

## Part 1

Summarize your response to the questions below in a 5-8 slide presentation intended for the Head of Product (who is not technical)

- What key metrics would you propose to monitor over time the success of the team's efforts (increasing the number of bookings, minimizing cancellations, etc.)? Clearly define your metric(s) and explain how each is computed.



Before starting the answers, there are some pre-processing steps that have been done in Python (the code will be provided), thereby, some of the solutions would be using the pre-processed data instead of the raw table.



Please note that for the ease of calculation, and since this is just a sample, based on the most active countries, I have kept 20 top countries that contain almost 118478 rows which is reasonably 93% of the whole data  $(118478/126676)*100= 0.935$ . Therefore, the rest of whole calculations are based on these 20 top countries.

- 1) The first key metric, is "Total Days" which means the number of days each country's people have booked. Here is how we calculate it:

(since there was no explanation provided in the assignment about a column called "Order\_Amount", I had to go ahead with the most probable case I could think of. I assumed this value shows the quantity of bookings per day per country.)

If the above assumption is correct, then it would be useful for us to know how many days each country has booked.

We can calculate it by:

`Order_Amount * Length_of_Stay_In_Nights`

and grouping them by country name.

Based on this calculation, we will later see what countries are top ranked customers based on their total booked days.

- 2) Another key metric that has to be monitored, is the Platforms' popularity and performance. It shows how much each platform gets used by customers comparatively and also how much each of the platforms has had booking versus cancellation.

The popularity among platforms simply comes from grouping data by Platform and aggregating sum of Total Days.

Also, the performance of the platforms can be seen by adding the count of "Active" and "Cancelled" values in the column called "Order\_Status".

Now we can calculate the proportion of each platform's cancelled and active deals and compare them. The formula will be like this:

$$(\text{count of Cancelled per platform} / \text{count of Active per platform}) * 100$$

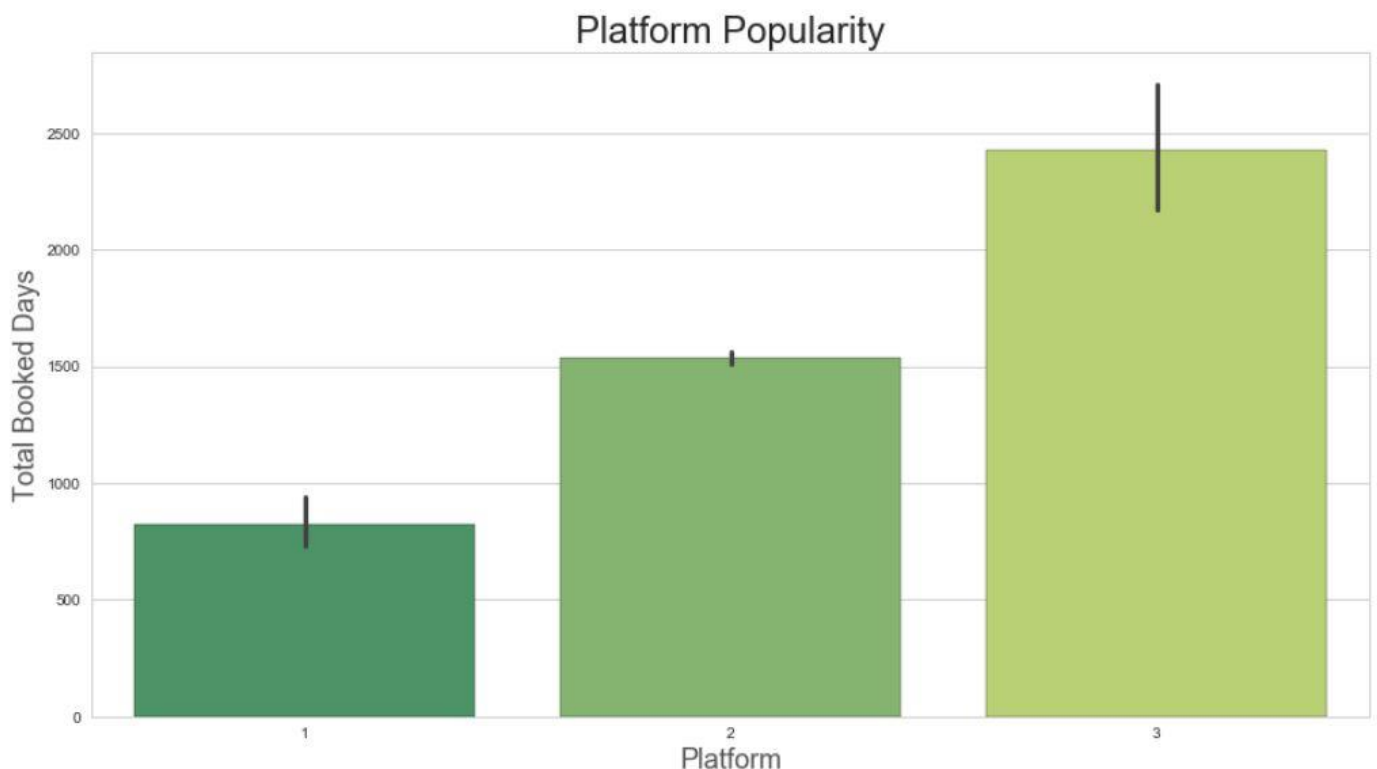
- 3) The third metric is how each country performs in terms of cancellations. I called it performance.

We calculate it by checking out how many percent of the whole bookings are cancelled. So, here would be how we calculate it:

$$(\text{count of Cancelled per country} / (\text{count of Cancelled} + \text{count of Active})) * 100$$

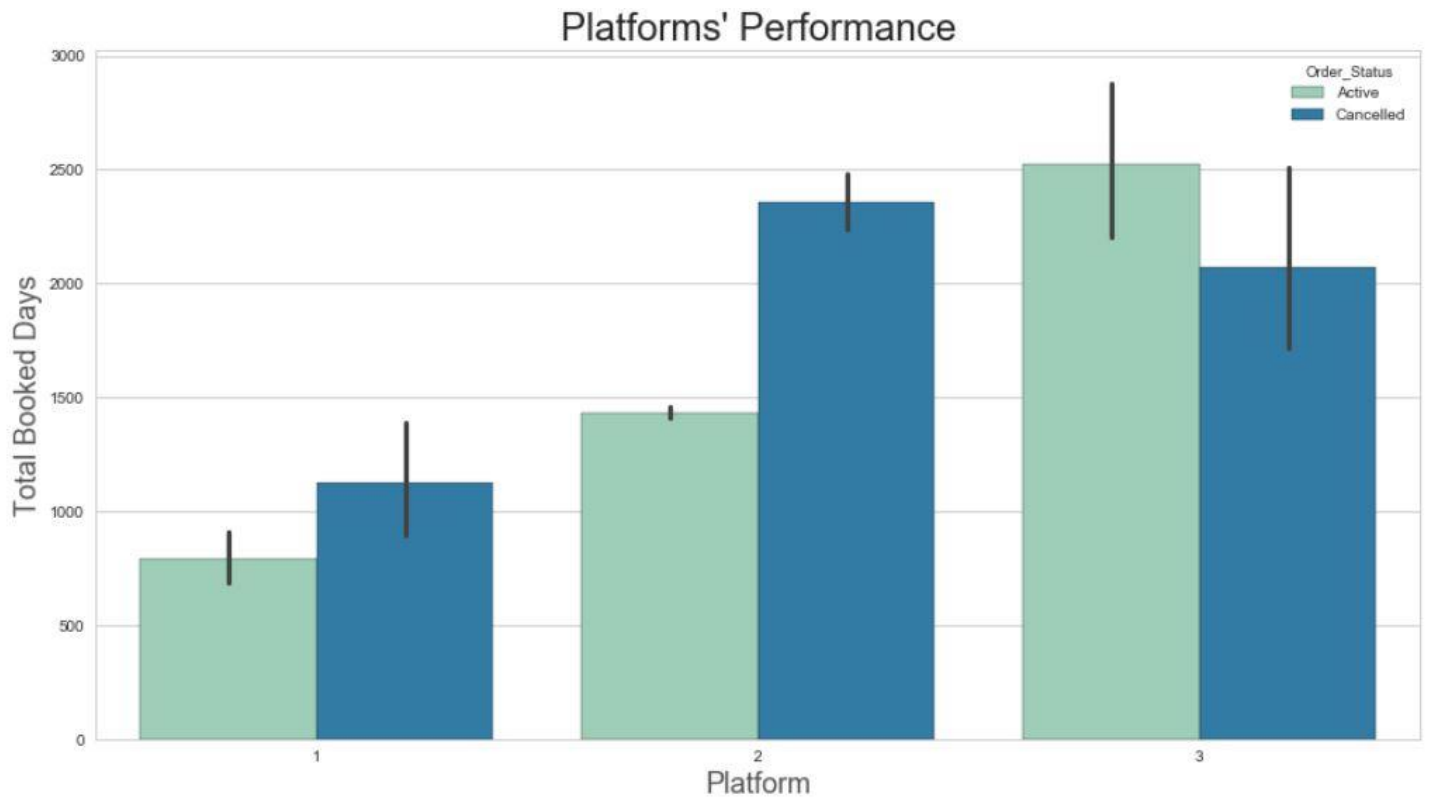
- Given the data in *hq\_bookings.csv* file, create some visualizations to display the information visually to the management. They want to know if there are any **interesting insights, existing problems, or changes that occurred recently that they are not yet aware of**. You can use Excel, R, Python, Tableau, or any other tool you wish.

- As mentioned, platform popularity and platform performance are important for the company to track. Here is a visualization using the given dataset:



Insight

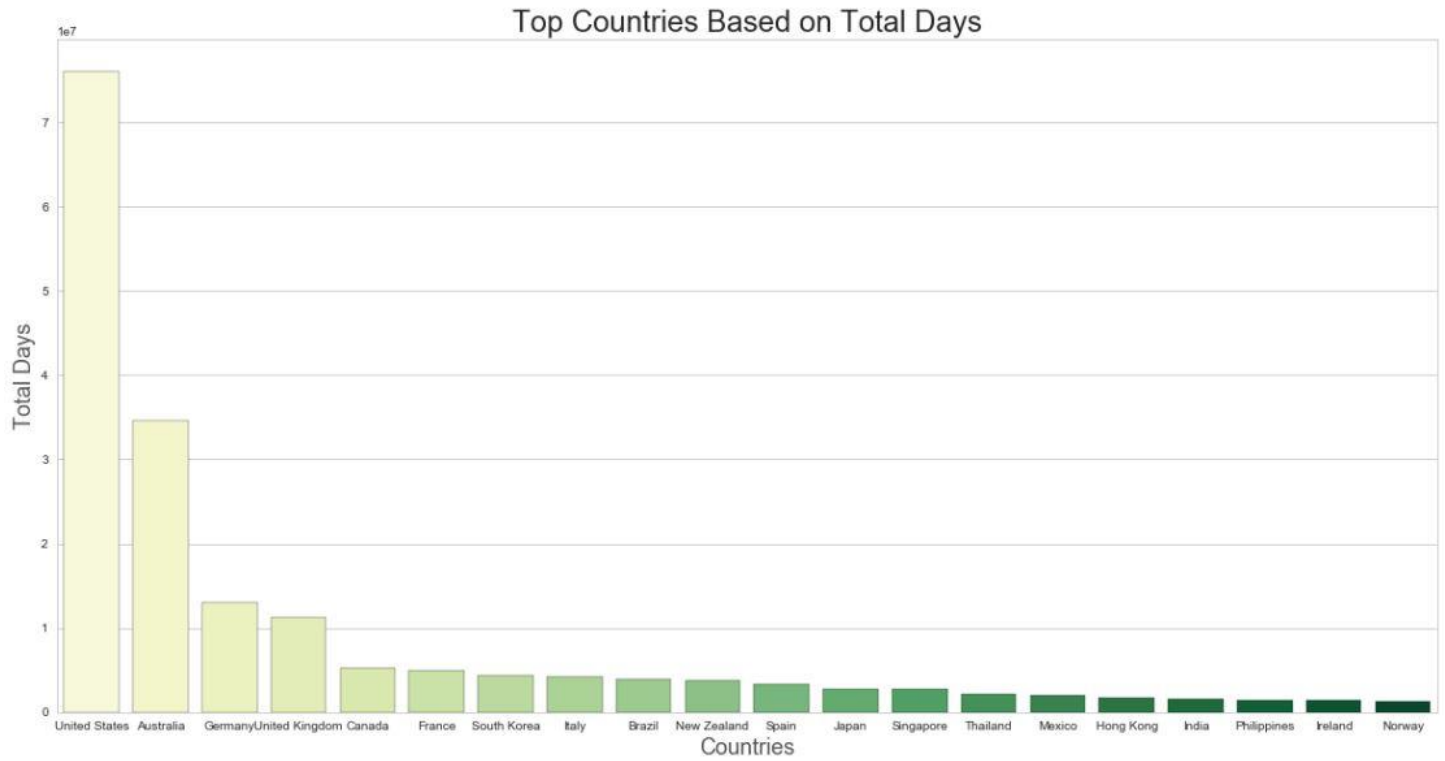
Platform three has more popularity than two, and two is more popular than one.



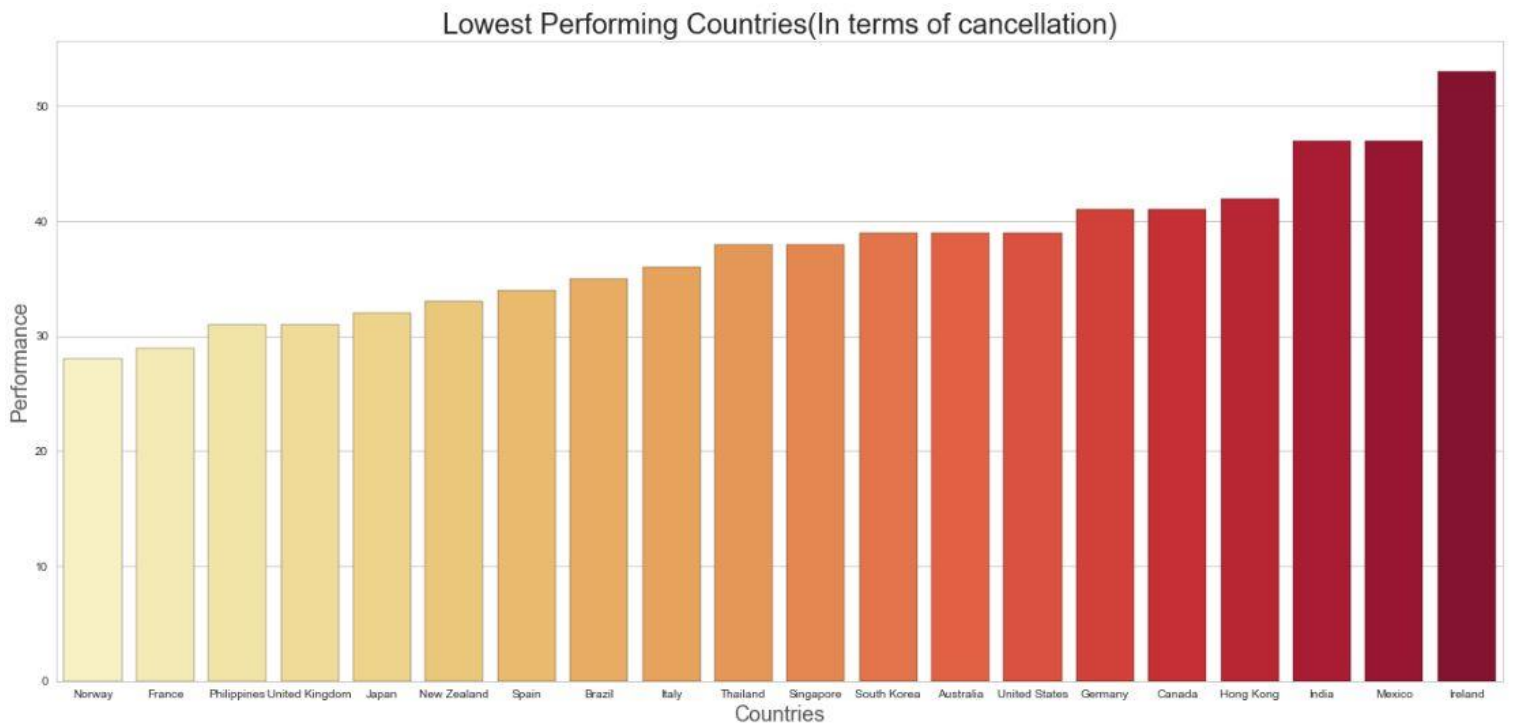
Insight

Platform one and two have their cancellation rate higher than platform three. This means there could be some technical problems with the first two platforms, such as misleading contents, low server responsibility rate, or none of them, just the company is replicating those platforms and doesn't promote them anymore.

- Countries' ranking based on their booking days (the first 20 countries), here is the table and the graph:



- Countries' ranked from lowest performing to highest performing in terms of cancellation.

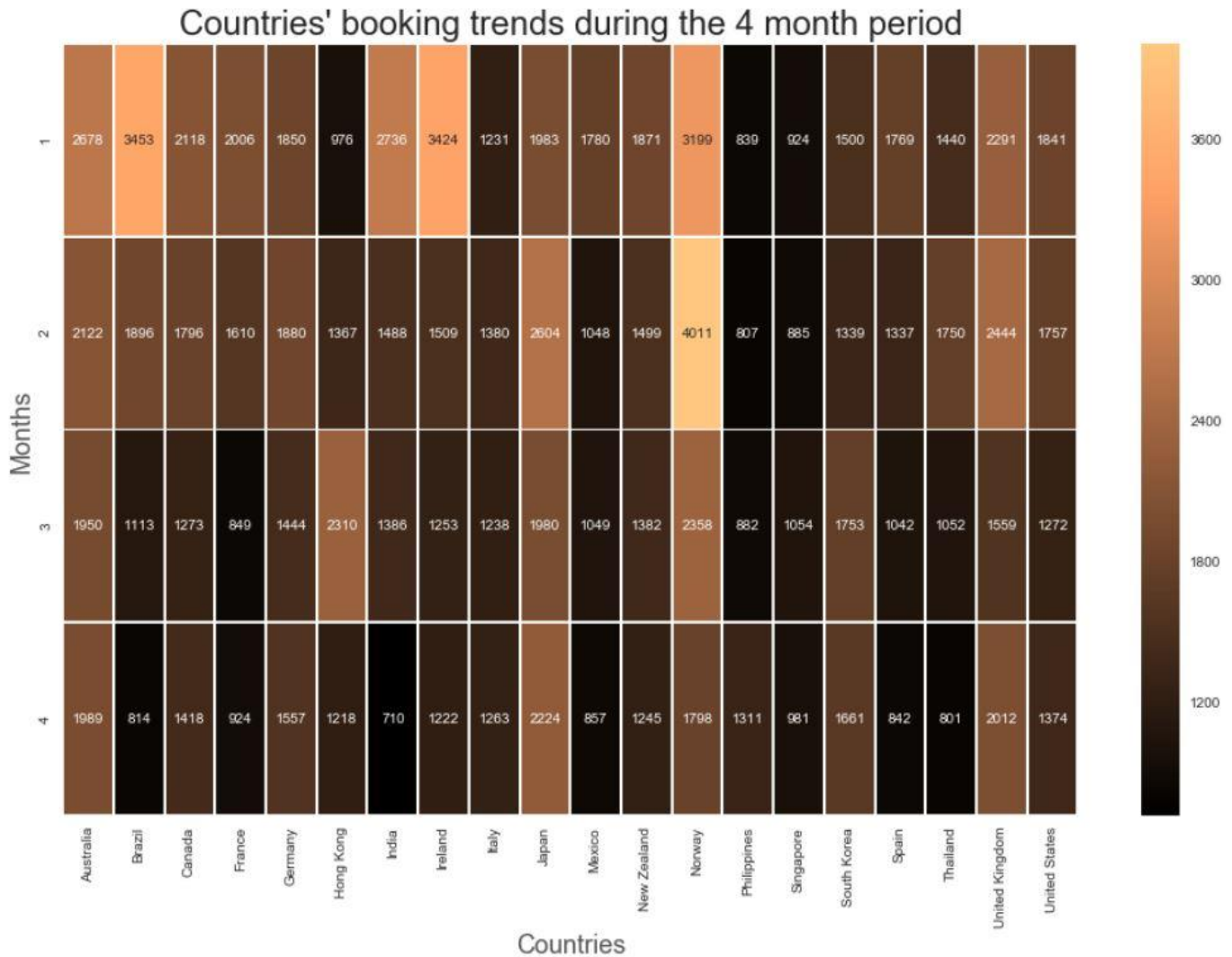


### Insight

As we can see for example in the UK which is among one of the highest countries in terms of booking days, there are a lot of booking days, however, so many of them are being cancelled, there must be a reason for it that the company should figure out. There is a similar situation in France more or less.

In case of the UK, we can compare it with Australia, that has a very good performance and low cancellation rate, we can check different features of both markets, since they have a lot of similarities in language and culture, this comparison can give us good insights.

- Heatmap graph of countries' bookings per month



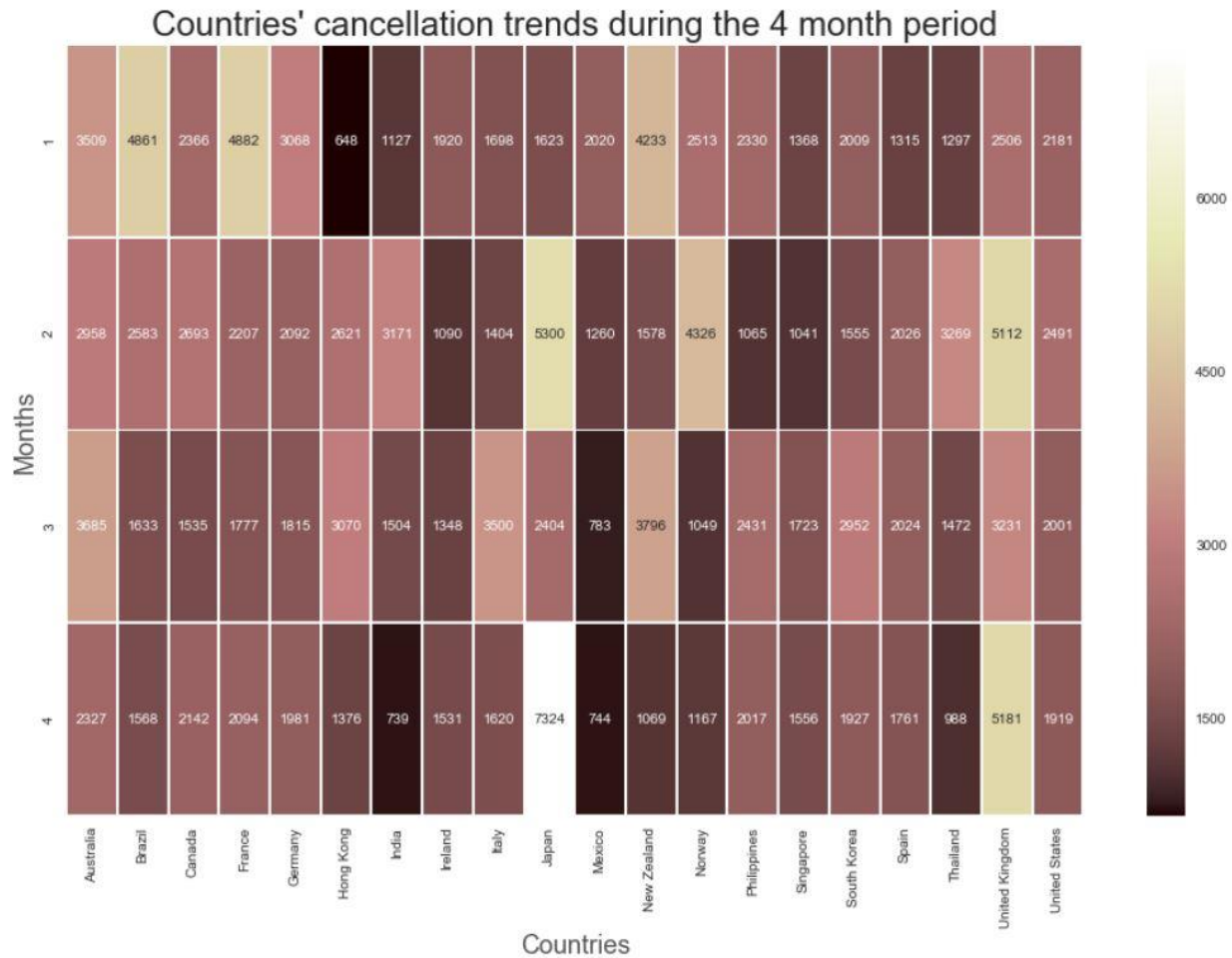
### Insight

Some countries have interesting insights. For example Norway had a very high rate of bookings at the second month which could be because of a promotion plan, or marketing investment.

Booking rates in Ireland, Brazil, and France have dropped significantly each month.

Booking in Philippines has increased. Hong Kong has a strange behavior. In January that most people travel, Singaporean didn't have much interest in booking but it has increased so much for the next two months, but again declined in the fourth month.

- Heatmap graph of countries' cancellations per month



Insight

This graph comes from a pivot table that shows the number of cancellations for each country per month. Interestingly, Norway in the second month has had an outlier value of 12576 as we can see in the pivot table here:

User_Country_Name	Australia	Brazil	Canada	France	Germany	Hong Kong	India	Ireland	Italy	Japan	Mexico	New Zealand	Norway	Philippines	Singapore	South Korea
Booking_Month																
1	3509	4861	2366	4882	3068	648	1127	1920	1698	1623	2020	4233	2513	2330	1368	2009
2	2958	2583	2693	2207	2092	2621	3171	1090	1404	5300	1260	1578	12576	1065	1041	1555
3	3685	1633	1535	1777	1815	3070	1504	1348	3500	2404	783	3796	1049	2431	1723	2952
4	2327	1568	2142	2094	1981	1376	739	1531	1620	7324	744	1069	1167	2017	1556	1927

In order to have a better idea about the other cancellation values, I replaced this Norway outlier, with the mean of all Norway value just to be able to see the other values better. The result is as follows:

User_Country_Name	Australia	Brazil	Canada	France	Germany	Hong Kong	India	Ireland	Italy	Japan	Mexico	New Zealand	Norway	Philippines	Singapore	South Korea
Booking_Month																
1	3509	4861	2366	4882	3068	648	1127	1920	1698	1623	2020	4233	2513	2330	1368	2009
2	2958	2583	2693	2207	2092	2621	3171	1090	1404	5300	1260	1578	4326	1065	1041	1555
3	3685	1633	1535	1777	1815	3070	1504	1348	3500	2404	783	3796	1049	2431	1723	2952
4	2327	1568	2142	2094	1981	1376	739	1531	1620	7324	744	1069	1167	2017	1556	1927

Cancellation in Japan has been so high especially in second and fourth months. There could be some serious issue that so many of the Japanese have done this. The UK has the same story as well.

The cancellation rate in Mexico has been decreased a lot recently. Also, Thailand has had good performance on cancellation on the fourth month.

- Based on your visualizations, what areas should we invest in to increase the number of successful bookings?
- Propose 2-3 specific recommendations (business initiatives and product changes) that could address these opportunities. Demonstrate rationale behind each recommendation.

- 1) The market in Norway sounds so promising. The company has to track it so closely, especially the high number of cancellations on the second month has to be carefully followed up and fixed. Those churned bookings could have been potentially converted to successful leads. The company has to make sure this never happens again in Norway. In my opinion, the case of Norway can be treated as a VIP account and the company can invest more on incentives to persuade customers, but parallel to it, technical issues or other reasons for cancellations in Norway have to be addressed immediately. As an idea, the company might have to use the best performing platform (Platform 3) in Norway.
- 2) There are some other special markets that the company can invest more on marketing there, because they have the potential to have better conversions. This group consists of India, Brazil, and Ireland. Obviously, these countries had great booking values in January but in the consequent three months the sales have dropped remarkably. On the other hand, we can think about how to improve the performance in some countries like Singapore, Hong Kong, and Thailand as they didn't have great sales in January. It sounds startling that these are all Asian countries. So maybe the marketing campaigns weren't so effective on Asian culture. For instance, Thailand has a very special taste of color and design on webpages, or the Chinese have to feel more engaged (As the majority of people in all mentioned three countries have Chinese roots). I believe so many multinationals get into the trap of using a unified taste for all their products and since the western culture is more common, they try to use the same recipe for Asian

market and this is where they miss the great opportunity of wealthy Asian customers.

- 3) The market in the US, Japan, and Australia seem to be more mature than others and the company has already made a good sale in these countries. Since the investment information about these markets aren't available, there is no way to track how much the company has profited in those markets, otherwise if the management feels the products are well known there, they might be considering to rely more on organic traffic, word of mouth, or referral conversion rather than paid marketing campaigns. In this way the company would save more funds to invest on growing markets.



## Part2

### Basic SQL Test

- Using the tables provided above, write a query to return all employees still working for the company with last names starting with "Baker" sorted by name then surname.

```
1 select * from hqEmployee where
2
3 (TerminationDate is NULL
4 OR
5 TerminationDate > now())
6 AND
7 Surname like "Baker%"
8
9 ORDER BY 'Name', 'Surname'
```

- As we know there is no TerminationDate which is bigger than the date of now in the given table, so for this dataset TerminationDate > now() is not necessary but it sounded more accurate to me to have this condition in case in the future someone adds a future date as TerminationDate so that we don't manually change the query.

- Given the `Employee` and `AnnualReviews` tables, write a query to return all employees who have never had a review sorted by the newest hire.

```
1 SELECT hqemployee.Number, hqemployee.Name, hqemployee.Surname, hqemployee.HireDate, annualreport.EmployeeID
2 FROM hqemployee
3 LEFT JOIN annualreport ON annualreport.EmployeeID = hqemployee.Number
4 WHERE annualreport.EmployeeID IS NULL
5 ORDER BY hqemployee.HireDate DESC
```

RESULT:

Result #1 (5x4)				
Number	Name	Surname	HireDate	EmployeeID
6	Frank	Nguyen	2015-04-10	(NULL)
4	Keith	Widjaja	2013-09-10	(NULL)
5	Kelly	Smalls	2013-09-10	(NULL)
3	Nancy	Soley	2012-03-14	(NULL)

- Write a query to calculate the difference (in days) between the most and least tenured employee still working for the company.

```

1 SELECT Name as Oldest_Member_name,
2 Surname as Oldest_Member_Surname,
3 Max(SUBQUERY.DateDiff) - Min(SUBQUERY.DateDiff) AS Diff_with_youngest
4 FROM
5 (
6   SELECT *,
7   DATEDIFF( now(), HireDate) AS DateDiff
8   from hqemployee
9   WHERE TerminationDate is NULL
10 OR
11 TerminationDate > now()
12 ORDER By DateDiff
13
14 ) AS SUBQUERY

```

RESULT:

Oldest_Member_name	Oldest_Member_Surname	Diff_with_youngest
Joe	Jarrod	1,306

- Given the employee table above, write a query to calculate the longest period (in days) that the company has gone without a hiring or firing anyone.

**Step 1:** concatenating both HireDate and TerminationDate as actionDate and sort them, then save them into a table called "results".

```

1 INSERT INTO results
2 select (@cnt := @cnt + 1) AS id,
3   actionDate
4
5 from (
6 (
7   select HireDate as actionDate
8   from hqemployee
9 )
10 UNION ALL
11 (
12   select TerminationDate as actionDate
13   from hqemployee
14 )
15 )as diff
16 CROSS JOIN (SELECT @cnt := 0) AS dummy
17
18 where actionDate is not null
19 ORDER BY actionDate
20

```

**Step 2:** Subtracting each row from the row above it in days.

```
1 select a.id,  
2 a.actionDate,  
3 (select b.actionDate from results b where b.id = a.id + 1)-a.actionDate as subtract  
4  
5 from results a
```

RESULT:

results (3×9)		
id	actionDate	subtract
1	2009-06-20	9,592
2	2010-02-12	20,102
3	2012-03-14	10,596
4	2013-09-10	0
5	2013-09-10	9,191
6	2014-01-01	10,309
7	2015-04-10	91
8	2015-05-01	9,600
9	2016-01-01	(NULL)