

The Las Vegas Strip is the heart of Sin City, with a 4.2 mile stretch full of hotels, casinos, restaurants, shops, performers, clubs, and more. No matter what attracts someone to Vegas, at the end of the night, all travellers need somewhere to stay comfortably and safely. Part of the allure of the entertainment capital of the world is that it is easy to find refuge in hotels within walking distance of all of the attractions to visit. For these hotels, their livelihood is based on customer service and providing a clean and relaxing stay for all guests, no matter where they come from or why. Online review platforms, such as TripAdvisor, Yelp, and Google Reviews, provide a forum for customers to rate their experience at a hotel and give specific comments. Reviews are critical to the likelihood of consumers choosing to book a particular hotel. If a hotel had a one star rating and the reviews say that it was dirty, the door did not lock, and the staff were rude, most people would avoid that hotel at all costs. On the contrary, a hotel with rave reviews and mentions of luxurious amenities and positive experiences will attract more new customers looking for a similar experience. Because of that, it is incredibly important for hotels to be aware of their reviews and ensure that they do everything possible to keep high reviews if they want a successful business. Good reviews tell hotel staff where to continue the practices they use and what customers like, and bad reviews highlight where the hotel can improve customer experience and consequently, ratings.

In the context of existing works relating to online consumer reviews, it is clear that there is a proven relationship between customer reviews and profitability. In a 2019 study of TripAdvisor ratings, it was found that every star in a given hotel's rating is equal to a \$280 per booking transaction (Jenq 2019), which financially rewards the hotel for positive reviews. Each review represented \$0.12 per booking transaction (Jenq 2019), which also indicates that not only does

the rating add value to a hotel, but the quantity of reviews available adds value as well, regardless of if the reviews are primarily positive or not.

It is also found in another 2019 study of Yelp reviews that restaurants, despite being in a different sector of the hospitality industry, have four main categories of guest review content: food/taste, value, location, and experience. Each category has associations with either positive or negative reviews, as determined by category-specific keywords. Though each category contained keywords that were both positive and negative, such as ‘expensive’ for a negative keyword and ‘reasonable’ for a positive keyword in the value category. The taste category is more highly associated with positive reviews and the value category is more highly associated with negative reviews, as determined by category-specific keywords (Luo 2019). Because of the nature of the restaurant industry, and how reviews have similar effects on restaurant success as to hotel success, a similar conclusion might be able to be drawn about the hotels on the Las Vegas Strip. This leads to the question that this paper aims to address, which asks what the best categories (features) and modeling type are in order to predict Las Vegas hotel ratings.

A model that would predict hotel ratings would be extremely beneficial to a hotel company. By knowing the features most correlated to both high scoring reviews and low scoring reviews, there is a way for hotel staff to know what things matter most to their guests for a fulfilling stay at their hotel. Hotels, investors, and forecasters/analysts alike are able to better understand the likelihood of random negative reviews, and know if the negative reviews that a hotel gets are statistically significant and need to be addressed appropriately. As opposed to a lexicon approach that focuses on the content of the review, this approach focuses more on the presence of hotel amenities and reviewer statistics to make successful predictions.

One such approach focused on the relationship of quantitative variables and review score of TripAdvisor reviews of Las Vegas hotels was done in a 2017 study by Moro et al, which shows that this question is worth pursuing if others see value in it as well. Their approach showed that the two most important features to determine hotel review score were related to 1) the statistics of the TripAdvisor user's helpfulness and longevity, and 2) the length of stay at the hotel (Moro et al 2017). Most studies focused around hotel reviews tend to use a lexicon approach to categorize and display the data in the content of the reviews, but a numerical approach has more potential to contribute since less has been done with it. Moro et al placed the groundwork for this project, which will continue identifying TripAdvisor features most associated with predicted review scores, but using a more simplistic model, which is hypothesized to create a more accessible prediction. By examining different classification and regression models and comparing their cross validation and F1 scores, this project will identify the most effective machine learning algorithm and features to predict hotel scores for Las Vegas hotels by TripAdvisor users.