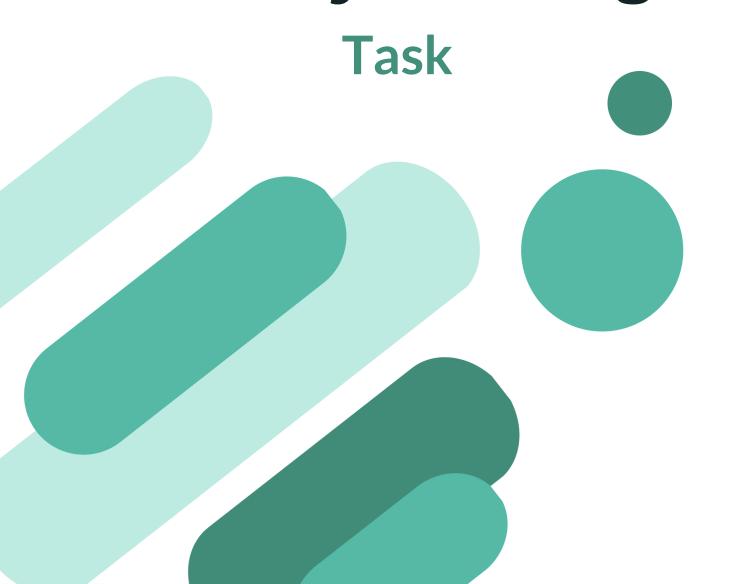
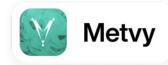


# Data Analytics Program





# **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**

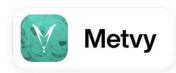
This dataset contains 515,000 customer reviews and scoring of 1493 luxury hotels across Europe. Meanwhile, the geographical location of hotels is also provided for further analysis.

### **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
- Ing: longitude of the hotel
- 1. Import 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.