One of Las Vegas' most famous assets is the Las Vegas Strip, located in the heart of the city and littered with hotels and casinos. Reviews are essential to running these businesses and attracting new visitors to "Sin City". Using the Las Vegas Strip data set from the Machine Learning Repository, our group aims to create a model to predict hotel reviews on TripAdvisor based on a column of data called "traveler type". This column is made up of five categorical variables that group each traveller by their travelling companions: families, friends, solo, business, or couples. Each category reflects a different demographic of travellers, and our research will focus on how each traveler type will review the twenty-one hotels in our data set, based from which traveler type they are classified as.

The hypothesis for our project is that certain reviewers that have a different traveller type, such as couples and friends, will have higher average reviews than other reviewers with different traveller types. Our main reasoning for forming this hypothesis is due to the general atmosphere of the Las Vegas strip and how it is mostly geared toward a certain clientele as opposed to being inclusive of various different traveller types. This hypothesis is important because it can provide lots of significant information for casinos/hotels on the Las Vegas strip. Some examples of this is by showing that certain groups tend to give lower reviews than others, then the respective casinos/hotels can take measures to properly compensate for those particular groups. For example, let's say that families tend to leave lower reviews than other traveller types. To properly respond to this, casinos/hotels can form certain amenities such as a children's area or more family friendly sections within their buildings.

Since we are trying to predict what kind of travelers will be more likely to leave better reviews, we will start with a boxplot in order to see if there are any differences between the traveler types. Then, to look for potential variables that predict it the best, we will check the

correlations between the variables along with a pair plot. When looking at the pair plot, we will separate it based upon the traveler types to see if there are any particular trends that may be of use to us. After doing all this, we will have a good idea of what variables we should focus on and will make a new set of data that just have those variables' entries listed. From there we will build the model and then measure the results by seeing if it returns the expected prediction review ratings based on what we plug in.