

Goal: What is the relationship between the Yelp rating of one’s experience and the positivity / negativity of their review for a restaurant in State College.

Identify the procedural steps in providing an *analytics* solution for the goal stated above.

Step 1	Obtain auxiliary data, i.e. list of positive / negative words, list of common words
Step 2	<del>Scrape each review provided on Yelp’s website</del> Use an API to pull reviews directly from the Yelp server
Step 3	Obtain a count of the number of words for each review
Step 4	Complete the necessary text processing steps for each review: 1) remove punctuation, 2) convert all text to lowercase, 3) remove common words
Step 5	Use a lookup process/function for each word in the reviews – obtain the value for each word found in the positive / negative list
Step 6	Sum the values for each review – adjust this sum for review length
Step 7	Summarize/visualize/correlate the scores from the reviews with the star ratings

What elements of Task #2 are

Statistical in nature	Data Science in nature
<ul style="list-style-type: none"> <li>• Obtaining summaries</li> <li>• Adjustment of measures to account for possible biases</li> </ul>	<ul style="list-style-type: none"> <li>• Setting up API connection to retrieve Yelp data</li> <li>• Text processing steps</li> <li>• Lookup function</li> </ul>