Kori Caldwell

Raamkumar Thiyagarajan

Serkan Comu

Shashank Goswami

### ***Business Objective***

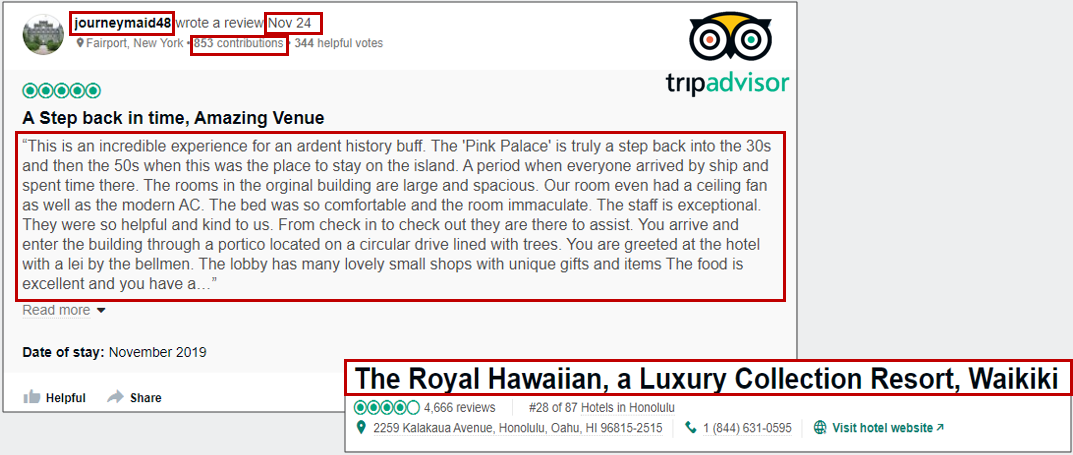
The motivation for this project is due to the overwhelming options of reviews consumers have on TripAdvisor. Consumers have come to heavily rely on reviews to choose a restaurant, hotel, activity, just about anything. However, sifting through all the reviews can be daunting. For instance, a popular destination such as New York, which has approximately 60 million tourists a year, has an abundance of hotels. Even by limiting a consumer’s hotel search by under $425, free Wi-Fi, and 4-star rating or above produces 182 results.

This is not surprising considering 390 million unique monthly visitors use TripAdvisor (2019), and more than 10,000 companies are added on TripAdvisor every week. As such, the primary goal of the analysis was to build a model that would provide a hotel recommendation in Honolulu, Hawaii, to Travelbug0082, a user-selected based on their TripAdvisor high contribution score. The suggestion was based on popular features (i.e., cleanliness, affordability, location, etc.) scraped from reviews written in the last three years by other travelers and then matched to the preferences indicated by Travelbug0082 in his/her 20+ reviews from hotels that they have stayed from around the world.

### ***Methodology***

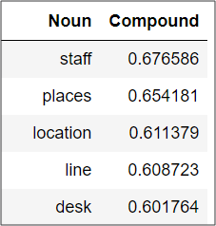
The first step to building the model was to use the package BeautifulSoup in Python to scrape reviews on TripAdvisor for at least eight 3-star hotels in Honolulu, Hawaii. The data extracted per review included the hotel name, user review, the date of the review, and the name and the total number of reviews per reviewer (as shown in Figure 1). The date of the review was used to limit our analysis to reviews from the last three years, and the total number of reviews per reviewer was used to select Travelbug0082 for whom we made the recommendation.

**Figure 1. Visualization of data scrapped**



Then, a data pre-processing technique was implemented to remove stop words or punctuations that did not add value to the overall analysis. After the pre-processing, the reviews were handled through a custom-built program created in Python to identify nouns and adjectives in a sentence. Finally, a sentiment analysis was conducted, and a composite score was calculated, ranging from -1 (negative sentiment) to 1 (positive sentiment) and was assigned to the features, which represented the polarity of that noun mentioned throughout all the reviews for one hotel ( as shown in Figure 2).

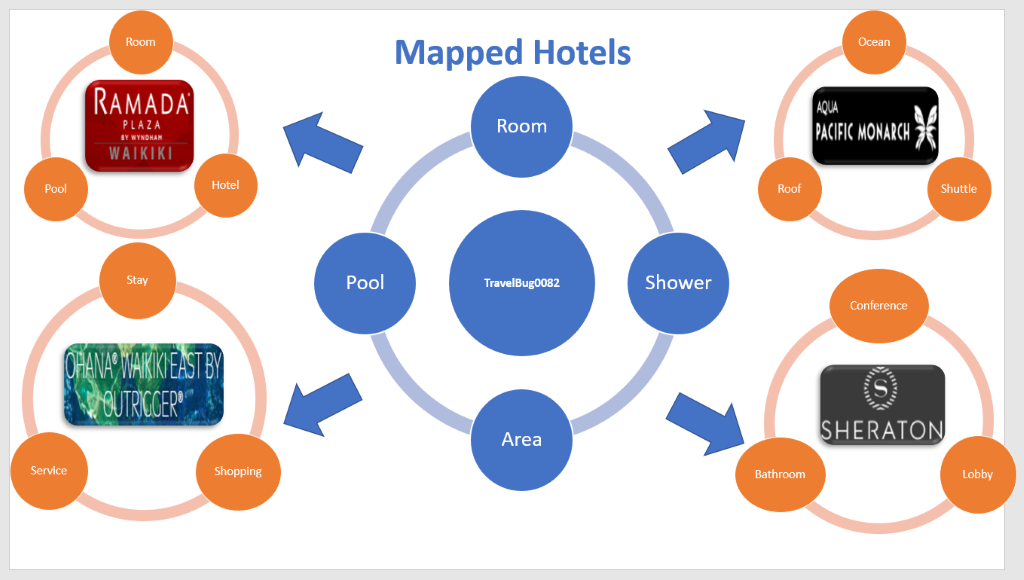
**Figure 2. Features of the hotel from the reviews ranked by high sentimental score**



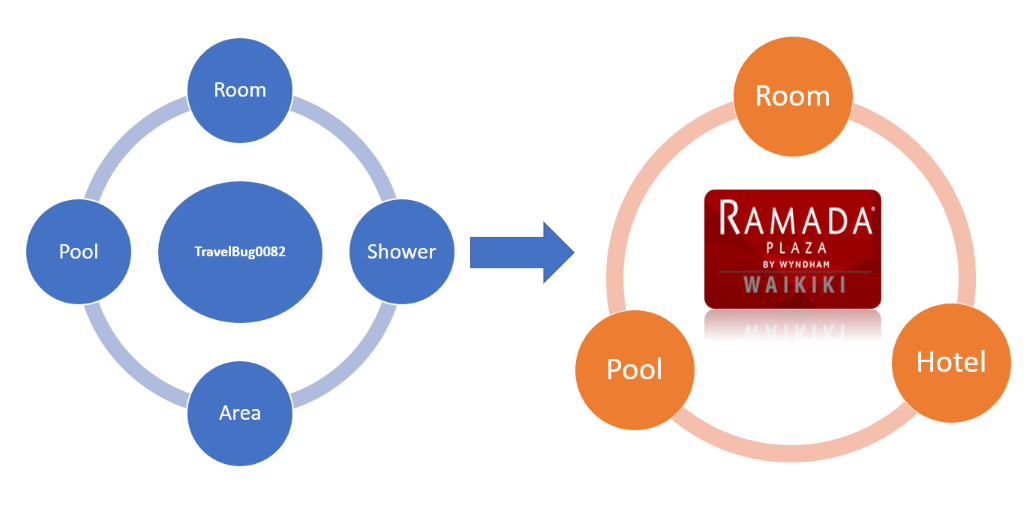
### ***Major Findings***

Based on the topics and sentiment extracted from approximately 4,000 reviews and eight hotels in Honolulu, Hawaii, the Ramada Plaza by Wyndham was recommended for Travelbug0082 (as shown in figure 3 and figure 4).

**Figure 3. Mapping of features of hotel to features most preferred by customer**

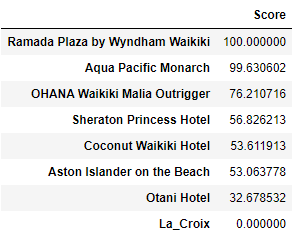


**Figure 4. Recommendation of hotel based on high sentimental scores of mapped features**



Also the recommendations were ranked from the most recommended to least recommended based on their cumulative sentimental scores for mapped features (as shown in figure 5).

**Figure 5. DataFrame showing the ranking of recommendation**



The hotel Ramada Plaza by Wyndham was most recommended due to its positive sentiment regarding the hotel’s pool area and helpful/accommodating staff, all of which were topics emphasized in the Travelbug0082’s reviews of other various hotels (see Figure 6).

**Figure 6. Word Cloud of topics emphasized by contributor Travelbug0082**



### ***Limitations and Recommendations***

A limitation of the model was the inability to identify word pairings confidently. For example, in a single sentence of a review, two nouns may have been recognized, each with a different sentiment (i.e., one positive the other negative). However, the current design of the model was unable to distinguish which noun was paired with which adjective that described it’s polarity. As such, a recommendation for future analysis is to improve the model to identify word pairings better. Once this limitation is refined and advanced, the model can be potentially be expanded from hotels to other industries such as restaurants.

**References**

1. TripAdvisor Network Effect and the Benefits of Total Engagement. (2019, February 11). Retrieved from https://www.tripadvisor.com/TripAdvisorInsights/w828.