

**ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
**TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**



**MÔN HỌC QUẢN TRỊ DỮ LIỆU VÀ TRỰC QUAN HÓA**

**ĐỀ TÀI: PHÂN TÍCH VÀ TRỰC QUAN HÓA DỮ LIỆU  
NHU CẦU ĐẶT PHÒNG KHÁCH SẠN**

**GVHD: TS. TRẦN VIỆT TRUNG**

**Nhóm 4:**

Nguyễn Đình Tuấn	20183849
Vũ Văn Trung	20183846
Lã Minh Toàn	20183841

**Hà Nội, 07/2022**

**MỤC LỤC**

**Lời nói đầu . . . . . 3**

**Giới thiệu chung . . . . . 4**

**Chương 1: Tổng quan về bộ dữ liệu . . . . . 5**

**Chương 2: Trực quan hóa dữ liệu . . . . . 7**

**Tổng kết . . . . . 14**

## Lời nói đầu

Hiện nay, đại dịch covid-19 vẫn còn đang diễn biến phức tạp với nhiều nguy cơ còn tiềm ẩn. Trong hai năm diễn ra đại dịch, cả thế giới nói chung và Việt Nam nói riêng, đã nỗ lực hết mình để có thể đương đầu với chủng virus mới này. Cùng với đó là nhu cầu du lịch cũng đã bị gián đoạn rất nhiều. Cho đến nay, dịch bệnh đã được nước ta kiểm soát khá tốt và các dịch vụ du lịch cũng đã được mở trở lại.

Tuy nhiên, nhiều người vẫn sẽ vướng mắc những câu hỏi như: thời điểm nào trong năm là thuận lợi trong việc đi du lịch cũng như đặt phòng, thời gian lưu trú tối ưu để có mức giá tốt nhất, hay chỉ đơn thuần là những khách sạn nào phù hợp với nhu cầu của mình. Vì vậy, trực quan hóa bộ dữ liệu “Nhu cầu đặt phòng khách sạn” sẽ hỗ trợ rất nhiều trong việc trả lời những câu hỏi đó.

Mặc dù chúng em đã cố gắng nhưng do thời gian và kiến thức có hạn nên sản phẩm và báo cáo chúng em có thể còn những thiếu sót,... Vì vậy, em mong nhận được sự thông cảm và những góp ý từ thầy để sản phẩm của bọn em có thể hoàn thiện hơn.

## Giới thiệu chung

- Phân công công việc

Sinh viên	Công việc
Nguyễn Đình Tuấn	Tìm kiếm dữ liệu, phân tích và trực quan hóa dữ liệu
Vũ Văn Trung	Làm báo cáo, trực quan hóa dữ liệu
Lã Minh Toàn	Làm slide, trực quan hóa dữ liệu

- Cấu trúc bài báo cáo

Bài báo cáo trình bày về quản trị, trực quan hóa và phân tích các số liệu phản ánh nhu cầu đặt phòng khách sạn trên toàn thế giới trong vòng 3 năm 2015-2016-2017.

# Chương 1: Tổng quan về bộ dữ liệu

Tập dữ liệu chứa thông tin đặt phòng của các khách sạn cũng như khu resort, đồng thời bao gồm các thông tin như thời điểm đặt phòng, thời gian lưu trú, số lượng người lớn, trẻ em,... Và tất nhiên để đảm bảo sự riêng tư, các thông tin cá nhân đã được loại bỏ.

Cấu trúc file dữ liệu như bên dưới

```
df = pd.read_csv("hotel_bookings.csv")
```

```
df.head()
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights
0	Resort Hotel	0	342	2015	July	27	1	0	0
1	Resort Hotel	0	737	2015	July	27	1	0	0
2	Resort Hotel	0	7	2015	July	27	1	0	0
3	Resort Hotel	0	13	2015	July	27	1	0	0
4	Resort Hotel	0	14	2015	July	27	1	0	0

5 rows × 32 columns

Độ lớn file dữ liệu cùng các trường thông tin:

```
print(df.shape)
df.columns
```

```
(119390, 32)
```

```
Index(['hotel', 'is_canceled', 'lead_time', 'arrival_date_year',
      'arrival_date_month', 'arrival_date_week_number',
      'arrival_date_day_of_month', 'stays_in_weekend_nights',
      'stays_in_week_nights', 'adults', 'children', 'babies', 'meal',
      'country', 'market_segment', 'distribution_channel',
      'is_repeated_guest', 'previous_cancellations',
      'previous_bookings_not_canceled', 'reserved_room_type',
      'assigned_room_type', 'booking_changes', 'deposit_type', 'agent',
      'company', 'days_in_waiting_list', 'customer_type', 'adr',
      'required_car_parking_spaces', 'total_of_special_requests',
      'reservation_status', 'reservation_status_date'],
      dtype='object')
```

# Định dạng dữ liệu:

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   hotel                                119390 non-null  object
1   is_canceled                          119390 non-null  int64
2   lead_time                            119390 non-null  int64
3   arrival_date_year                    119390 non-null  int64
4   arrival_date_month                  119390 non-null  object
5   arrival_date_week_number            119390 non-null  int64
6   arrival_date_day_of_month           119390 non-null  int64
7   stays_in_weekend_nights             119390 non-null  int64
8   stays_in_week_nights                119390 non-null  int64
9   adults                              119390 non-null  int64
10  children                             119386 non-null  float64
11  babies                              119390 non-null  int64
12  meal                                119390 non-null  object
13  country                             118902 non-null  object
14  market_segment                      119390 non-null  object
15  distribution_channel                 119390 non-null  object
16  is_repeated_guest                   119390 non-null  int64
17  previous_cancellations               119390 non-null  int64
18  previous_bookings_not_canceled       119390 non-null  int64
19  reserved_room_type                   119390 non-null  object
20  assigned_room_type                   119390 non-null  object
21  booking_changes                      119390 non-null  int64
22  deposit_type                         119390 non-null  object
23  agent                                103050 non-null  float64
24  company                              6797 non-null   float64
25  days_in_waiting_list                 119390 non-null  int64
26  customer_type                        119390 non-null  object
27  adr                                  119390 non-null  float64
28  required_car_parking_spaces          119390 non-null  int64
29  total_of_special_requests            119390 non-null  int64
30  reservation_status                   119390 non-null  object
31  reservation_status_date              119390 non-null  object
dtypes: float64(4), int64(16), object(12)
memory usage: 29.1+ MB
```

## Chương 2: Trực quan hóa dữ liệu

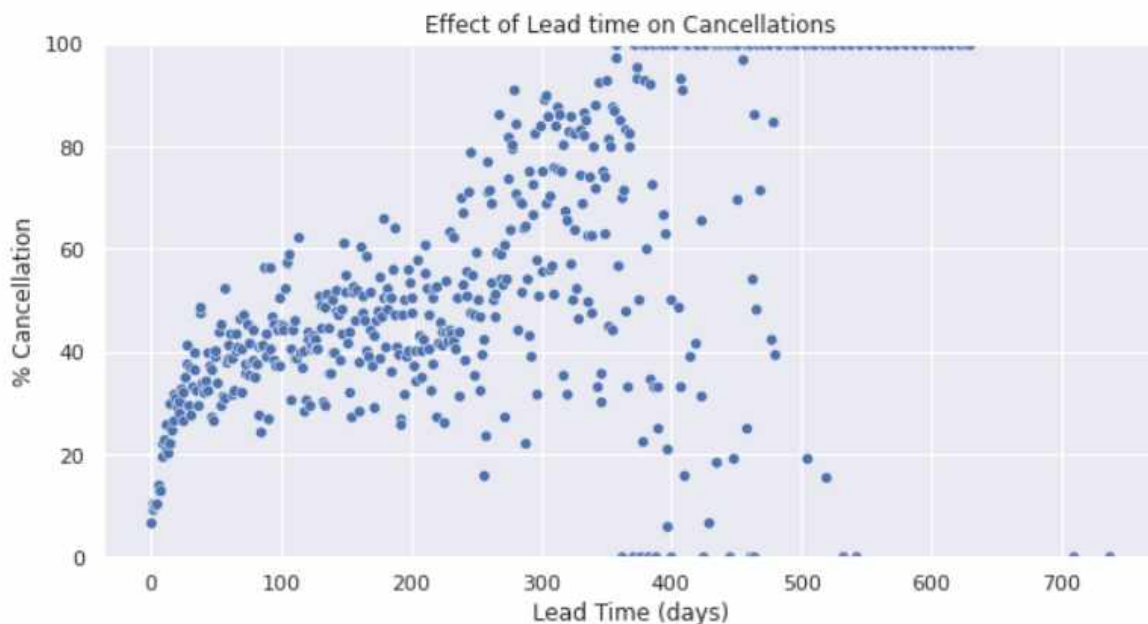
Sau khi trực quan hóa dữ liệu, có thể thấy nhu cầu đặt phòng trong những khách sạn còn khá sao so với những khu resort. Có thể kể đến vài nguyên nhân dẫn đến tình trạng này như thói quen của khách du lịch hướng đến những khách sạn nhiều hơn, giá tiền các khu resort còn khá cao so với bình quân thu nhập,...



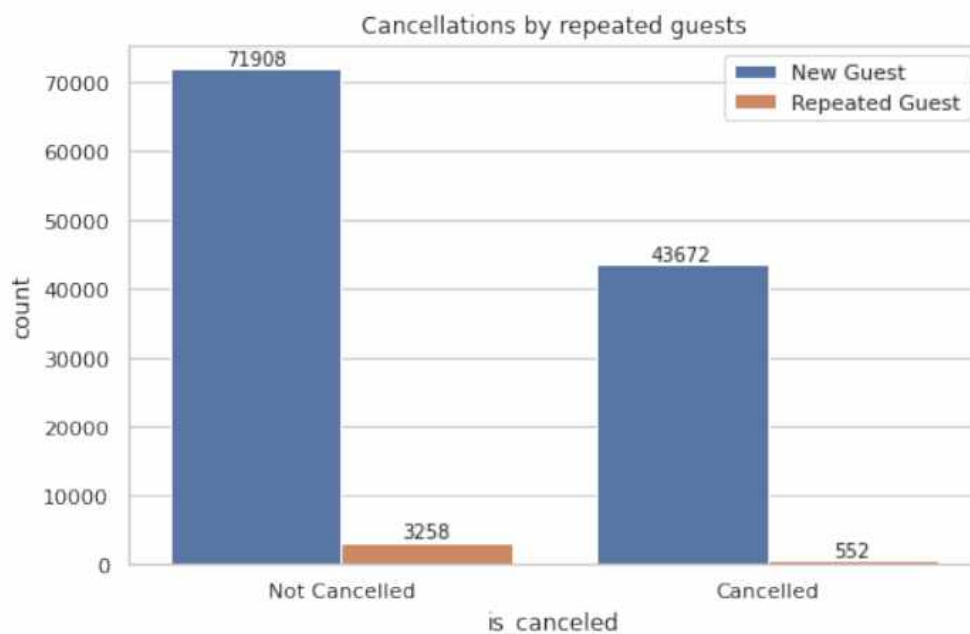
Lượng đặt phòng cao cũng đồng nghĩa với việc tỉ lệ phòng bị hủy cao. Điều đó đã đúng trong trường hợp này khi lượng hủy phòng ở các khách sạn là cao hơn nhiều so với các khu resort.



Đi sâu hơn vào vấn đề hủy phòng, chúng ta có thể suy nghĩ rằng lượng phòng bị hủy sẽ có liên quan đến thời gian từ lúc đặt phòng đến lúc ở hoặc theo phương diện khách đặt phòng là khách mới hay cũ. Đầu tiên là sự ảnh hưởng của khoảng thời gian đặt phòng đến việc phòng bị hủy:



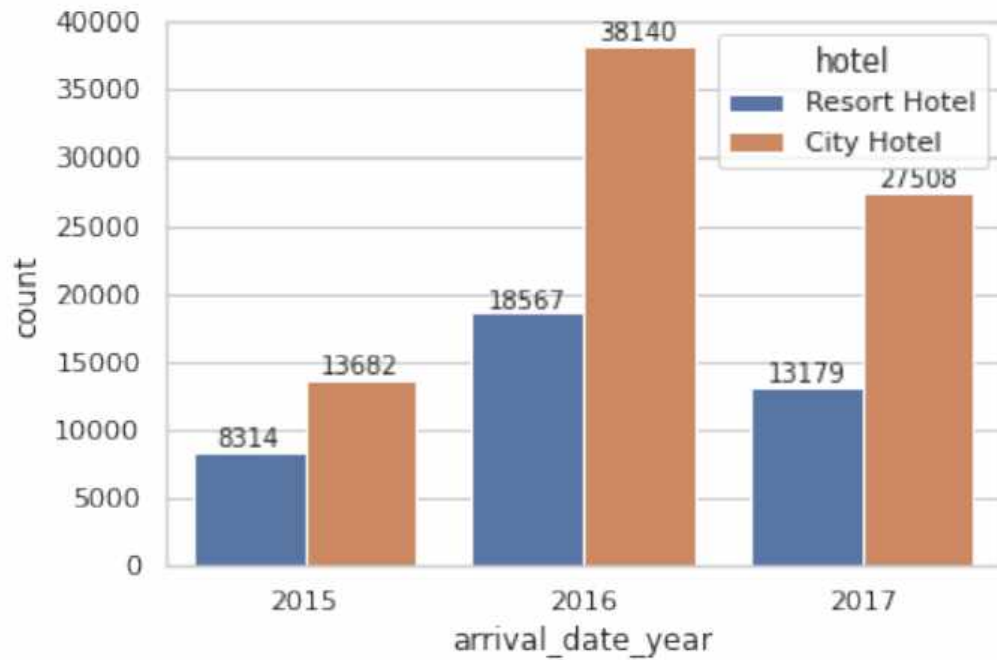
Có thể thấy rằng thời gian kể từ lúc đặt phòng đến lúc ở càng kéo dài thì tỉ lệ phòng bị hủy là rất cao. Tiếp theo là dựa theo khách mới hay cũ:



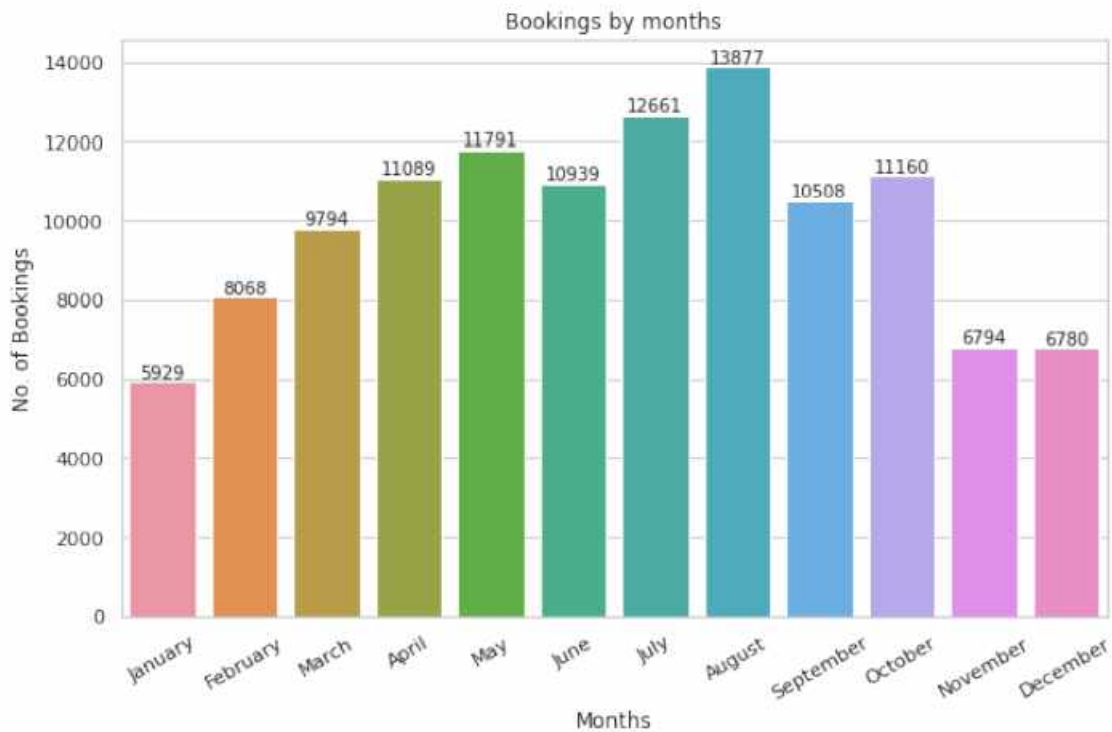
Chắc chắn rằng nếu là khách cũ của khách sạn đặt lại thì họ sẽ không dễ dàng hủy phòng khi đã quyết định quay lại khách sạn đó. Tuy vậy vẫn có 1 số ít hủy phòng trong khi mình là khách cũ.



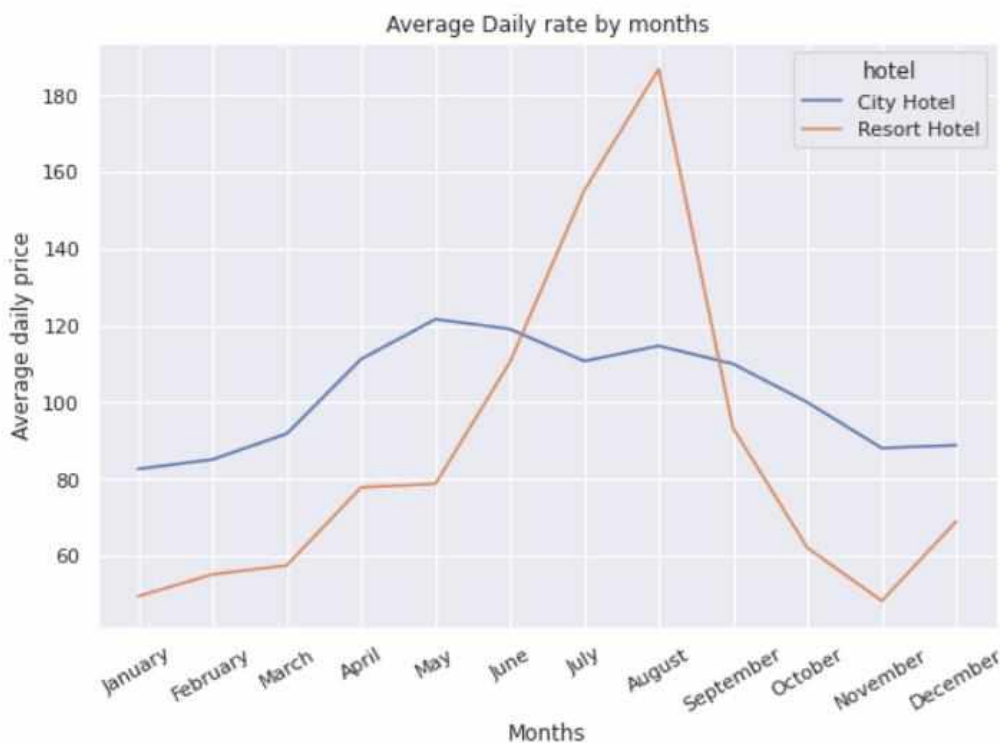
Lượng đặt phòng được trực quan hóa theo từng năm:



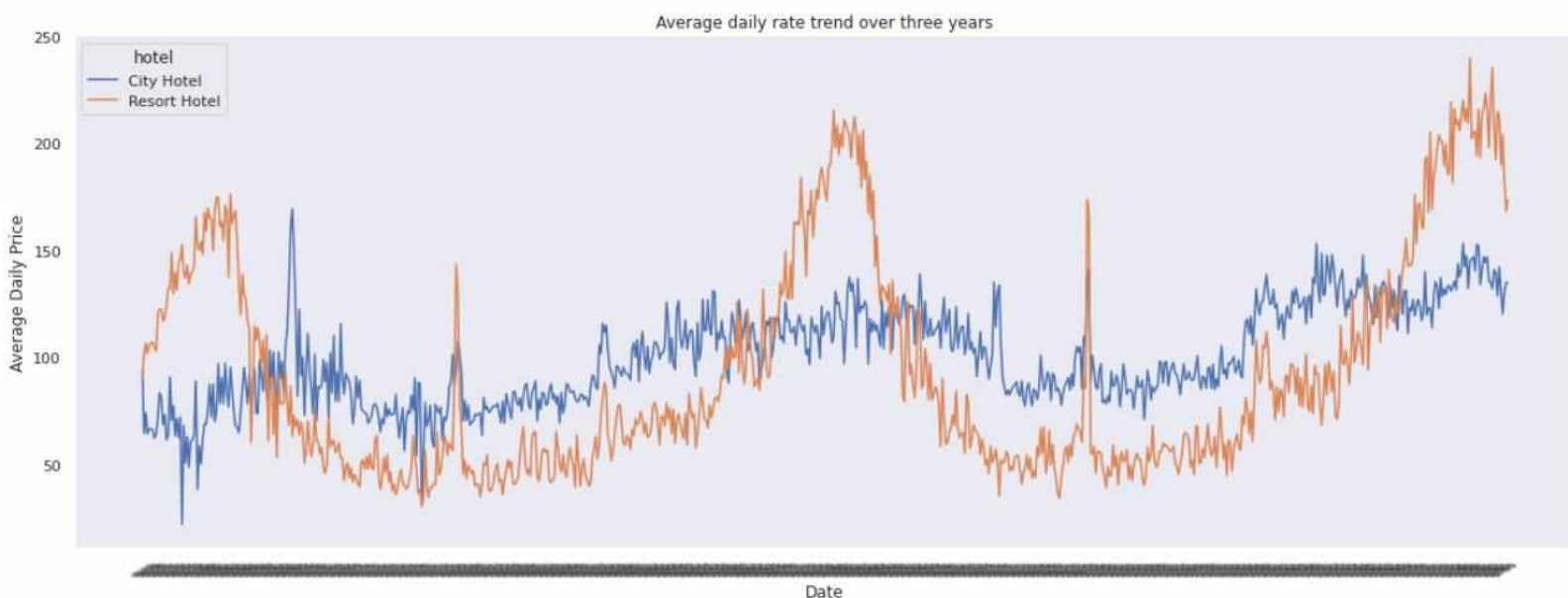
Và theo từng tháng. Có thể thấy khách du lịch có xu hướng đi nhiều vào giữa năm – thời gian mà thời tiết khá đẹp và thuận lợi ở những khu du lịch.



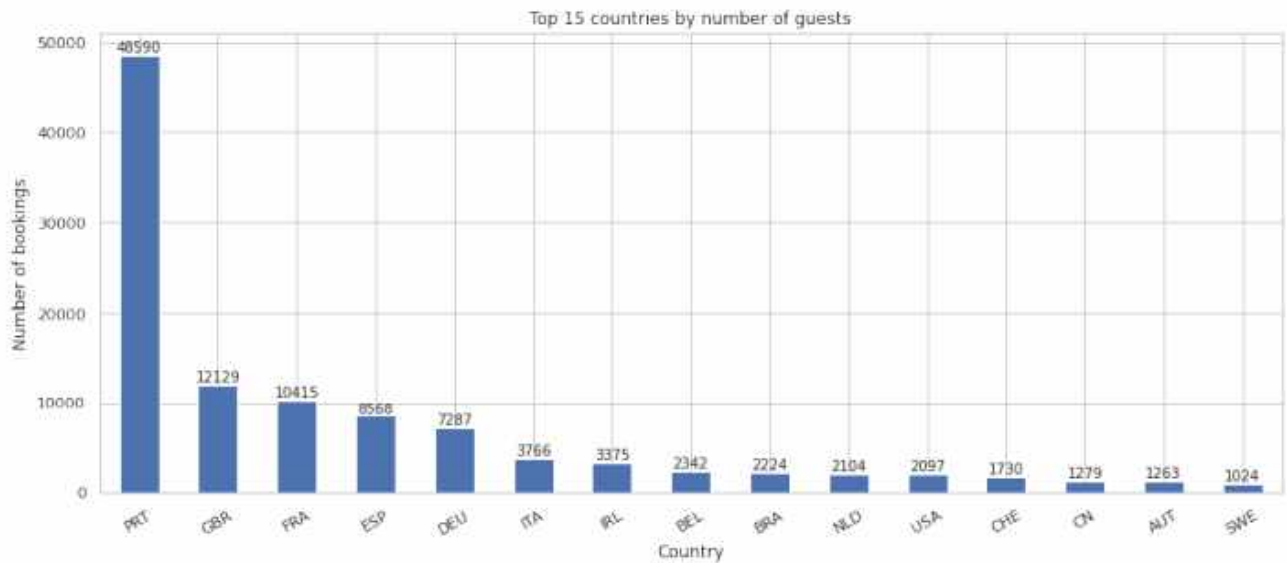
Tiếp theo là trực quan hóa giá đặt phòng theo từng tháng trong năm. Có thể dễ dàng nhận thấy giá của các khách sạn chỉ dao động nhẹ vào những tháng cao điểm trong khi các khu resort có giá phòng vào tháng 7, 8 là cực kỳ cao.



Trực quan chi tiết hơn về giá phòng từng ngày trong vòng 3 năm:



Top 15 nước có lượng khách du lịch đặt phòng nhiều nhất. Có thể thấy việc đặt trước phòng đã khá phổ biến ở Bồ Đào Nha khi lượng đặt trước tại quốc gia này lớn hơn rất nhiều so với các quốc gia khác xếp sau như GBR – Liên hiệp Anh và Bắc Ireland hay FRA – Pháp.

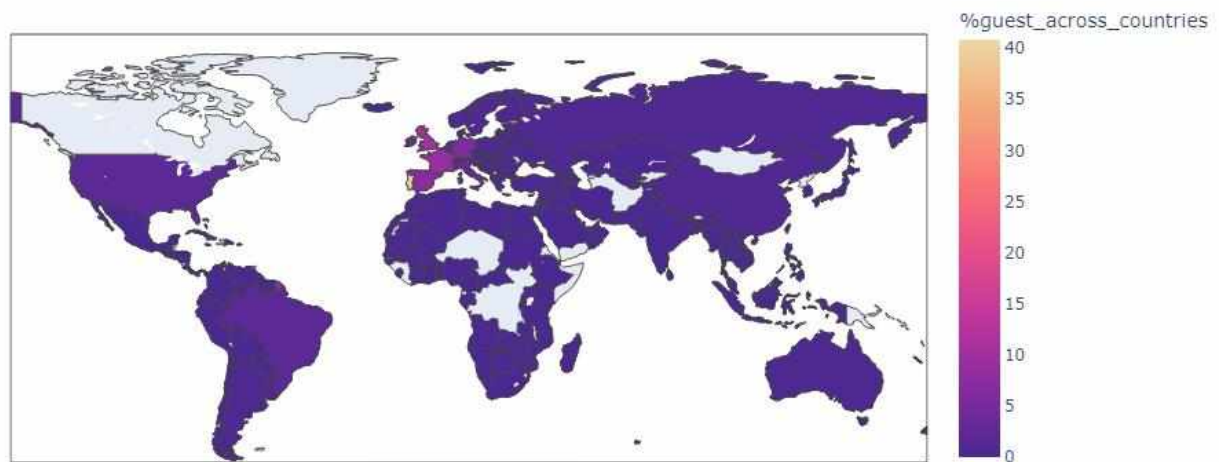


Lượng đặt phòng được phân chia theo khách sạn và resort.



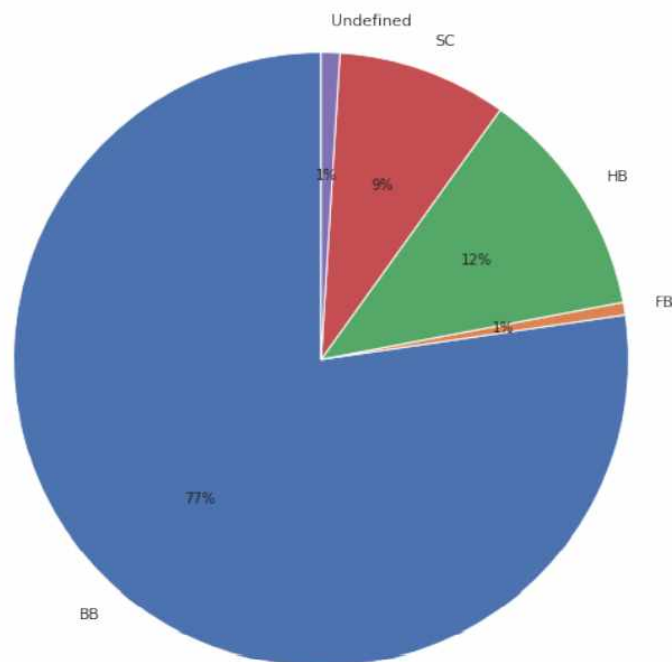
Trực quan tỉ lệ đặt phòng của các nước trên bản đồ thế giới. Có thể thấy xu hướng đến các nước Tây Âu là rất cao.

Total Bookings Across Countries

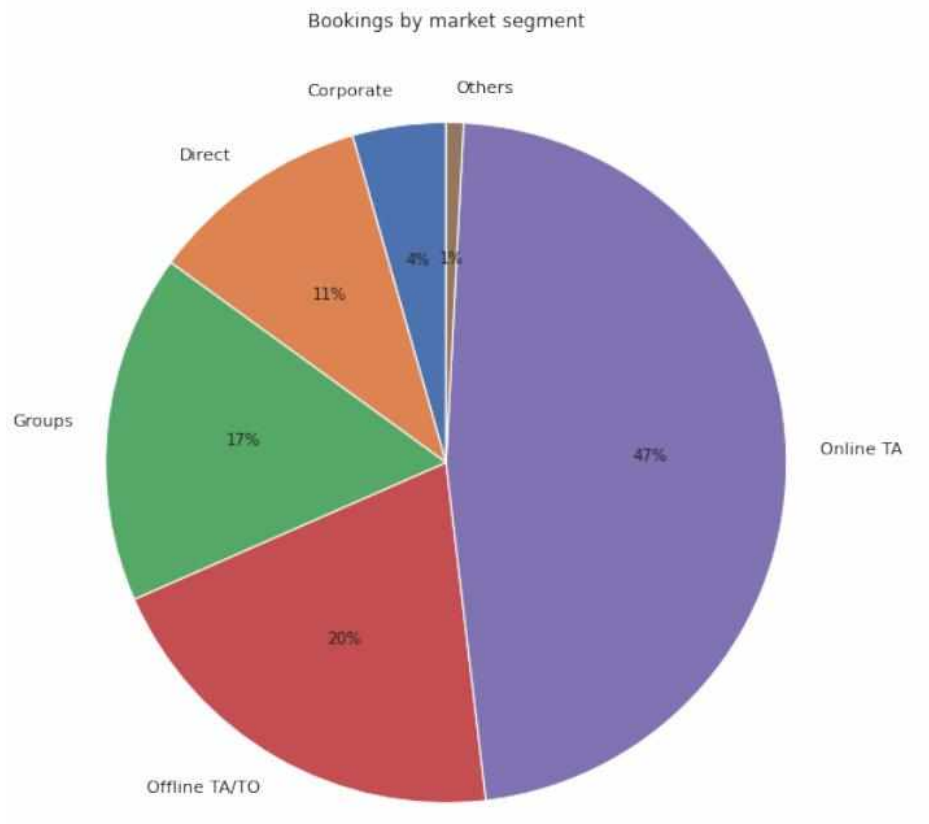


Đi cùng với đặt phòng thì mọi người sẽ có thêm xu hướng đặt kèm các bữa ăn. Nhu cầu đặt bữa ăn được trực qua hóa như hình dưới. Người dùng có xu hướng chỉ đặt phòng kèm bữa sáng (BB) khi tỉ lệ này chiếm tới 77%, tiếp đó là đặt phòng kèm bữa sáng và tối chiếm 12% và 9% người dùng không đặt kèm bữa ăn.

Portion of bookings with meals and its type



Xu hướng đặt phòng thông qua các kênh thông tin khác nhau cũng có sự phân hóa rất lớn. Gần 1 nửa số phòng được đặt thông qua các đại lý du lịch trực tuyến – Online Travel Agent. Và có hơn 90% số phòng được đặt thông qua các hình thức môi giới trung gian.



## **Tổng kết**

Vậy qua những phân tích trên, chúng ta đã nhận ra được những xu hướng đặt phòng của khách du lịch trong những năm. Vì vậy, khi các nơi trên thế giới phần nào đã mở cửa du lịch trở lại sau khi đã khắc phục được hầu hết các hậu quả của đại dịch Covid19, những phân tích ấy sẽ góp phần rất lớn để gợi ý cho khách du lịch nên đặt phòng khách sạn nào phù hợp nhất với bản thân mình. Và cuối cùng, chúng ta vẫn mong muốn thế giới chấm dứt dịch bệnh để có thể mở cửa hoàn toàn và đón khách du lịch nhằm quảng bá những địa điểm đẹp hay các công trình danh lam thắng cảnh của mỗi đất nước. Điều đó cũng góp phần rất lớn vào việc phát triển kinh tế đất nước cũng như khắc phục lại những khó khăn khi nền kinh tế bị kìm hãm khi đại dịch diễn ra.