Metvy

Data Analytics Program

Task

Metvy

Task

Acknowledgments:

The data was scraped from Booking.com. All data in the file is publicly available to everyone already. Please be noted that data is originally owned by Booking.com.

Data Context

This dataset contains 515,000 customer reviews and scoring of 1493 luxury hotels across Europe. Meanwhile, the geographical location of hotels is also provided for further analysis.

Data Content

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

- Hotel_Address: Address of hotel.
- Review_Date: Date when the reviewer posted the corresponding review.
- Average_Score: Average Score of the hotel, calculated based on the latest comment in the last year.
- Hotel_Name: Name of Hotel
- Reviewer_Nationality: Nationality of Reviewer
- Negative_Review: Negative Review the reviewer gave to the hotel. If the reviewer does not give a negative review, then it should be: 'No Negative'
- Review_Total_Negative_Word_Counts: Total number of words in the negative review.

- Positive_Review: Positive Review the reviewer gave to the hotel. If the reviewer does not give a negative review, then it should be: 'No Positive'
- Review_Total_Positive_Word_Counts: Total number of words in the positive review.
- Reviewer_Score: Score the reviewer has given to the hotel, based on his/her experience
- Total_Number_of_Reviews_Reviewer_Has_Given: Number of Reviews the reviewers have given in the past.
- Total_Number_of_Reviews: Total number of valid reviews the hotel has.
- Tags: Tags reviewer gave the hotel.
- days_since_review: Duration between the review date and scrape date.
- Additional_Number_of_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.
- lat: Latitude of the hotel
- lng: longitude of the hotel

1. I**mport the necessary modules.**
2. **Load the data set. Find the number of rows and columns. Look for missing values.**
3. **Drop the unnecessary columns.**
4. **Look for any anomalies in the dataset. Clean the data.**
5. **Create a correlation heatmap and write the inference.**
6. **Plot the worst top 10 hotels according to the reviewer score and top best hotel according to the reviewer score. Create a bar plot**.