

ĐẠI HỌC KINH TẾ THÀNH PHỐ HỒ CHÍ MINH



Trường Công nghệ và Thiết kế
KHOA CÔNG NGHỆ THÔNG TIN KINH DOANH

BÁO CÁO CUỐI KỲ

TRỰC QUAN BỘ DỮ LIỆU

515K HOTEL REVIEWS IN EUROPE

Môn học: Biểu diễn trực quan dữ liệu

Giảng viên hướng dẫn: TS. Nguyễn An Tế

Mã lớp học phần: 24C1INF50908201

Nhóm sinh viên thực hiện:

- Nguyễn Phúc Minh Trâm - 31221024796
- Trần Vọng Triển - 31221021725
- Trầm Thái Tú - 31221022394
- Nguyễn Thành Vinh - 31221025662
- Nguyễn Văn Phi Yến - 31221021785

Hồ Chí Minh, ngày 08 tháng 12 năm 2024

LỜI CẢM ƠN

Để có thể hoàn thiện bài luận báo cáo cuối kỳ môn học *Biểu diễn trực quan dữ liệu* với đề tài “*Phân tích bộ dữ liệu 515K Hotel Reviews Data in Europe*”, nhóm chúng em xin gửi lời tri ân sâu sắc đến *Thầy TS. Nguyễn An Té*, giảng viên hướng dẫn trực thuộc khoa Công nghệ Thông tin Kinh doanh của Đại học Kinh tế Thành phố Hồ Chí Minh.

Nhóm chúng em chân thành cảm ơn Thầy vì sự tận tâm trong giảng dạy và những kiến thức Thầy đã truyền đạt trong suốt quá trình học tập môn học này. Thầy đã giúp chúng em nắm vững các kiến thức nền tảng về việc trực quan hóa dữ liệu, đồng thời khuyến khích chúng em nghiên cứu sâu hơn để ứng dụng các phương pháp trực quan hóa vào thực tế.

Trong quá trình thực hiện bài báo cáo, nhóm chúng em đã cố gắng vận dụng những kiến thức đã được học và nghiên cứu thêm nhiều tài liệu để hoàn thiện bài tiểu luận này. Tuy nhiên, do kiến thức còn hạn chế và thiếu kinh nghiệm thực tiễn nên nội dung bài tiểu luận khó tránh khỏi những thiếu sót. Chúng em rất mong nhận được những ý kiến đóng góp quý báu từ Thầy để nhóm có thể hoàn thiện bài làm hơn.

Nhóm chúng em xin chân thành cảm ơn Thầy vì sự đồng hành và hỗ trợ trong suốt quá trình học tập và nghiên cứu.

Trân trọng,

Nhóm thực hiện

PHÂN CÔNG CÔNG VIỆC

Nhiệm vụ	Thành viên	Mức độ hoàn thành
<ul style="list-style-type: none"> - Phụ trách nội dung: Chương 1, 5 - Tìm insight và biểu diễn trực quan dữ liệu - Phụ trách làm slide thuyết trình 	Nguyễn Phúc Minh Trâm	100%
<ul style="list-style-type: none"> - Phụ trách nội dung: Chương 2, 5 - Tìm insight và biểu diễn trực quan dữ liệu 	Trầm Thái Tú	100%
<ul style="list-style-type: none"> - Phụ trách nội dung: Chương 3, 5 - Tìm insight và biểu diễn trực quan dữ liệu 	Trần Vọng Triển	100%
<ul style="list-style-type: none"> - Phụ trách nội dung: Chương 4, 5 - Xử lý ngôn ngữ tự nhiên các đánh giá - Tìm insight và biểu diễn trực quan dữ liệu 	Nguyễn Thành Vinh	100%
<ul style="list-style-type: none"> - Phụ trách nội dung: Chương 3, 5 - Tìm insight và biểu diễn trực quan dữ liệu - Phụ trách làm slide thuyết trình 	Nguyễn Văn Phi Yến <i>(Nhóm trưởng)</i>	100%

MỤC LỤC

LỜI CẢM ƠN.....	1
PHÂN CÔNG CÔNG VIỆC.....	2
MỤC LỤC.....	3
DANH MỤC HÌNH ẢNH.....	5
DANH MỤC BIỂU ĐỒ.....	6
CHƯƠNG 1 - TỔNG QUAN ĐỀ TÀI.....	8
1.1. Giới thiệu đề tài.....	8
1.2. Mục tiêu nghiên cứu.....	8
1.3. Phương pháp nghiên cứu.....	8
1.4. Tài nguyên sử dụng.....	9
CHƯƠNG 2 - TỔNG QUAN BỘ DỮ LIỆU.....	10
2.1. Sơ lược về bộ dữ liệu.....	10
2.2. Mô tả thuộc tính.....	10
CHƯƠNG 3 - TIỀN XỬ LÝ DỮ LIỆU.....	12
3.1. Tổng quan bộ dữ liệu ban đầu.....	12
3.2. Chính dạng dữ liệu.....	12
3.2.1. Thêm thuộc tính.....	13
3.2.2. Xóa các cột không cần thiết.....	15
3.3. Kiểm tra dữ liệu.....	16
3.3.1. Kiểm tra giá trị bị thiếu (missing values) và giá trị bất thường (abnormal values).....	16
3.3.2. Kiểm tra và xử lý giá trị trùng lặp (duplicate values).....	17
3.4. Phân tích thống kê tổng quát.....	18
CHƯƠNG 4 - ĐÁNH GIÁ SỰ TƯƠNG QUAN GIỮA ĐIỂM SỐ VÀ BÌNH LUẬN... 23	23
4.1. Phân phối điểm đánh giá của khách hàng.....	23
4.2. Tương quan giữa điểm số và bình luận.....	24
CHƯƠNG 5 - BIỂU DIỄN TRỰC QUAN DỮ LIỆU..... 27	27

5.1. Tổng quan tình hình lưu trú khách sạn tại châu Âu.....	27
5.1.1. Số lượt đánh giá theo các quốc gia ở châu Âu.....	27
5.1.2. Điểm trung bình của từng quốc gia.....	28
5.1.3. Số khách du lịch đến các quốc gia theo từng tháng năm 2016.....	29
5.1.4. Hình thức và nhóm theo đánh giá khi đến các quốc gia.....	30
5.1.5. Top 3 khách sạn theo đánh giá ở mỗi quốc gia.....	33
5.1.6. Top 3 khách sạn có số lượng đánh giá cao nhất về dịch vụ cho phép mang thú cưng theo của từng quốc gia.....	34
5.2. Hành vi người châu Á khi đến các quốc gia châu Âu.....	35
5.2.1. Lượt đánh giá của khách châu Á theo các quốc gia.....	35
5.2.2. Khoảng thời gian khách châu Á đánh giá khi đến các quốc gia năm 2016.....	36
5.2.3. Phân tích loại hình chuyến đi của du khách châu Á và các khu vực khác..	37
5.2.4. Phân tích thông tin về loại khách (Guests) của du khách châu Á (Asia) và không phải châu Á (Non-Asia).....	38
5.2.5. Histogram điểm đánh giá của người châu Á.....	40
5.2.6. Top 3 khách sạn ở mỗi quốc gia theo đánh giá của người châu Á.....	42
5.2.7. Tương quan giữa điểm trung bình và số đêm ở lại.....	43
5.2.8. Những từ phổ biến nhất trong các đánh giá của khách châu Á cho từng quốc gia.....	45
5.2.9. Những từ phổ biến nhất trong các đánh giá tại 3 khách sạn có nhiều đánh giá nhất theo từng quốc gia của khách châu Á.....	46
CHƯƠNG 6 - TỔNG KẾT ĐỀ TÀI.....	52
TÀI LIỆU THAM KHẢO.....	53

DANH MỤC HÌNH ẢNH

Hình 3.1. Mô tả thông tin cơ bản của bộ dữ liệu nguyên bản.....	12
Hình 3.2. Kết quả phân bố đánh giá của khách du lịch theo quốc gia.....	13
Hình 3.3. Kết quả phân bố số lượng đánh giá theo khu vực Reviewer_Regions.....	14
Hình 3.4. Bộ dữ liệu sau khi tách ‘Tags’ thành các cột ‘Guests’, ‘Trip_Type’, ‘Stayed’, ‘Submit’, ‘With a pet’, ‘Room_Type’	15
Hình 3.5. Quan sát Missing Values.....	17
Hình 3.6. Kết quả kiểm tra giá trị trùng lặp.....	17
Hình 3.7. Mô tả thông tin cơ bản của bộ dữ liệu sau khi tiền xử lý.....	19
Hình 3.8. Thống kê tổng quát với các thuộc tính kiểu numerical.....	19
Hình 3.9. Thống kê tổng quát với các thuộc tính kiểu categorical.....	22
Hình 4.1. Kiểm định Spearman’s rank correlation.....	26
Hình 5.1. Top 3 khách sạn ở mỗi quốc gia.....	34
Hình 5.2. Top 3 khách sạn ở mỗi quốc gia đối với người châu Á.....	42

DANH MỤC BIỂU ĐỒ

Biểu đồ 4.1. Phân phối điểm đánh giá chung.....	23
Biểu đồ 4.2. Phân phối điểm đánh giá của khách Châu Á.....	23
Biểu đồ 4.3. Tương quan giữa số lượng từ tích cực, tiêu cực trong bình luận với điểm đánh giá.....	24
Biểu đồ 5.1. Số lượt đánh giá theo các quốc gia ở Châu Âu.....	27
Biểu đồ 5.2. Biểu đồ thể hiện điểm trung bình của từng quốc gia.....	28
Biểu đồ 5.3. Điểm đánh giá trung bình các khách sạn theo từng quốc gia (khách hàng Châu Á).....	29
Biểu đồ 5.4. Số lượt đánh giá theo tháng của từng quốc gia năm 2016.....	29
Biểu đồ 5.5. Hình thức chuyến đi theo từng quốc gia.....	30
Biểu đồ 5.6. Nhóm theo hình thức chuyến đi của từng quốc gia.....	32
Biểu đồ 5.7. Top 3 khách sạn có số lượng đánh giá cao nhất về dịch vụ cho phép mang thú cưng theo của từng quốc gia.....	34
Biểu đồ 5.8. Số đánh giá khách Châu Á theo từng quốc gia.....	35
Biểu đồ 5.9. Số đánh giá của người Châu Á theo tháng của từng quốc gia năm 2016.....	36
Biểu đồ 5.10. Biểu đồ thể hiện số lượt đánh giá của du khách Châu Á và không phải Châu Á theo loại hình chuyến đi (Trip Type).....	38
Biểu đồ 5.11. Biểu đồ thể hiện số lượt đánh giá của du khách châu Á (Asia) và không phải châu Á (Non-Asia) theo thông tin về loại khách (Guests).....	39
Biểu đồ 5.12. Biểu đồ thể hiện phân phối Điểm đánh giá (Reviewer Score) của Du khách Châu Á (Asia).....	40
Biểu đồ 5.13. Biểu đồ thể hiện phân phối Điểm đánh giá (Reviewer Score) của Du khách từ khu vực khác (Non-Asia).....	41
Biểu đồ 5.14. So sánh điểm đánh giá trung bình dựa trên số đêm khách hàng ở lại khách sạn.....	43
Biểu đồ 5.15. So sánh điểm đánh giá trung bình dựa trên số đêm khách hàng Châu Á ở lại khách sạn theo từng quốc gia.....	44
Biểu đồ 5.16. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại các quốc	

gia.....	45
Biểu đồ 5.17. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Tây Ban Nha.....	46
Biểu đồ 5.18. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Pháp...	47
Biểu đồ 5.19. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Ý.....	48
Biểu đồ 5.20. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Vương Quốc Anh.....	49
Biểu đồ 5.21. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Hà Lan	50
Biểu đồ 5.22. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Áo.....	51

CHƯƠNG 1 - TỔNG QUAN ĐỀ TÀI

1.1. Giới thiệu đề tài

Theo báo cáo từ UN Tourism, tính đến tháng 7 năm 2024, ngành du lịch quốc tế đã phục hồi được 96% so với mức trước đại dịch COVID-19. Cụ thể, theo World Tourism Barometer của UN Tourism, khoảng 790 triệu du khách đã đi du lịch quốc tế trong bảy tháng đầu năm 2024, tăng khoảng 11% so với năm 2023, và chỉ giảm 4% so với năm 2019. Đặc biệt, dữ liệu cho thấy khu vực châu Âu và châu Mỹ đã phục hồi lần lượt 99% và 97% lượng khách đến trước đại dịch trong bảy tháng vừa qua. Theo Tổ chức Du lịch Thế giới (UNWTO), năm 2023, châu Âu chiếm hơn 50% tổng số lượt khách du lịch toàn cầu, cùng với hàng triệu lượt đặt phòng khách sạn mỗi năm.

Cùng với đó, các nền tảng đặt phòng trực tuyến như Booking.com, TripAdvisor hay Agoda dần trở nên phổ biến và trở thành công cụ chính để khách hàng lựa chọn tham khảo và đánh giá các dịch vụ đặt phòng. Các đánh giá từ khách hàng không chỉ giúp khách du lịch tham khảo để lựa chọn dịch vụ, mà còn cung cấp những thông tin quý giá giúp các khách sạn cải thiện chất lượng dịch vụ. Tuy nhiên, với bộ dữ liệu hàng trăm nghìn đánh giá được ghi nhận, việc phân tích dữ liệu này theo phương pháp truyền thống sẽ gặp nhiều khó khăn. Do đó, nhóm đã sử dụng biểu diễn trực quan nhằm chuyển đổi bộ dữ liệu lớn này thành các biểu đồ thể hiện nhiều thông tin trực quan và dễ hiểu hơn.

Đề tài “*Biểu diễn trực quan bộ dữ liệu 515K Hotel Reviews Data in Europe*” khai thác bộ dữ liệu bao gồm 515.738 đánh giá từ khách hàng đã trải nghiệm dịch vụ tại 1.493 khách sạn cao cấp tại châu Âu, được lấy từ nền tảng Booking.com.

1.2. Mục tiêu nghiên cứu

Mục tiêu tổng quát của đề tài nhằm phân tích và biểu diễn trực quan dữ liệu đánh giá các khách sạn cao cấp tại khu vực châu Âu dựa trên bộ dữ liệu ‘515K Hotel Reviews Data in Europe’ từ Kaggle. Thông qua đó, đưa đến cho người đọc một cái nhìn tổng quan, dễ hiểu và trực quan hóa các thông tin hữu ích nhằm hỗ trợ du khách khi muốn đến thăm châu Âu, đặc biệt là khách du lịch từ châu Á, trong việc chọn lựa khách sạn phù hợp.

1.3. Phương pháp nghiên cứu

Các phương pháp nghiên cứu được sử dụng trong bài báo cáo bao gồm:

- Tiền xử lý dữ liệu
- Phân tích tổng quan dữ liệu

- Exploratory Data Analysis
- Biểu diễn trực quan dữ liệu
- Đánh giá kết quả

1.4. Tài nguyên sử dụng

Đề tài sử dụng các tài nguyên và công cụ như sau:

- *Dữ liệu*: Bộ dữ liệu *515K Hotel Reviews Data in Europe* từ Kaggle, bao gồm 515.738 đánh giá từ khách hàng đã trải nghiệm dịch vụ tại 1.493 khách sạn cao cấp tại châu Âu, được lấy từ nền tảng Booking.com, với các thông tin như điểm số, bình luận và vị trí địa lý của khách sạn.
- *Ngôn ngữ và thư viện lập trình*:
 - Ngôn ngữ lập trình Python: Dùng để xử lý dữ liệu, phân tích và biểu diễn trực quan
 - Các thư viện hỗ trợ:
 - pandas
 - numpy
 - matplotlib
 - seaborn
 - nltk

CHƯƠNG 2 - TỔNG QUAN VỀ BỘ DỮ LIỆU

2.1. Sơ lược về bộ dữ liệu

Bộ dữ liệu “515K Hotel Reviews Data in Europe” được đăng tải trên Kaggle bởi tác giả Jiashen Liu. Bộ dữ liệu được thu thập từ các đánh giá công khai trên nền tảng Booking.com và đã qua xử lý văn bản để loại bỏ các ký tự không hợp lệ, chuyển đổi toàn bộ văn bản sang dạng chữ thường.

Bộ dữ liệu có tổng cộng 515,738 quan sát, trong đó mỗi quan sát tương ứng với một đánh giá của khách hàng đã đến và trải nghiệm dịch vụ tại khách sạn cao cấp trong khu vực châu Âu. Bên cạnh đó, bộ dữ liệu còn cung cấp các thông tin về điểm đánh giá trung bình và vị trí địa lý của khách sạn, quốc tịch của người đánh giá, và nhiều thuộc tính khác liên quan đến trải nghiệm của khách hàng.

2.2. Mô tả thuộc tính

Dữ liệu bao gồm 515,738 dòng, với 17 thuộc tính được mô tả chi tiết như sau:

- ***Hotel_Address* (object)**: Địa chỉ của khách sạn
- ***Review_Date* (object)**: Ngày khách hàng đăng đánh giá
- ***Average_Score* (float 64)**: Điểm trung bình của khách sạn, được tính dựa trên các đánh giá gần nhất trong năm qua
- ***Hotel_Name* (object)**: Tên của khách sạn
- ***Reviewer_Nationality* (object)**: Quốc tịch của người đánh giá
- ***Negative_Review* (object)**: Nội dung đánh giá tiêu cực mà khách hàng dành cho khách sạn. Nếu không có đánh giá tiêu cực, sẽ được gán giá trị “No Negative”
- ***Review_Total_Negative_Word_Counts* (int64)**: Tổng số từ trong đánh giá tiêu cực
- ***Positive_Review* (object)**: Nội dung đánh giá tích cực mà khách hàng dành cho khách sạn. Nếu không có đánh giá tích cực, sẽ được gán giá trị “No Positive”
- ***Review_Total_Positive_Word_Counts* (int64)**: Tổng số từ trong đánh giá tích cực
- ***Reviewer_Score* (float64)**: Điểm số khách hàng chấm cho khách sạn dựa trên trải nghiệm thực tế
- ***Total_Number_of_Reviews_Reviewer_Has_Given* (int64)**: Tổng số lượng đánh giá mà khách hàng đã thực hiện trước đây
- ***Total_Number_of_Reviews* (int64)**: Tổng số lượng đánh giá hợp lệ của khách sạn

- **Tags (object)**: Các thẻ mô tả mà người dùng gắn cho khách sạn, tương ứng với trải nghiệm của họ (Ví dụ: “Leisure Trip”, “Couple”, “Stayed 2 nights”)
- **days_since_review (int64)**: Khoảng thời gian (tính bằng ngày) giữa ngày đăng đánh giá và ngày thu thập dữ liệu
- **Additional_Number_of_Scoring (int64)**: Số lượng đánh giá của khách hàng chỉ cho điểm, mà không có nội dung văn bản.
- **lat (float64)**: Vĩ độ của khách sạn
- **lng (float64)**: Kinh độ của khách sạn

CHƯƠNG 3 - TIỀN XỬ LÝ DỮ LIỆU

3.1. Tổng quan bộ dữ liệu ban đầu

Thể hiện các thông tin cơ bản của bộ dữ liệu như số lượng quan sát (số dòng), số lượng biến (số cột), và các mô tả chi tiết về các biến của bộ dữ liệu.

Kết quả kiểm tra thông tin cơ bản:

- *Tổng số dòng:* 515,738 (số lượng quan sát)
- *Tổng số cột:* 17 (biến thuộc tính)
- *Không chứa giá trị null*

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 515738 entries, 0 to 515737
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Hotel_Address    515738 non-null   object  
 1   Additional_Number_of_Scoring  515738 non-null   int64  
 2   Review_Date       515738 non-null   object  
 3   Average_Score     515738 non-null   float64 
 4   Hotel_Name        515738 non-null   object  
 5   Reviewer_Nationality  515738 non-null   object  
 6   Negative_Review   515738 non-null   object  
 7   Review_Total_Negative_Word_Counts  515738 non-null   int64  
 8   Total_Number_of_Reviews  515738 non-null   int64  
 9   Positive_Review    515738 non-null   object  
 10  Review_Total_Positive_Word_Counts  515738 non-null   int64  
 11  Total_Number_of_Reviews_Reviewer_Has_Given  515738 non-null   int64  
 12  Reviewer_Score      515738 non-null   float64 
 13  Tags               515738 non-null   object  
 14  days_since_review  515738 non-null   object  
 15  lat                512470 non-null   float64 
 16  lng                512470 non-null   float64 

dtypes: float64(4), int64(5), object(8)
memory usage: 66.9+ MB
```

Hình 3.1. Mô tả thông tin cơ bản của bộ dữ liệu nguyên bản

3.2. Chính dạng dữ liệu

Bối cảnh: Du lịch châu Âu đang là xu hướng phổ biến đối với du khách từ các nước châu Á. Tuy nhiên, thông tin chi tiết về chất lượng khách sạn tại mỗi quốc gia, trải nghiệm thực tế từ các khách hàng trước vẫn còn hạn chế. Việc phân tích dữ liệu này sẽ giúp cung cấp cái nhìn trực quan và toàn diện về tình hình du lịch tại châu Âu. Thông qua đó, hỗ trợ du khách châu Á trong việc lựa chọn điểm đến và khách sạn phù hợp với nhu cầu.

Mục đích: Cung cấp thông tin tổng quát hỗ trợ du khách Châu Á chọn lựa các khách sạn phù hợp dựa trên đánh giá và trải nghiệm thực tế của khách hàng.

3.2.1. Thêm thuộc tính

a) Cột 'Country'

Thêm cột 'Country' vào bộ dữ liệu nhằm phân tích dữ liệu theo từng quốc gia mà các khách sạn tọa lạc dễ dàng hơn. Thuộc tính 'Country' được trích xuất bằng cách tách quốc gia từ địa chỉ khách sạn (*Hotel_Address*), giúp dễ dàng nhóm dữ liệu và thực hiện phân tích theo khu vực địa lý.

```
# Tạo cột Country
def extract_country(address):
    words = address.split()
    two_word_countries = ['United Kingdom']
    if len(words) >= 2 and ' '.join(words[-2:]) in two_word_countries:
        return ' '.join(words[-2:])
    return words[-1]

df_hotel['Country'] = df_hotel['Hotel_Address'].apply(extract_country)
```

Sau khi tách, nhóm đã thu được kết quả phân bố số lượng đánh giá của khách du lịch đối với các khách sạn tại châu Âu, tập trung chủ yếu vào các khách sạn ở 6 quốc gia, cụ thể như sau:

Country	count
United Kingdom	262301
Spain	60149
France	59928
Netherlands	57214
Austria	38939
Italy	37207
Name: count, dtype: int64	

Hình 3.2. Kết quả phân bố đánh giá của khách du lịch theo quốc gia

b) Cột 'Reviewer_Regions'

Để hỗ trợ cho việc phân tích sự khác biệt về xu hướng và đánh giá của khách du lịch từ các khu vực khác nhau, đặc biệt tập trung vào du khách từ châu Á, nhóm đã tiến hành tạo thêm thuộc tính mới 'Reviewer_Regions' nhằm phân nhóm quốc tịch của người đánh giá 'Reviewer_Nationality' thành hai khu vực: 'Asia' và 'Non-Asia'.

```
df_copy=df_hotel.copy()
df_copy['Reviewer_Nationality']=df_copy['Reviewer_Nationality'].str.strip()

# Chuẩn hóa tên quốc gia
asia_list = [x.replace('China', 'People\'s Republic of China') for x in asia_list]

# Thêm cột Reviewer_Regions với hai giá trị 'Asia' và 'Non-Asia'
df_copy['Reviewer_Regions'] = df_copy['Reviewer_Nationality'].apply(lambda x: 'Asia' if x in asia_list else 'Non-Asia')
```

Reviewer_Regions	
Non-Asia	483184
Asia	32554
Name: count, dtype: int64	

Hình 3.3. Kết quả phân bố số lượng đánh giá theo khu vực Reviewer_Regions

- c) Tách cột ‘Tags’ thành các cột ‘Guests’, ‘Trip_Type’, ‘Stayed’, ‘Submit’, ‘With a pet’, ‘Room_Type’

Cột ‘Tags’ chứa thông tin văn bản chi tiết về các đặc điểm của chuyến đi, được người đánh giá cung cấp nhằm bổ sung tính chính xác và độ tin cậy cho đánh giá của họ.

	Tags
0	['Leisure trip', 'Couple', 'Duplex Double Room', 'Stayed 6 nights']
1	['Leisure trip', 'Couple', 'Duplex Double Room', 'Stayed 4 nights']
2	['Leisure trip', 'Family with young children', 'Duplex Double Room', 'Stayed 3 nights', 'Submitted from a mobile device']
...	

Kiểu dữ liệu của biến ‘Tags’ khá phức tạp do kết hợp nhiều thông tin khác nhau trong một chuỗi duy nhất. Điều này gây khó khăn trong việc phân tích và trực quan hóa dữ liệu, bởi mỗi thông tin cần được xử lý riêng biệt để đưa ra kết quả đánh giá chính xác. Do đó, nhóm tiến hành việc tách cột ‘Tags’ thành các thuộc tính cụ thể và riêng biệt, giúp đơn giản hóa quá trình phân tích dữ liệu.

Guests	Trip_Type	Stayed	Submit	With a pet	Room_Type
Couple	Leisure trip	6	Not Mention	Not Mention	Duplex Double Room
Couple	Leisure trip	4	Not Mention	Not Mention	Duplex Double Room
Family with young children	Leisure trip	3	from mobile	Not Mention	Duplex Double Room
Solo traveler	Leisure trip	3	Not Mention	Not Mention	Duplex Double Room
Couple	Leisure trip	2	from mobile	Not Mention	Suite
...

Hình 3.4. Bộ dữ liệu sau khi tách ‘Tags’ thành các cột ‘Guests’, ‘Trip_Type’, ‘Stayed’, ‘Submit’, ‘With a pet’, ‘Room_Type’

Chi tiết về các cột:

- **Guests (object)**: Nhóm khách tham gia chuyến đi
- **Trip_Type (object)**: Loại chuyến đi
- **Stayed (object)**: Thời gian lưu trú
- **Submit (object)**: Nền tảng gửi đánh giá
- **With a pet (object)**: Có mang theo thú cưng hay không
- **Room_Type (object)**: Loại phòng khách sạn

3.2.2. Xóa các cột không cần thiết

Trong quá trình xử lý dữ liệu, nhóm nhận thấy một số cột trong bộ dữ liệu không đóng góp vào việc phân tích. Do đó, nhóm tiến hành loại bỏ các cột này nhằm giảm kích thước dữ liệu, và tập trung vào các thông tin quan trọng hơn.

Cụ thể, nhóm đã loại bỏ các cột sau:

- **Average_Score (int64)**: Điểm đánh giá trung bình của khách sạn
- **Total_Number_of_Reviews_Reviewer_Has_Given (int64)**: Tổng số lượng đánh giá mà khách hàng đã thực hiện trước đây
- **Total_Number_of_Reviews (int64)**: Tổng số lượng đánh giá hợp lệ của khách sạn

- ***days_since_review*** (*int64*): Khoảng thời gian (tính bằng ngày) giữa ngày đăng đánh giá và ngày thu thập dữ liệu
 - ***Additional_Number_of_Scoring*** (*int64*): Số lượng đánh giá của khách hàng chỉ cho điểm, mà không có nội dung văn bản.
 - ***lat*** (*float64*): Vĩ độ của khách sạn
 - ***lng*** (*float64*): Kinh độ của khách sạn

```
df_copy.drop(columns=['Average_Score', 'Total_Number_of_Reviews_Reviewer_Has_Given',  
                     'Total_Number_of_Reviews', 'days_since_review', 'Additional_Number_of_Scoring', 'lat', 'lng'], inplace=True)
```

3.3. Kiểm tra dữ liệu

Trước khi tiến hành phân tích, nhóm sẽ tiến hành kiểm tra chất lượng của bộ dữ liệu nhằm kịp thời phát hiện và xử lý các vấn đề tiềm ẩn.

3.3.1. Kiểm tra giá trị bị thiếu (*missing values*) và giá trị bất thường (*abnormal values*)

Trong quá trình tiền xử lý dữ liệu, kiểm tra sự tồn tại của giá trị bị thiếu (*missing values*) và các giá trị bất thường (*abnormal values*) là một bước quan trọng, bởi những giá trị này có thể làm sai lệch kết quả phân tích. Việc kiểm tra giá trị bị thiếu nhằm đảm bảo bộ dữ liệu đầy đủ, đồng thời giảm ảnh hưởng của các lỗi do thiếu dữ liệu.

Sau khi kiểm tra, nhóm nhận thấy rằng bộ dữ liệu không chứa giá trị thiếu (*missing values*). Tất cả các cột đều có đầy đủ giá trị cho tất cả các dòng quan sát.

	Missing Values	Percentage (%)
Hotel_Address	0	0.0
Review_Date	0	0.0
Hotel_Name	0	0.0
Reviewer_Nationality	0	0.0
Negative_Review	0	0.0
Review_Total_Negative_Word_Counts	0	0.0
Positive_Review	0	0.0
Review_Total_Positive_Word_Counts	0	0.0
Reviewer_Score	0	0.0
Country	0	0.0
Reviewer_Regions	0	0.0
Guests	0	0.0
Trip_Type	0	0.0
Stayed	0	0.0
Submit	0	0.0
With a pet	0	0.0
Room_Type	0	0.0

Hình 3.5. Quan sát Missing Values

Ngoài ra, đối với các thuộc tính kiểu số trong bộ dữ liệu đều yêu cầu phải mang giá trị dương. Vì vậy, nhóm đã thực hiện kiểm tra để xác định xem có thuộc tính kiểu số nào mang giá trị âm, và nếu phát hiện, những giá trị này sẽ được xem là giá trị bất thường (*abnormal values*). Kết quả kiểm tra cho thấy không có giá trị nào bất thường.

3.3.2. Kiểm tra và xử lý giá trị trùng lặp (*duplicate values*)

Nhóm tiến hành kiểm tra các giá trị trùng lặp trong bộ dữ liệu và kết quả cho thấy có **527 giá trị trùng lặp**.

```
# Kiểm tra giá trị trùng lặp
df_copy.duplicated().sum()
✓ 1.5s
527
```

Hình 3.6. Kết quả kiểm tra giá trị trùng lặp

Nếu không xử lý các giá trị này thì có thể gây ảnh hưởng đến chất lượng phân tích và kết quả biểu diễn trực quan dữ liệu. Do đó, để khắc phục, nhóm đã loại bỏ các giá trị trùng lặp trên bộ dữ liệu, nhằm đảm bảo rằng mỗi quan sát trong tập dữ liệu là duy nhất, giúp nâng cao độ tin cậy của các kết quả phân tích.

3.4. Phân tích thống kê tổng quát

Trước khi bắt đầu phân tích thống kê tổng quát, nhóm sẽ thực hiện việc quan sát tổng quan bộ dữ liệu đã qua bước tiền xử lý. Cấu trúc và đặc điểm chính của tập dữ liệu sau khi xử lý được mô tả như sau:

- *Tổng số dòng*: 515,211 (số lượng quan sát)
- *Tổng số cột*: 17 (biến thuộc tính)
- Bộ dữ liệu mới đã được loại bỏ 8 cột so với nguyên bản và bổ sung thêm 8 cột dữ liệu mới. Dưới đây là thông tin chi tiết về các cột mới trong tập dữ liệu:

- **Country (object)**: Quốc gia nơi khách sạn tọa lạc
- **Reviewer_Regions (object)**: Người đánh giá có đến từ châu Á (Asia) hay không (Non-Asia)
- **Guests (object)**: Thông tin về loại khách (Ví dụ: Nhóm, Gia đình, Cặp đôi)
- **Trip_Type (object)**: Loại chuyến đi (Ví dụ: Công tác, Du lịch, Nghỉ dưỡng)
- **Stayed (object)**: Khoảng thời gian lưu trú tại khách sạn
- **Submit (object)**: Nền tảng gửi đánh giá
- **With a pet (object)**: Thông tin khách có đi kèm thú cưng hay không
- **Room_Type (object)**: Loại phòng được sử dụng

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 515211 entries, 0 to 515210
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   Hotel_Address    515211 non-null   object 
 1   Review_Date      515211 non-null   object 
 2   Hotel_Name       515211 non-null   object 
 3   Reviewer_Nationality 515211 non-null   object 
 4   Negative_Review  515211 non-null   object 
 5   Review_Total_Negative_Word_Counts 515211 non-null   int64  
 6   Positive_Review  515211 non-null   object 
 7   Review_Total_Positive_Word_Counts 515211 non-null   int64  
 8   Reviewer_Score    515211 non-null   float64
 9   Country          515211 non-null   object 
 10  Reviewer_Regions 515211 non-null   object 
 11  Guests           515211 non-null   object 
 12  Trip_Type        515211 non-null   object 
 13  Stayed           515211 non-null   object 
 14  Submit            515211 non-null   object 
 15  With a pet       515211 non-null   object 
 16  Room_Type         515211 non-null   object 
dtypes: float64(1), int64(2), object(14)
memory usage: 66.8+ MB
```

Hình 3.7. Mô tả thông tin cơ bản của bộ dữ liệu sau khi tiền xử lý

Với tập dữ liệu đã được tiền xử lý, tiếp theo, nhóm sẽ tiến hành phân tích thống kê tổng quát hai kiểu thuộc tính của bộ dữ liệu để hiểu rõ hơn về các đặc điểm chính của dữ liệu.

a) Thuộc tính kiểu số (Numerical)

Thuộc tính kiểu số (Numerical) bao gồm các cột dữ liệu có kiểu dữ liệu *int64* và *float64*. Cụ thể, bao gồm các cột như *Review_Total_Negative_Word_Counts*, *Review_Total_Positive_Word_Counts* và *Reviewer_Score*.

➤ ***Review_Total_Negative_Word_Counts*** (*Tổng số từ trong đánh giá tiêu cực*):

- Trung bình trong mỗi đánh giá tiêu cực chứa khoảng 18.54 từ, với độ lệch chuẩn khá cao (29.69).
- Giá trị min và max lần lượt bằng 0 và bằng 408, cho thấy có những đánh giá không chứa từ tiêu cực nào và có đánh giá chứa nhiều nhất tới 408 từ.

➤ ***Review_Total_Positive_Word_Counts*** (*Tổng số từ trong đánh giá tích cực*):

- Trung bình trong mỗi đánh giá tích cực chứa khoảng 17.78 từ, với độ lệch chuẩn là 21.80.
- Giá trị min và max lần lượt bằng 0 và bằng 395, cho thấy có những đánh giá không chứa từ tích cực nào và có đánh giá chứa nhiều nhất tới 395 từ.

➤ ***Reviewer_Score*** (*Điểm đánh giá khách hàng chấm cho khách sạn dựa trên trải nghiệm thực tế*):

- Điểm trung bình mà các khách hàng đánh giá là 8.40, với độ lệch chuẩn là 1.64.
- Điểm thấp nhất là 2.50, trong khi điểm cao nhất được đánh giá đạt 10.

	Rows	Mean	Standard Deviation	Min	Max
Review_Total_Negative_Word_Counts	515211	18.540811	29.694018	0.0	408.0
Review_Total_Positive_Word_Counts	515211	17.778268	21.804561	0.0	395.0
Reviewer_Score	515211	8.395530	1.637468	2.5	10.0

Hình 3.8. Thống kê tổng quát với các thuộc tính kiểu numerical

b) *Thuộc tính phân loại (Categorical)*

Thuộc tính phân loại (*Categorical*) bao gồm các cột dữ liệu có kiểu dữ liệu *object*. Cụ thể, bao gồm các cột như *Hotel_Address*, *Review_Date*, *Hotel_Name*, *Reviewer_Nationality*, *Negative_Review*, *Positive_Review*, *Country*, *Reviewer_Regions*, *Guests*, *Trip_Type*, *Stayed*, *Submit*, *With a pet*, *Room_Type*.

➤ ***Hotel_Address*** (*Địa chỉ của khách sạn*) và ***Hotel_Name*** (*Tên khách sạn*):

- Có 1493 địa chỉ khách sạn và 1492 khách sạn khác nhau trong bộ dữ liệu
- Địa chỉ được xuất hiện nhiều nhất là ‘163 Marsh Wall Docklands Tower Hamlets London’, tương ứng với khách sạn ‘Britannia International Hotel Canary Wharf’, với 4789 lần xuất hiện.

➤ ***Country*** (*Quốc gia nơi khách sạn tọa lạc*):

- Được tách ra và phân loại từ thuộc tính *Hotel_Address*, bao gồm 6 quốc gia: ‘United Kingdom’, ‘Australia’, ‘Netherlands’, ‘France’, ‘Italy’, ‘Spain’.
- Trong đó, phần lớn các khách sạn tọa lạc ở ‘United Kingdom’, chiếm 50.9% tổng số quan sát (262,297 đánh giá)

➤ ***Reviewer_Nationality*** (*Quốc tịch của người đánh giá*):

- Có 227 quốc tịch khác nhau
- Du khách đến từ ‘United Kingdom’ chiếm số lượng áp đảo với 245,110 đánh giá.

➤ ***Reviewer_Regions*** (*Người đánh giá có đến từ châu Á (Asia) hay không (Non-Asia)*):

- Phân loại dữ liệu của thuộc tính *Reviewer_Nationality* thành 2 khu vực: ‘Asia’ và ‘Non-Asia’
- Du khách ‘Non-Asia’ chiếm đa số với 482,694 đánh giá (gần 93.7% tổng số quan sát)

➤ ***Negative_Review*** (*Đánh giá tiêu cực*):

- Có 330,011 đánh giá khác nhau, trong đó nội dung phổ biến nhất là ‘No Negative’ với 127,757 lần xuất hiện.

➤ ***Positive_Review*** (*Đánh giá tích cực*):

- Có 412,601 đánh giá khác nhau, trong đó nội dung phổ biến nhất là ‘No Positive’ với 35,904 lần xuất hiện.

➤ ***Guests*** (*Thông tin về loại khách*):

- Bao gồm 6 nhóm khách: ‘Couple’, ‘Group’, ‘Solo traveler’, ‘Travelers with friends’, ‘Family with older children’, ‘Family with young children’.
- Các cặp đôi là khách hàng chính, chiếm 252,005 đánh giá.

➤ ***Trip_Type*** (*Loại chuyến đi*):

- Bao gồm 3 loại chuyến đi: ‘Leisure trip’ (*Nghỉ dưỡng*), ‘Business trip’ (*Công tác*) và ‘Not mention’ (*Không đề cập*)
- Chủ yếu là các du khách với mục đích nghỉ dưỡng, chiếm áp đảo với hơn 81% (417,355 đánh giá).

➤ ***Stayed*** (*Khoảng thời gian lưu trú tại khách sạn*):

- Chủ yếu du khách lưu trú 1 đêm, chiếm 193,497 đánh giá (khoảng 37.5% tổng số dữ liệu)

➤ ***Submit*** (*Nền tảng gửi đánh giá*):

- Du khách dùng 2 loại nền tảng chính để gửi đánh giá là ‘from mobile’ (*từ thiết bị di động*) và ‘from web’ (*từ trang web*).
- Đa số, du khách gửi đánh giá từ thiết bị di động, với 307,355 đánh giá.

➤ ***With a pet*** (*Thông tin khách có đi kèm thú cưng hay không*):

- Có 2 giá trị: ‘Not Mention’ (*Không đi kèm thú cưng*) và ‘With pet’ (*Đi kèm thú cưng*).
- Phần lớn các du khách không đi kèm thú cưng, chiếm 513,806 đánh giá (khoảng 99.7%).

➤ ***Room_Type*** (*Loại phòng*):

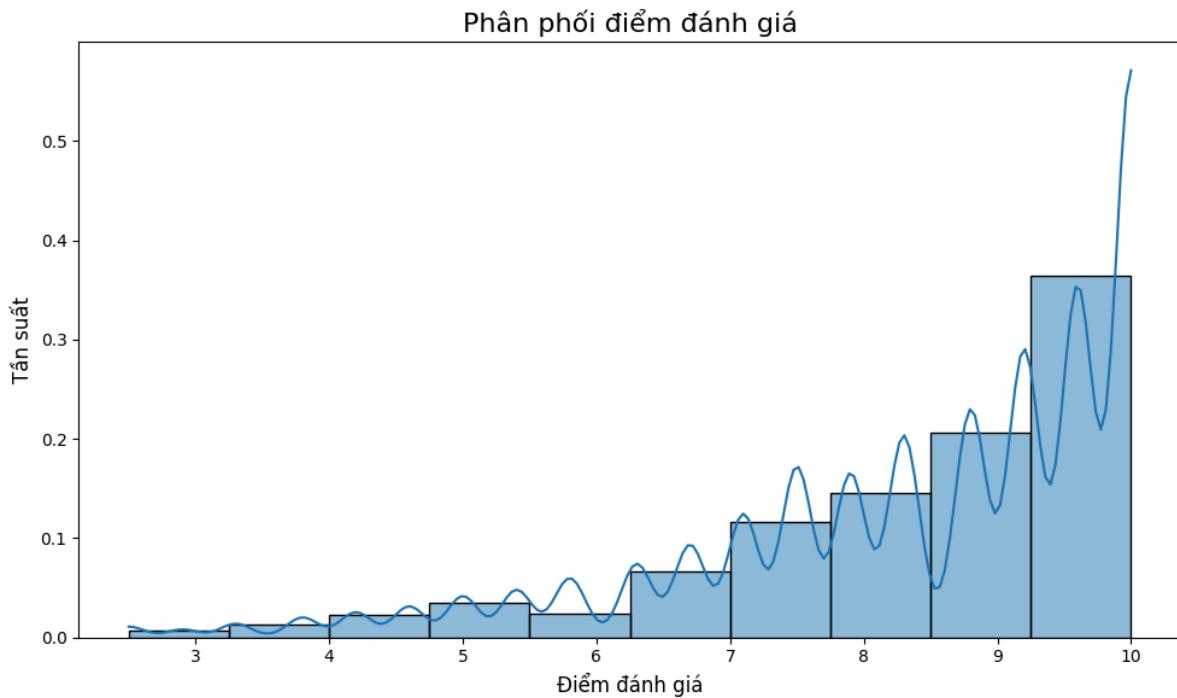
- Có hơn 2,382 loại phòng khác nhau.
- Loại phòng được du khách ưu chuộng nhất là ‘Double Room’ (*Phòng đôi*), với 34,027 đánh giá.

		count	unique		top	freq
Hotel_Address	515211	1493	163 Marsh Wall Docklands Tower Hamlets London ...		4789	
Review_Date	515211	731		8/2/2017	2584	
Hotel_Name	515211	1492	Britannia International Hotel Canary Wharf		4789	
Reviewer_Nationality	515211	227		United Kingdom	245110	
Negative_Review	515211	330011		No Negative	127757	
Positive_Review	515211	412601		No Positive	35904	
Country	515211	6		United Kingdom	262297	
Reviewer_Regions	515211	2		Non-Asia	482694	
Guests	515211	6		Couple	252005	
Trip_Type	515211	3		Leisure trip	417355	
Stayed	515211	32		1	193497	
Submit	515211	2		from mobile	307355	
With a pet	515211	2		Not Mention	513806	
Room_Type	515211	2382		Double Room	34027	

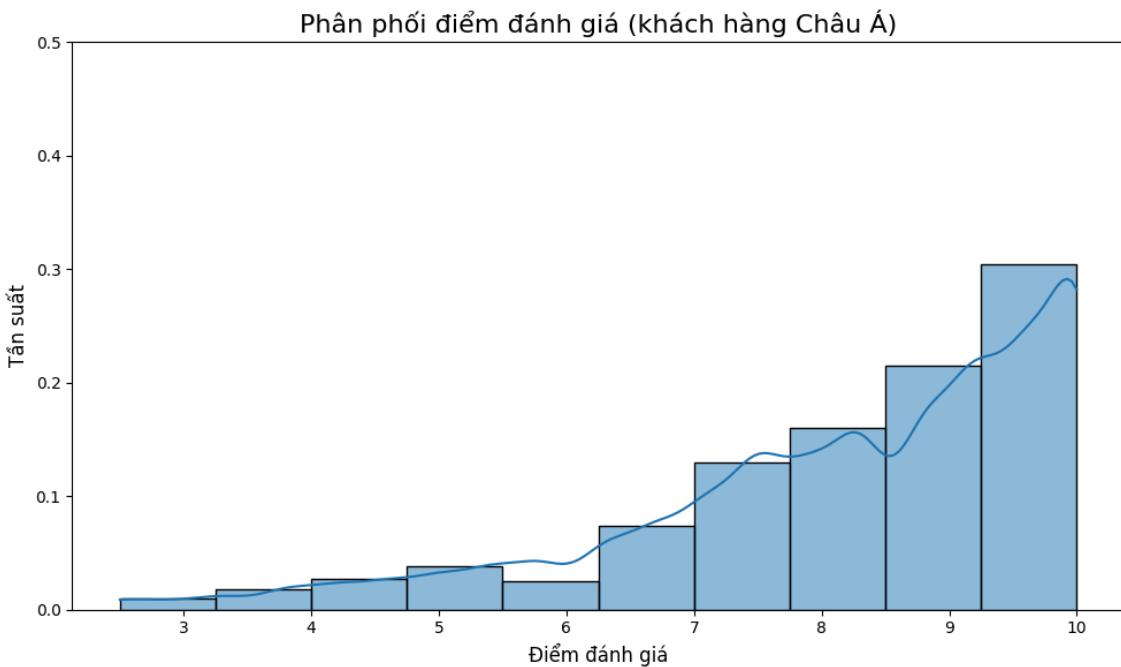
Hình 3.9. Thống kê tổng quát với các thuộc tính kiểu categorical

CHƯƠNG 4 - ĐÁNH GIÁ SỰ TƯƠNG QUAN GIỮA ĐIỂM SỐ VÀ BÌNH LUẬN

4.1. Phân phối điểm đánh giá của khách hàng



Biểu đồ 4.1. Phân phối điểm đánh giá chung

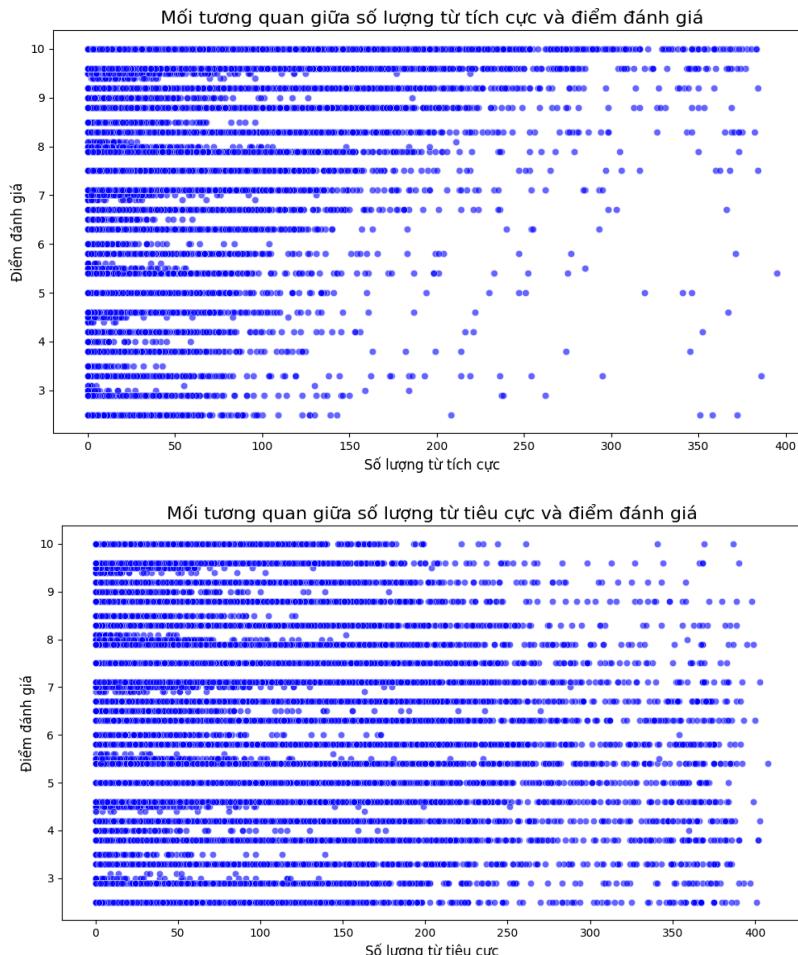


Biểu đồ 4.2. Phân phối điểm đánh giá của khách hàng Châu Á

Nhìn vào bảng phân phối điểm đánh giá của khách hàng chung và khách hàng Châu Á nói riêng đối với các khách sạn Châu Âu, ta thấy không có sự khác biệt nhiều về phổ điểm. Phần lớn được đánh giá từ ở mức điểm trên trung bình. Do vậy, nhìn chung, phần lớn

khách hàng có vẻ hài lòng với dịch vụ của khách sạn. Đồng thời, không có sự khác biệt trong việc cho điểm của khách hàng Châu Á và phần còn lại.

4.2. Tương quan giữa điểm số và bình luận



Biểu đồ 4.3. Tương quan giữa số lượng từ tích cực, tiêu cực trong bình luận với điểm đánh giá

Một mối quan tâm được nhóm đặt ra rằng, liệu rằng điểm đánh giá của khách hàng có thể hiện được những đánh giá của họ đối với dịch vụ khách sạn. Để xác minh điều đó, nhóm tiến hành kiểm định giả thuyết thống kê mối tương quan giữa điểm đánh giá của khách hàng và số lượng từ bình luận tiêu cực, tích cực trong mỗi bình luận của họ. Nếu thật sự có mối tương quan như mong đợi, chúng ta có thể kết luận điểm đánh giá đã nói lên phần nào đó những đánh giá của khách hàng đối với dịch vụ của khách sạn.

Nhóm tiến hành kiểm định giả thuyết mối tương quan:

- **H0:** Không có mối tương quan giữa điểm đánh giá và số lượng từ tích cực/tiêu cực trong câu bình luận

➤ **H1:** *Tồn tại mối tương quan giữa điểm đánh giá và số lượng từ tích cực/tiêu cực trong câu bình luận.*

Sau khi kiểm tra phân phối của các biến, nhóm nhận thấy các biến không có phân phối chuẩn, do đó, nhóm sử dụng phương pháp Spearman để kiểm định giả thuyết trên. Bên dưới là kết quả kiểm định.

```
mean_score_p = shapiro(df_hotel['Reviewer_Score']).pvalue
positive_ratio_p = shapiro(df_hotel['Review_Total_Positive_Word_Counts']).pvalue

if mean_score_p > 0.05 and positive_ratio_p > 0.05:
    corr, p_value = pearsonr(df_hotel['Reviewer_Score'], df_hotel['Review_Total_Positive_Word_Counts'])
    method = "Pearson's correlation"
else:
    corr, p_value = spearmanr(df_hotel['Reviewer_Score'], df_hotel['Review_Total_Positive_Word_Counts'])
    method = "Spearman's rank correlation"

print(f"Kiểm định: {method}")
print(f"Hệ số tương quan: {corr:.2f}, P-value: {p_value:.4f}")

if p_value < 0.05:
    print("Có mối tương quan có ý nghĩa thống kê giữa điểm đánh giá và số lượng từ tích cực.")
    if corr > 0:
        print("Mối tương quan thuận: khi số lượng từ tích cực tăng, điểm đánh giá cũng tăng.")
    elif corr < 0:
        print("Mối tương quan nghịch: khi số lượng từ tích cực tăng, điểm đánh giá giảm.")
    else:
        print("Mối tương quan gần bằng 0: không có mối liên hệ rõ rệt.")
else:
    print("Không có mối tương quan có ý nghĩa thống kê giữa điểm đánh giá trung bình và số lượng từ tích cực.")
```

```
mean_score_p = shapiro(df_hotel['Reviewer_Score']).pvalue
positive_ratio_p = shapiro(df_hotel['Review_Total_Positive_Word_Counts']).pvalue

if mean_score_p > 0.05 and positive_ratio_p > 0.05:
    corr, p_value = pearsonr(df_hotel['Reviewer_Score'], df_hotel['Review_Total_Negative_Word_Counts'])
    method = "Pearson's correlation"
else:
    corr, p_value = spearmanr(df_hotel['Reviewer_Score'], df_hotel['Review_Total_Negative_Word_Counts'])
    method = "Spearman's rank correlation"

print(f"Kiểm định: {method}")
print(f"Hệ số tương quan: {corr:.2f}, P-value: {p_value:.4f}")

if p_value < 0.05:
    print("Có mối tương quan có ý nghĩa thống kê giữa điểm đánh giá và số lượng từ tiêu cực.")
    if corr > 0:
        print("Mối tương quan thuận: khi số lượng từ tiêu cực tăng, điểm đánh giá cũng tăng.")
    elif corr < 0:
        print("Mối tương quan nghịch: khi số lượng từ tiêu cực tăng, điểm đánh giá giảm.")
    else:
        print("Mối tương quan gần bằng 0: không có mối liên hệ rõ rệt.")
else:
    print("Không có mối tương quan có ý nghĩa thống kê giữa điểm đánh giá trung bình và số lượng từ tích cực.")
```

Kiểm định: Spearman's rank correlation

Hệ số tương quan: 0.31, P-value: 0.0000

Có mối tương quan có ý nghĩa thống kê giữa điểm đánh giá và số lượng từ tích cực.

Mối tương quan thuận: khi số lượng từ tích cực tăng, điểm đánh giá cũng tăng.

Kiểm định: Spearman's rank correlation

Hệ số tương quan: -0.47, P-value: 0.0000

Có mối tương quan có ý nghĩa thống kê giữa điểm đánh giá và số lượng từ tiêu cực.

Mối tương quan nghịch: khi số lượng từ tiêu cực tăng, điểm đánh giá giảm.

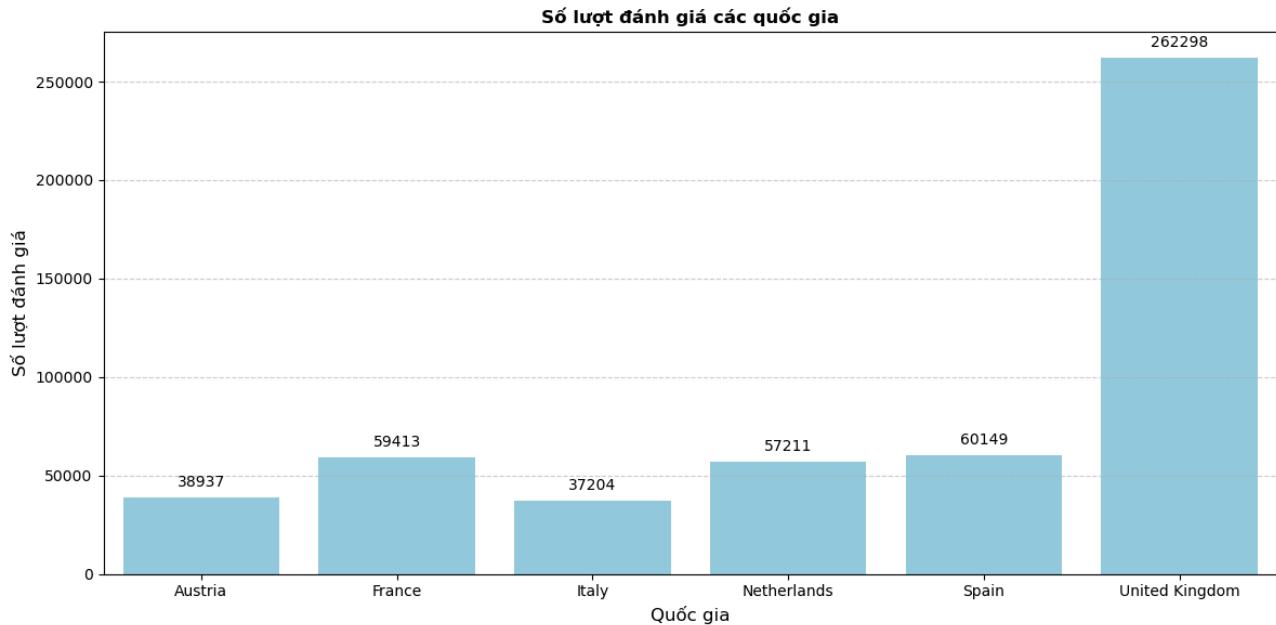
Hình 4.1. Kiểm định Spearman's rank correlation

Như vậy, có thể kết luận rằng điểm đánh giá nói lên một phần nào đó những đánh giá của khách hàng đối với dịch vụ của khách sạn.

CHƯƠNG 5 - BIỂU ĐIỂN TRỰC QUAN DŨ LIỆU

5.1. Tổng quan tình hình lưu trú khách sạn tại châu Âu

5.1.1. Số lượt đánh giá theo các quốc gia ở châu Âu

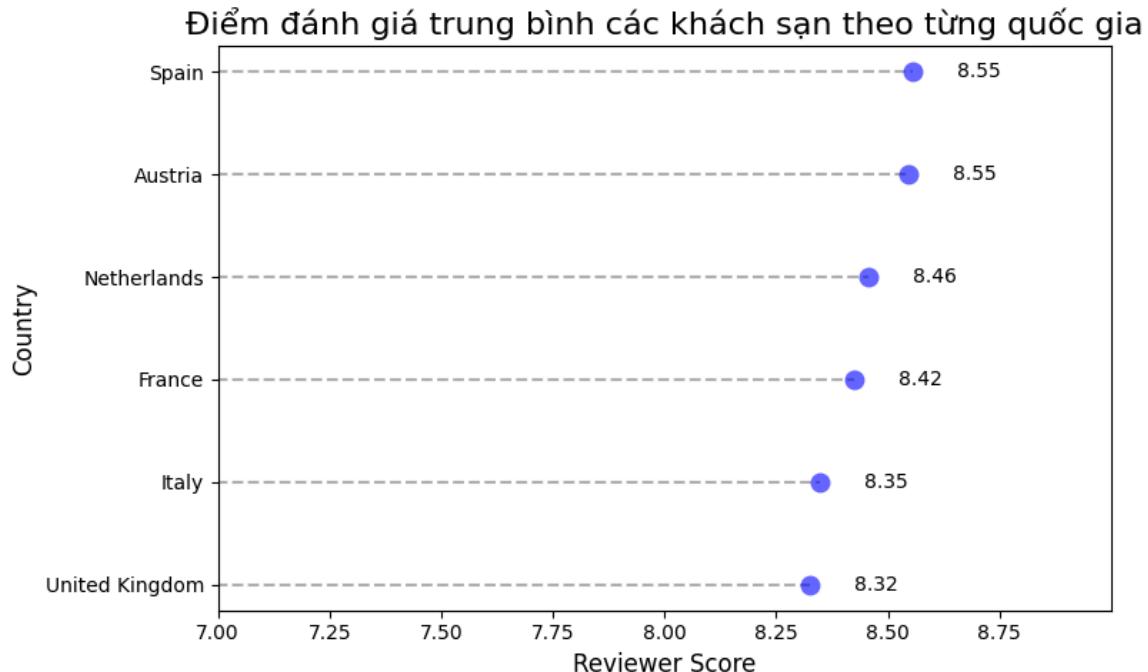


Biểu đồ 5.1. Số lượt đánh giá theo các quốc gia ở Châu Âu

Châu Âu tiếp tục là điểm đến du lịch hấp dẫn, với các quốc gia thu hút lượng lớn khách du lịch nhờ vào những địa danh nổi tiếng và văn hóa đặc sắc. Anh dẫn đầu với số lượt đánh giá cao nhất là 262,298, nổi bật với các địa điểm như Tháp London, Cung điện Buckingham và Tháp Stonehenge. Tây Ban Nha đứng thứ hai với 60,149 lượt đánh giá, thu hút du khách đến với Sagrada Familia ở Barcelona, Cung điện Hoàng gia Madrid và Tháp Giralda ở Seville. Pháp cũng là điểm đến không thể bỏ qua với 59,413 lượt đánh giá, với những địa điểm hấp dẫn như Tháp Eiffel, Bảo tàng Louvre và Cung điện Versailles.

Các quốc gia khác như Hà Lan, với 57,211 lượt đánh giá, nổi bật nhờ Kênh đào Amsterdam, Bảo tàng Van Gogh và vườn hoa tulip Keukenhof, trong khi Ý có 37,204 lượt đánh giá, với các di tích lịch sử như Đầu trường Colosseum, Nhà thờ St. Peter và Quảng trường St. Mark. Cuối cùng, Áo thu hút 38,937 lượt đánh giá, với các điểm đến nổi tiếng như Cung điện Hofburg ở Vienna và Pháo đài Hohensalzburg ở Salzburg.

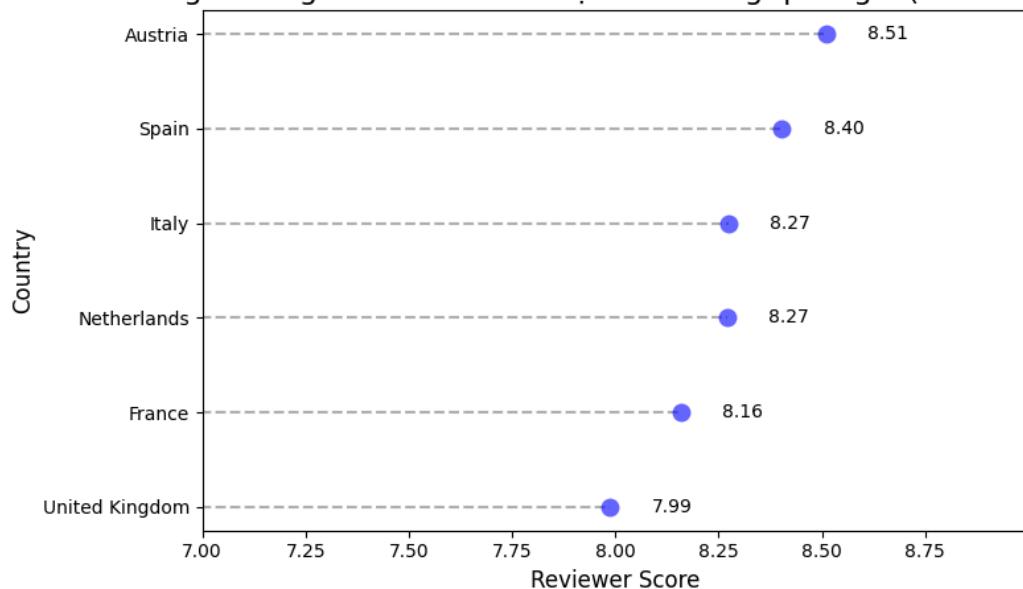
5.1.2. Điểm trung bình của từng quốc gia



Biểu đồ 5.2. Biểu đồ thể hiện điểm trung bình của từng quốc gia

Với dữ liệu điểm đánh giá của các khách du lịch với các khách sạn tại 6 quốc gia, biểu đồ trên liệt kê theo thứ tự các quốc gia có điểm số từ cao đến thấp. Đầu tiên là Tây Ban Nha với 8.55, theo sau lần lượt là Áo, Hà Lan, Pháp, Ý và cuối cùng là Anh. Tuy có sự sắp xếp giảm dần, nhưng các điểm số này không chênh nhau quá lớn, đều trên 8.3.

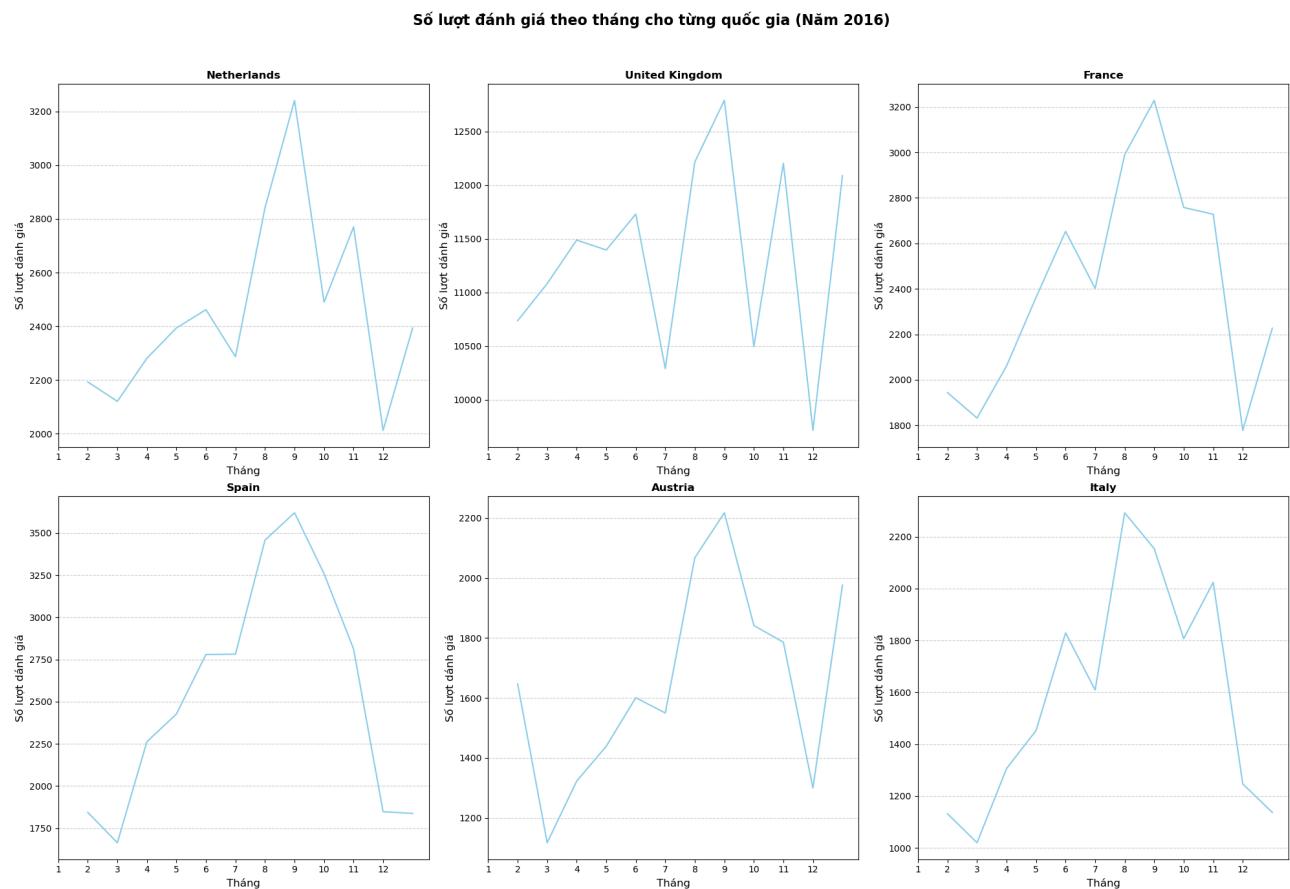
Điểm đánh giá trung bình các khách sạn theo từng quốc gia (Khách hàng Châu Á)



Biểu đồ 5.3. Điểm đánh giá trung bình các khách sạn theo từng quốc gia (khách hàng Châu Á)

Để có góc nhìn cho những khách là người Châu Á, nhóm phân tích điểm đánh giá trung bình của các khách người Châu Á trước đó đối với các khách sạn tại 6 quốc gia và thấy thứ tự các quốc gia đã có sự thay đổi. Lúc này Áo đứng đầu với 8.51, sau đó là Tây Ban Nha, Ý, Hà Lan, Pháp và Anh. Trong trường hợp này, điểm số đã có sự biến động lớn hơn so với lúc nãy, điểm trung bình thấp nhất là 7.99 và cao nhất là 8.51.

5.1.3. Số khách du lịch đến các quốc gia theo từng tháng năm 2016

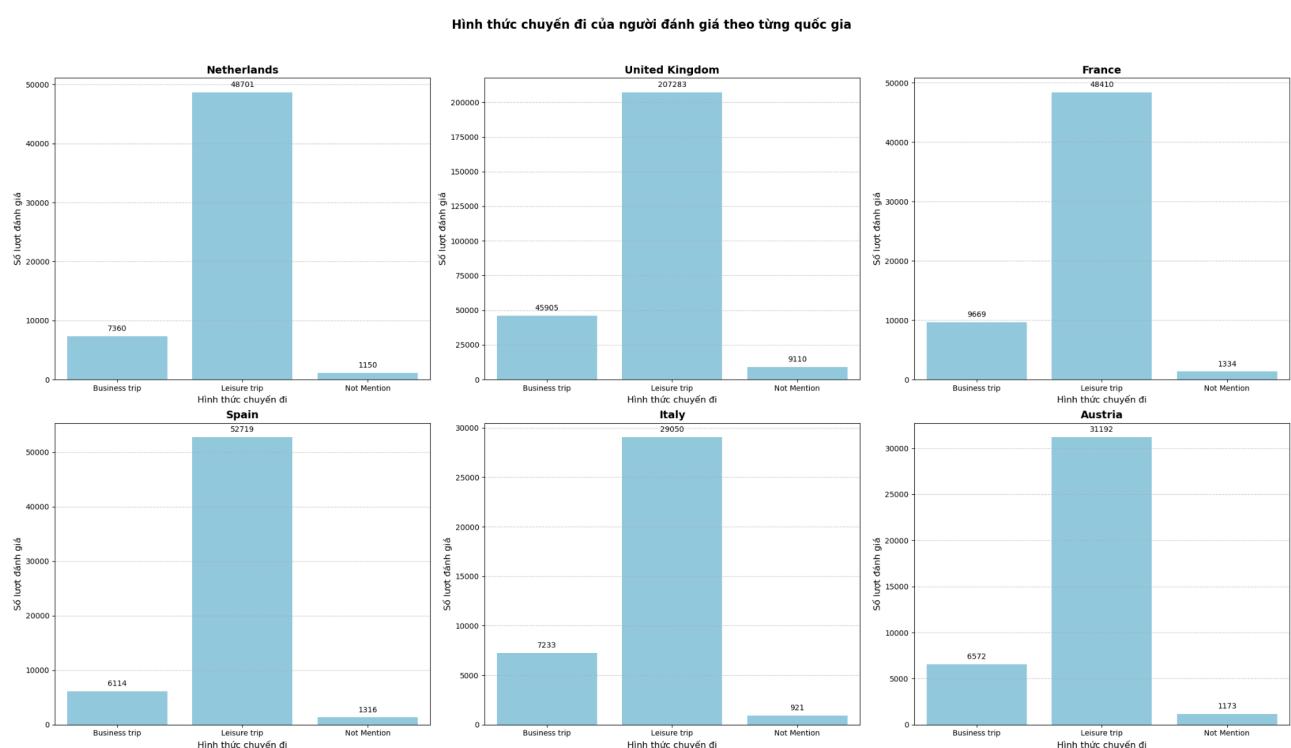


Biểu đồ 5.4. Số lượt đánh giá theo tháng của từng quốc gia năm 2016

Dựa trên số liệu lượt đánh giá theo tháng của các quốc gia ở Châu Âu trong năm 2016, có thể thấy rằng mùa hè (tháng 6 đến tháng 8) là thời điểm lý tưởng để du lịch đến các quốc gia này. Cụ thể, Hà Lan đạt đỉnh lượt đánh giá vào tháng 8 với 3,241 lượt, cho thấy đây là thời gian tuyệt vời để thăm các kênh đào Amsterdam và vườn hoa tulip Keukenhof, những địa điểm nổi tiếng thu hút du khách quốc tế vào mùa hè. Vương quốc Anh cũng ghi nhận lượng du khách cao nhất vào tháng 8 với 12,790 lượt đánh giá, là thời điểm lý tưởng để tham quan các địa điểm nổi bật như Tháp London, Cung điện Buckingham, hay tham gia

các lễ hội văn hóa lớn tại Edinburgh. Tương tự, Pháp và Tây Ban Nha cũng có lượng khách du lịch tăng mạnh trong mùa hè, với đỉnh điểm là tháng 8, khi lần lượt đạt 3,229 và 3,620 lượt đánh giá, thích hợp cho việc khám phá các lâu đài Loire, Paris, Nice, các bãi biển Costa Brava, và tham gia các sự kiện ngoài trời như Lễ hội San Fermin ở Tây Ban Nha. Áo và Ý cũng ghi nhận sự gia tăng lượt đánh giá vào giữa năm, đặc biệt là trong tháng 8, với các điểm đến nổi tiếng như Cung điện Schönbrunn ở Vienna, Nhà hát Opera Vienna, hay các lâu đài và vườn hoa ở Salzburg. Tại Ý, du khách có thể khám phá các địa điểm như Colosseum ở Rome, Thành Venice và các vườn nho Tuscany.

5.1.4. Hình thức và nhóm theo đánh giá khi đến các quốc gia



Biểu đồ 5.5. Hình thức chuyến đi theo từng quốc gia

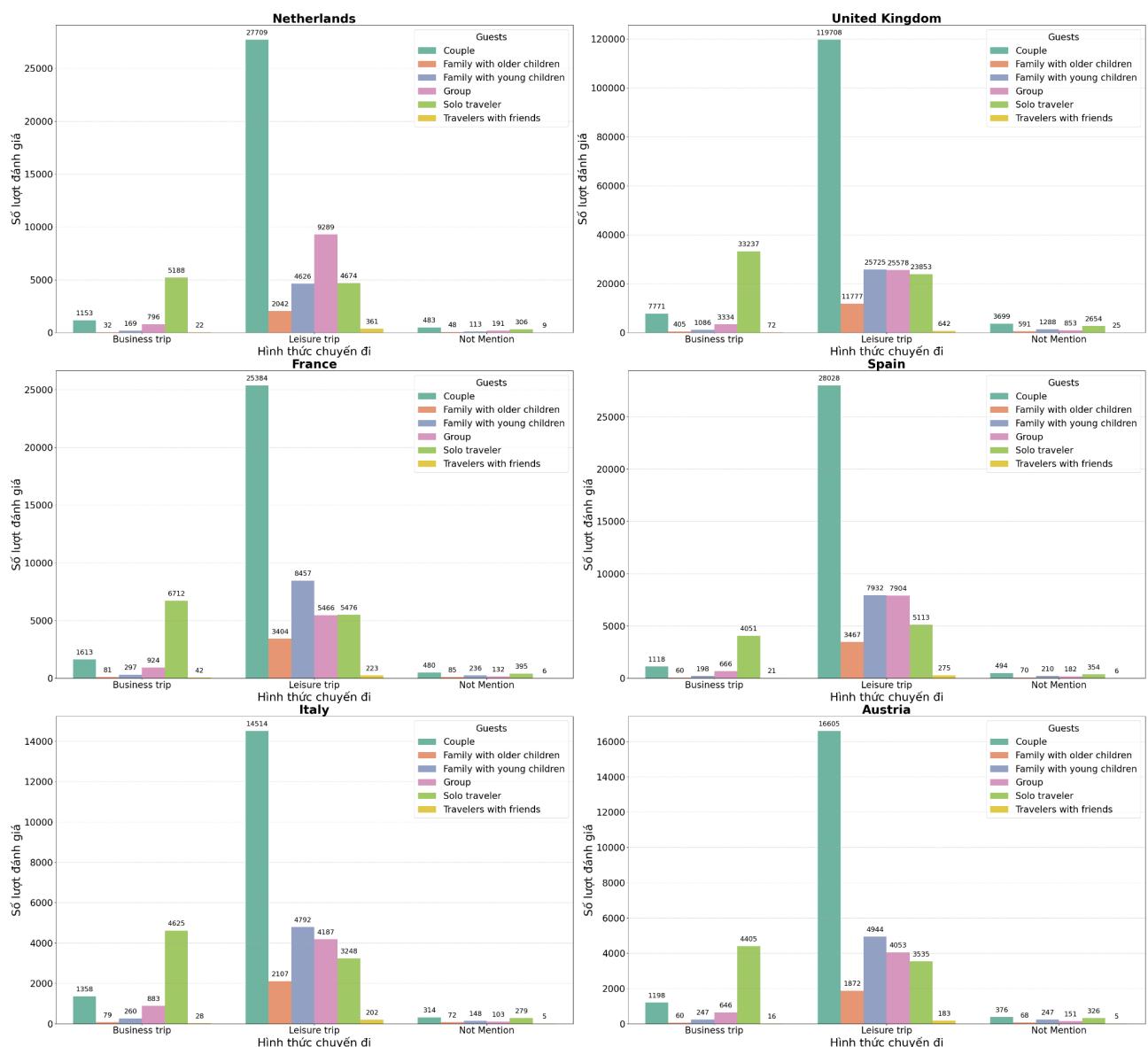
Dựa trên biểu đồ phân tích hình thức chuyến đi đến các quốc gia châu Âu, có thể thấy rằng loại hình du lịch nghỉ dưỡng (Leisure Trip) chiếm ưu thế vượt trội tại tất cả các quốc gia. Vương quốc Anh dẫn đầu với 207,283 lượt khách cho hình thức này, tiếp theo là Hà Lan (48,701), Pháp (48,410), Tây Ban Nha (52,719), Ý (29,050) và Áo (31,192). Điều này cho thấy các quốc gia này là điểm đến lý tưởng cho các chuyến du lịch nghỉ dưỡng nhờ cảnh quan phong phú, có lịch sử và văn hóa đa dạng và hấp dẫn.

Loại hình chuyến đi công tác (Business Trip) cũng đóng vai trò quan trọng, đặc biệt tại Vương quốc Anh với 45,905 khách và Hà Lan với 7,360 khách. Các quốc gia khác như

Pháp (9,669), Tây Ban Nha (6,114), Ý (7,233) và Áo (6,572) cũng thu hút một lượng lớn khách đến công tác.

Điều này phản ánh vị thế của các quốc gia châu Âu không chỉ là những điểm đến du lịch hấp dẫn mà còn là các trung tâm kinh tế và thương mại quan trọng. Đặc biệt, Vương quốc Anh và Pháp đã khẳng định sức hút của mình nhờ sự phát triển mạnh mẽ về kinh tế, cơ sở hạ tầng và tiềm năng hợp tác kinh doanh, trở thành những điểm đến hàng đầu cho các chuyến đi công tác quốc tế.

Nhóm theo hình thức chuyến đi của người đánh giá cho từng quốc gia



Biểu đồ 5.6. Nhóm theo hình thức chuyến đi của từng quốc gia

Dựa trên biểu đồ, với loại hình du lịch nghỉ dưỡng (Leisure Trip) chiếm ưu thế tại tất cả các quốc gia, hầu hết khách du lịch đi theo cặp đôi, đặc biệt ở những điểm đến nổi tiếng như Anh, Tây Ban Nha và Pháp. Các cặp đôi thường tìm kiếm những trải nghiệm sang trọng, lãng mạn và riêng tư, điều này mở ra tiềm năng phát triển các dịch vụ cao cấp như resort dành riêng cho hai người, các nhà hàng lãng mạn, hoặc các hoạt động dành cho cặp đôi.

Đối với nhóm bạn bè, họ có xu hướng chọn các điểm đến có nhiều hoạt động vui chơi giải trí hoặc các sự kiện sôi động. Tây Ban Nha, Ý và Hà Lan là những quốc gia thu hút nhóm khách này bởi có nhiều bờ biển đẹp, lễ hội, và cuộc sống về đêm sôi động. Đây là cơ hội để phát triển các gói dịch vụ tập trung vào hoạt động nhóm như đi phiêu lưu, cắm trại, các sự kiện ngoài trời.

Nhóm gia đình có con nhỏ cũng đóng góp một phần quan trọng vào loại hình du lịch nghỉ dưỡng. Những gia đình này thường ưu tiên các địa điểm an toàn, thân thiện với trẻ em, và có nhiều tiện ích phù hợp như công viên giải trí, khu vui chơi trẻ em hoặc dịch vụ trông trẻ tại khách sạn. Áo, Ý và Pháp đều là những điểm đến lý tưởng cho các gia đình có con nhỏ. Ở Áo, Vienna với các công viên giải trí và bảo tàng là lựa chọn tuyệt vời, trong khi các khu vực như Salzburg cung cấp không gian yên bình và các hoạt động ngoài trời. Ý, đặc biệt là Rome và Venice, kết hợp giữa lịch sử và các khu vui chơi cho trẻ em, như công viên nước và các hoạt động thám hiểm. Pháp cũng không kém phần hấp dẫn với Paris và Disneyland, cùng các công viên và khu vui chơi dành cho trẻ em. Các bãi biển ở Côte d'Azur hay làng quê ở Provence đều mang lại trải nghiệm tuyệt vời và thú vị cho gia đình. Những địa điểm này đều cung cấp các dịch vụ thân thiện với trẻ em, giúp các gia đình tận hưởng kỳ nghỉ một cách thoải mái.

Cuối cùng, nhóm du khách đi một mình đang ngày càng trở thành một phân khúc tiềm năng. Những du khách này thường tìm kiếm các trải nghiệm cá nhân hóa, khám phá văn hóa, thiên nhiên, cũng như sự kết nối qua các hoạt động xã hội, tour khám phá địa phương, đi làm việc kết hợp với nghỉ dưỡng.

Đối với chuyến đi công tác (Business Trip), Anh dẫn đầu với 33,237 đánh giá, tiếp theo là Pháp (6,712 đánh giá) và Hà Lan (5,188 đánh giá). Điều này khẳng định vai trò quan

trọng của các trung tâm kinh tế lớn tại châu Âu trong việc thu hút khách du lịch công vụ. Loại hình này chủ yếu thu hút du khách đi một mình (Solo Traveler), phản ánh tính chất độc lập của công việc.

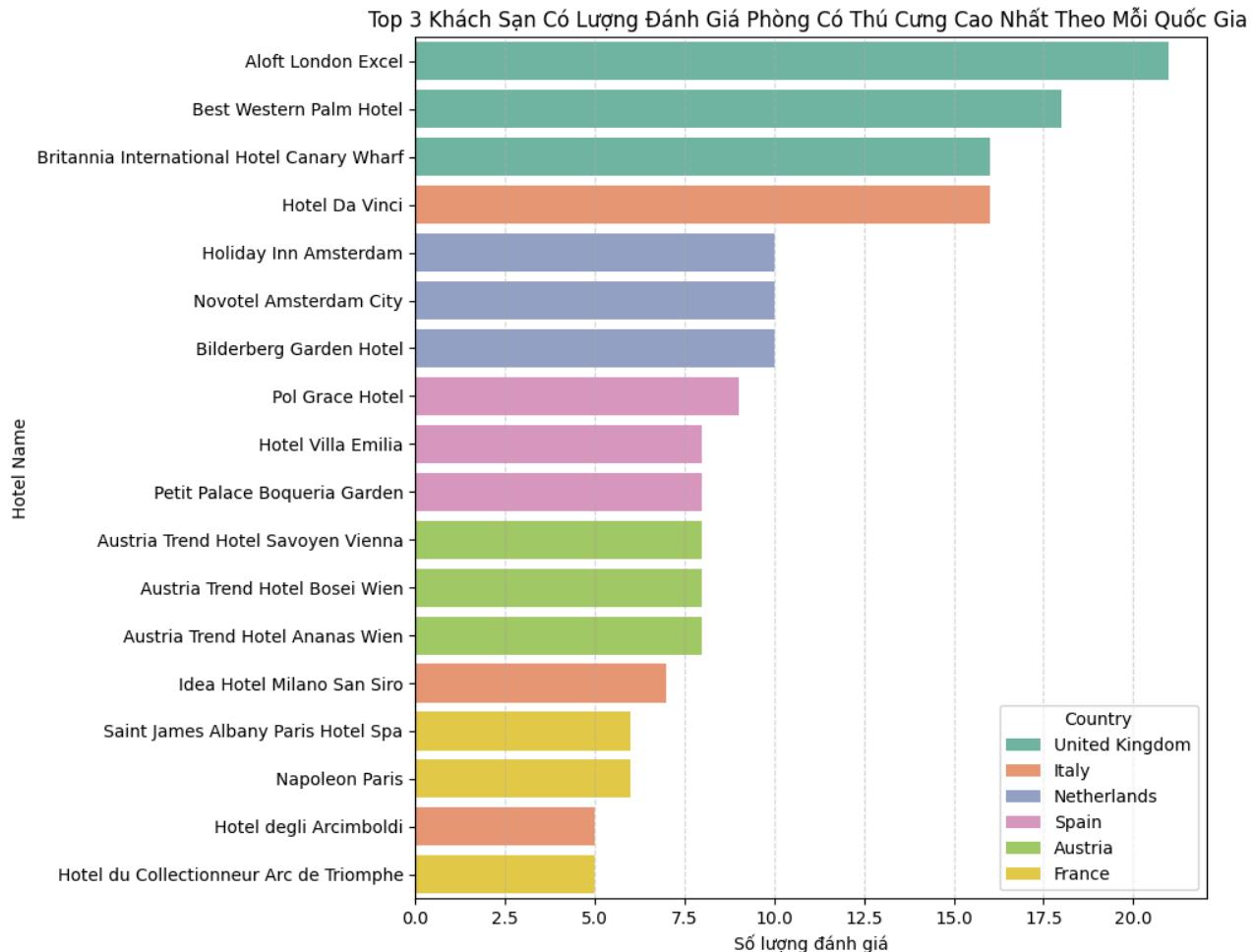
5.1.5. Top 3 khách sạn theo đánh giá ở mỗi quốc gia

Bảng xếp hạng này được dựa trên 2 tiêu chí là điểm đánh giá trung bình và số lượng đánh giá của mỗi khách sạn. Chỉ có các khách sạn có 700 lượt đánh giá trở lên mới được đưa vào bảng xếp hạng. Việc đặt ra ngưỡng tối thiểu cho số lượng đánh giá giúp loại bỏ các khách sạn cho điểm trung bình cao nhưng lượt đánh giá rất thấp. 700 là con số giúp đảm bảo lấy được top 3 khách sạn ở mỗi nước, nếu lấy ngưỡng cao hơn thì top 3 khách sạn ở một số nước sẽ không đáp ứng được tiêu chí này.

Quốc gia	Khách sạn	Điểm trung bình	Số đánh giá
Áo	Hotel Stefanie	9.2	884
Áo	Austria Trend Hotel Savoyen Vienna	8.9	713
Áo	Hotel de France Wien	8.6	956
Pháp	Pullman Paris Tour Eiffel	8.6	846
Pháp	Saint James Albany Paris Hotel Spa	8.0	700
Pháp	Mercure Paris Centre Tour Eiffel	7.9	770
Ý	Hotel Berna	9.3	1052
Ý	The Square Milano Duomo	9.1	898
Ý	Glam Milano	8.7	1335
Hà Lan	Hotel Esther a	9.3	1837
Hà Lan	Jaz Amsterdam	9.0	1157
Hà Lan	INK Hotel Amsterdam MGallery by Sofitel	9.0	876
Tây Ban Nha	Olivia Plaza Hotel	9.4	773
Tây Ban Nha	Hotel Barcelona Catedral	9.2	870
Tây Ban Nha	Ayre Hotel Gran V a	8.8	797
Anh	Intercontinental London The O2	9.5	2551
Anh	Shangri La Hotel at The Shard London	9.4	1113
Anh	Amber Hotel Charing Cross	9.4	785

Hình 5.1. Top 3 khách sạn ở mỗi quốc gia

5.1.6. Top 3 khách sạn có số lượng đánh giá cao nhất về dịch vụ cho phép mang thú cưng theo của từng quốc gia

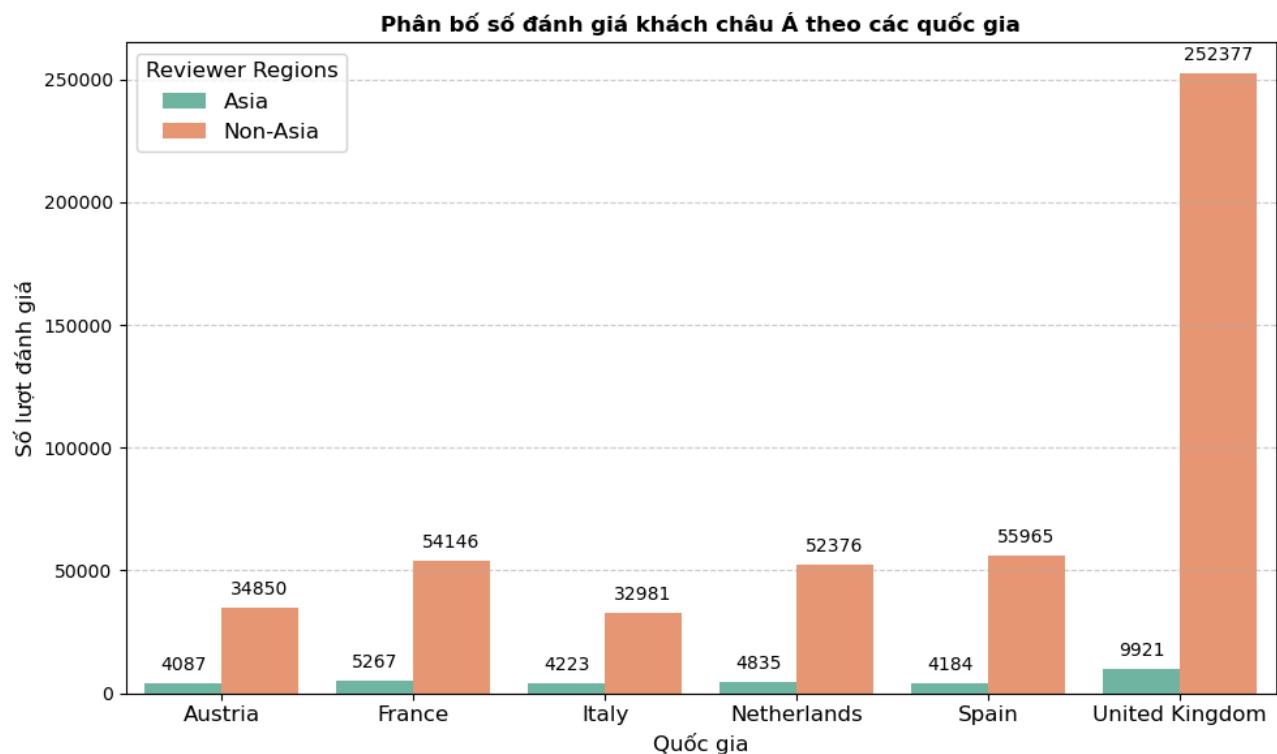


Biểu đồ 5.7. Top 3 khách sạn có số lượng đánh giá cao nhất về dịch vụ cho phép mang thú cưng theo của từng quốc gia

Biểu đồ thể hiện Top 3 khách sạn có số lượng đánh giá cao nhất về dịch vụ cho phép mang thú cưng theo của từng quốc gia. Các du khách khi đi du lịch có thể hoàn toàn yên tâm đặt phòng tại những khách sạn này cùng với thú cưng của mình.

5.2. Hành vi người châu Á khi đến các quốc gia châu Âu

5.2.1. Lượt đánh giá của khách châu Á theo các quốc gia

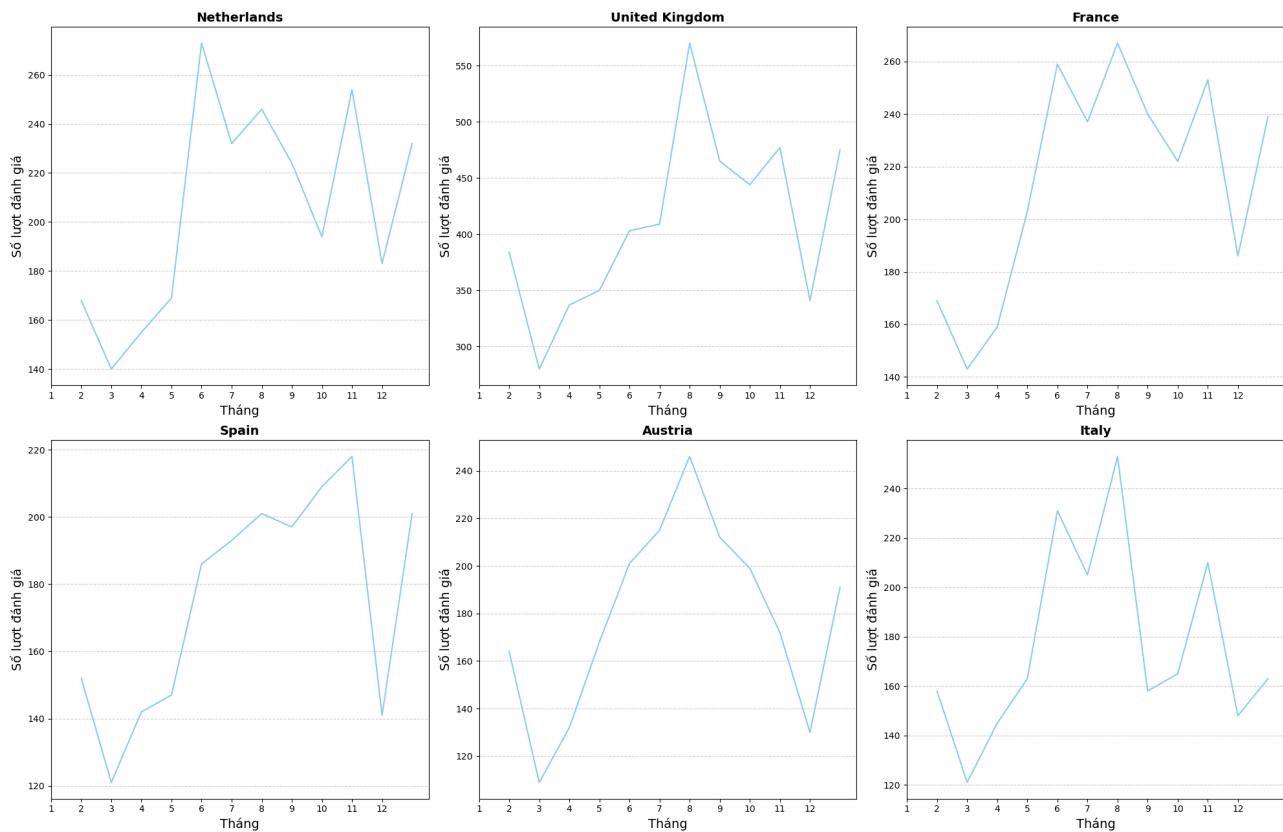


Biểu đồ 5.8. Số đánh giá khách Châu Á theo từng quốc gia

Dựa trên biểu đồ, có thể thấy Vương quốc Anh (United Kingdom) là điểm đến được người châu Á ưa chuộng nhất tại châu Âu, với số lượng đánh giá cao nhất (9,921). Tiếp theo là Pháp (France) với 5,267 đánh giá, cho thấy đây cũng là một điểm đến hấp dẫn nhờ vào sự nổi tiếng về văn hóa, nghệ thuật và thời trang. Các quốc gia khác như Áo (Austria), Hà Lan (Netherlands), Ý (Italy) và Tây Ban Nha (Spain) thu hút một lượng du khách châu Á đáng kể nhờ vào các điểm tham quan nổi tiếng và văn hóa đặc sắc.

5.2.2. Khoảng thời gian khách châu Á đánh giá khi đến các quốc gia năm 2016

Số lượt đánh giá của người Châu Á theo tháng cho từng quốc gia (Năm 2016)



Biểu đồ 5.9. Số đánh giá của người Châu Á theo tháng của từng quốc gia năm 2016

Thời gian từ tháng 5 đến tháng 8 là thời điểm cao điểm du lịch của người châu Á tới hầu hết các quốc gia. Trong khoảng thời gian này, số lượng đánh giá của họ thường cao hơn so với các tháng khác. Đặc biệt, tháng 6 và tháng 7 là các tháng được ưa chuộng nhất. Điều này có thể do thời tiết đẹp, nhiều sự kiện và lễ hội hấp dẫn diễn ra tại các quốc gia châu Âu trong mùa hè, khi đây là thời gian để đi nghỉ dưỡng và vui chơi.

Tháng 9 đến tháng 11 cũng là giai đoạn được nhiều người châu Á lựa chọn để du lịch châu Âu. Mặc dù số lượng đánh giá không nhiều bằng mùa hè, nhưng vẫn ở mức khá ổn định. Đây có thể là thời điểm ít đông đúc hơn so với mùa hè, tận hưởng các địa điểm du lịch với ít người hơn.

Ngược lại, tháng 12 đến tháng 3 là giai đoạn ít người châu Á đến thăm các quốc gia châu Âu. Số lượng đánh giá thường thấp hơn các thời điểm khác trong năm. Điều này có thể do thời tiết lạnh giá, ít sự kiện và không khí Giáng sinh đã qua tại các nước châu Âu. Đối với một số nước ở châu Á, đây còn là thời điểm để chuẩn bị cho Tết âm lịch hàng năm.

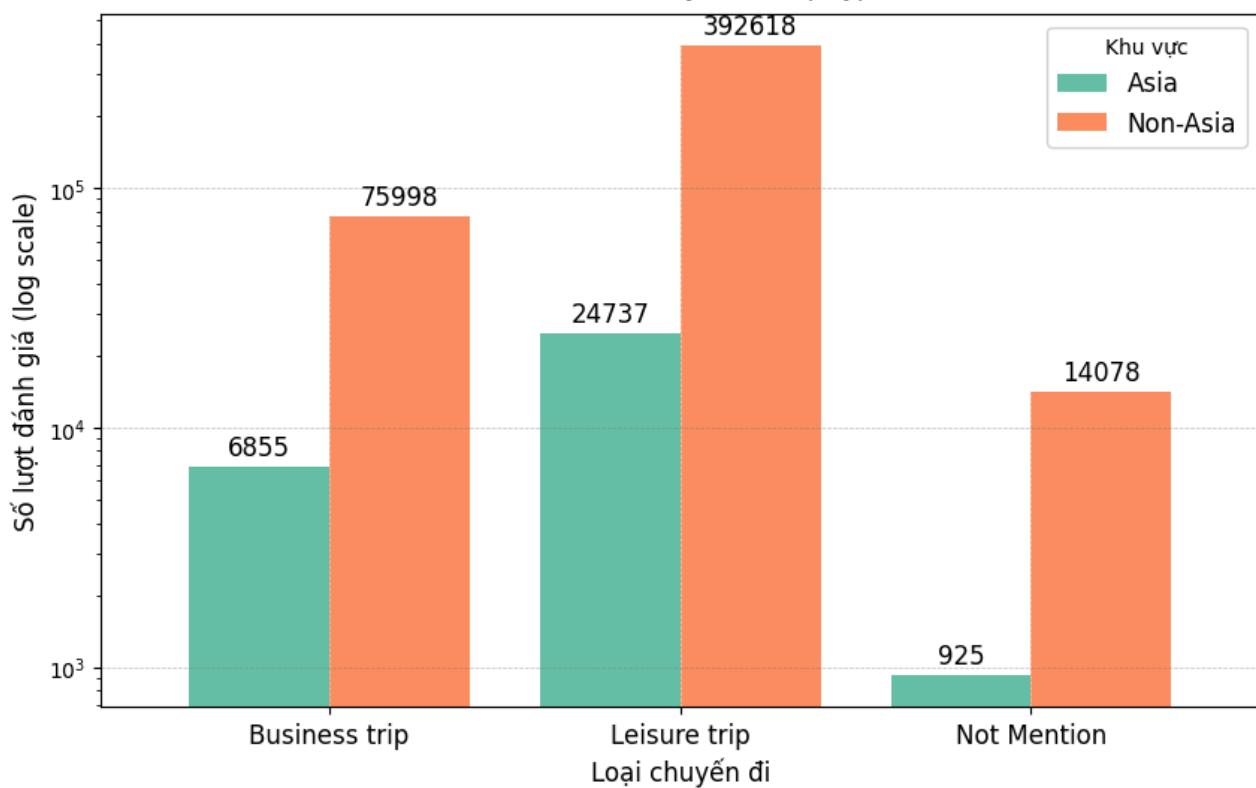
5.2.3. Phân tích loại hình chuyến đi của du khách châu Á và các khu vực khác

Nhằm hiểu rõ hơn về mục đích chuyến đi của các du khách khi đến châu Âu, nhóm lựa chọn biểu đồ bar chart để thể hiện số lượt đánh giá của du khách từ khu vực châu Á (Asia) và từ các khu vực khác (Non-Asia) theo từng loại hình chuyến đi (Trip Type), bao gồm: *Business Trip (Công tác)*, *Leisure Trip (Nghỉ dưỡng)* và *Not Mention (Không được đề cập)*.

Kết quả thu được từ *Biểu đồ 5.10* cho thấy:

- *Loại hình Business Trip (Công tác)* được du khách từ các khu vực khác (Non-Asia) ưu tiên lựa chọn (với 75,998 lượt), cao hơn đáng kể so với du khách từ châu Á (Asia) (6,855 lượt). Điều này thể hiện được du khách từ các khu vực khác ngoài châu Á (Non-Asia) có xu hướng đến các khách sạn cao cấp tại châu Âu trong những chuyến công tác.
- *Loại hình Leisure Trip (Nghỉ dưỡng)* là loại hình du lịch có số lượt đánh giá cao nhất từ cả hai khu vực, chỉ ra được phần lớn du khách đã đánh giá trải nghiệm dịch vụ của mình tại châu Âu đều với mục đích nghỉ dưỡng.
- *Không được đề cập (Not Mention)* có số lượt đánh giá thấp nhất trong 3 loại hình, đồng nghĩa với việc phần lớn du khách thường chỉ định rõ mục đích chuyến đi của họ khi đánh giá trải nghiệm, thay vì để trống thông tin.

Biểu đồ thể hiện số lượt đánh giá của Du khách từ Châu Á (Asia) và từ khu vực khác (Non-Asia) theo loại hình chuyến đi (Trip Type)



Biểu đồ 5.10. Biểu đồ thể hiện số lượt đánh giá của du khách Châu Á và không phải Châu Á theo loại hình chuyến đi (Trip Type)

Nhận xét tổng quan:

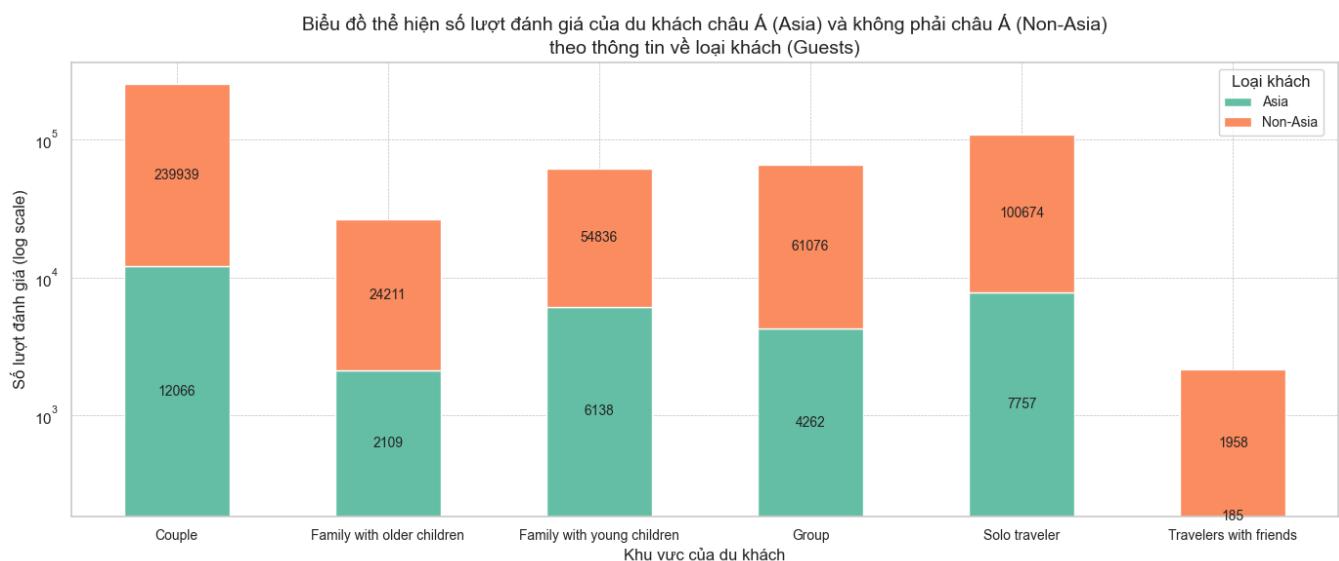
- *Dối với du khách từ châu Á:* Loại hình nghỉ dưỡng (Leisure Trip) vẫn là mục đích chính của du khách châu Á khi ghé thăm các nước tại châu Âu.

5.2.4. Phân tích thông tin về loại khách (Guests) của du khách châu Á (Asia) và không phải châu Á (Non-Asia)

Để hiểu rõ hơn về hành vi và thói quen của du khách châu Á khi lưu trú tại khách sạn trong khu vực châu Âu, dựa trên loại khách du lịch (Guests). Việc phân tích này nhằm làm rõ các nhóm đối tượng khách hàng chính của các khách sạn tại châu Âu, họ thường đi theo các diện như cặp đôi (Couple), đi một mình (Solo traveler), hay đi cùng gia đình (Family). Thông qua đó cung cấp thông tin cho các doanh nghiệp khách sạn để có thể cải thiện dịch vụ và xây dựng chiến lược tiếp cận khách hàng hiệu quả hơn. Đồng thời, kết quả cũng hỗ trợ du khách châu Á giúp dễ dàng tìm kiếm các khách sạn phù hợp hơn với nhu cầu của bản thân.

Phân tích *Biểu đồ 5.11* cho thấy khách du lịch đến từ châu Á chủ yếu đi du lịch đến các nước châu Âu theo dạng cặp đôi (Couple) và du lịch một mình (Solo traveler), vì đây là

hai nhóm có số lượt đánh giá cao nhất từ du khách châu Á với kết quả lần lượt là 12,066 lượt và 7,757 lượt đánh giá. Điều này cho thấy xu hướng du lịch theo cặp đôi và du lịch cá nhân của du khách châu Á khi đến châu Âu, mục đích của chuyến đi của họ có thể là đi hưởng tuần trăng mật, hay đi phượt một mình. Trong khi đó, du khách châu Á ít khi đi du lịch theo dạng gia đình có con lớn (*Family with older children*) (2,109 lượt đánh giá) hay đi du lịch theo nhóm (*Group*) (4,262 lượt đánh giá).



Biểu đồ 5.11. Biểu đồ thể hiện số lượt đánh giá của du khách châu Á (Asia) và không phải châu Á (Non-Asia) theo thông tin về loại khách (Guests)

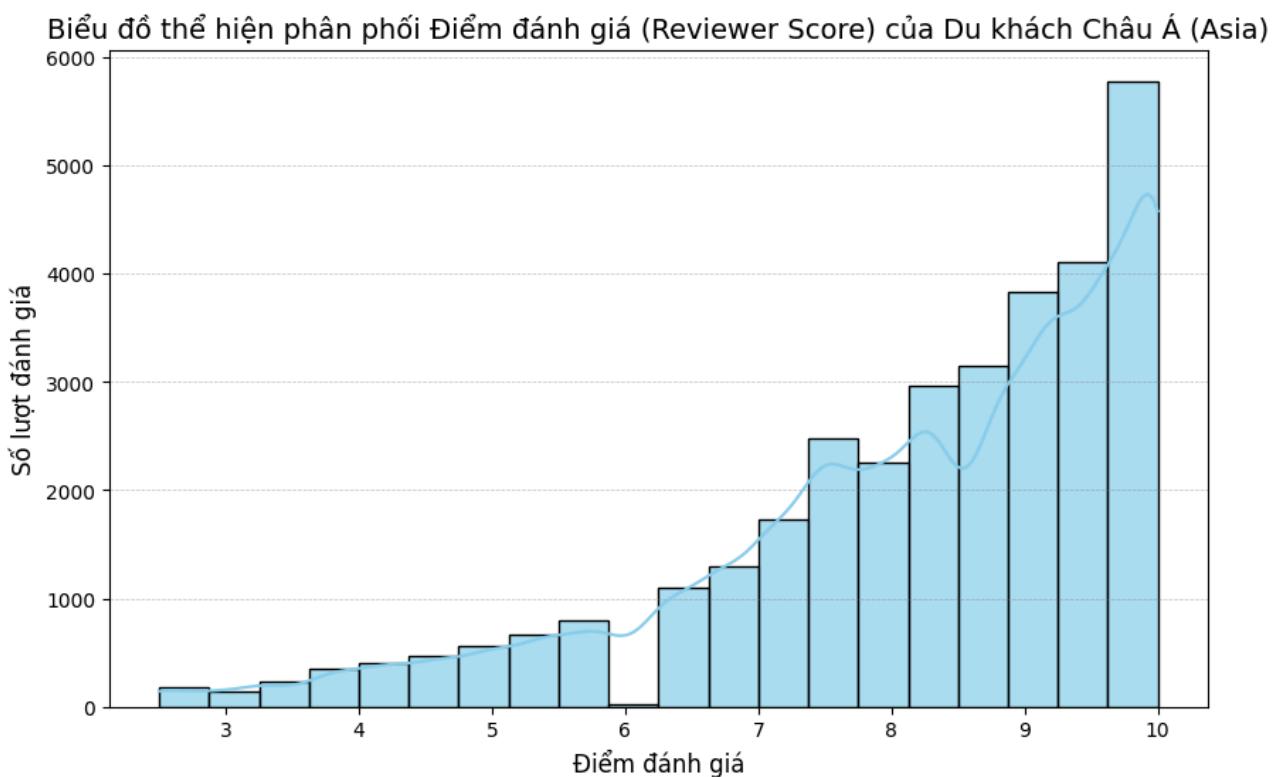
5.2.5. Histogram điểm đánh giá của người châu Á

Nhóm sử dụng biểu đồ histogram để đánh giá mối quan hệ giữa điểm đánh giá của du khách (*Reviewer_Score*) và vùng xuất xứ của du khách (*Reviewer_Regions*). Thông qua đó, nhận diện được sự khác biệt trong hành vi và xu hướng cho điểm đánh giá của du khách từ các khu vực khác nhau sau khi trải nghiệm dịch vụ tại các khách sạn trong khu vực châu Á.

a) Xu hướng đánh giá của du khách châu Á (Asia)

Biểu đồ 5.12 cho thấy phần lớn điểm đánh giá của du khách châu Á tập trung từ mức 6 đến 10, với đỉnh cao nhất ở mức 10. Cụ thể, số lượt đánh giá cho điểm ở mức cao (9 - 10) vượt trội hơn hẳn so với các mức điểm thấp (3 - 6). Điều này cho thấy nhóm du khách từ châu Á có xu hướng đánh giá tích cực, và rất ít khi đưa ra nhận xét tiêu cực.

- Số lượt đánh giá bắt đầu thấp ở khoảng điểm từ 3 đến 6, dưới 1000 lượt đánh giá tại mỗi mức điểm.
- Từ điểm 6 trở đi, số lượt đánh giá tăng rõ rệt và đạt đỉnh tại mức điểm 10, với gần 6000 lượt đánh giá.



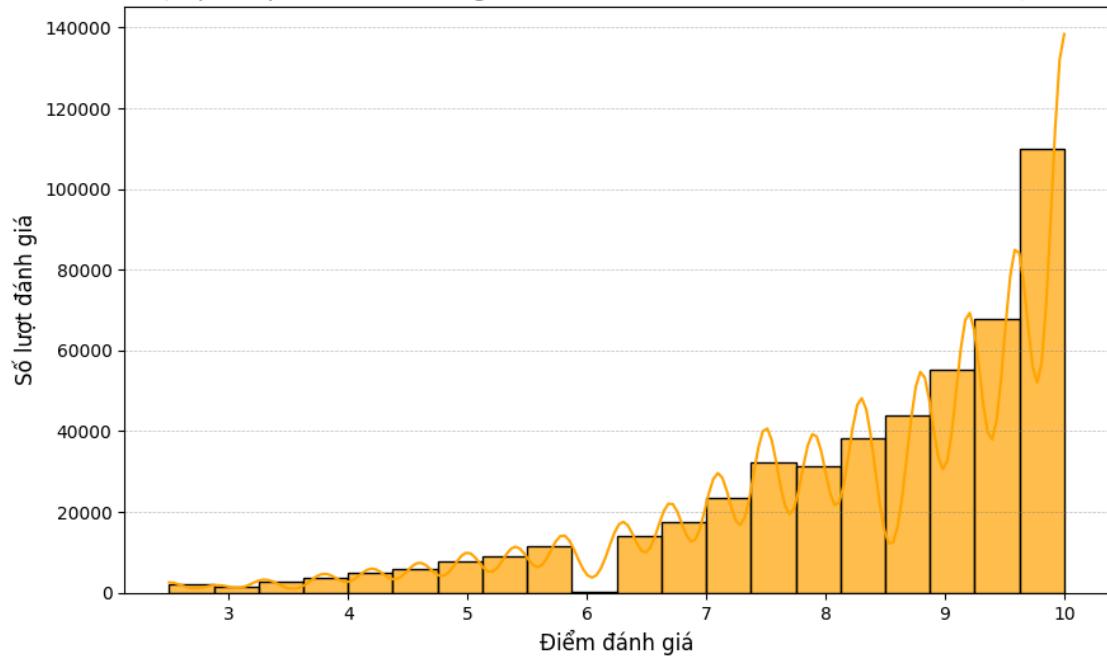
Biểu đồ 5.12. Biểu đồ thể hiện phân phối Điểm đánh giá (Reviewer Score) của Du khách Châu Á (Asia)

b) Xu hướng đánh giá của du khách không phải là người châu Á (Non-Asia)

Biểu đồ 5.13 cho thấy số lượt đánh giá của nhóm du khách từ các vùng không phải châu Á (*Non-Asia*) trải rộng hơn từ mức điểm 3 đến điểm 10, và dao động mạnh mẽ ở các

mức trung bình (6 - 8). Điểm đỉnh cũng nằm ở mức 10, nhưng sự biến động ở các mức thấp và trung bình (3 - 8) lớn hơn nhiều so với nhóm du khách từ châu Á.

Biểu đồ thể hiện phân phối Điểm đánh giá (Reviewer Score) của Du khách từ khu vực khác (Non-Asia)



Biểu đồ 5.13. Biểu đồ thể hiện phân phối Điểm đánh giá (Reviewer Score) của Du khách từ khu vực khác (Non-Asia)

Nhận xét tổng quan:

- Cả hai nhóm du khách đều có xu hướng đánh giá tập trung vào các mức điểm cao (8 - 10), cho thấy trải nghiệm của họ tại các khách sạn ở châu Âu nhìn chung là khá tốt, đáp ứng được đầy đủ kỳ vọng của du khách.
- Mức điểm 10 chiếm tỷ lệ cao nhất trong cả hai biểu đồ, điều này thể hiện được phần lớn du khách, bất kể khu vực nào để cảm thấy hài lòng hoặc vượt mong đợi với trải nghiệm tại các khách sạn ở khu vực châu Âu

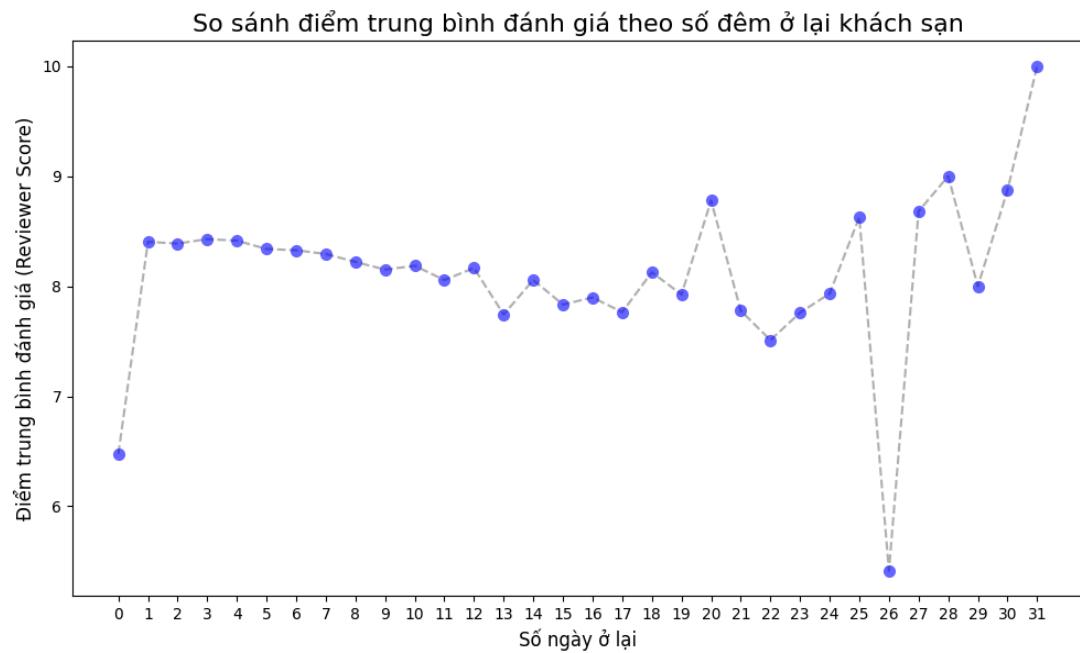
5.2.6. Top 3 khách sạn ở mỗi quốc gia theo đánh giá của người châu Á

Bảng xếp hạng này được dựa trên 2 tiêu chí là điểm đánh giá trung bình và số lượng đánh giá của khách là người châu Á. Chỉ có các khách sạn có 150 lượt đánh giá trở lên mới được đưa vào bảng xếp hạng. Việc đặt ra ngưỡng tối thiểu cho số lượng đánh giá giúp loại bỏ các khách sạn cho điểm trung bình cao nhưng lượt đánh giá rất thấp. 150 là con số giúp đảm bảo lấy được top 3 khách sạn ở mỗi nước, nếu lấy ngưỡng cao hơn thì top 3 khách sạn ở một số nước sẽ không đáp ứng được tiêu chí này.

Quốc gia	Khách sạn	Điểm trung bình	Số đánh giá
Áo	Austria Trend Hotel Europa Wien	8.6	151
Áo	Hotel de France Wien	8.2	168
Áo	Hilton Vienna	7.6	166
Pháp	Pullman Paris Tour Eiffel	8.3	179
Pháp	Warwick Paris Former Warwick Champs Elysees	8.0	248
Pháp	H tel California Champs Elys es	7.9	150
Ý	Room Mate Giulia	9.2	206
Ý	Hotel Berna	9.2	150
Ý	The Square Milano Duomo	8.8	265
Hà Lan	art otel Amsterdam	8.9	173
Hà Lan	Urban Lodge Hotel	8.7	151
Hà Lan	The Student Hotel Amsterdam City	8.6	227
Tây Ban Nha	Olivia Plaza Hotel	9.3	219
Tây Ban Nha	Majestic Hotel Spa Barcelona GL	8.7	155
Tây Ban Nha	Eurostars Grand Marina Hotel GL	8.4	290
Anh	The Montcalm Marble Arch	8.7	366
Anh	Amber Hotel Marble Arch	8.3	572
Anh	Royal Garden Hotel	8.3	200

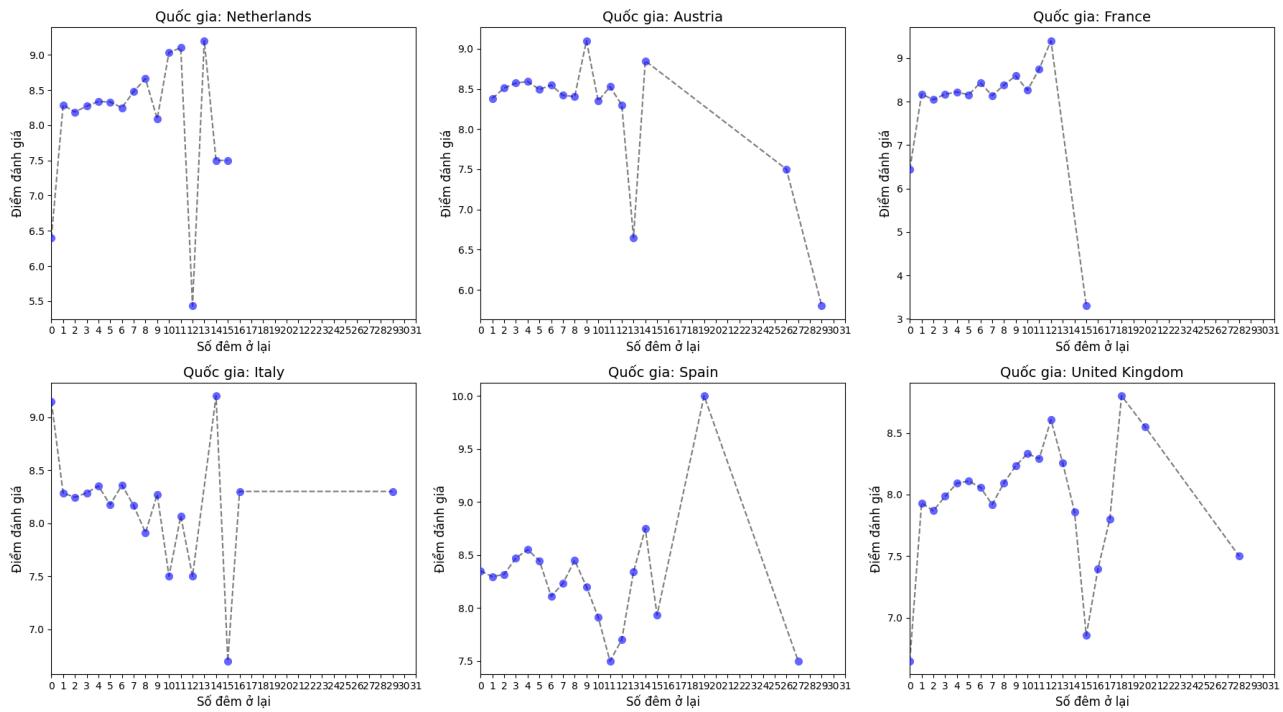
Hình 5.2. Top 3 khách sạn ở mỗi quốc gia đối với người châu Á

5.2.7. Tương quan giữa điểm trung bình và số đêm ở lại



Biểu đồ 5.14. So sánh điểm đánh giá trung bình dựa trên số đêm khách hàng ở lại khách sạn

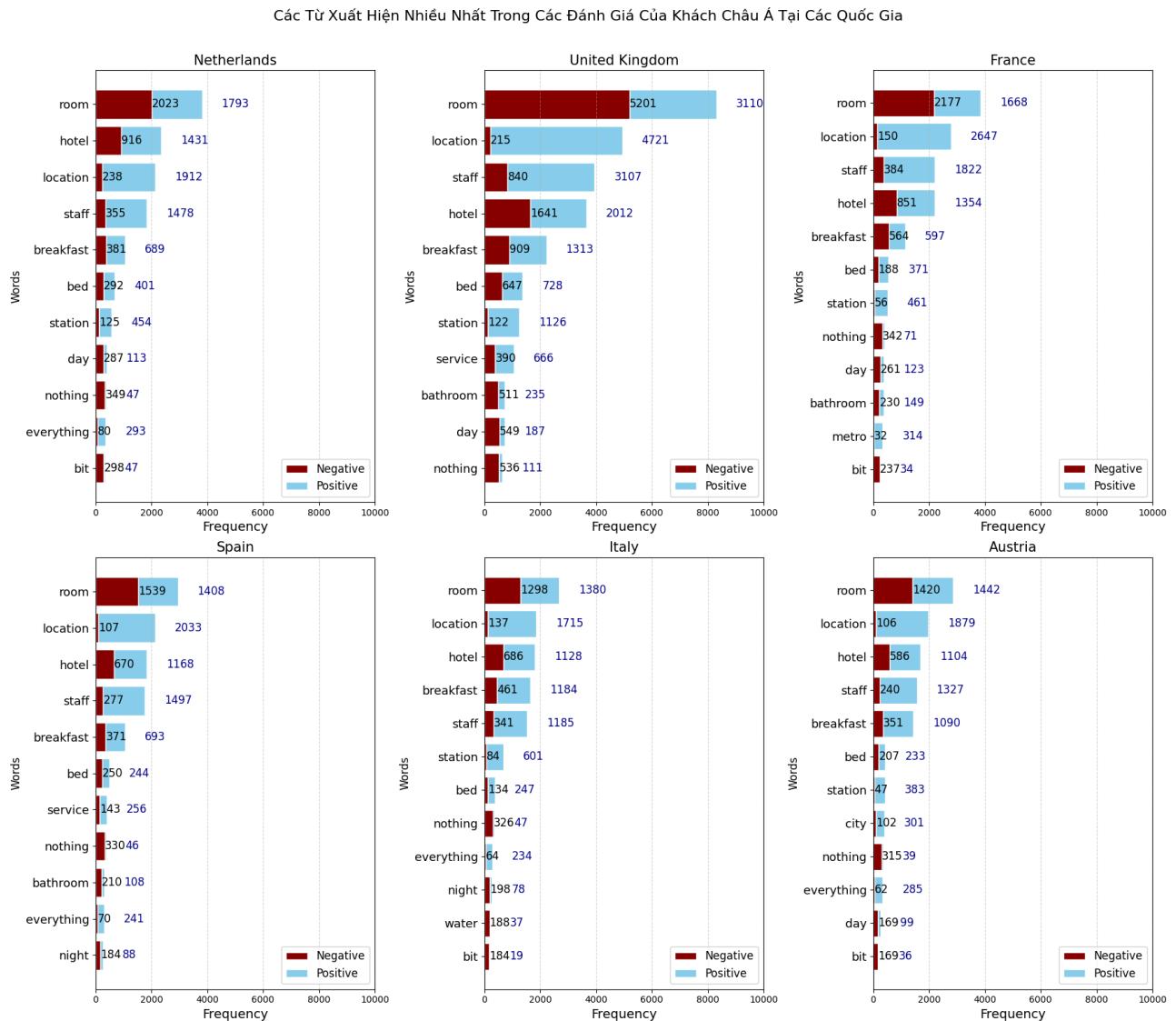
Dựa vào *Biểu đồ 5.14*, nhóm nhận thấy khách hàng có điểm đánh giá ổn định đối với số đêm ở lại khách sạn 1 - 10 ngày. Đối với các chuyến đi trong ngày (không qua đêm), điểm số khá thấp, chỉ ở mức trên 6 điểm. Các chuyến đi trên 10 đêm có điểm đánh giá trung bình không ổn định, nhất là đối với các chuyến đi 26 đêm, điểm đánh giá thấp bão động, dưới 6 điểm. Những chuyến đi ở lại 31 đêm cho điểm gần như tuyệt đối.



Biểu đồ 5.15. So sánh điểm đánh giá trung bình dựa trên số đêm khách hàng Châu Á ở lại khách sạn theo từng quốc gia

Dựa vào *Biểu đồ 5.15*, nhóm nhận thấy có sự khác biệt rõ ràng giữa số đêm ở lại của các khách Châu Á. Đối với các quốc gia như Hà Lan, Pháp, không có chuyến đi nào quá dài, trên 17 đêm. Nhìn chung, người Châu Á không thường ở lại quá lâu (trên 17 ngày) tại các khách sạn ở 6 nước trên. Điểm đánh giá cũng có sự dao động mạnh, đối với các khách Châu Á tại Hà Lan và Anh, họ đánh giá rất thấp đối với các chuyến đi ở lại trong ngày (không qua đêm). Dường như các nước này đều có xu hướng chung các chuyến đi của người Châu Á dài trên 17 ngày có điểm đánh giá khá thấp.

5.2.8. Những từ phổ biến nhất trong các đánh giá của khách châu Á cho từng quốc gia



Biểu đồ 5.16. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại các quốc gia

Có những yếu tố chung nổi bật trong các đánh giá của khách du lịch châu Á, bao gồm phòng (room), vị trí (location), khách sạn (hotel), nhân viên (staff), bữa sáng (breakfast) và nhà ga (station), những yếu tố này cho thấy đây là mối quan tâm hàng đầu của khách du lịch khi lựa chọn nơi lưu trú. Tại Pháp, sự xuất hiện của từ "metro" (tàu điện ngầm) đặc biệt chỉ ra rằng hệ thống giao thông công cộng là yếu tố quan trọng trong trải nghiệm của du khách.

Ở Anh và Tây Ban Nha, từ "dịch vụ" (service) được đề cập nhiều, cho thấy chất lượng phục vụ của nhân viên là yếu tố then chốt ảnh hưởng đến sự hài lòng của khách châu

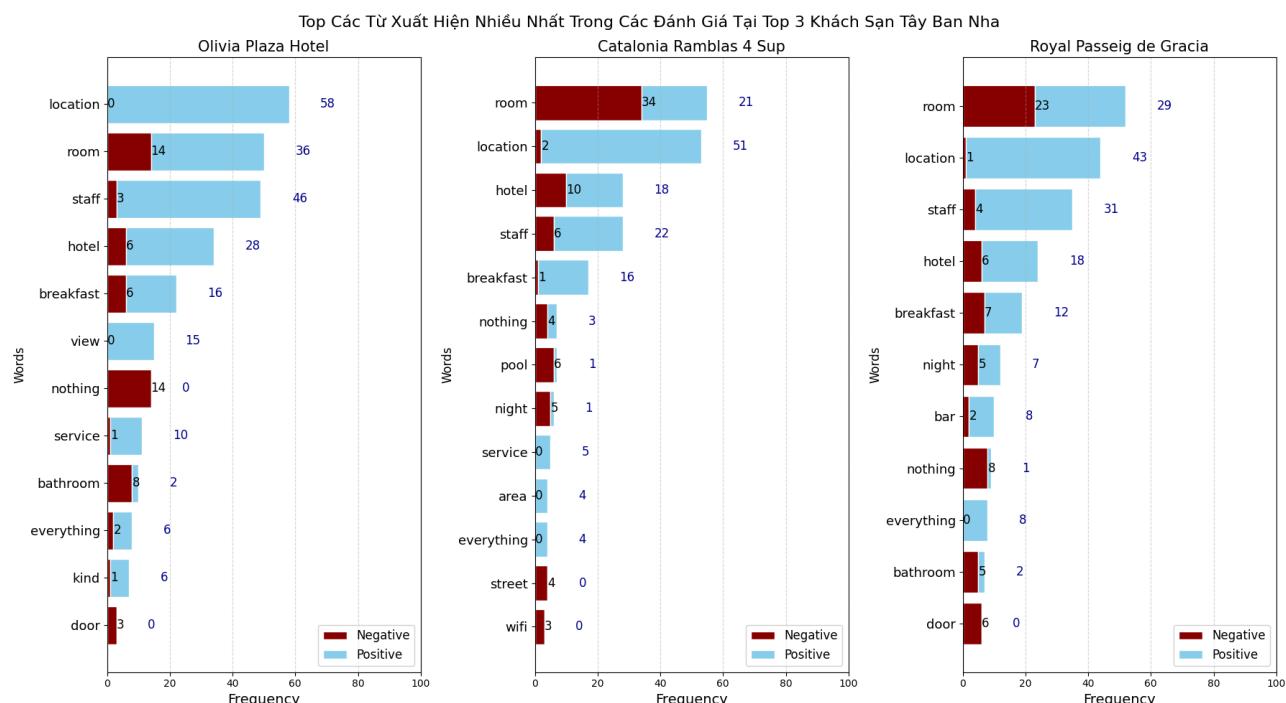
Á. Trong khi đó, tại Ý, các yếu tố cơ sở vật chất như "nước" (water) lại nhắc lại nhiều lần trong các đánh giá tiêu cực, phản ánh sự quan tâm đến các vấn đề liên quan đến nước.

Tại Vương Quốc Anh, Pháp, Hà Lan và Áo, từ ngày (day) có thể được sử dụng để miêu tả trải nghiệm trong ngày vì các quốc gia này nổi tiếng với các thành phố mang tính lịch sử và văn hóa, nơi du khách thường dành cả ngày khám phá các địa điểm tham quan hoặc danh lam thắng cảnh.

Ngược lại, ở Ý và Tây Ban Nha đều nổi tiếng với các trải nghiệm ban đêm như ăn uống, tiệc tùng, điều này có thể khiến từ đêm (night) trở nên quan trọng hơn trong đánh giá của họ thay vì ngày (day).

5.2.9. Những từ phổ biến nhất trong các đánh giá tại 3 khách sạn có nhiều đánh giá nhất theo từng quốc gia của khách châu Á

Với mỗi quốc gia, chọn ra 3 khách sạn có nhiều đánh giá của khách châu Á nhất, sau đó tính tần số xuất hiện của các từ trong các đánh giá tích cực và đánh giá tiêu cực.



Biểu đồ 5.17. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Tây Ban Nha

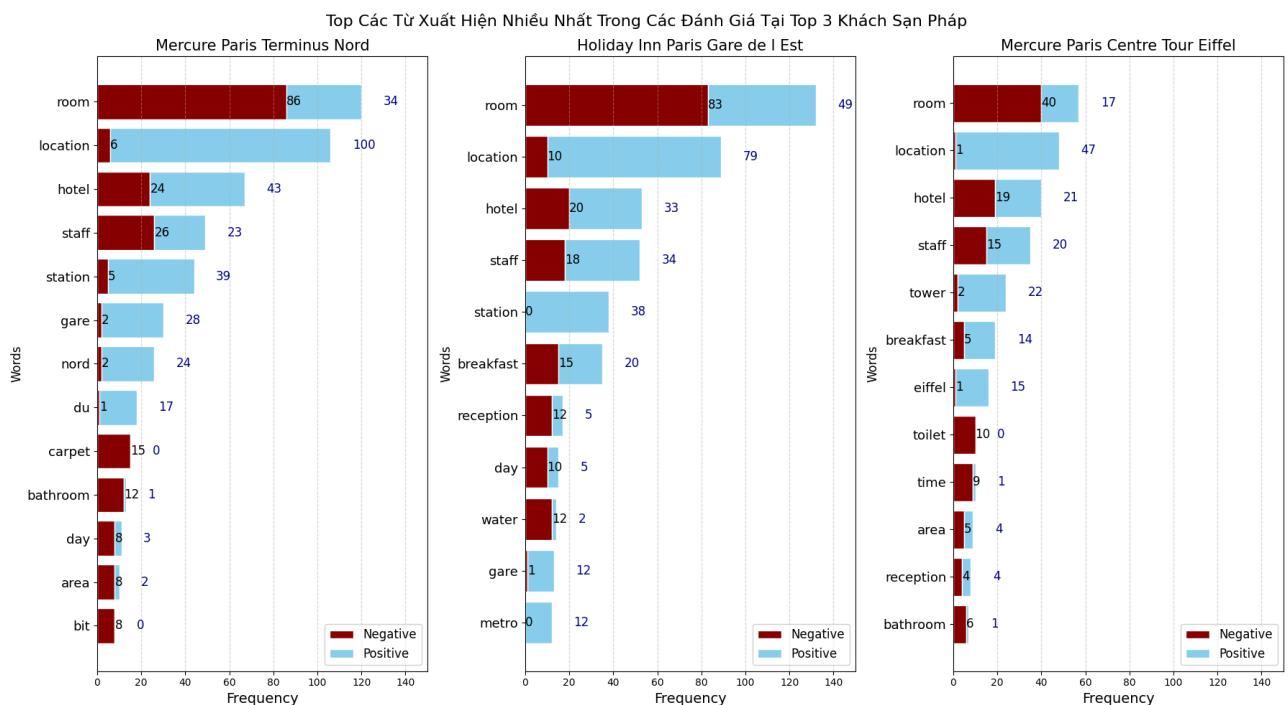
Biểu đồ thể hiện ba khách sạn nhận được nhiều lượt đánh giá nhất từ du khách châu Á tại Tây Ban Nha. Trong đó:

Olivia Plaza Hotel nhận được điểm trung bình 9.49 từ 86 du khách châu Á. Dựa trên các từ khóa xuất hiện trong các đánh giá, khách sạn này nổi bật với vị trí (location) và tầm nhìn (view) đẹp, dịch vụ (service) và nhân viên (staff) thân thiện, rất phù hợp với sở thích của du khách châu Á. Tuy nhiên, cơ sở vật chất như phòng (room), phòng tắm (bathroom) và cửa (door) là những yếu tố được nhắc đến nhiều trong các đánh giá tiêu cực.

Catalonia Ramblas 4 Sup đạt điểm 8.46 từ 78 đánh giá. Khách sạn này cũng được khen ngợi về vị trí, không gian (area), dịch vụ và nhân viên tốt, đặc biệt là bữa sáng (breakfast) phù hợp với khẩu vị của du khách châu Á. Tuy nhiên, các dịch vụ phụ như hồ bơi (pool), dịch vụ ban đêm (night) và wifi đang nhận nhiều phản hồi tiêu cực.

Royal Passeig de Gracia đạt điểm 8.58 từ 75 du khách. Khách sạn này có vị trí và nhân viên tốt, nhưng các yếu tố về cơ sở vật chất và dịch vụ khác nhận được đánh giá ở mức khá.

Du khách châu Á muốn du lịch tại Tây Ban Nha có thể cân nhắc chọn Olivia Plaza Hotel hoặc Catalonia Ramblas 4 Sup nếu ưu tiên tiện nghi và dịch vụ tốt.

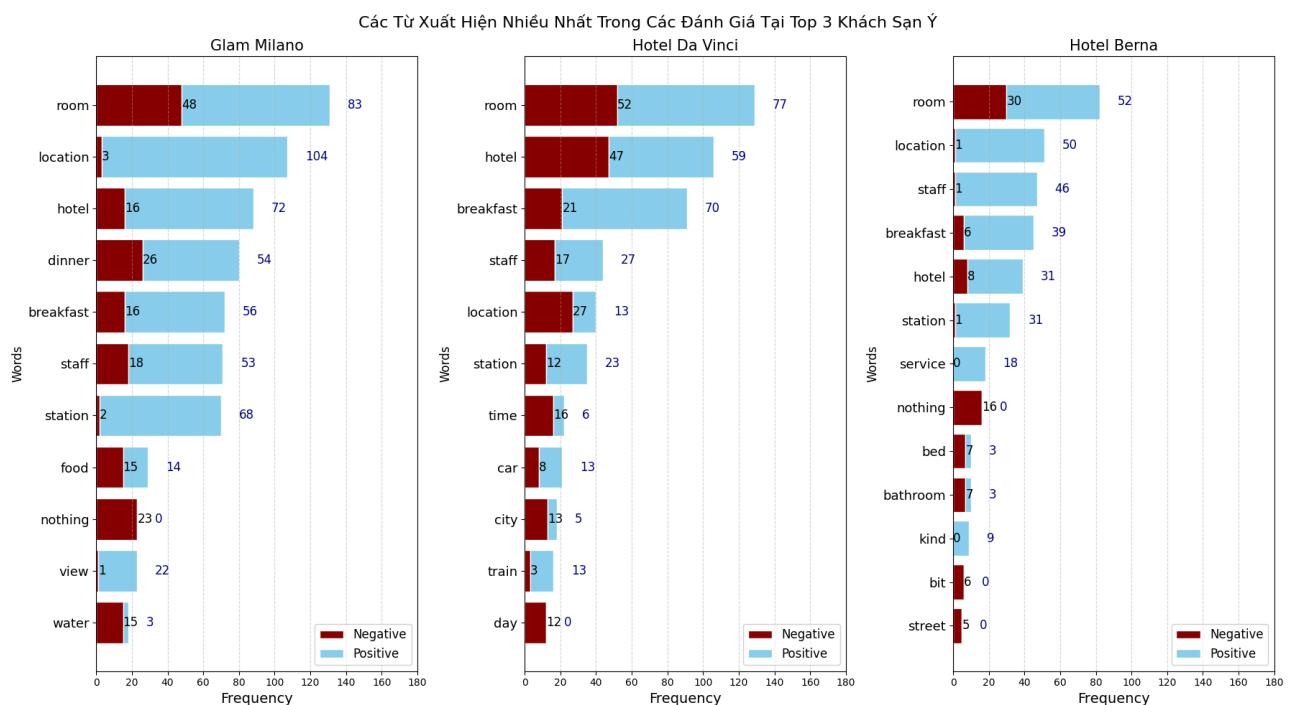


Biểu đồ 5.18. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Pháp

Khách sạn Mercure Paris Terminus Nord nhận điểm trung bình 7.36 từ 188 đánh giá của du khách châu Á. Các đánh giá tiêu cực chủ yếu nhắm vào cơ sở vật chất của khách sạn, nhiều hơn là các đánh giá tích cực. Mặc dù vậy, vị trí và sự thuận tiện trong di chuyển, đặc biệt là gần ga tàu (station, gare), là những yếu tố quan trọng cần cân nhắc khi lựa chọn khách sạn này.

Holiday Inn Paris Gare de l'Est đạt điểm 7.75 từ 180 đánh giá. Các vấn đề về cơ sở vật chất và một số yếu tố quan trọng như nước (water), thanh toán (reception) và ngày (day) đã được nhắc đến trong các đánh giá tiêu cực, cho thấy du khách châu Á không hoàn toàn hài lòng với khách sạn này. Tuy nhiên, các đánh giá về yếu tố di chuyển, bao gồm ga tàu (station, gare) và metro, chủ yếu là tích cực.

Mercure Paris Centre Tour Eiffel, với điểm 7.67 từ 90 đánh giá, nằm ở vị trí thuận lợi và cung cấp tầm nhìn (view) ra tháp Eiffel. Nếu du khách châu Á muốn lựa chọn khách sạn này, họ cần xem xét kỹ các yếu tố khác như toilet, thời gian (time) và thanh toán.

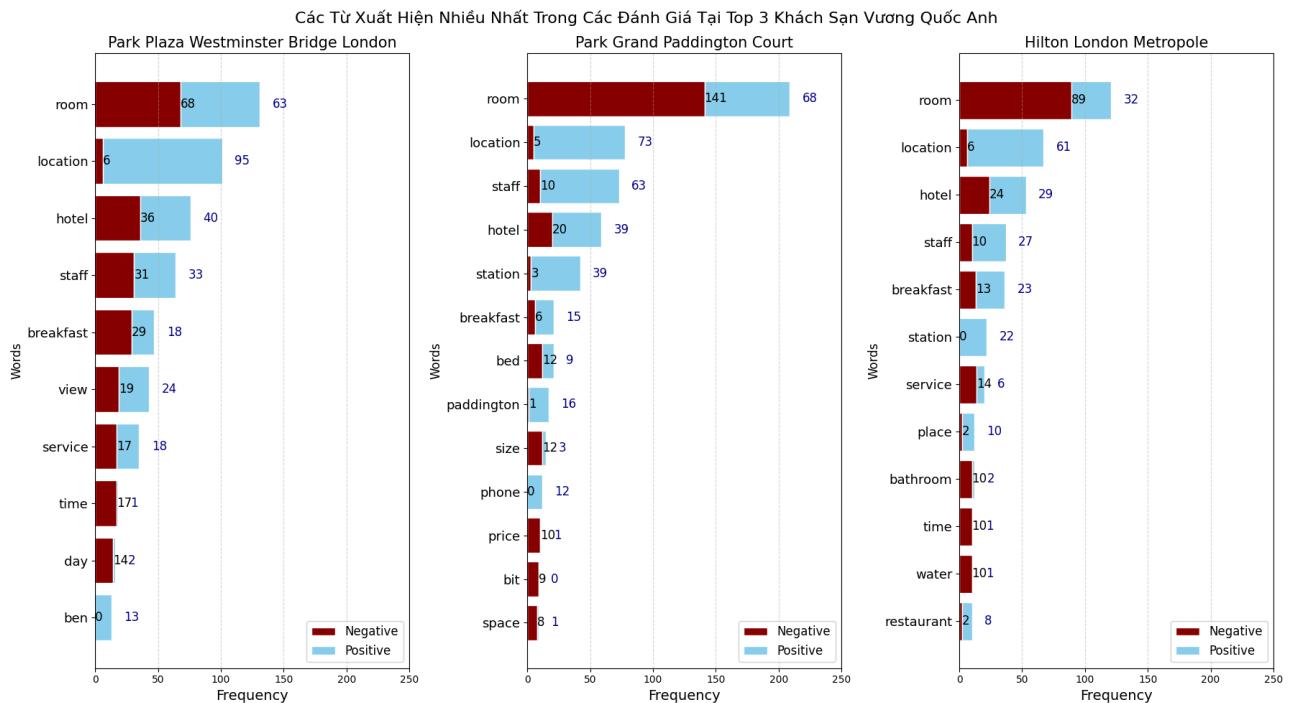


Biểu đồ 5.19. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Ý

Khách du lịch châu Á có thể cân nhắc lựa chọn Hotel Berna (9.18, 114 đánh giá) làm điểm dừng chân khi đến Ý. Với vị trí thuận tiện, gần các điểm tham quan và các phương tiện giao thông công cộng, cùng dịch vụ thân thiện, khách sạn này thực sự là một sự lựa chọn lý tưởng cho những ai muốn khám phá thành phố.

Một lựa chọn khác cũng rất đáng xem xét là Glam Milano (8.77, 196 đánh giá). Với vị trí lý tưởng gần ga tàu và tầm nhìn đẹp, khách sạn này mang lại cảm giác dễ chịu cho du khách. Tuy nhiên, một vấn đề cần lưu ý là chất lượng nước tại khách sạn, điều này có thể gây khó chịu cho đa số khách hàng châu Á, đặc biệt là đối với những người du lịch dài ngày.

Còn Hotel Da Vinci (7.73, 169 đánh giá), khách sạn này hầu như không nổi bật với bất kỳ đặc điểm đặc trưng nào.



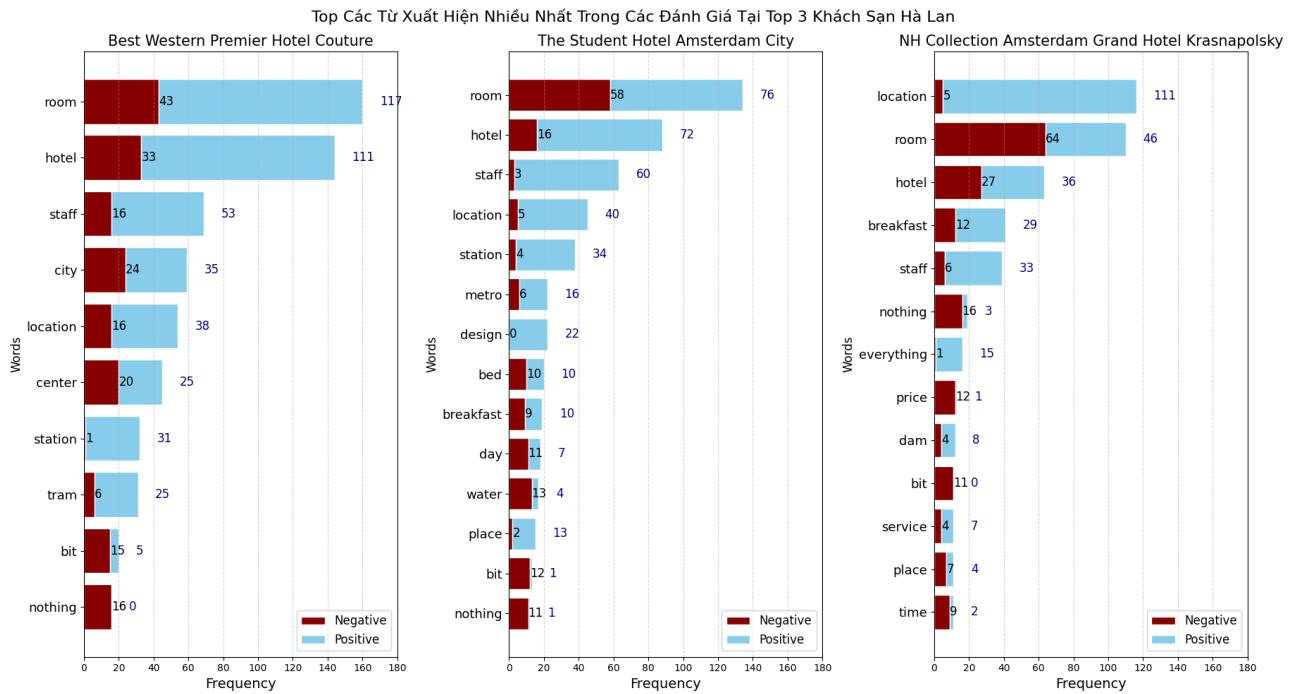
Biểu đồ 5.20. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Vương Quốc Anh

Park Plaza Westminster Bridge London (điểm 8.08 từ 183 đánh giá) có vị trí rất thuận tiện, nhưng các yếu tố khác như chất lượng phòng và dịch vụ không có gì nổi bật.

Park Grand Paddington Court (điểm 7.69 từ 172 đánh giá) có lợi thế về vị trí, gần các trạm tàu, thuận tiện cho việc di chuyển. Tuy nhiên, giá cả, chất lượng phòng và không gian (space) lại không đáp ứng được kỳ vọng.

Hilton London Metropole (điểm 7.11 từ 165 đánh giá) cũng gặp phải vấn đề tương tự, với cơ sở vật chất chỉ ở mức trung bình. Đặc biệt, chất lượng nước và tình trạng phòng ốc là những điểm được phê bình khá nhiều trong các đánh giá, cho thấy khách sạn này không thể đáp ứng được nhu cầu của du khách châu Á.

Cá ba khách sạn này đều không phải là những lựa chọn tốt cho du khách châu Á khi du lịch tại Vương Quốc Anh, dù vị trí có thể thuận lợi, nhưng chất lượng dịch vụ và cơ sở vật chất đều có nhiều hạn chế.

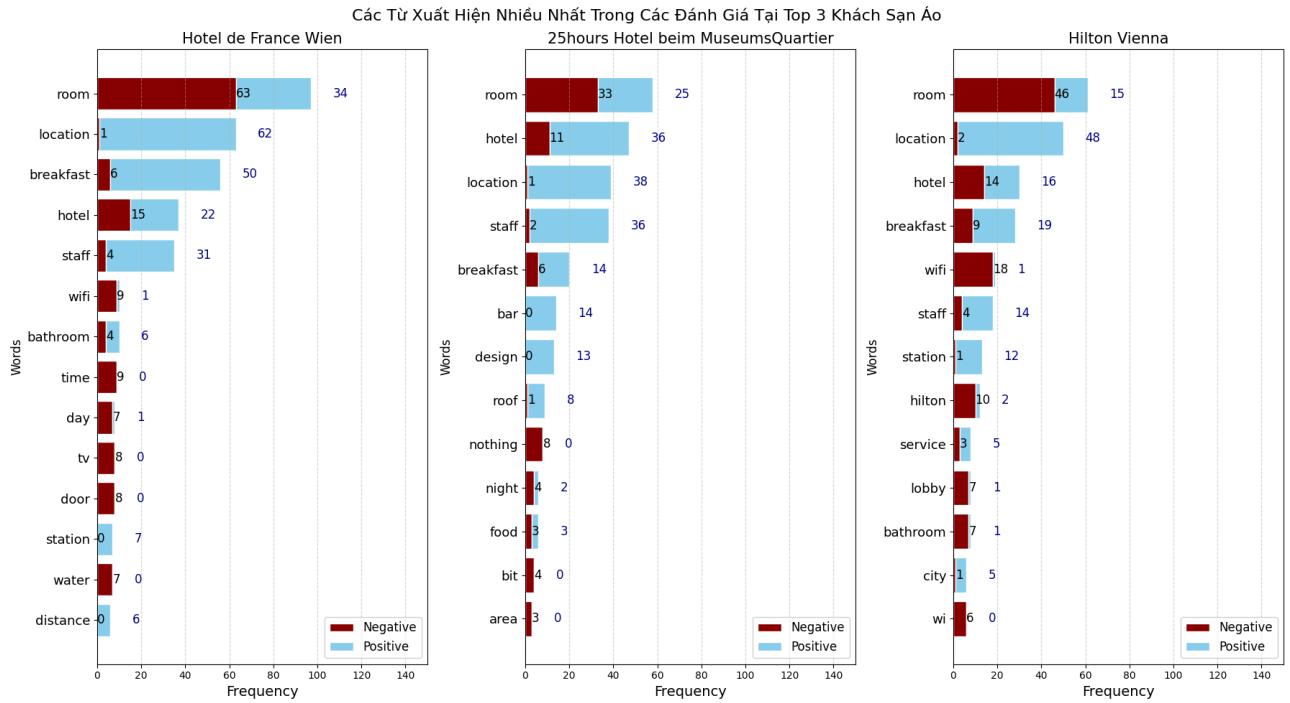


Biểu đồ 5.21. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Hà Lan

Best Western Premier Hotel Couture (điểm 8.56 từ 188 đánh giá) là một lựa chọn tốt với phần lớn các đánh giá tích cực. Những từ khóa như dịch vụ, vị trí thuận tiện được nhắc đến nhiều trong các phản hồi tích cực. Từ nothing (không có gì) là từ xuất hiện trong các đánh giá tiêu cực, điều này cho thấy các du khách cảm thấy hài lòng và không có điều gì để phàn nàn.

The Student Hotel Amsterdam City (điểm 8.72 từ 165 đánh giá) là một lựa chọn thú vị nhờ vào thiết kế (design) và vị trí. Tuy nhiên, một số đánh giá chỉ ra rằng chất lượng nước khá kém và cơ sở vật chất chỉ ở mức trung bình, điều này có thể khiến một số du khách cảm thấy không hài lòng, đặc biệt đối với những ai coi trọng các yếu tố tiện nghi cơ bản.

NH Collection Amsterdam Grand Hotel Krasnapolsky (điểm 8.53 từ 164 đánh giá) có vị trí đẹp là điểm cộng lớn đối với du khách. Tuy nhiên, các đánh giá về giá cả và thời gian chủ yếu mang tính tiêu cực, cho thấy khách sạn này có thể không mang lại giá trị tốt so với chi phí bỏ ra, có thể làm du khách cảm thấy không thoải mái với những chi phí không tương xứng với chất lượng dịch vụ.



Biểu đồ 5.22. Các từ xuất hiện nhiều nhất trong các đánh giá của người châu Á tại Áo

Hotel de France Wien (điểm 8.53 từ 106 đánh giá), 25hours Hotel beim MuseumsQuartier (điểm 9.1 từ 88 đánh giá) và Hilton Vienna (điểm 8.21 từ 87 đánh giá) đều có vị trí thuận tiện, tuy nhiên, chất lượng phòng ốc lại là yếu tố được đề cập nhiều trong các đánh giá tiêu cực hơn là tích cực, cho thấy các khách sạn này chưa hoàn toàn đáp ứng được kỳ vọng về tiện nghi và không gian phòng.

Trong ba khách sạn, 25hours Hotel beim MuseumsQuartier nổi bật hơn một chút nhờ vào yếu tố thiết kế và dịch vụ bar được khen ngợi nhiều trong các đánh giá, đây sẽ là một lựa chọn đáng cân nhắc cho du khách châu Á khi dừng chân tại Áo.

CHƯƠNG 6 - TỔNG KẾT ĐỀ TÀI

Báo cáo cung cấp cái nhìn tổng quan về các đánh giá của du khách khi lưu trú tại các khách sạn ở châu Âu, bao gồm những quốc gia mà du khách đến thăm nhiều nhất, hình thức du lịch phổ biến, nhóm du khách chủ yếu, thời điểm du lịch, và các khách sạn cho phép mang thú cưng. Theo đó, báo cáo phân tích sự phổ biến của các quốc gia châu Âu đối với du khách như là Anh, Pháp, Tây Ban Nha... Du khách chủ yếu đi du lịch theo hình thức đi nghỉ dưỡng, công tác và thường đi theo các nhóm gia đình, bạn bè, cặp đôi, đi một mình. Thời gian du lịch chủ yếu rơi vào mùa cao điểm, đặc biệt là vào mùa hè và các kỳ nghỉ lễ, những khách sạn chấp nhận thú cưng cũng thu hút một lượng du khách nhất định.

Phân tích hành vi của khách du lịch châu Á cho thấy nhóm khách này đến Anh nhiều nhất, các quốc gia còn lại không có sự chênh lệch nhiều. Họ chủ yếu du lịch vào các kỳ nghỉ hè, cuối năm hoặc các dịp lễ lớn, và thường chọn đi theo cặp đôi và du lịch một mình là chủ yếu. Các đánh giá của du khách châu Á về các quốc gia này phần lớn tập trung vào sự hài lòng về cảnh quan, văn hóa và dịch vụ, tuy nhiên cũng không thiếu những nhận xét tiêu cực liên quan đến chất lượng phòng, wifi.... Đối với các khách sạn, du khách châu Á thường đánh giá cao sự thân thiện của nhân viên và chất lượng phòng nghỉ, nhưng cũng có những phản nàn về các yếu tố như phòng nhỏ, thiếu tiện nghi, hoặc dịch vụ chưa đáp ứng kỳ vọng.

Tuy nhiên, một hạn chế đó là dữ liệu chỉ đầy đủ các tháng trong năm 2016, riêng năm 2015 và 2017 thì chỉ có 2-3 quý, làm ảnh hưởng đến việc phân tích xu hướng và nhận định tổng quát.

TÀI LIỆU THAM KHẢO

UN Tourism. (2024). International Tourist Arrivals hit 96% of pre-pandemic levels through July 2024. Unwto.org.

<https://www.unwto.org/news/international-tourist-arrivals-hit-96-of-pre-pandemic-levels-through-july-2024>

Topic: Travel and tourism in Europe. (2019). Www.statista.com; Statista.

<https://www.statista.com/topics/3848/travel-and-tourism-in-europe/>

Matplotlib. (2024). Matplotlib: Python plotting — Matplotlib 3.3.4 documentation. Matplotlib.org. <https://matplotlib.org/stable/index.html>

Liu, J. (2017). 515K Hotel Reviews Data in Europe. Kaggle.com.

<https://www.kaggle.com/datasets/jiashenliu/515k-hotel-reviews-data-in-europe/data>

Chocharah, N. (2022, June). Data_preprocessing_&_EDA_Hotel_Reviews. Kaggle.com; Kaggle.

https://www.kaggle.com/code/nourchocharah/data-preprocessing-eda-hotel-reviews/notebook?fbclid=IwZXh0bgNhZW0CMTEAAR3KWC3W0qGuzreQ5_pI1F9MNdN_kft7JXfQOS4Z2EcWwxqK7R4qATcdCSA_aem_YyBTUN1hEJ4e1XgPKBEOrA