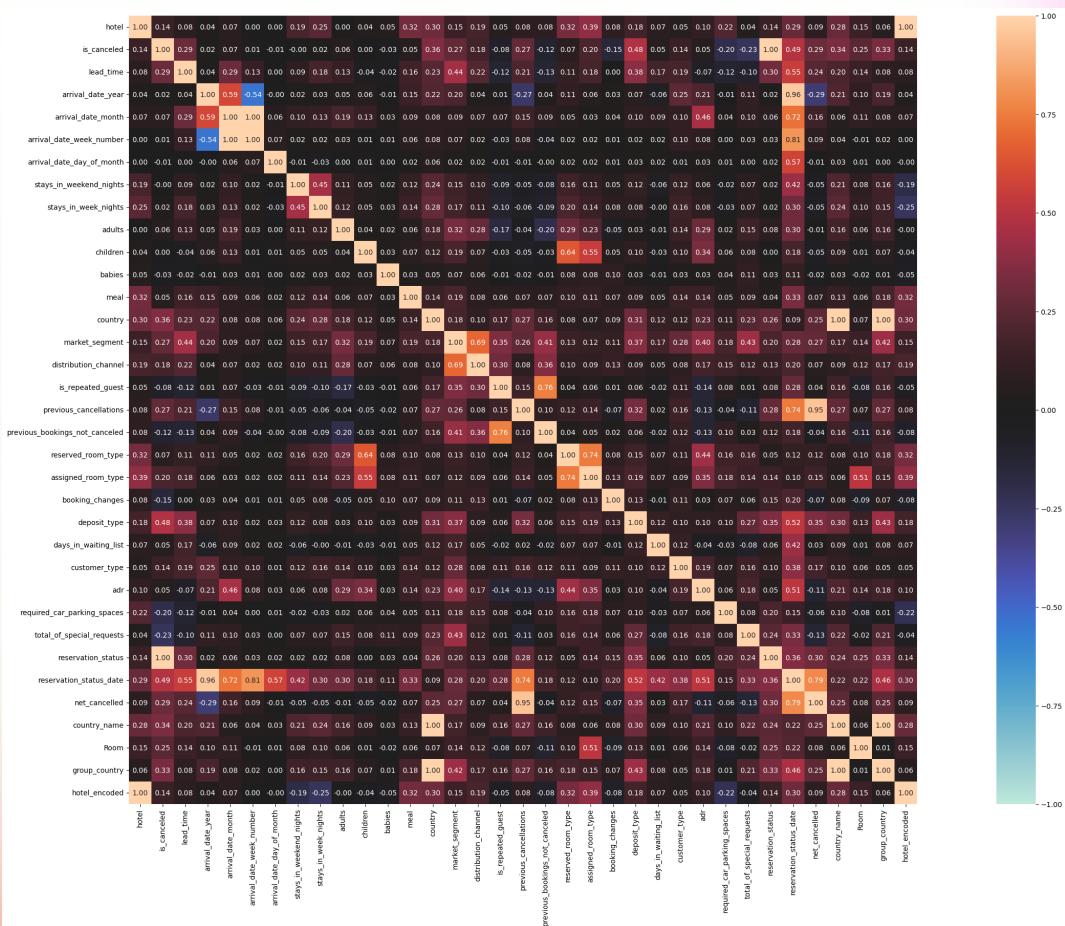
## PROBLEM STATEMENT

## CANCELLATION INSIGHT &amp; ANALYSIS

## EDA &amp; RECOMMENDATION

## MACHINE LEARNING

## SUMMARY



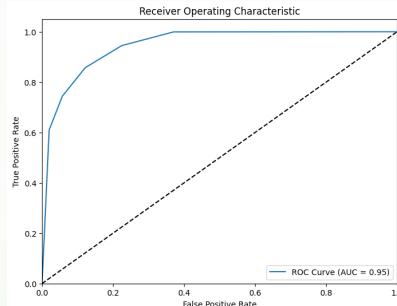
# MACHINE Learning

## CÁC MÔ HÌNH SỬ DỤNG TRONG DỰ ÁN:

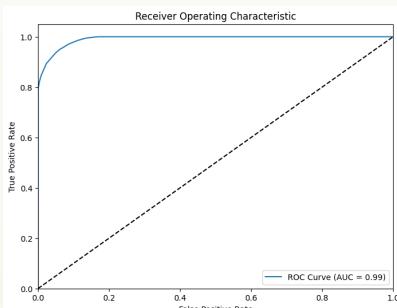
1. KNeighbors Classifier
2. Logistic Regression
3. Random Forest Classifier
4. XGB Classifier
5. AdaBoost Classifier
6. Decision Tree Classifier

# MACHINE Learning

## KNeighbors Classifier

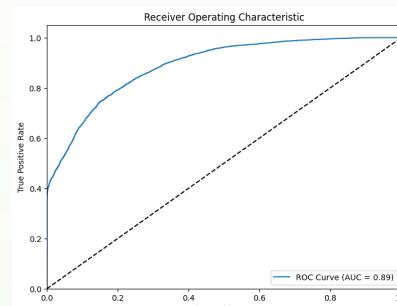


## AdaBoost Classifier

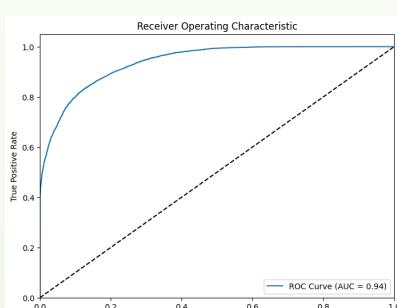


## RECEIVER OPERATING CHARACTERISTIC

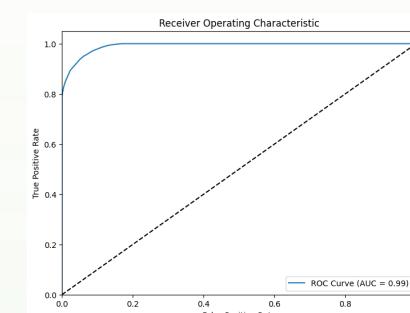
## Logistic Regression



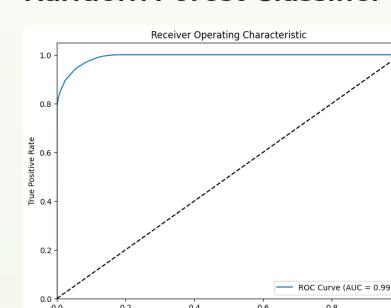
## XGB Classifier



## Decision Tree Classifier



## Random Forest Classifier



Nhìn chung tất cả các mô hình đều có **AUC > 0,89** -> Mô hình tốt và có thể sử dụng.

# MACHINE Learning

## EVALUATION MATRIX

### KNeighbors Classifier

Training Accuracy: 0.8722737186477645

Test Accuracy: 0.8705226071638286

	precision	recall	f1-score	support
0	0.91	0.88	0.90	14982
1	0.81	0.86	0.83	8860
accuracy		0.87	0.87	23842
macro avg	0.86	0.87	0.86	23842
weighted avg	0.87	0.87	0.87	23842

### Logistic Regression

Training Accuracy: 0.8091078768559684

Test Accuracy: 0.8082795067527891

	precision	recall	f1-score	support
0	0.81	0.92	0.86	14982
1	0.81	0.63	0.71	8860
accuracy		0.81	0.81	23842
macro avg	0.81	0.77	0.78	23842
weighted avg	0.81	0.81	0.80	23842

### Decision Tree Classifier

Training Accuracy: 0.9427376059055448

Test Accuracy: 0.9457679724855297

	precision	recall	f1-score	support
0	0.94	0.98	0.96	14982
1	0.96	0.89	0.92	8860
accuracy		0.95	0.95	23842
macro avg	0.95	0.94	0.94	23842
weighted avg	0.95	0.95	0.95	23842

### AdaBoost Classifier

Training Accuracy: 0.9427376059055448

Test Accuracy: 0.9457679724855297

	precision	recall	f1-score	support
0	0.94	0.98	0.96	14982
1	0.96	0.89	0.92	8860
accuracy		0.95	0.95	23842
macro avg	0.95	0.94	0.94	23842
weighted avg	0.95	0.95	0.95	23842

### XGB Classifier

Training Accuracy: 0.8637907893633084

Test Accuracy: 0.8641473030786008

	precision	recall	f1-score	support
0	0.87	0.92	0.89	14982
1	0.85	0.78	0.81	8860
accuracy		0.86	0.86	23842
macro avg	0.86	0.85	0.85	23842
weighted avg	0.86	0.86	0.86	23842

### Random Forest Classifier

Training Accuracy: 0.9427376059055448

Test Accuracy: 0.9457679724855297

	precision	recall	f1-score	support
0	0.94	0.98	0.96	14982
1	0.96	0.89	0.92	8860
accuracy		0.95	0.95	23842
macro avg	0.95	0.94	0.94	23842
weighted avg	0.95	0.95	0.95	23842

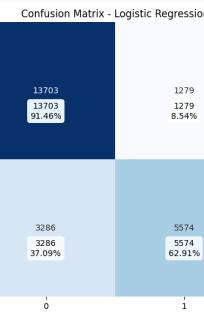
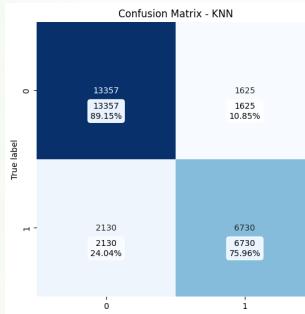
Random Forest Classifier, AdaBoost, Decision Tree Classifier có accuracy tốt nhất. Tuy nhiên ta phải xét thêm các yếu tố khác

Dataset cân bằng: 63% Confirmed, 37% Canceled  
 Mọi class đều quan trọng như nhau

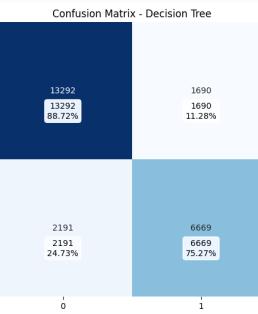
# MACHINE Learning

## CONFUSION MATRIX

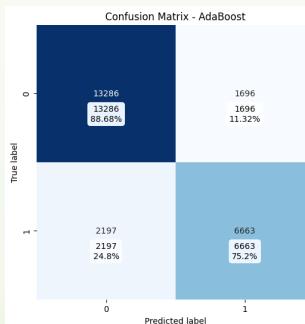
KNeighbors Classifier   Logistic Regression



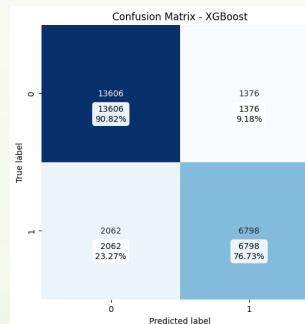
Decision Tree Classifier



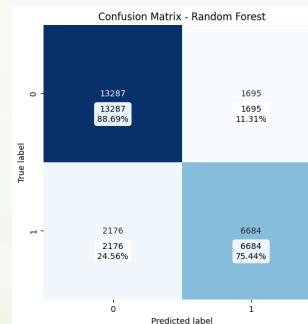
AdaBoost Classifier



XGB Classifier



Random Forest Classifier



- TPR cao nhất nằm ở mô hình XGB với **76,73%** sau đó là KNN với **75,96%** và Random Forest **75,44%**.

- Kết hợp với Accuracy và AUC ở trên thì chúng ta có các mô hình tốt: **XGB Classifier, Random Forest Classifier, AdaBoost Classifier, KNN Classifier** là mô hình tốt.

# MACHINE Learning

## TUNED MODEL EVALUATION MATRIX

### XGB Classifier

Training Accuracy: 0.8629309621676033

Test Accuracy: 0.8558006878617566

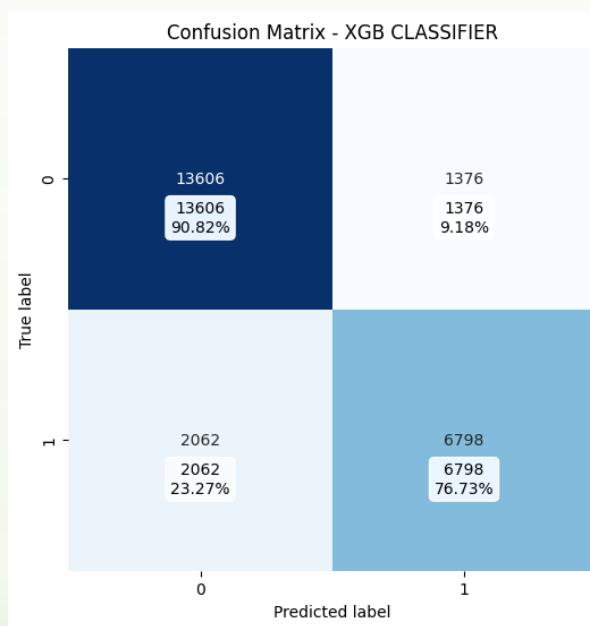
	precision	recall	f1-score	support
0	0.87	0.91	0.89	14982
1	0.83	0.77	0.80	8860
accuracy			0.86	23842
macro avg	0.85	0.84	0.84	23842
weighted avg	0.85	0.86	0.85	23842

- Sau khi sử dụng Hyperparameter để tuning model thì XGB Classifier có accuracy tốt nhất
- Không xảy ra hiện tượng Overfitting

# MACHINE Learning

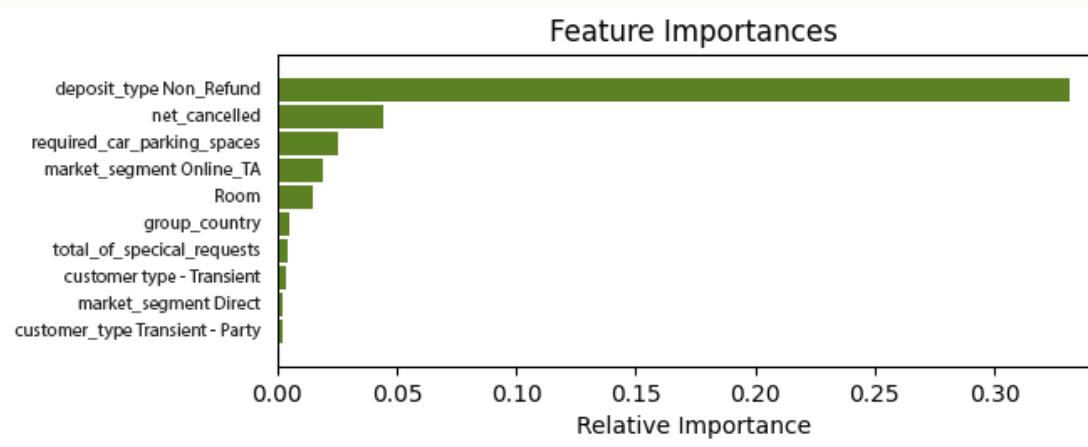
## TUNED MODEL CONFUSION MATRIX

### XGB Classifier



# MACHINE Learning

## FEATURE IMPORTANCE OF XGB CLASSIFIER



# MACHINE Learning

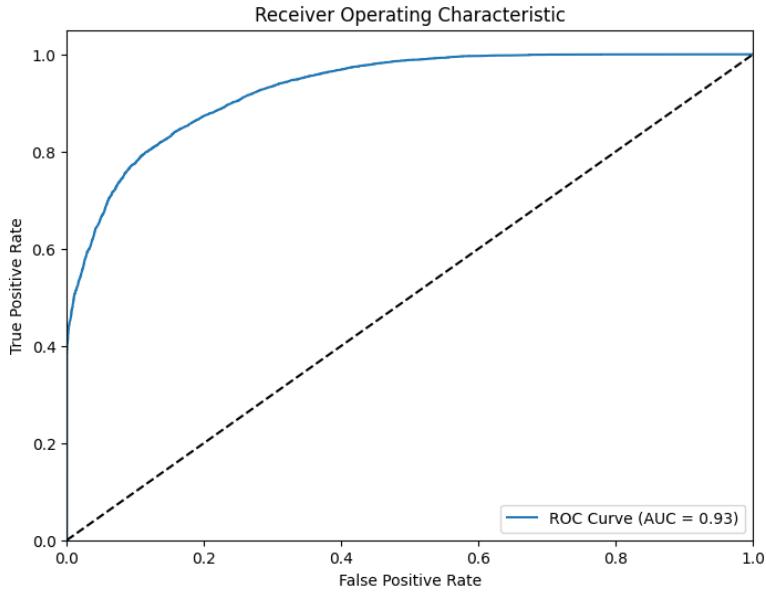
## FEATURE IMPORTANCE

Dựa trên biểu đồ ta thấy hoàn toàn phù hợp với những finding ta tìm ra trước đó:

- Đặt phòng không hoàn tiền có tỷ lệ hủy cao nhất so với loại đặt phòng khác, tiếp đến là không cọc. Dựa trên phân tích về thời gian chờ và hủy, ta thấy rằng thời gian chờ dài hơn có tỷ lệ hủy cao hơn so với đặt trước ngắn hơn.
- Những người đã huỷ phòng ở những lần trước hoặc có số lượng huỷ những lần trước nhiều hơn số lần thành công thì sẽ huỷ tiếp.
- Những người có yêu cầu đặc biệt và bãi giữ xe thì sẽ có tỷ lệ huỷ phòng cực thấp
- Số lượng huỷ phòng ở địa phương cao hơn quốc tế.
- Đặt phòng A nhưng nhận được phòng B sẽ huỷ phòng
- Nhóm khách hàng Groups có tỷ lệ huỷ cao, nhóm TA/TO Offline và Online cũng có tỷ lệ huỷ bằng  $\frac{1}{2}$  tỷ lệ đặt phòng thành công
- Kiểu khách hàng ngắn hạn có tỉ lệ huỷ cao

# MACHINE Learning

## ROC SCORE SAU TUNNING CỦA XGB



SUMMARY

SUMMARY

SUMMARY

SUMMARY

SUMMARY

SUMMARY

SUMMARY

SUMMARY



# SUMMARY SUMMARY



SUMMARY

SUMMARY

SUMMARY

SUMMARY

SUMMARY

SUMMARY

SUMMARY

SUMMARY

## KẾT LUẬN RÚT RA

Trong dự án này, mục đích đưa ra là để phát triển một mô hình dự đoán việc hủy phòng khách sạn. Sau khi đánh giá nhiều mô hình và xem xét các chỉ số hiệu suất khác nhau thì mô hình XGB được nhận thấy là mô hình phù hợp và tốt nhất. XGB Classifier đã thể hiện:

**Độ chính xác** là **85%** trên tập test, **AUC** là **0,93** và **True Positive Rate (TPR)** là **76,73%**.

Phân tích cho thấy rằng một số tính năng đóng vai trò quan trọng trong việc dự đoán hủy phòng khách sạn. Các tính năng có ảnh hưởng nhất bao gồm các kiểu đặt cọc, quốc gia, các yêu cầu đặc biệt của khách hàng, phòng được nhận và phòng đặt, nhóm khách hàng, nhóm khách hàng yêu cầu có bãi giữ xe, việc huỷ phòng nhiều hơn đặt ở những lần trước, các kiểu khách hàng, thời gian đặt phòng. *Các tính năng này cung cấp thông tin chi tiết có giá trị về hành vi và sở thích của khách hàng, cho phép các khách sạn tối ưu hóa hoạt động của họ và giảm thiểu việc hủy phòng.*

## Ý NGHĨA THỰC TIỄN

Mô hình đã phát triển có thể được các khách sạn tận dụng để nâng cao chiến lược quản lý hủy đặt phòng và tối ưu hóa doanh thu. Bằng cách sử dụng các đặc điểm có ảnh hưởng đã xác định, các khách sạn có thể thực hiện **các biện pháp chủ động để giảm lượng hủy đặt phòng và cải thiện sự hài lòng của khách hàng**. Ví dụ:

- **Loại đặt cọc:** Khách sạn có thể điều chỉnh chính sách đặt cọc dựa trên phân khúc khách hàng và xu hướng hủy đặt phòng của họ. Cách tiếp cận này có thể khuyến khích khách tiếp tục đặt phòng, giảm khả năng hủy đặt phòng.
- **Phân khúc quốc gia và thị trường:** Bằng cách hiểu các kiểu hủy đặt phòng ở các quốc gia và phân khúc thị trường khác nhau, các khách sạn có thể nhắm mục tiêu nỗ lực tiếp thị của mình và cung cấp các chương trình khuyến mãi được cá nhân hóa cho các nhóm khách hàng cụ thể, do đó tăng cam kết đặt phòng của họ.
- **Tổng số yêu cầu đặc biệt và bãi đỗ xe:** Bằng cách ưu tiên sở thích của khách và đáp ứng các yêu cầu đặc biệt, các khách sạn có thể nâng cao trải nghiệm tổng thể của khách và tăng lòng trung thành của khách, giúp giảm khả năng hủy đặt phòng.

## HẠN CHẾ VÀ CẢI TIẾN

Một số **hạn chế và cơ hội** để cải thiện trong dự án này:

- **Hạn chế về dữ liệu:** Độ chính xác của dự đoán của mô hình phụ thuộc rất nhiều vào chất lượng và tính đại diện của dữ liệu có sẵn. Việc thu thập thêm dữ liệu và bao gồm các tính năng bổ sung có liên quan có thể cải thiện hiệu suất của mô hình.
- **Kỹ thuật đối với feature:** Khám phá các kỹ thuật kỹ thuật nâng cao hơn, chẳng hạn như hiệu ứng kết hợp và chuyển đổi feature, có khả năng phát hiện ra các mẫu dự đoán bổ sung và cải thiện độ chính xác của mô hình.
- **Các yếu tố bên ngoài:** Mô hình này chỉ tập trung vào các feature có sẵn trong tập dữ liệu. Tuy nhiên, các yếu tố bên ngoài như điều kiện kinh tế, sự kiện công cộng hoặc tư vấn du lịch cũng có thể ảnh hưởng đến việc hủy khách sạn. Tích hợp các nguồn dữ liệu bên ngoài hoặc xem xét các yếu tố này trong phân tích có thể nâng cao hơn nữa khả năng dự đoán của mô hình.

# CUỐI CÙNG

Thông qua dự án này đã chứng minh **tính hiệu quả của Công cụ phân loại XGB trong việc dự đoán hủy phòng khách sạn**. Các tính năng có ảnh hưởng đã xác định cung cấp thông tin chi tiết có giá trị có thể định hướng cho các khách sạn trong quá trình ra quyết định nhằm **giảm thiểu việc hủy đặt phòng và tối ưu hóa hoạt động**. Các khách sạn nên **kết hợp mô hình đã phát triển vào hệ thống đặt phòng hiện tại của họ và theo dõi hiệu suất của nó theo thời gian**. Cập nhật và cải tiến mô hình thường xuyên dựa trên dữ liệu mới có thể cải thiện hơn nữa độ chính xác và khả năng dự đoán của mô hình. Bằng cách tận dụng những hiểu biết thu được từ phân tích này, các khách sạn có thể **cải thiện các chiến lược quản lý doanh thu và nâng cao trải nghiệm tổng thể của khách**.

A dark, atmospheric night scene of a resort. In the foreground, there's a large, calm body of water reflecting the surrounding lights. Along the water's edge, several lounge chairs with white umbrellas are arranged. In the background, there's a multi-story building with many windows, likely a hotel or resort. Several palm trees are scattered throughout the scene, their silhouettes standing out against the dark sky. The overall mood is peaceful and sophisticated.

THANK YOU