Yelp Business Ratings Prediction:

Objective: Identify which factors contribute most to whether a business has a high or low rating.
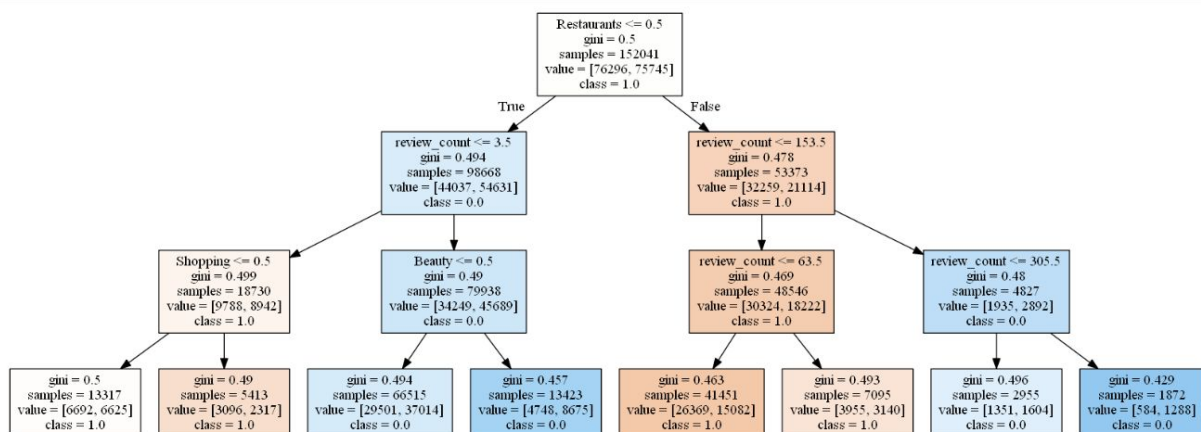
Dataset Description: Dataset was obtained from Kaggle.It includes information regarding business attributes to check ins and reviews on 174,000 businesses .Used 4 of the 7 provided datasets;Business,Business Attributes,Tips, Checkins.

Project Outline:
1) Data Preparation
2) Exploratory Analysis
3) Detailed Analysis

Findings & Recommendation:
Finding:1

1. Whether the business is a restaurant or not is the most defining factor for ratings, since it splits the tree at the first level.
2. A more interesting finding is that if the business is not a restaurant, the data is split by review count <= 3.5. This is a really low number of reviews, suggesting that people tend to review restaurants much more than they review other types of businesses on Yelp.

Thus, restaurants, coffee shops, etc. leverage Yelp reviews as a way to boost their online presence and attract more customers.

Finding 2:
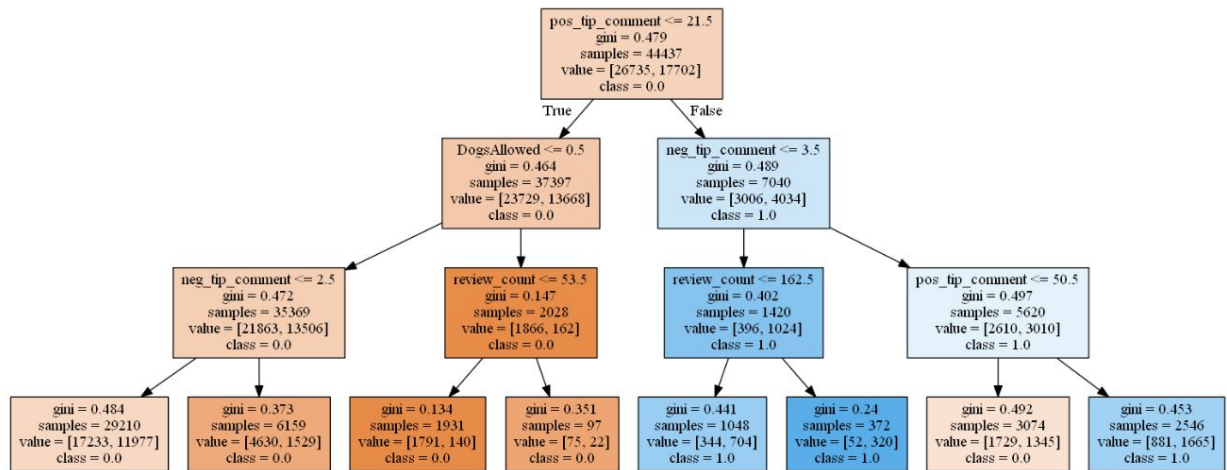
1. Number of positive tip mentions is the most defining factor… if number of positive count is less than 21.5 and number of negative tip count is less than 3.5, then there is a high chance of high rating for the restaurant. If the number of negative tip count is higher but number of positive tip comment is higher than number of negative count than there is high chance of high rating.
2. Surprisingly, for restaurants with a low number of reviews, Dogs Allowed is an important defining factor for whether they have a high or lower rating.

So it's very important that managers give special attention to the comments that customers give. In case they are negative, they should talk to customers immediately to know their reasons of dissatisfaction and improve their restaurant and rating.

Finding 3:

cluster	stars	review_count	stars_dum	US_states	is_open	BusinessAcceptsCreditCards	BikeParking	Alcohol	DogsAllo
0	3.138767	36.017621	0.114537	0.916300	0.964758	0.004405	0.986784	1.000000	
1	3.319785	9.345092	0.292945	0.544479	0.915644	0.000000	0.995399	0.984663	
2	2.963235	79.602941	0.029412	1.000000	1.000000	0.000000	1.000000	1.000000	

The finding based on the above is that the Starbucks in the US is rated lower than those outside of the US (which valideates our finding). So, for the cluster 2 where Starbucks is in the US we see that it has the lowest star rating. Also, we can see that those that have the highest star

ratings(cluster 1) are reviewed the least. We can see that the results for Starbucks stayed almost the same as for McDonald's and differes from Subway.

Customer perceptions of a restaurant can be very different within the US versus outside the US. Chain restaurants can do some market research so that they can tailor their food offerings and their marketing strategy to each particular country/region.