Analysis Hotels Reviews in Europe (EDA) Report

Abstract:

This is the first project for the data science bootcamp T5, which is an exploratory data analysis of the hotels reviews in Europe data set using Python and its libraries Pandas, Numby, Matplotlib, seaborn, word cloud and sklearn. This project discusses the work I did to explore the customers experiences in visiting the Europe Hotels in order to help the Hotel's owners to enhance their facilities.

Design:

The dataset is large and informative, So, the Hotel's Owners will get benefit from it and will help them improve their services and facilities and to point out the customers complaints. While working on this problem, I answered the questions below.

1. the Average rating of all hotels?
2. Distribution of reviews to highly rated hotels?
3. Rating of the worst hotels?
4. What are the services that get the most complaints?

The value for the hotel's owners is that it will save the Hotel's and improve their work quality and reputation as a hotel.
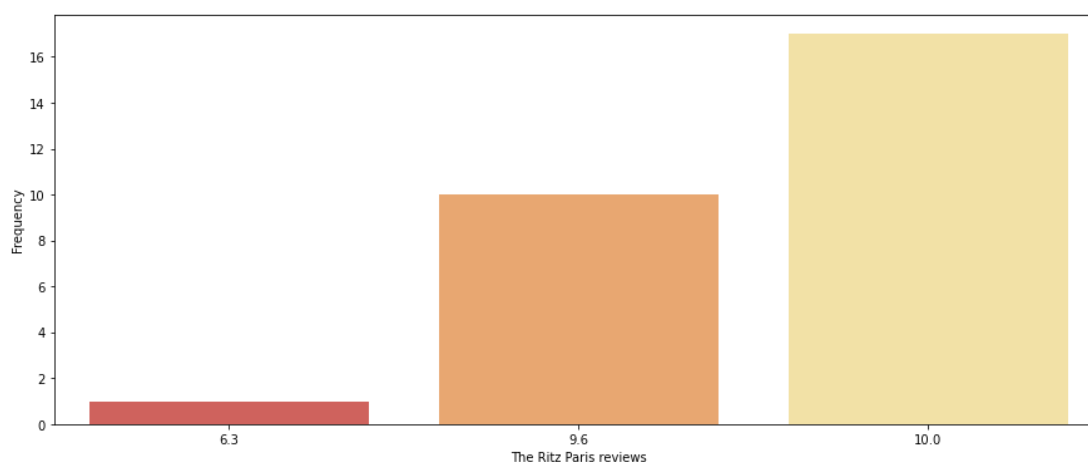
Dataset:

The csv file contains 17 fields. The description of each field is as below:

1. Hotel_Address: Address of hotel.
2. Review_Date: Date when reviewer posted the corresponding review.
3. Average_Score: Average Score of the hotel, calculated based on the latest comment in the last year.
4. Hotel_Name: Name of Hotel

5. Reviewer_Nationality: Nationality of Reviewer
6. Negative_Review: Negative Review the reviewer gave to the hotel. If the reviewer does not give the negative review, then it should be: 'No Negative'
7. ReviewTotalNegativeWordCounts: Total number of words in the negative review.
8. Positive_Review: Positive Review the reviewer gave to the hotel. If the reviewer does not give the negative review, then it should be: 'No Positive'
9. ReviewTotalPositiveWordCounts: Total number of words in the positive review.
10. Reviewer_Score: Score the reviewer has given to the hotel, based on his/her experience
11. TotalNumberofReviewsReviewerHasGiven: Number of Reviews the reviewers has given in the past.
12. TotalNumberof_Reviews: Total number of valid reviews the hotel has.
13. Tags: Tags reviewer gave the hotel.
14. dayssincereview: Duration between the review date and scrape date.
15. AdditionalNumberof_Scoring: There are also some guests who just made a scoring on the service rather than a review. This number indicates how many valid scores without review in there.
16. lat: Latitude of the hotel
17. lng: longtitude of the hotel

Communication:

The Distribution of the reviews of Ritz Paris

The Distribution of the reviews of Hotel Ritz_Paris

Distribution of reviews to highly rated hotels

The Distribution of the reviews of Hotel Liberty

Rating of the worst hotels and Why did it get this rating

Conclusion:

After data cleaning and analysis (EDA), we recommend that you choose one of the hotels with high five ratings; for your security, comfort and excellent services.