

Hotel Booking Analysis Report

Problem Statement

Trong vài năm qua, cả Khách sạn Thành phố và Khách sạn Khu nghỉ dưỡng đã trải qua sự gia tăng đáng kể trong tỷ lệ hủy đặt phòng. Do đó, cả hai khách sạn hiện đang đối mặt với một loạt các thách thức, chẳng hạn như doanh thu giảm và phòng khách sạn bị sử dụng không hiệu quả. Vì vậy, ưu tiên hàng đầu của cả hai khách sạn là giảm tỷ lệ hủy đặt phòng, điều này sẽ nâng cao hiệu quả của họ trong việc tạo ra doanh thu. Báo cáo này tập trung vào phân tích việc hủy đặt phòng khách sạn và các yếu tố khác không ảnh hưởng trực tiếp đến hoạt động kinh doanh và việc tạo ra doanh thu hàng năm của họ.

Research Question

1. Có bao nhiêu đặt phòng bị hủy?
2. Tỷ lệ đặt phòng giữa Khách sạn Nghỉ dưỡng và Khách sạn Thành phố là bao nhiêu?
3. Tỷ lệ đặt phòng cho từng năm là bao nhiêu?
4. Tháng nào là bận rộn nhất cho khách sạn?
5. Khách đến từ quốc gia nào nhiều nhất?
6. Mọi người lưu trú bao lâu tại khách sạn?
7. Loại chỗ ở nào được đặt nhiều nhất (Đơn, Cặp đôi, Gia đình)?

Sau đó, tôi đã xây dựng mô hình dự đoán để dự đoán liệu đặt phòng có bị hủy hay không. Tôi sẽ:

- Thực hiện Feature Engineering để tạo ra các đặc trưng mới.
- Thực hiện Data Selection để chọn chỉ các đặc trưng liên quan.
- Trực quan hóa dữ liệu bằng cách sử dụng các biểu đồ và đồ thị.
- Cuối cùng, đưa ra một số hiểu biết sâu sắc bằng cách phân tích chúng.

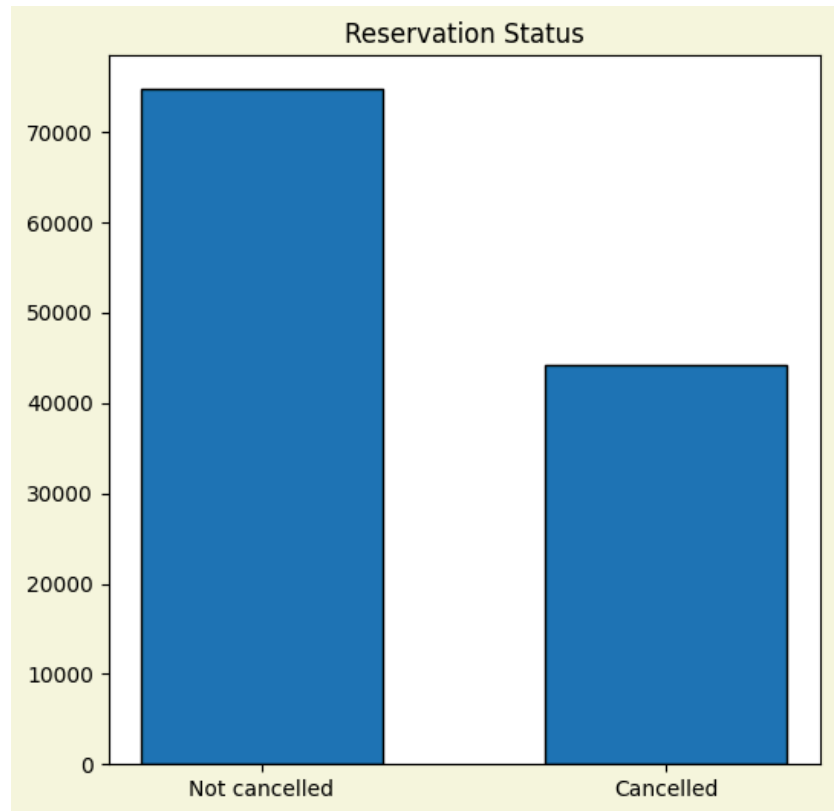
Analysis and Findings

1. Tỷ lệ hủy phòng

75.166 khách hàng thực tế đã làm thủ tục trả phòng tại khách sạn.

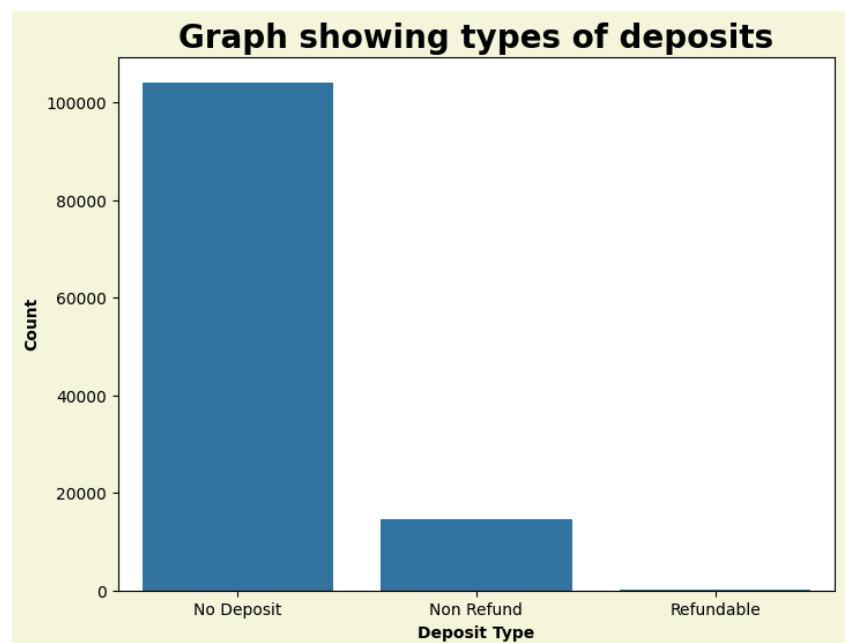
43.017 khách hàng đã hủy đặt phòng của họ.

Do các đặt phòng bị hủy này, sẽ có tác động tiêu cực đến hoạt động kinh doanh của khách sạn, có nghĩa là khách sạn không thể tạo ra nhiều lợi nhuận hơn, họ đang mất đi khách hàng của mình.

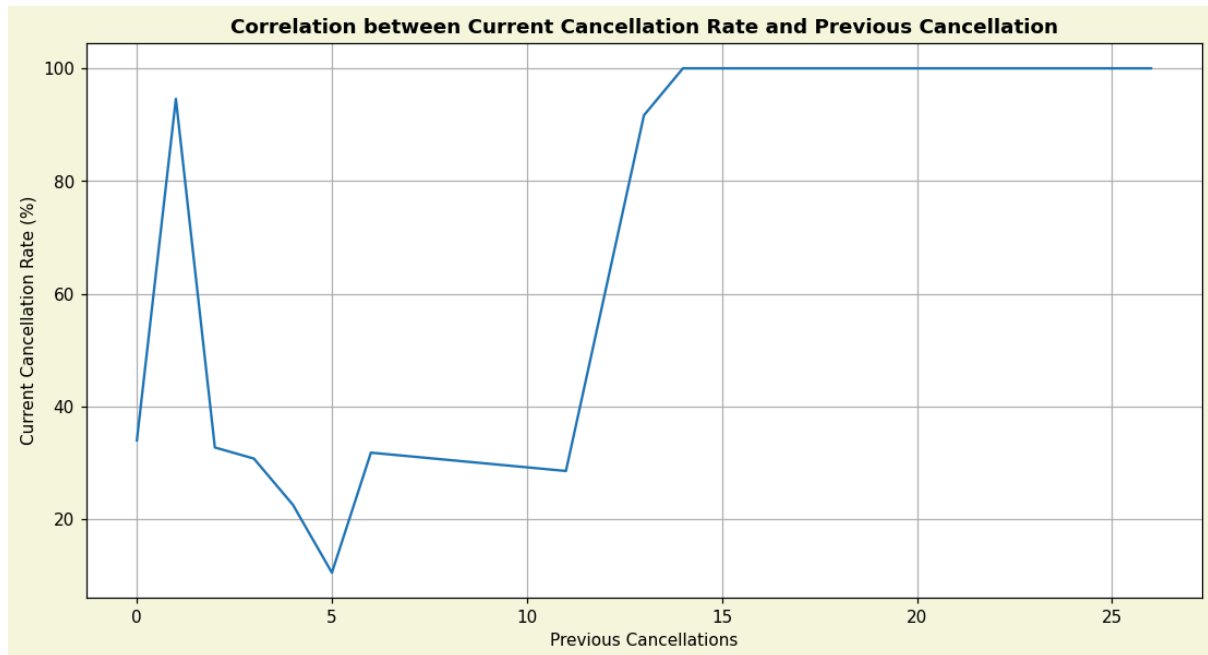


2. Chính sách đặt cọc của khách sạn

Hơn 85% khách sạn có chính sách “Không đặt cọc”.

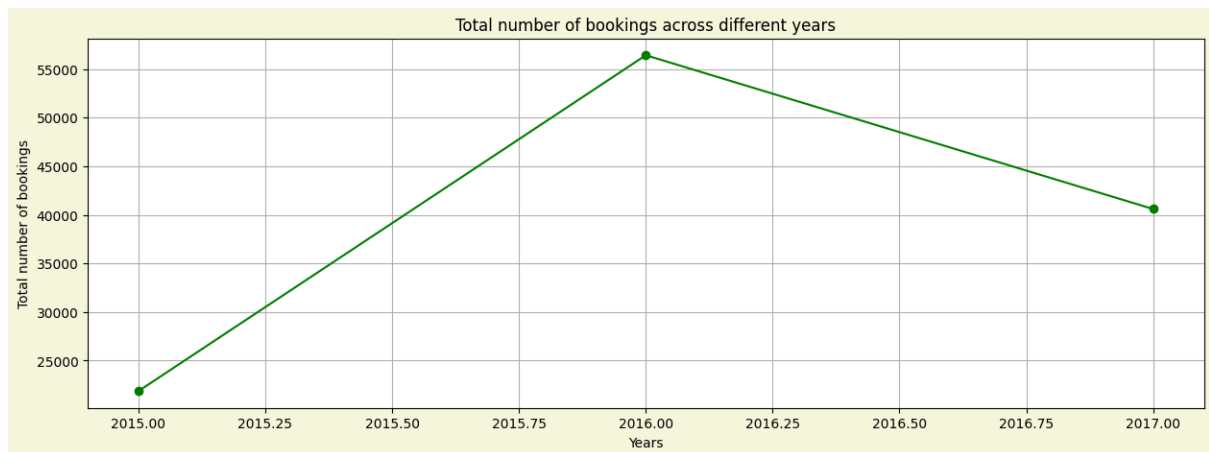


3. Mối quan hệ giữa các lần hủy trước đây và các lần hủy hiện tại

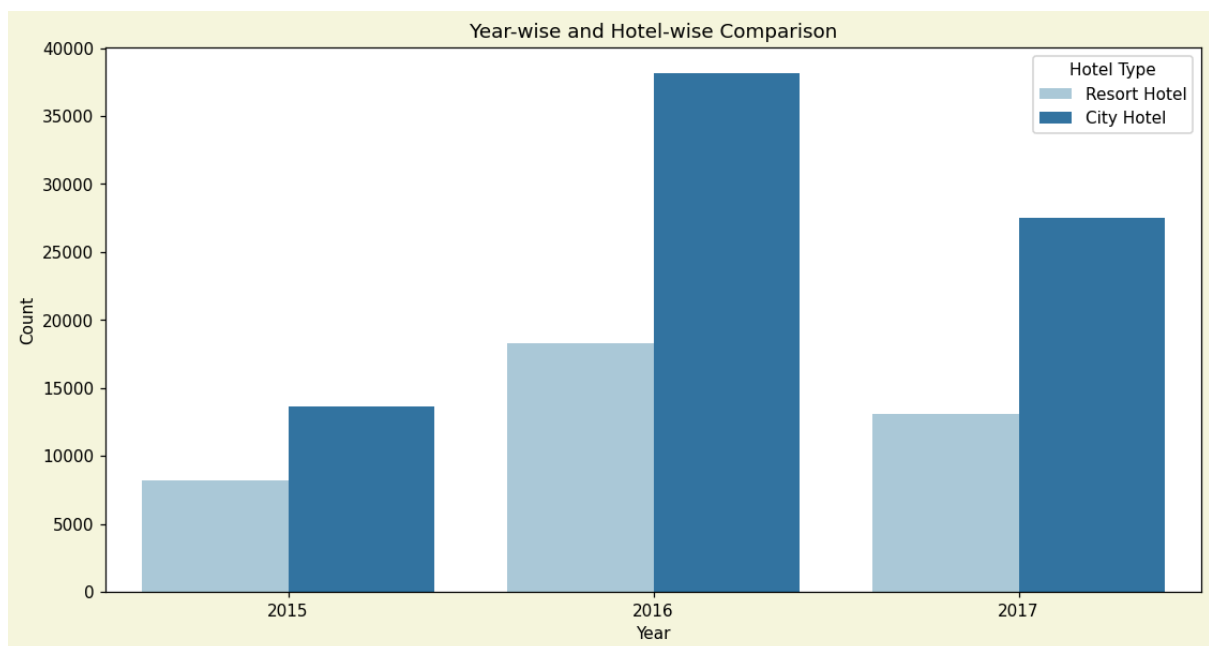


- Theo biểu đồ, khi khách hàng đã từng hủy đặt phòng khách sạn một lần trước đó, 94,58% trong số họ lại hủy đặt phòng lần nữa.
- Tỷ lệ hủy của những khách hàng đã hủy hơn 10 lần trước đó là 85,56%.

4. Tỷ lệ đặt phòng qua từng năm

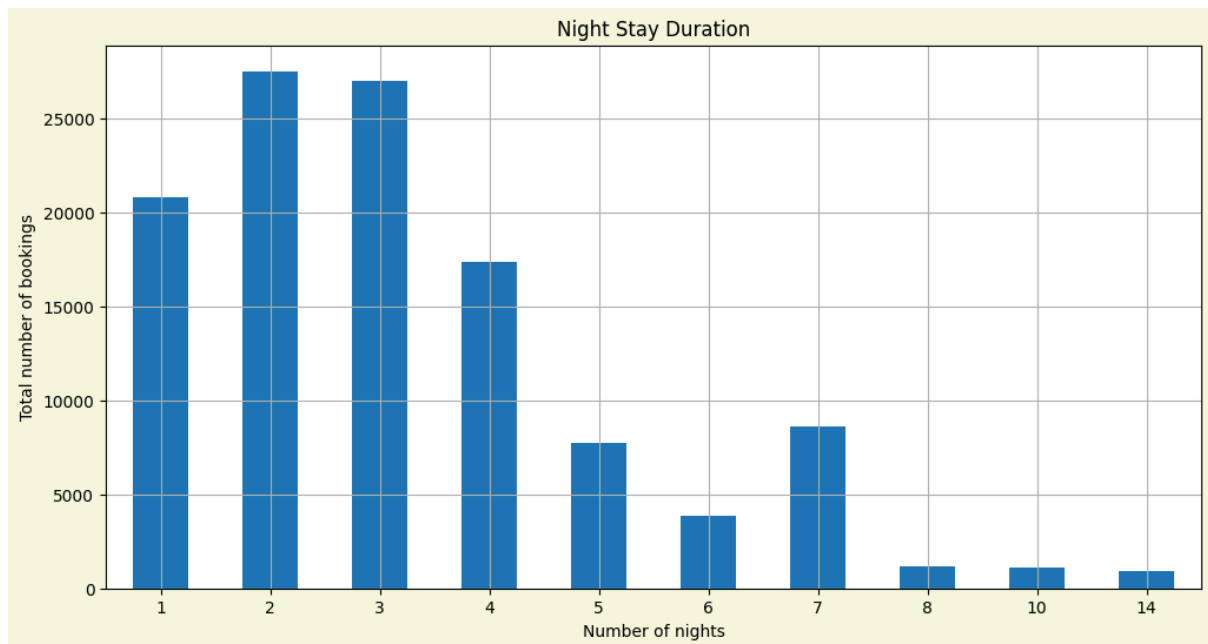


Biểu đồ này rõ ràng cho thấy rằng số lượng đặt phòng trong năm 2016 là cao nhất so với các năm khác, với con số khoảng hơn 56.500. Vào năm 2017, số lượng đặt phòng giảm xuống so với năm trước, cụ thể là 40.600.

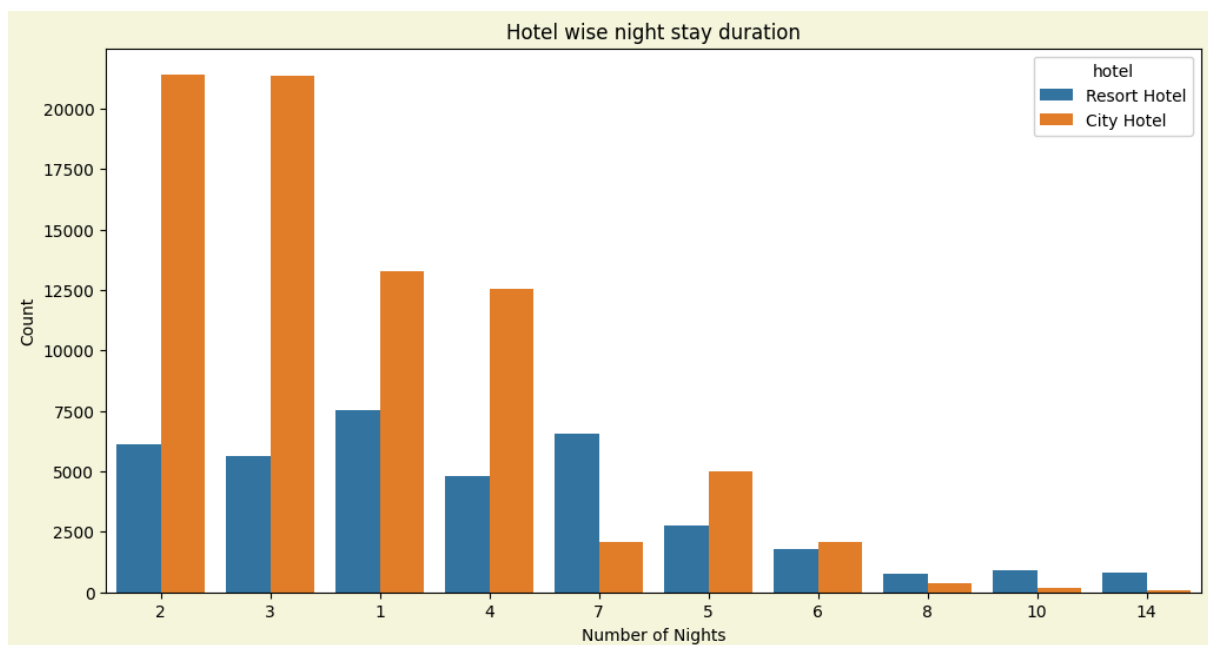


Nhu cầu thuê khách sạn thành phố (City Hotel) luôn cao hơn so với khách sạn nghỉ dưỡng (Resort Hotel) trong 3 năm liên tiếp.

5. Thời gian lưu trú tại khách sạn



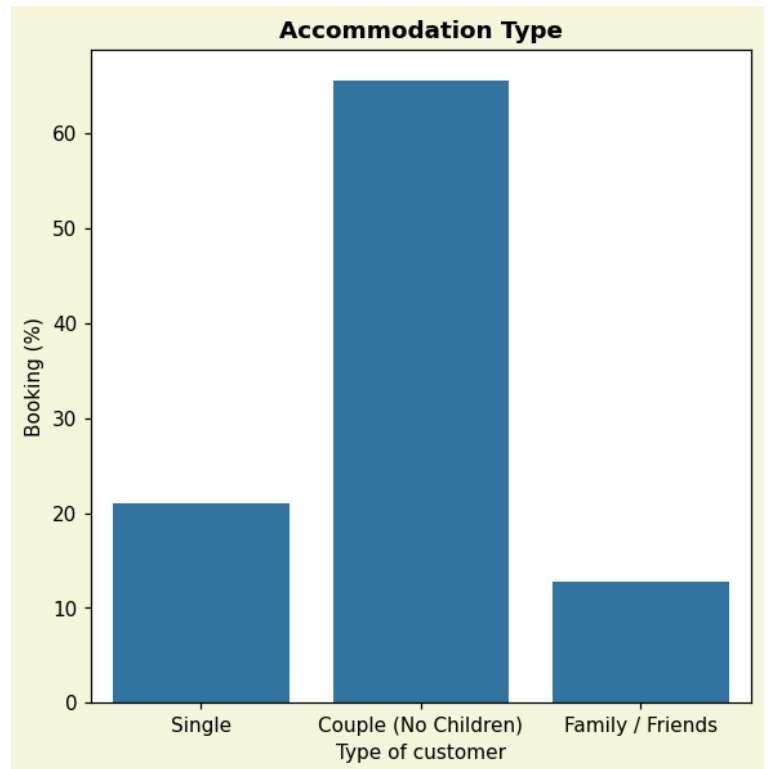
- Hầu hết khách hàng thích ở lại một, hai, ba và bốn đêm, trong đó số lượng đặt phòng cho hai và ba đêm có nhu cầu cao nhất.
- Nếu chúng ta cố gắng phân tích thời gian lưu trú tối ưu để có được giá hàng ngày tốt nhất, các khách sạn cũng nên đưa ra các ưu đãi cho thời gian lưu trú trung hạn vì các ưu đãi cho thời gian lưu trú dài hạn sẽ tiêu tốn phòng của họ và khách hàng mới không thể đặt phòng, dẫn đến khách sạn có thể không tạo ra được lợi nhuận thêm.



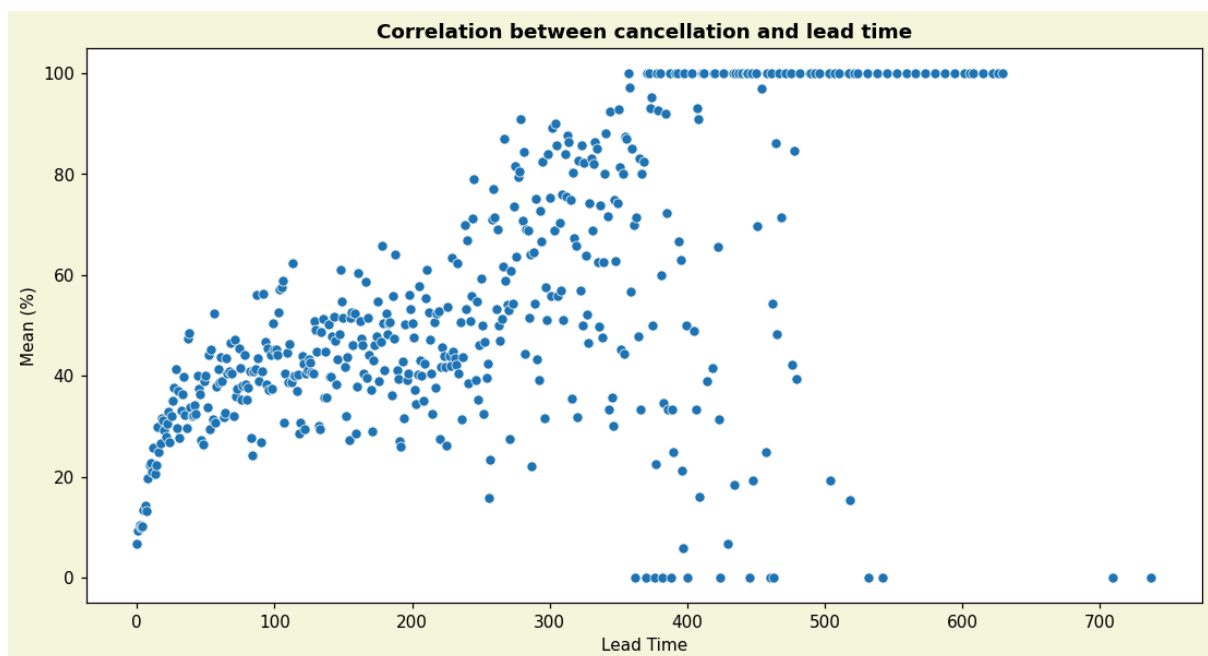
Theo biểu đồ trực quan này, mọi người quan tâm đến việc lưu trú tại khách sạn thành phố trong hai, ba, một và bốn ngày, trong khi đối với khách sạn nghỉ dưỡng, mọi người quan tâm đến việc lưu trú trong một, bảy, hai và ba ngày.

6. Loại chỗ ở được đặt nhiều nhất - Đơn, Cặp đôi, Gia đình

Từ biểu đồ, có thể thấy rằng số lượng khách hàng lớn nhất là những khách hàng cặp đôi (không có trẻ em). Những khách hàng gia đình hoặc nhóm bạn có số lượng khách hàng thấp nhất.

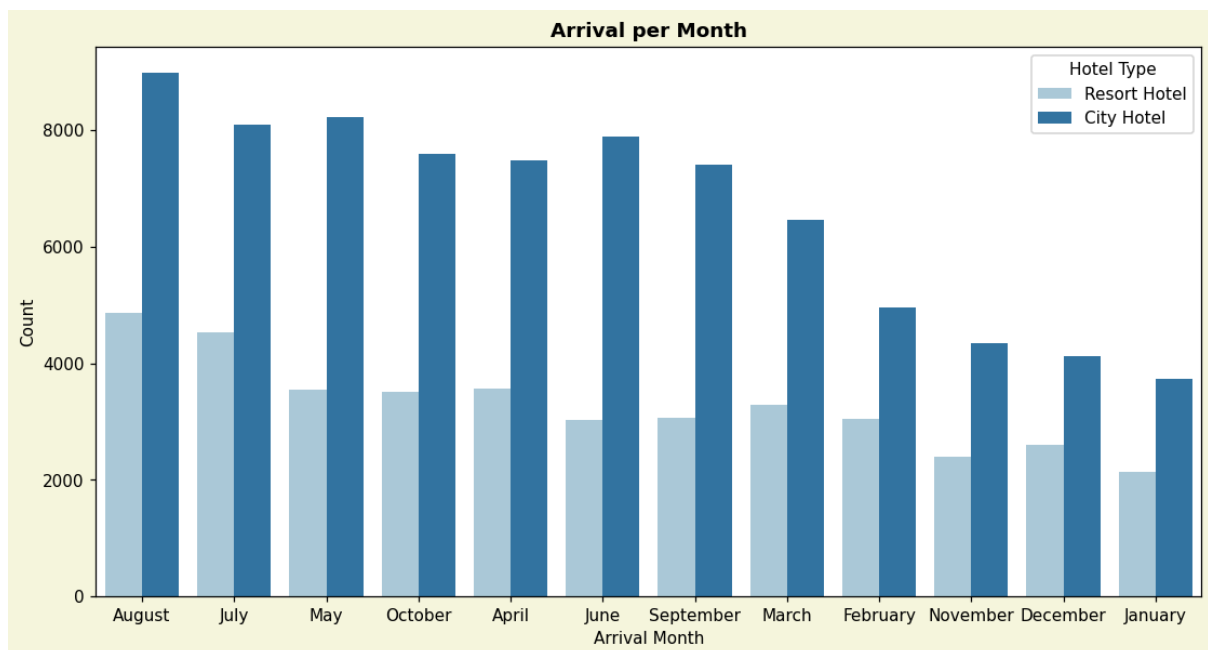
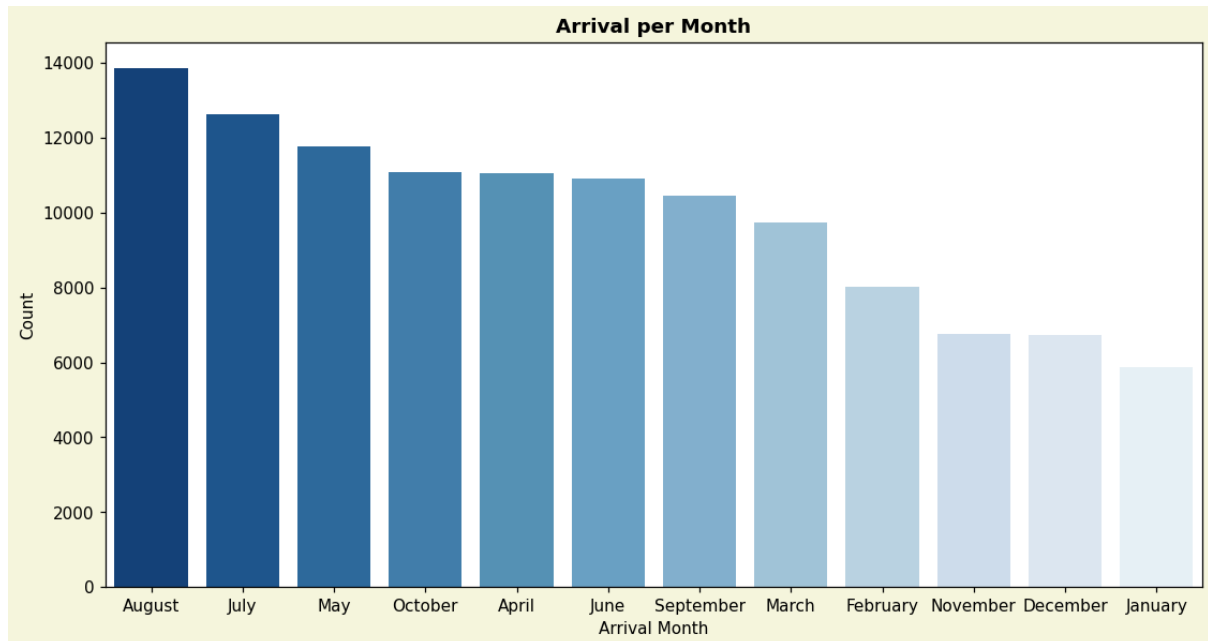


7. Mối tương quan giữa việc hủy đặt phòng và thời gian đặt trước



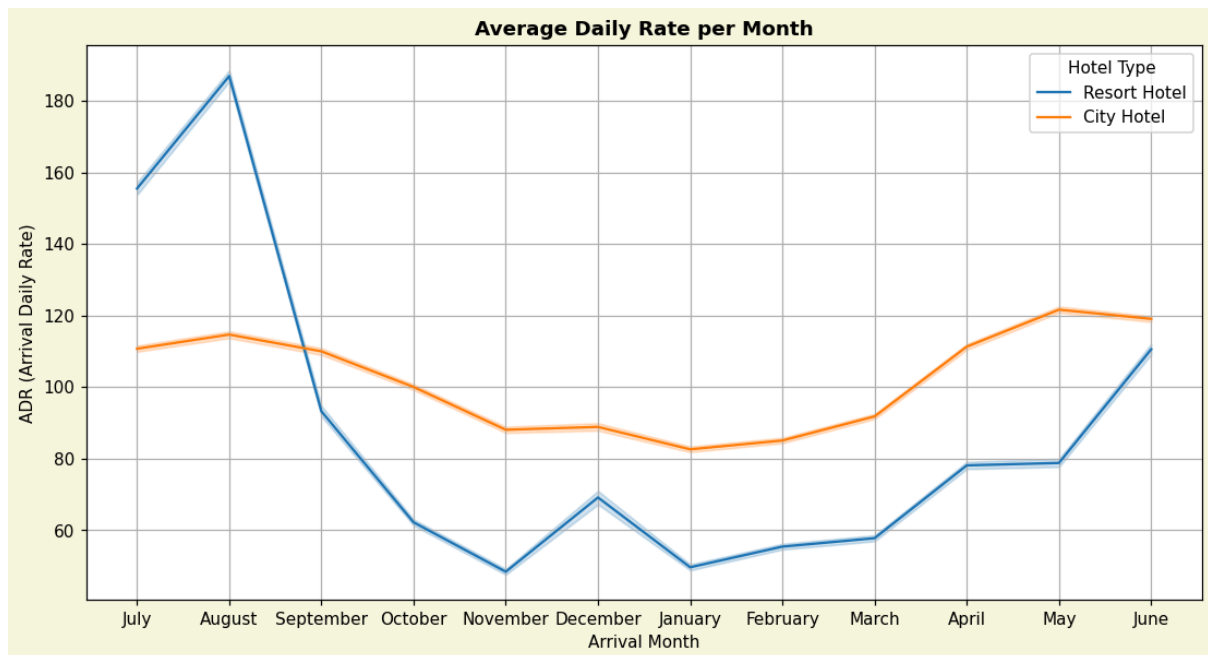
- Thời gian đặt trước càng dài thì tỷ lệ hủy đặt chỗ càng cao
- Theo biểu đồ, có thể thấy rằng thời gian đặt trước có mối tương quan dương với việc hủy đặt chỗ.

8. Xu hướng đặt phòng trong năm



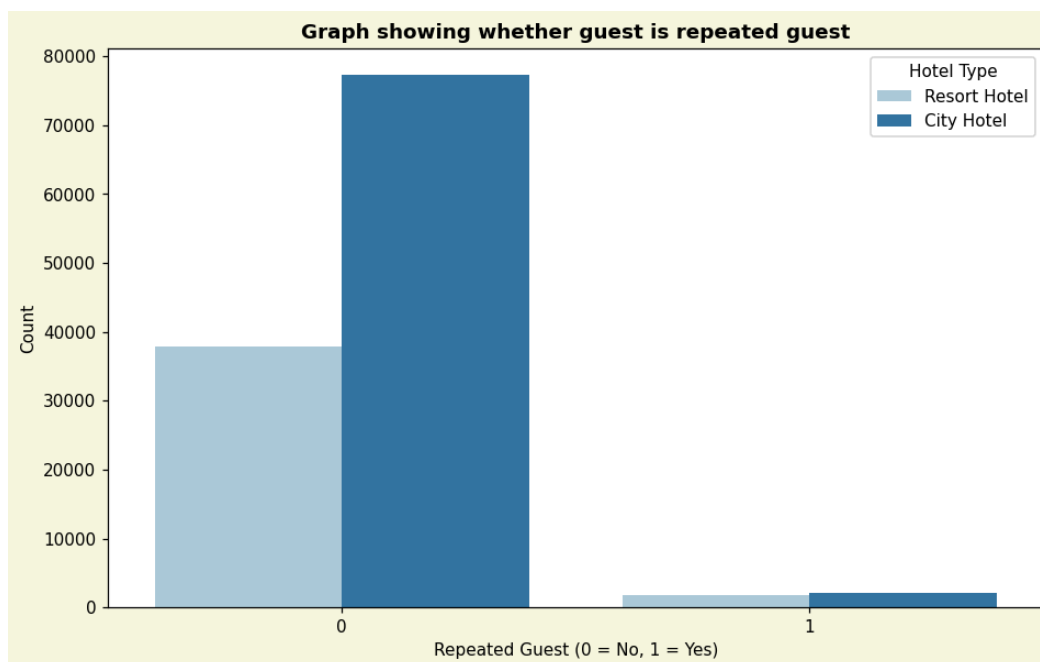
- Số lượng khách hàng cao nhất vào tháng Tám.
- Vì tháng Một có lượng đặt phòng thấp nhất, nên đây có thể là thời điểm tốt nhất trong năm để đặt phòng khách sạn. Do nhu cầu về phòng thấp, chi phí cho phòng hàng ngày cũng thấp hơn so với các tháng khác, trong khi tháng Tám có nhu cầu về phòng cao nên rõ ràng chi phí cho phòng cũng ở mức cao nhất.

9. Biến động ADR trong năm



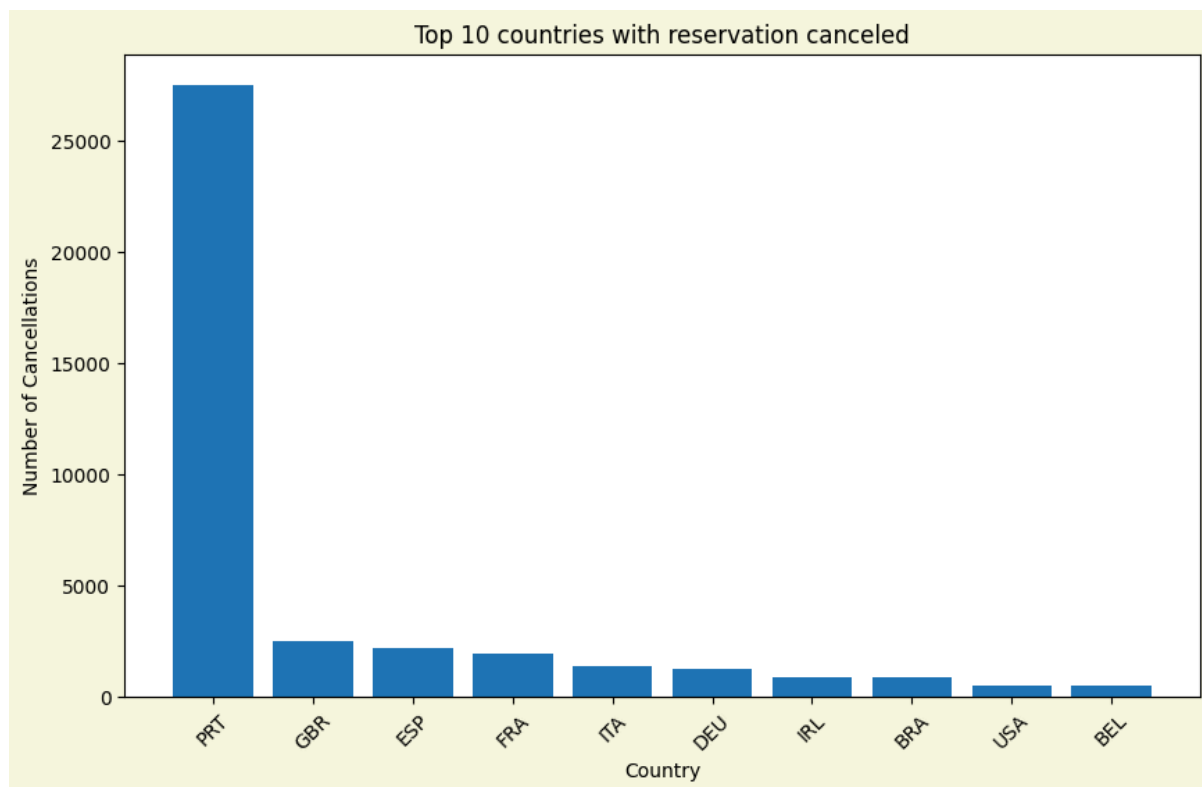
- Giá trung bình hàng ngày (ADR) của khách sạn nghỉ dưỡng cao ở các tháng 8, 7, 6, 9 trong khi có giá rẻ hơn vào tháng 1 và tháng 11.
- Đối với khách sạn thành phố, giá trung bình hàng ngày cao ở tháng 6, tháng 5.
- Giá tại khách sạn nghỉ dưỡng (Resort Hotel) cao hơn trong tháng Tám và Tháng bảy, các tháng còn lại thấp hơn khách sạn thành phố (City Hotel)
- Giá tại khách sạn thành phố (City Hotel) biến động ít hơn so với khách sạn nghỉ dưỡng (Resort Hotel)

10. Khách hàng đặt phòng là khách hàng cũ hay mới?



Có khoảng 3.2% là khách hàng tái đặt phòng và khoảng 96.8% là khách hàng mới

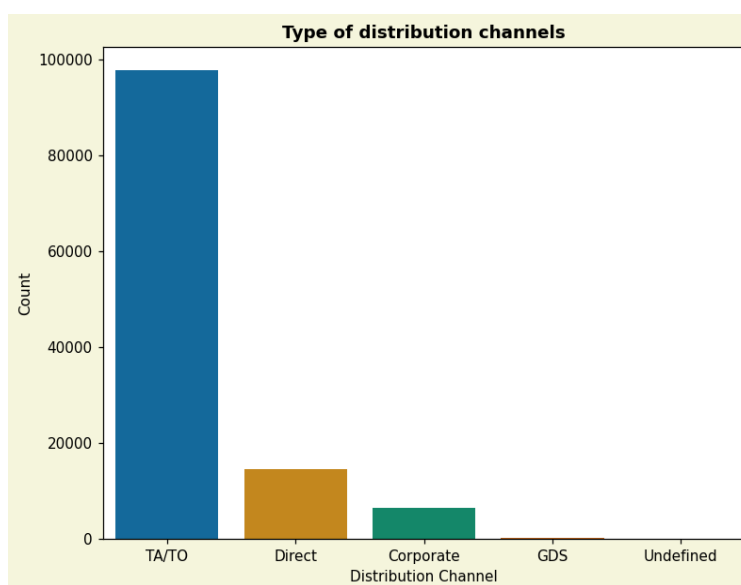
11. Top 10 quốc gia có khách hàng nhiều nhất



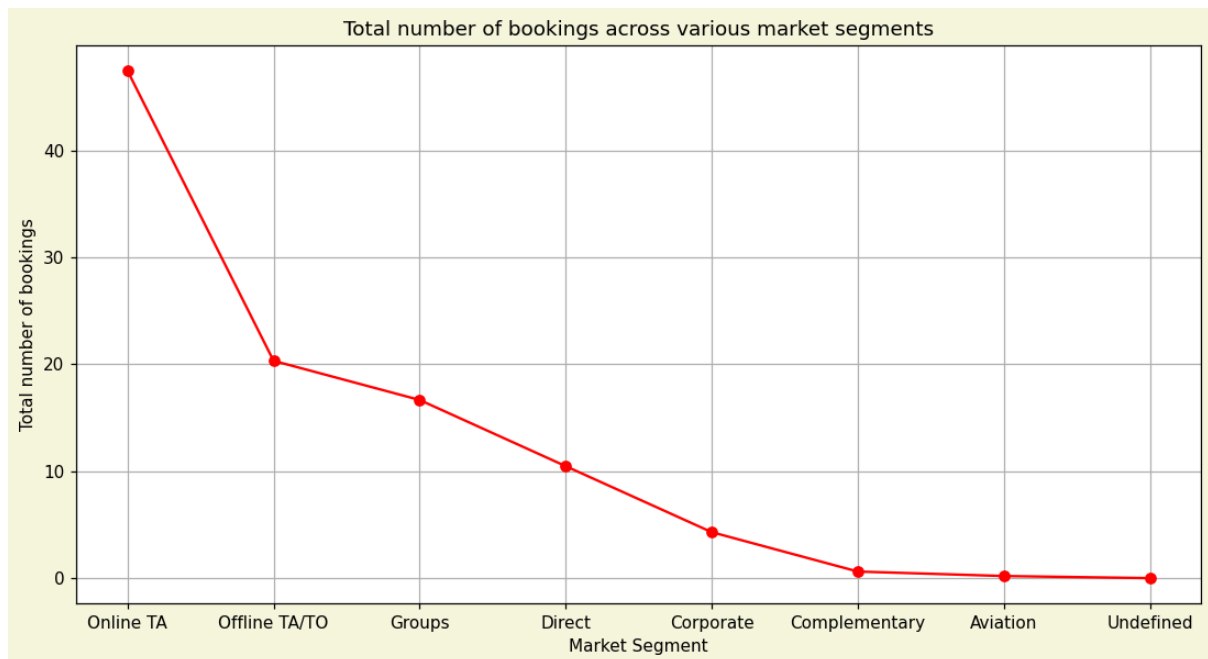
Bồ Đào Nha là quốc gia có số lượng đặt phòng rất cao với khoảng 40.9%, tiếp theo là Vương Quốc Anh và Tây Ban Nha.

12. Kênh phân phối

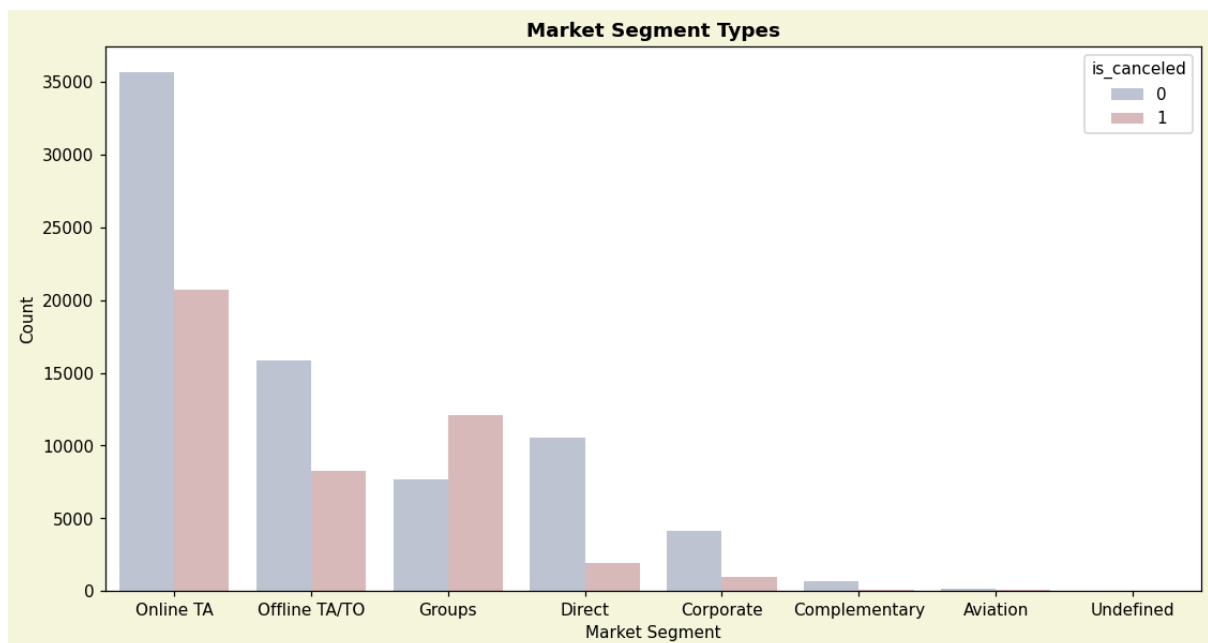
Hầu hết khách hàng đều thích đặt phòng khách sạn thông qua TA/TO (Travel agent / Tour operators) hay (đại lý du lịch / nhà điều hành tour).



13. Tổng số đặt phòng trên các phân khúc thị trường



Phân khúc Online TA (Travel Agency) mang lại số lượng khách hàng cao. Vì vậy, chúng ta có thể tập trung khu vực tiếp thị của mình vào các trang web của các đại lý du lịch này và làm việc với họ vì phần lớn khách truy cập có xu hướng tìm đến họ.



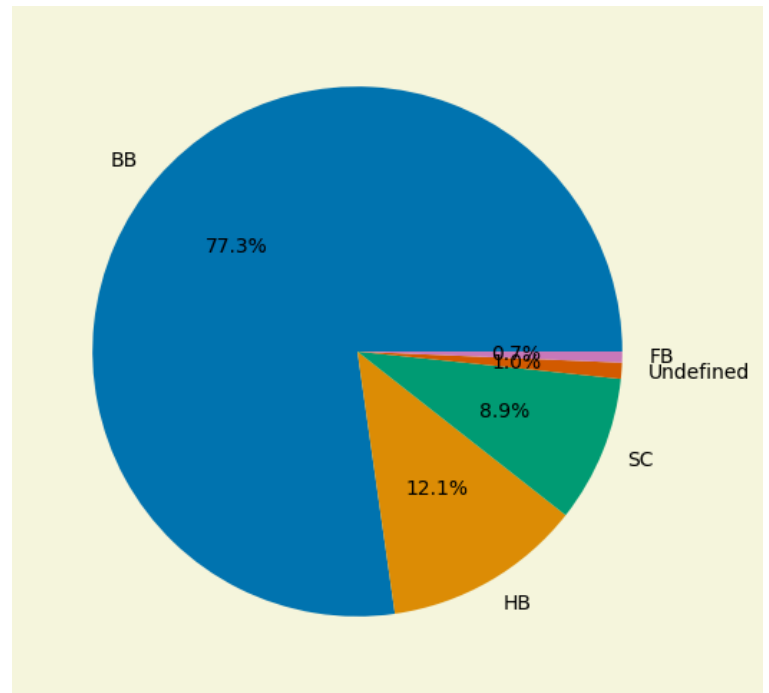
- Phân khúc Nhóm có tỷ lệ hủy đặt phòng trên 50%
- Phân khúc Online TA và Offline TA có tỷ lệ hủy đặt phòng trên 33%
- Tỷ lệ hủy đặt phòng ở phân khúc trực tiếp (Direct Segment) dưới 20%

14. Loại đồ ăn

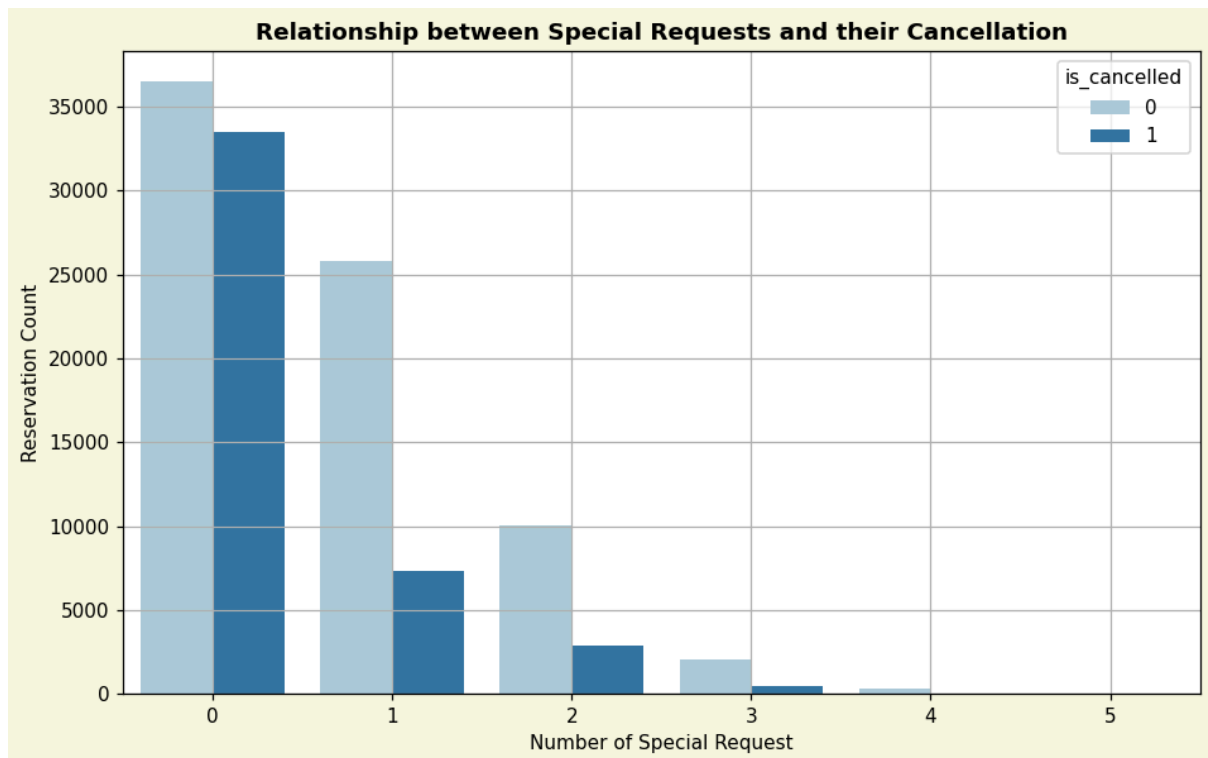
Loại đồ ăn được đặt:

- BB: Chỗ ở và bữa sáng
- Undefined/SC : không có gói bữa ăn
- HB :Nửa ngày (bữa sáng và một bữa khác thường là bữa tối)
- FB : Đầy đủ bữa ăn (bữa sáng, bữa trưa và bữa tối)

Loại bữa ăn BB (Chỗ ở và bữa sáng) là loại được ưa chuộng nhất trong tất cả các loại khách hàng.

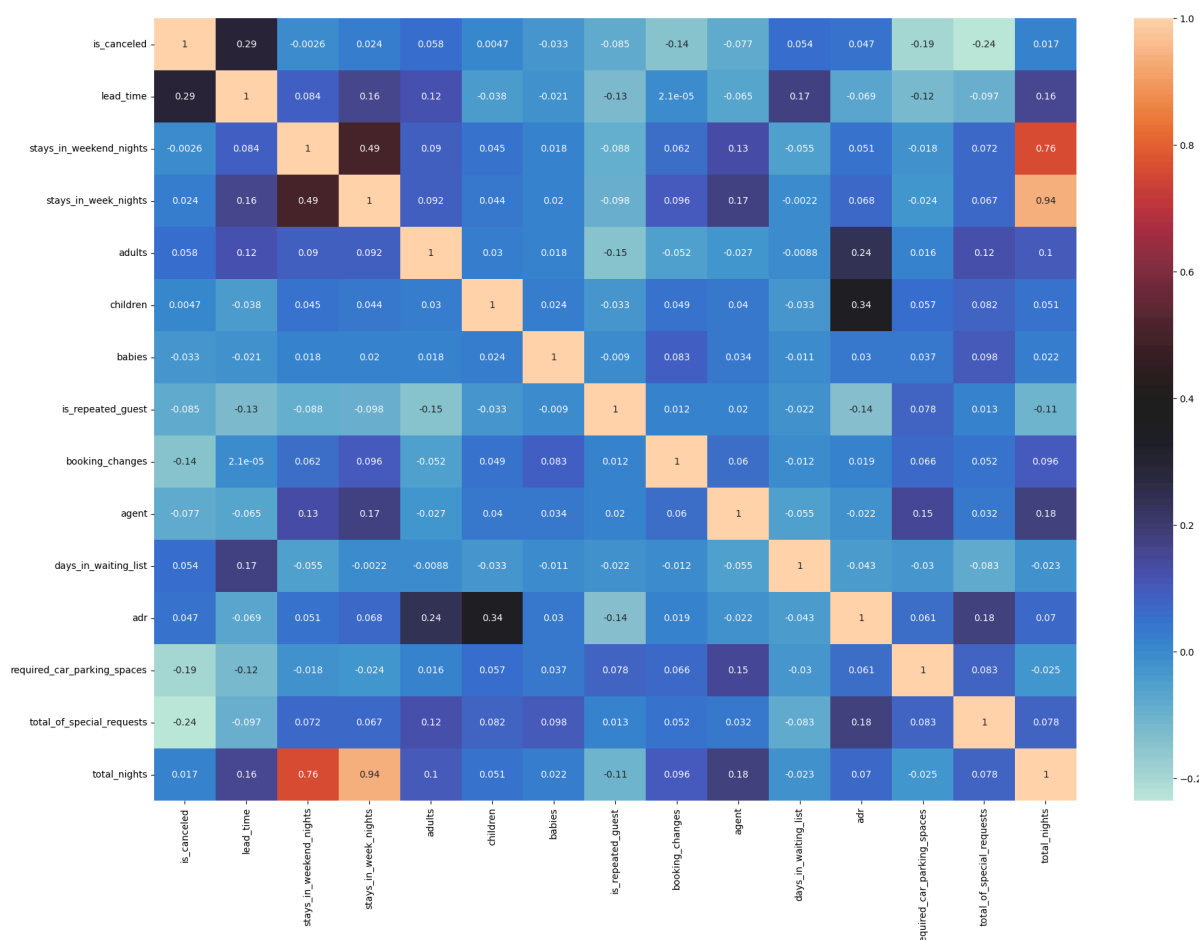


15. Yêu cầu đặc biệt



Biểu đồ trên mô tả mối quan hệ giữa tổng số yêu cầu đặc biệt và tình trạng hủy đặt phòng. Khoảng 28% số đặt phòng bị hủy mà không có bất kỳ yêu cầu đặc biệt nào từ khách, tiếp theo là 6% số đặt phòng bị hủy có một yêu cầu đặc biệt từ khách.

Bảng tương quan giữa các biến



Ta có thể đưa ra một vài nhận xét đặc trưng sau:

1. **is_canceled** và **lead_time**: Có một mối tương quan dương (+0.29) giữa việc hủy đặt phòng và thời gian đặt trước, điều này cho thấy khi thời gian đặt trước càng dài, khả năng hủy đặt phòng càng cao.
2. **is_canceled** và **required_car_parking_spaces**: Có một mối tương quan âm (-0.19) giữa việc hủy đặt phòng và số lượng chỗ đậu xe yêu cầu, cho thấy rằng khách hàng yêu cầu chỗ đậu xe ít có khả năng hủy đặt phòng.
3. **is_canceled** và **total_of_special_requests**: Có một mối tương quan âm (-0.24) giữa việc hủy đặt phòng và tổng số yêu cầu đặc biệt, cho thấy rằng khách hàng có nhiều yêu cầu đặc biệt ít có khả năng hủy đặt phòng.
4. **stays_in_week_nights** và **stays_in_weekend_nights**: Có một mối tương quan rất cao (+0.49) giữa số đêm ở lại vào các ngày trong tuần và các ngày cuối tuần, cho thấy khách hàng thường có xu hướng lưu trú lâu hơn, không phân biệt là ngày trong tuần hay cuối tuần.
5. **adr** và **total_of_special_requests**: Có một mối tương quan dương (+0.18) giữa giá trung bình hàng ngày (ADR) và tổng số yêu cầu đặc biệt, cho thấy rằng khách hàng có nhiều yêu cầu đặc biệt có xu hướng chi tiêu nhiều hơn.
6. **total_nights** và **stays_in_week_nights**: Có một mối tương quan rất cao (+0.94) giữa tổng số đêm lưu trú và số đêm lưu trú vào các ngày trong tuần, cho thấy phần lớn các lưu trú dài ngày là trong tuần.

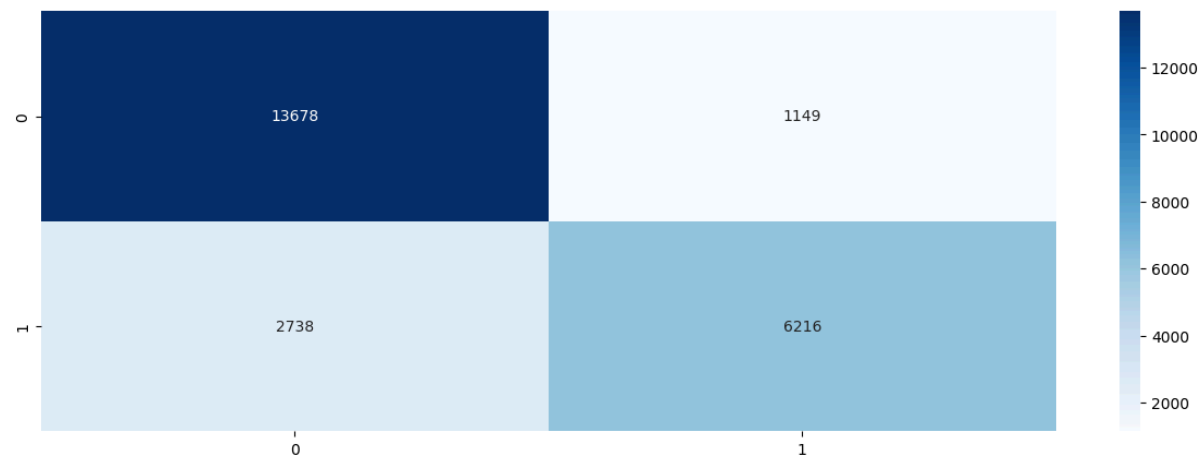
Model Prediction

Tôi xây dựng 3 mô hình dự đoán là Random Forest Classifier, KNN Classifier và Logistic Regression

Random Forest Classifier

	precision	recall	f1-score	support
0	0.83	0.92	0.88	14827
1	0.84	0.69	0.76	8954
accuracy			0.84	23781
macro avg	0.84	0.81	0.82	23781
weighted avg	0.84	0.84	0.83	23781

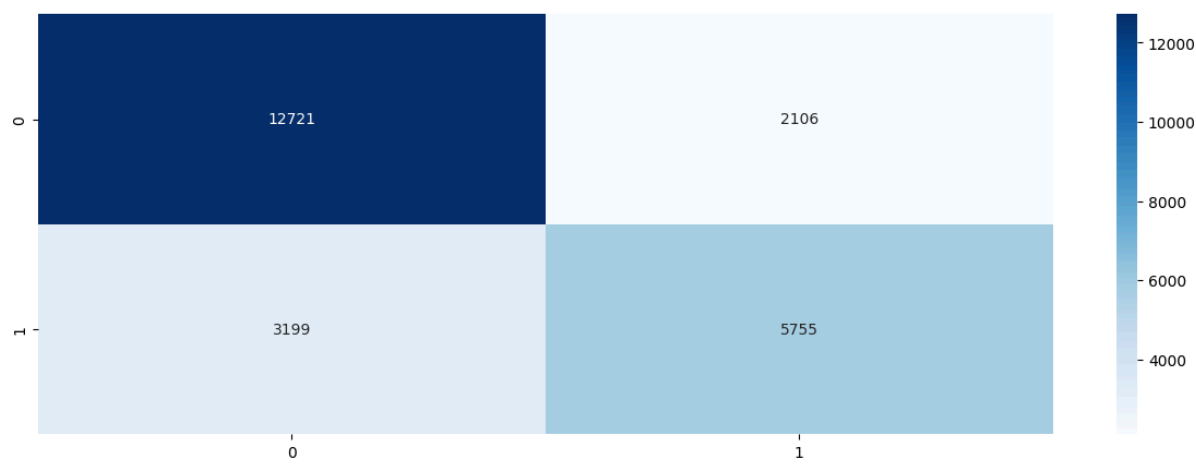
Mô hình có độ chính xác 84%



KNN Classifier

	precision	recall	f1-score	support
0	0.80	0.86	0.83	14827
1	0.73	0.64	0.68	8954
accuracy			0.78	23781
macro avg	0.77	0.75	0.76	23781
weighted avg	0.77	0.78	0.77	23781

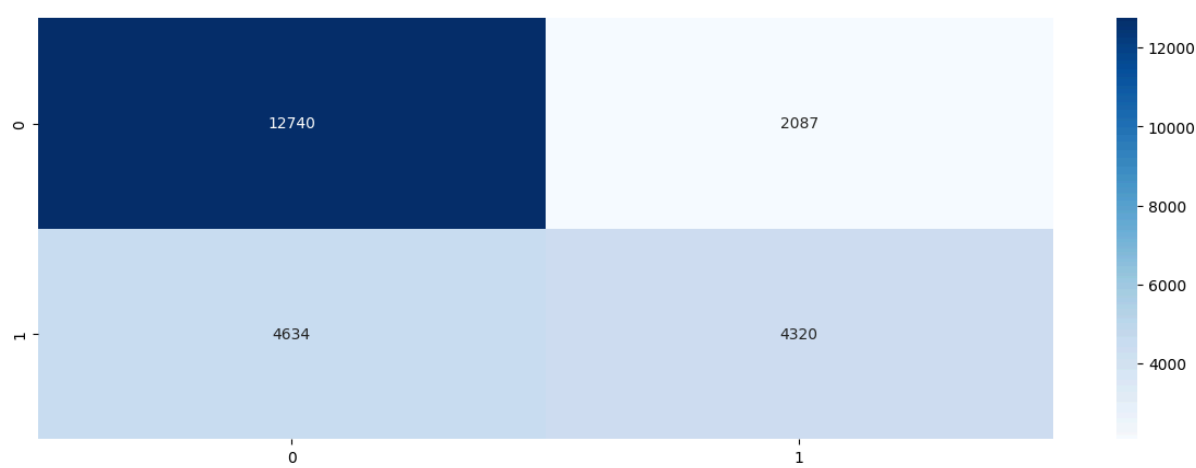
Mô hình có độ chính xác 78%



Logistic Regression

	precision	recall	f1-score	support
0	0.73	0.86	0.79	14827
1	0.67	0.48	0.56	8954
accuracy			0.72	23781
macro avg	0.70	0.67	0.68	23781
weighted avg	0.71	0.72	0.71	23781

Mô hình có độ chính xác 72%



Conclusion

1. Trong tổng số 119.000 khách sạn, 75.000 khách sạn đã làm thủ tục nhận phòng tại khách sạn trong khi 44.000 khách sạn đã hủy đặt phòng. Nếu xét về tỷ lệ phần trăm, 37% các đặt phòng đã bị hủy trong khi 63% khách sạn đã làm thủ tục nhận phòng. Vì vậy, chúng ta nhận thấy rằng tỷ lệ hủy cao có thể là do chính sách **không đặt cọc**.
2. Ta có thể thấy rằng năm 2016 dường như là năm có số lượng đặt phòng khách sạn cao nhất. Chúng ta cũng thấy xu hướng tăng trong việc đặt phòng vào giữa năm, với tháng Tám là cao nhất, tiếp theo là tháng Bảy và tháng Năm. Do đó **mùa hè** là giai đoạn **cao điểm** cho việc đặt phòng khách sạn. Tháng Một có số lượng khách sạn thấp nhất, tiếp theo là tháng Mười Một và tháng Mười Hai. Tương tự, **mùa đông** là giai đoạn **thấp điểm nhất** cho việc đặt phòng khách sạn.

Vì vậy, trong mùa đông, giá đặt phòng khách sạn nên được giảm hoặc cung cấp một số ưu đãi đặc biệt để thu hút nhiều khách sạn hơn và cũng có thể là thời điểm tốt nhất trong năm để đặt phòng khách sạn. Do nhu cầu phòng ít hơn, chi phí cho phòng theo ngày cũng thấp hơn so với các tháng khác, trong khi tháng Tám có nhu cầu phòng cao nên dễ hiểu rằng chi phí phòng cũng ở mức cao nhất.

3. Các khách sạn trong thành phố luôn có nhu cầu cao hơn so với các khu nghỉ dưỡng. Điều này có nghĩa là phần lớn các khách sạn được đặt là khách sạn trong thành phố. Vì vậy, chắc chắn cần phải dành **phần lớn ngân sách tiếp thị** để nhắm mục tiêu vào các khách sạn này.
4. Phần lớn khách sạn đến từ Bồ Đào Nha. Dễ thấy, các đặt phòng chủ yếu đến từ các nước châu Âu, trong đó Bồ Đào Nha là cao nhất với 48,59 nghìn lượt đặt phòng.

Do đó, chúng ta nên nhắm mục tiêu khu vực này để thu hút thêm khách sạn. Chúng ta nên dành một phần ngân sách đáng kể cho khu vực này.

5. Phần lớn khách sạn thích lưu trú một, hai, ba và bốn đêm, trong đó số lượng đặt phòng cho hai và ba đêm có nhu cầu cao.

Các khách sạn nên đưa ra các ưu đãi cho lưu trú trung hạn vì các ưu đãi lưu trú dài hạn sẽ tiêu thụ phòng của họ, và khách sạn mới sẽ không thể đặt phòng, dẫn đến việc khách sạn có thể không tạo ra thêm lợi nhuận.

6. Phân khúc Đại lý Du lịch Trực tuyến (Online TA) mang lại lượng khách sạn lớn, sau đó lần lượt là Đại lý Du lịch/Đơn vị Tổ chức Ngoại tuyến (Offline TA/TO), nhóm khách, đặt phòng trực tiếp, v.v.

Vì vậy, từ đây chúng ta có thể kết luận rằng nên tập trung khu vực tiếp thị của chúng ta vào các trang web của các đại lý du lịch này và làm việc với họ vì phần lớn khách du lịch có xu hướng tìm đến các đại lý này.

7. Phần lớn các đặt phòng được thực hiện với gói giường và bữa sáng (bed and breakfast). Vì vậy, loại hình bữa ăn BB là lựa chọn ưa thích nhất của tất cả các loại khách sạn.

8. Phần lớn các đặt phòng là khách vắng lai. Điều này có nghĩa là đặt phòng không thuộc về một nhóm hoặc hợp đồng nào. Với sự tiện lợi của việc đặt phòng trực tiếp từ trang web, hầu hết mọi người có xu hướng bỏ qua các trung gian để đảm bảo phản hồi nhanh chóng từ việc đặt phòng của họ.
9. Đối với các khu nghỉ dưỡng, Giá Trung Bình Hàng Ngày (ADR) đắt hơn vào các tháng Tám, Bảy, Sáu và Chín. Đối với các khách sạn trong thành phố, Giá Trung Bình Hàng Ngày (ADR) đắt hơn vào các tháng Tám, Bảy, Năm và Sáu.

Vì vậy, nhìn chung, Giá Trung Bình Hàng Ngày của cả khách sạn trong thành phố và khu nghỉ dưỡng đều đắt hơn trong khoảng từ tháng Năm đến tháng Chín.

10. Sử dụng mô hình dự đoán với độ chính xác cao nhất để dự đoán khách hàng có hủy đặt phòng khách sạn hay không.