



# A LOOK INTO EUROPEAN HOTEL REVIEWS

---

## Spark Report

Begoña Frigolet de la Vega

IE BUSINESS SCHOOL  
MBD 2020

# A look into Hotels Reviews

## Background & Methodology

### INTRODUCTION:

The Business world has had a drastically change since the introduction of online companies and their digitalization model in which the largest taxi company does not own cars, the most popular owner does not create their own content and the largest accommodation providers owns no real state.

For the traditional companies this digitalization holds a challenge on how to keep loyal clients and increase their added value towards their costumer's perception, a solution to this problem is data, since information is power the key factor to compete is know what customers perceive as added value and transform data into insights.

According to an article in Forbes released in 2018, Airbnb effect on Hotels resulted on a "1.3% fewer hotel nights booked and a 1.5 percent loss in hotel revenue." <sup>1</sup>In the US alone, having an even bigger impact worldwide. Not only relying on price but the different benefits on quality, intensifying the competition among traditional hotels and their listings.

The porpoise of this analysis is to aid the traditional hotels, by having a closer look on the review data from hotels in four main locations: **United Kingdom, Paris, Amsterdam and Barcelona**, as a result of the European Tourism increase in 2016 of 3.5%<sup>2</sup>, and the quality of the data that is available.

### DATA & METHODOLOGY

The original data was scraped by Booking.com and they are the owners of the data, for this analysis 2017 was taken into consideration.

**Methodology:** Through a Virtual Machine and using Spark, data transformations were done with the help of the "PySpark" library.

<sup>1</sup> <https://www.forbes.com/sites/hbsworkingknowledge/2018/02/27/the-airbnb-effect-cheaper-rooms-for-travelers-less-revenue-for-hotels/>

<sup>2</sup> Euromonitor: Passport Portal

## Goal of the Analysis:

**MAIN GOAL:** Gain a perspective on how the customers review the different hotels, which are the **best in class and the lowest rated**, by location and further factors that can **influence** this decision: Being the main focus: **Location and Reviewer Nationality**.

**BUSINESS QUESTIONS:** Through the analysis there are **three Business questions** that help the reader understand the behavior of the **consumers and achieve the needed actions to increase positive reviews**:

1. Which hotels are the ones that have the best and worst score devised by location?
2. Which location is the one that has to watch out the most and what are my worst hotels based on the classification of goodness?
3. What other factor besides location and hotel are taken into consideration for giving a good or bad review?

## Analysis Deep Dive:

The first step into analyzing the data is to do a basic profiling on the data understanding **how many different hotels there are, categorization of cities and their importance towards the user** which ones are the most visited and less visited, how many different nationalities and which ones are the most vocal?

**This pre analysis** of the data will aid the reader to get a clear knowledge of the data set and begin gathering the **necessary information on the users and their clustering**.

### 1. Hotels & Location Analysis

- a. Most Visited Hotel, Ratings & Visits:** There are **1,170 hotels**, the most visited one has been visited **1,232 times** and the least visited **has been visited 1 time**, for **getting the average of rating the threshold occupied was of 60** to eliminate bias from the analysis.

Hotel_Name	location	avg(Reviewer_Score)	count(Hotel_Name)
Britannia International Hotel Canary Wharf	United Kingdom	6.769967532467543	1232
Park Plaza Westminster Bridge London	United Kingdom	8.641539763113405	1182
Strand Palace Hotel	United Kingdom	7.951236399604377	1011
Copthorne Tara Hotel London Kensington	United Kingdom	8.129572836801769	913
DoubleTree by Hilton Hotel London Tower of London	United Kingdom	8.619907940161129	869
Intercontinental London The O2	United Kingdom	9.519401041666697	768
citizenM Tower of London	United Kingdom	9.167014341590628	767
Holiday Inn London Kensington	United Kingdom	8.062162162162178	740
Park Plaza London Riverbank	United Kingdom	8.357103064066866	718
Grand Royale London Hyde Park	United Kingdom	7.742792109256459	659

- b. Most Visited Location:** As previously mentioned there are 4 cities, UK being the most visited one, followed by Amsterdam, the last three have similar visits.

Summary of columns Hotel Names: Most Visited Cities	
Location	NumLocation_Name
United Kingdom	65349
Amsterdam Netherl...	15648
Barcelona Spain	15214
Paris France	14811

**[ Learning: United Kingdom is the location with the hotels that have more visits, and the ratings are mixed, also by looking at the table of most visited cities UK has 4 times more visits that Amsterdam which is the second one, which means UK is the most important location to be aware of the impact that it may have above hotels, especially for those who have hotel chains across Europe ]**

## 2. Reviews & Nationality Analysis:

- a. Reviews Analysis:** The threshold of reviews is from 10 to 2.5, which has to be taken into consideration for further categorization, also there is an Average Score, which takes into account all historical data, for that reason the Reviewer Score has more accountability.

Summary of columns Review_Total_Negative_Word_Counts, Review_Total_Positive_Word_Counts, Reviewer_Score				
summary	Review_Total_Negative_Word_Counts	Review_Total_Positive_Word_Counts	Reviewer_Score	Average_Score
count	111022	111022	111022	111022
mean	20.289384086037003	18.853983895083857	8.374934697626752	8.388800417936194
stddev	32.4170026242823	24.25433818381829	1.6824347060653249	0.564340692822694
min	0	0	2.5	6.4
25%	2	5	7.5	8.1
50%	10	11	8.8	8.4
75%	25	23	9.6	8.8
max	402	395	10.0	9.8

- b. Nationality Analysis:** There are 201, to take into account which nationalities are the keenest to making a review and the less participative with the account of how many reviews have they made, and their impact on the final reviews.

- i. **Nationality Classification:** With this analysis the nationalities of UK, USA, Australia and Ireland are the ones that give the most reviews.

Reviewer_Nationality	Num_mostvocal_Nationalities
United Kingdom	57670
United States of...	8212
Australia	4416
Ireland	3919
Netherlands	1798
United Arab Emir...	1698
Canada	1671
Germany	1505
France	1471
Switzerland	1467
Saudi Arabia	1340
Israel	1251
Belgium	1224
Spain	1040
South Africa	1025
Italy	986
Turkey	971
Kuwait	907
India	874
Russia	737

Reviewer_Nationality	Num_leastvocal_Nationalities
Jersey	194
Taiwan	189
Nigeria	182
Iran	178
Ukraine	172
Croatia	164
Guernsey	136
Slovakia	129
Mexico	126
Jordan	125
Latvia	123
Estonia	122
Lithuania	122
Isle of Man	100
Argentina	95
Serbia	87
Chile	85

**[ Learning:** Again, **UK is on the top of Nationalities that make the most word of mouth**, leaving 7 times more reviews than USA, which is the second Nationality with most reviews made and **678 more than the least vocal Nationality**.

Top Nationalities are quite important and further analysis will be done to know how they influence on overall rating]

After the pre analysis, there is a clear profiling of which hotels and cities are the most important ones, as well as nationalities, to dig into the further impact on reviews by answering the previously set **Business questions:**

## BUSINESS QUESTIONS:

1. Which hotels are the ones that have the **best and worst score** devised by location?

To answer this question a filter was created to ensure equity between the reviews taking into consideration hotels with more than 60 reviews which is an average of 5 per month

**[ Learning:** Even though in the **pre analysis Barcelona was the second to last of most visited cities**, there is a clear trend that the hotels in **this location have the best reviews, even though they are not among the most visited, there are some factors that they are providing a food services to their customers**.

Following Barcelona, the best rated hotels are in United Kingdom, Amsterdam is not between the top best hotels]



## Best in class Hotels

Hotel_Name	location	avg(Reviewer_Score)	count(Hotel_Name)
Hotel Casa Camper	Barcelona Spain	9.74705882352942	102
H10 Casa Mimosa 4 Sup	Barcelona Spain	9.63714285714286	70
Catalonia Magdalenes	Barcelona Spain	9.611764705882356	85
Hotel Palace GL	Barcelona Spain	9.601587301587303	63
Nolinski Paris	Paris France	9.585245901639349	61
H tel D Aubusson	Paris France	9.561904761904765	63
Intercontinental London The O2	United Kingdom	9.519401041666697	768
Rosewood London	United Kingdom	9.504166666666667	72
Batty Langley s	United Kingdom	9.49540229885058	87
The Ritz London	United Kingdom	9.495000000000003	100

only showing top 10 rows

**[ Learning:** Barcelona is nowhere in the least favorite hotels, Amsterdam and United Kingdom are predominant within the top 10, which means further sentiment analysis can give out the reason why the rating of this hotels is among the lowest.]

## Least favorite Hotels

Hotel_Name	location	avg(Reviewer_Score)	count(Hotel_Name)
Savoy Hotel Amsterdam	Amsterdam Netherlands	5.881250000000002	64
Hotel Cavendish	United Kingdom	6.257021276595739	235
H tel De Castiglione	Paris France	6.426666666666667	75
Commodore Hotel	United Kingdom	6.48116591928251	223
Gainsborough Hotel	United Kingdom	6.490697674418598	215
Hallmark Hotel London Chigwell Prince Regent	United Kingdom	6.555555555555555	108
London Elizabeth Hotel	United Kingdom	6.564102564102564	78
Ozo Hotel	Amsterdam Netherlands	6.5813084112149545	107
Henry VIII	United Kingdom	6.582848837209296	344
Bloomsbury Palace Hotel	United Kingdom	6.616566265060234	332

only showing top 10 rows

- Which **location** is the one that has to **watch out the most** and what are the worst rated hotels based on the classification of goodness?

For this analysis the categorization used was based on the previous seen statistics of the average review per customer:

- Exceptional:** Reviews Above 9.6
- Very Good:** Reviews Between 9.5 -8.8
- Below expectations:** Reviews Between 8.7 - 7.5
- Alarming:** Reviews Below 7.5

### Categorization per location:

location	1.Alarming	2.Below expectations	3.Very Good	4.Exceptional
United Kingdom	91903.3000000059	108846.80000000211	108733.79999999708	233793.60000003438
Paris France	18521.69999999991	23545.09999999708	26707.600000000137	55463.5999999981
Barcelona Spain	16446.200000000004	23439.29999999979	28379.200000000237	62107.9999999977
Amsterdam Netherl...	19359.899999999896	24910.29999999843	29960.000000000193	57683.59999999774

**[ Learning:** Even though **UK** is the location with most visits and most, the overall reviews are “Exceptional”, if reviewed by category, **Amsterdam** and **Paris** are the locations which have the worst reviews: **Amsterdam** in “Alarming” and **Paris** in “Below Expectations”, there is a learning possibility in those locations, but further depth is needed to determine which hotels are the worst rated.

**Barcelona’s** hotels are **quite well reviewed**, for more depth an analysis of the best hotels is also needed aside from the ones that were already identified ]

### Hotels are the one that has to watch out the most:<sup>3</sup>

PerformanceReview	hotel_name	location	Reviewer_Score	avg(Reviewer_Score)
1.Alarming	Best Western Premier Kapital Op ra	Paris France	2.5	2.5
1.Alarming	Hotel le Lapin Blanc	Paris France	2.5	2.5
1.Alarming	Novotel Amsterdam City	Amsterdam Netherlands	2.5	2.5
1.Alarming	Marlin Waterloo	United Kingdom	2.5	2.5
1.Alarming	The Colonnade	United Kingdom	2.5	2.5
1.Alarming	Holiday Inn Paris Gare de l Est	Paris France	2.5	2.5
1.Alarming	Hotel Xenia Autograph Collection	United Kingdom	2.5	2.5
1.Alarming	Le Mathurin Hotel Spa	Paris France	2.5	2.5
1.Alarming	Elys es R gencia	Paris France	2.5	2.5
1.Alarming	NH Collection Amsterdam Barbizon Palace	Amsterdam Netherlands	2.5	2.5
1.Alarming	The Royal Park Hotel	United Kingdom	2.5	2.5
1.Alarming	Alexandra Barcelona A DoubleTree by Hilton	Barcelona Spain	2.5	2.5
1.Alarming	London Marriott Maida Vale	United Kingdom	2.5	2.5
1.Alarming	Mandarin Oriental Hyde Park London	United Kingdom	2.5	2.5
1.Alarming	COMO Metropolitan London	United Kingdom	2.5	2.5

**[ Learning:** The identified hotels are alarmingly below average (7.4), learning from these hotels is crucial to increase the average rating and the word of mouth.

Even though UK holds the best average reviews, in this analysis it is concluded that UK holds 54% of the worst rated.]

### 3. What other factor besides location and hotel are taken into consideration for giving a good bad review?

Summary of Nationalities & Reviews				
Reviewer_Nationality	1.Alarming	2.Below expectations	3.Very Good	4.Exceptional
United Kingdom	7.1	8.5	9.5	10.0
India	7.1	8.5	9.5	10.0
Qatar	7.1	8.5	9.5	10.0
Australia	7.1	8.5	9.5	10.0
France	7.1	8.5	9.5	10.0
Ireland	7.1	8.5	9.5	10.0
Lebanon	7.1	8.5	9.5	10.0
Belgium	7.1	8.5	9.5	10.0
Denmark	7.1	8.5	9.5	10.0
United States of America	7.1	8.5	9.5	10.0
Israel	7.1	8.5	9.5	10.0
Canada	7.1	8.5	9.5	10.0
New Zealand	7.1	8.5	9.4	10.0
Iran	7.1	8.5	9.2	10.0
Cyprus	7.1	8.5	9.2	10.0
Papua New Guinea	6.3	null	8.8	9.6
Syria	6.3	null	null	10.0
Brunei	5.8	8.3	9.2	10.0
Dominican Republic	5.8	8.3	9.2	10.0
Ecuador	5.8	7.9	9.2	10.0
Bahamas	5.8	null	9.2	10.0
Democratic Republic of the Congo	5.8	null	null	null
Northern Mariana Islands	5.8	null	null	null
Nepal	5.8	null	null	null
El Salvador	5.4	null	null	10.0
Malawi	5.4	null	null	10.0
French Polynesia	5.4	null	null	null
Guatemala	5.0	8.3	9.2	10.0
Turks Caicos Islands	5.0	7.9	null	10.0
Cook Islands	5.0	null	null	null
Paraguay	4.2	null	null	10.0
Dominica	3.8	null	null	null
Mali	2.9	8.5	null	null
Cameroon	2.9	null	null	10.0

**[ Learning:** Making a further analysis, even though UK, US, Australia and Ireland are the most vocal they are mostly leaving good reviews since they are within the top 10, being the main insight that this nationalities will give a good review if they like the service provided to them.

Nationalities like Mali, Dominican, Paraguay etc., are more inclined to have less word of mouth, but provide fewer good reviews, a conclusion could be the distance they travel, total vacation time in those countries that generate a higher expectancy.]



## Conclusions & Recommendations:

1. United Kingdom the key location to keep track on since is the most visited one and the one that is between the best and worst hotel reviews.
2. The location with above best reviews is Barcelona, a further analysis based on the content of the reviews is advised to get the best qualities the location has and replicate them among other chains within and outside the city.
3. It is recommended to make the same sentiment analysis in the hotels categorized as “Alarming” since they are alarmingly below the expected average, and in order to increase the average rating and provide a good service the analysis can provide the most used words by the customer and find patterns.
4. Since the most vocal nationalities (UK, US, Australia and Ireland) give above all good reviews, a loyalty program is recommended towards these nationalities to promote word of mouth on the hotels, as well as an online marketing campaign targeting these nationalities.