

Yelp Reviews: Las Vegas

We were given the data set containing Yelp reviews for restaurants located in Las Vegas. Specifically, we wanted to make observations between certain business types using unsupervised learning. Some of the observations included comparing the average rating for each restaurant type, dividing the business types into clusters, comparing unique lemmas used in positive and negative reviews, and utilizing principle component analysis. As a result, we were able to discover some interesting observations, particularly centered around Casino-type businesses since Las Vegas is known for its gambling.

Firstly, we found the average rating and the standard error between each type of business. The highest average ratings came from Korean restaurants and wine bars. The lowest average ratings came from primarily American food (diners and burgers). We also separated the business types into 12 different clusters. The casino business type was found to have the most similar attributes to the resort business type. Additionally, we wanted to see the frequency of the term “cheap” in comparison to the frequency of the term “nice.” This was to see if cheap prices affected the quality of food. Sushi bars had reviews that used the term “nice” used frequently, but the term “cheap” was not present that often. On the other hand, Casinos had the terms “cheap” and “nice” occurring in many reviews.

What lemmas suggest that a review is positive vs. negative and what can that tell us? A negative Casino review may mention: Lasagna, miso, and oyster. Perhaps lasagna, miso soup, and oysters are not good food options at Las Vegas casinos. A neutral Casino review may mention: Starbucks. Having Starbucks at a casino is usually tied to a neutral experience. A positive Casino review may mention duck, roasted, treat, barkeep. Perhaps duck is enjoyable at Las Vegas casinos, as may be roasted meat, treats, and helpful barkeeps.

By analyzing angle distance between documents, we learn that casinos are most similar to resorts in Las Vegas. What can we learn by looking at lemmas associated with positive and negative reviews for resorts? A negative review for a resort may mention: sushi. It may be best to avoid sushi the next time you visit a Las Vegas resort. A neutral review for a resort may mention: butter, bartender, chocolate. It seems butter, the presence of a bar/bartender, and chocolate in a review predict a neutral experience. A positive review for a resort may mention: dish, doughnut, spicy, cappuccino. The next time you visit a Las Vegas resort, have a dish at a restaurant, a doughnut, spicy food, or a cappuccino, if you are looking for an enjoyable experience.

Along with the insight from the previous analysis, we decided to take a separate graphical approach to gaining some insight into this dataset. We decided to focus on the combined categories of Casinos and Resorts (as previously mentioned, these are similar categories, and it seems reasonable to combine them for a task like this) and plot their locations on a graph using the included longitude and latitude values. After having that graph, we split the data into reviews that had 1/2 and 4/5 star reviews, to signify negative and positive reviews respectively. What we found was the majority of the reviews for Resorts/Casinos were positive, which we knew from earlier. However, it seemed that a large majority of the positive reviews seemed to be contained on the strip, whereas the negative reviews seemed to be relatively uniformly spread out. We investigated this further, by applying clustering to the points, creating automated geographical clusters. The clusters confirmed our results, showing that the majority of the positively rated resorts and casinos happen to be in the strip.

