A PROJECT REPORT

ON

“YELP TIPS EXTRACTION FOR MENU RECOMMENDATION”

BY

|  |  |
| --- | --- |
| JITENDRA PUROHIT | CWID – 10412835 |
| KEVAL JAIN | CWID – 10416387 |
| PRATHAMESH PARCHURE | CWID – 10417041 |
| SIDDHARTH GANDHI | CWID – 10429142 |

UNDER THE GUIDANCE OF

PROF. RONG LIU



**Motivation**

Restaurant goers often face the problem of picking a restaurant. They often go through multiple reviews to come to a decision. The restaurant owners too must go through each and every long review to understand what the customers are thinking about their restaurants. We device a model where we generate relevant tips about the restaurants.

**Objective**

To design a model for generating relevant tips which could be used for recommending the restaurant owners and potential customers about the restaurants.

**Introduction**

To facilitate the restaurant owners and potential customers, we have developed a model for recommending both the owner and customers about the positives and negatives of that restaurants. As far as customers are concerned, they will be informed about all the positives tips about the restaurants i.e., food dishes, ambience, location, etc. As far as restaurants owners are concerned, the will be given tips about all the positive as well as negative tips about the restaurants so that they can improve upon the negatives of the restaurants.

**Methodology**

We built a web scrapper to scrape reviews of 50 restaurants across New York. The total reviews scrapped were close to 30,000. We used this data as testing dataset to generate tips and derive sentiments out of them.

We built another scrapper to scrape the review highlights provided by Yelp.com. The review highlights are nothing but those reviews that are talked about frequently by many customers. We used these highlights to prepare a training lexicon.

The lexicon was prepared using Natural Language Tool Kit (NLTK) package’s bigrams and unigrams. These unigrams and bigrams used adverb and adjective to prepare the lexicon.

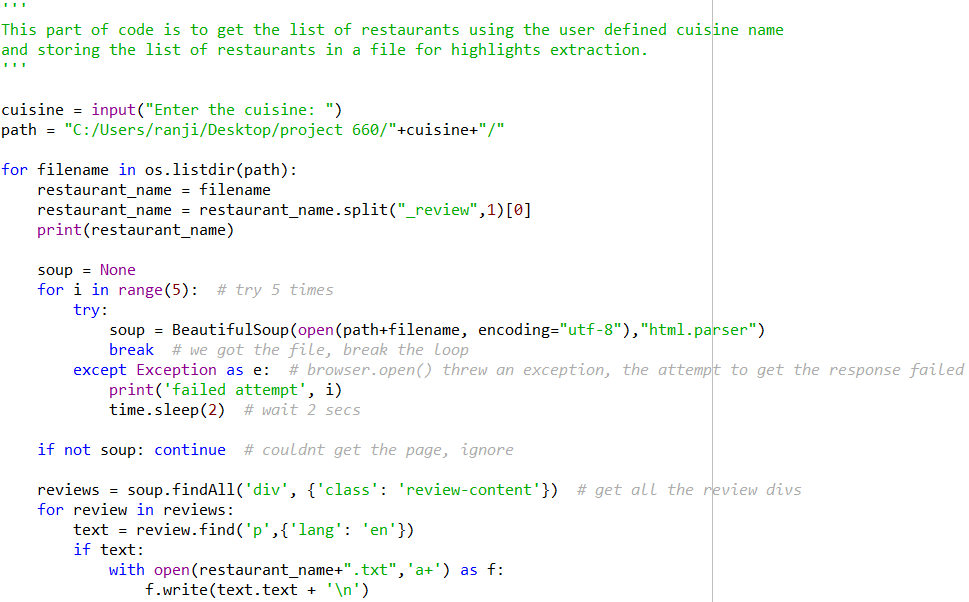
Using the generated lexicon, which would contain expressions like ‘excellent’, ‘very good’, ‘poor’, ‘bad’ etc. we filter out the highlights to get only the relevant highlights, which are actual customer reviews about the place.

Having filtered out the relevant restaurant reviews, we put another filter to get restaurant reviews that talk only about relevant keywords like ‘food’, ‘restaurant’. To achieve this, we create a list that contains only relevant nouns. We use textblob package to extract this list of nouns. Using this newly generated list, using word2vec model we find out the similar words to ‘food’, ‘restaurant’ or similar words of interest.

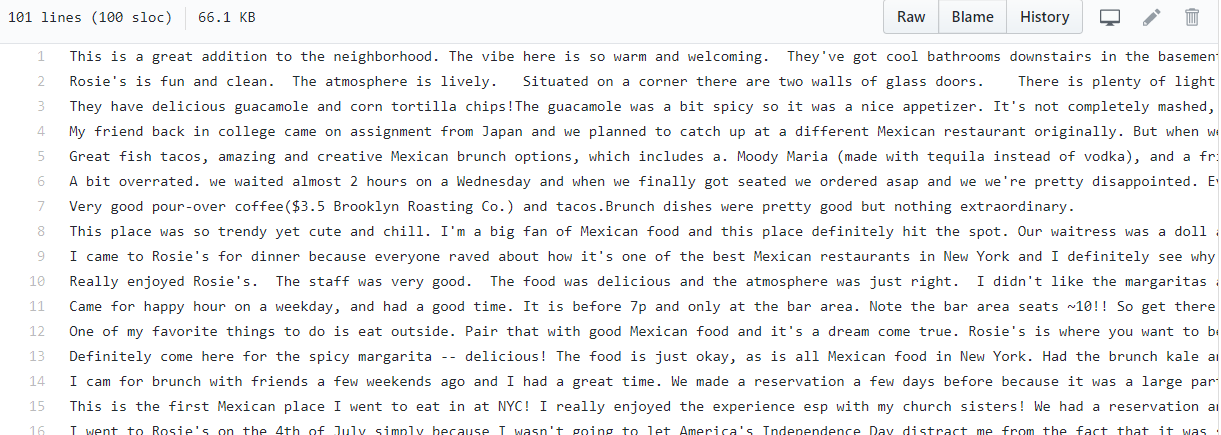
Using the list of similar and relevant words’ list we extract the relevant tips from the testing dataset of reviews.

It is important to understand whether a particular tip is positive or negative. Restaurant owners would work on the negative reviews and try to improve upon those, and the customers would want to go to a restaurant after looking at the positive reviews. We perform sentiment analysis on these relevant extracted tips to find out the polarity of the tips.

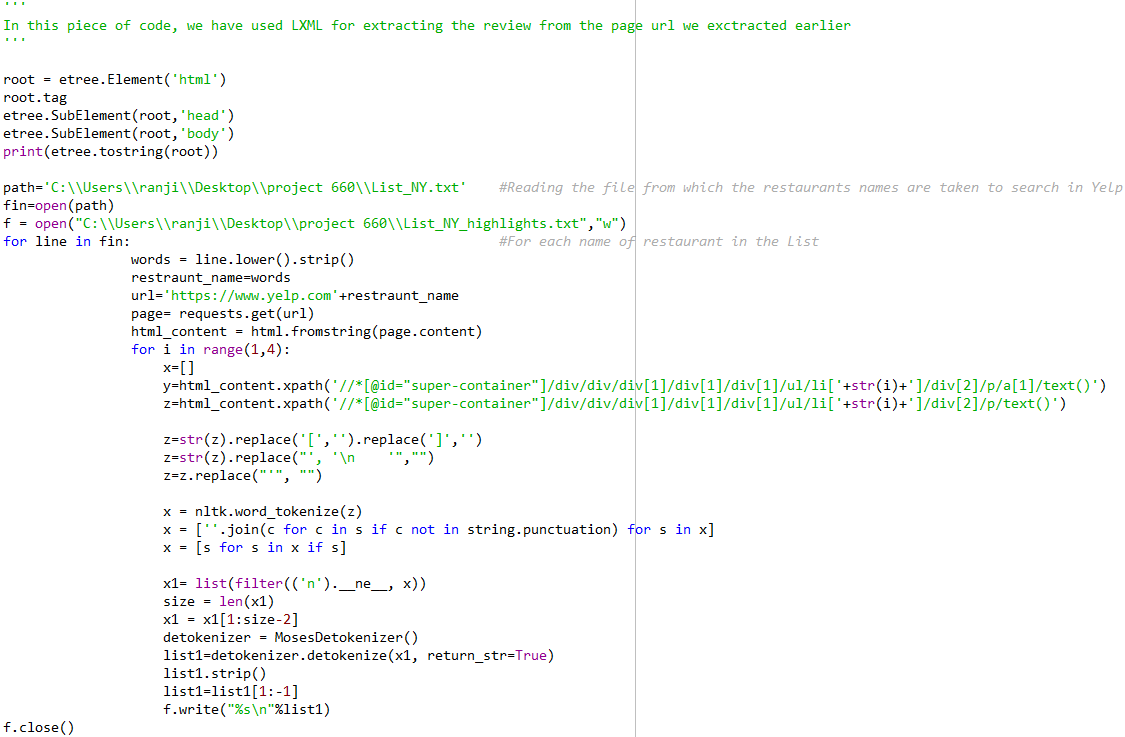
Step 1: Scape restaurant reviews



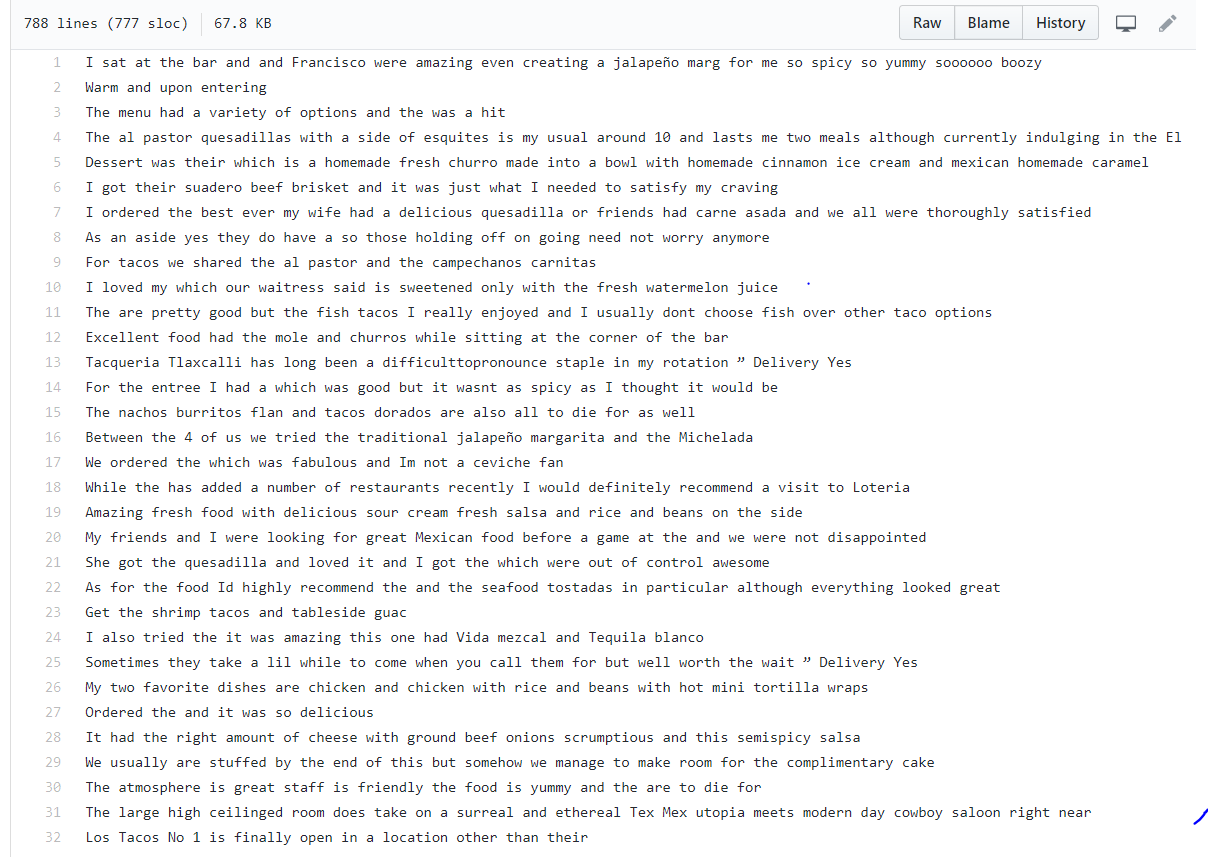
Output:



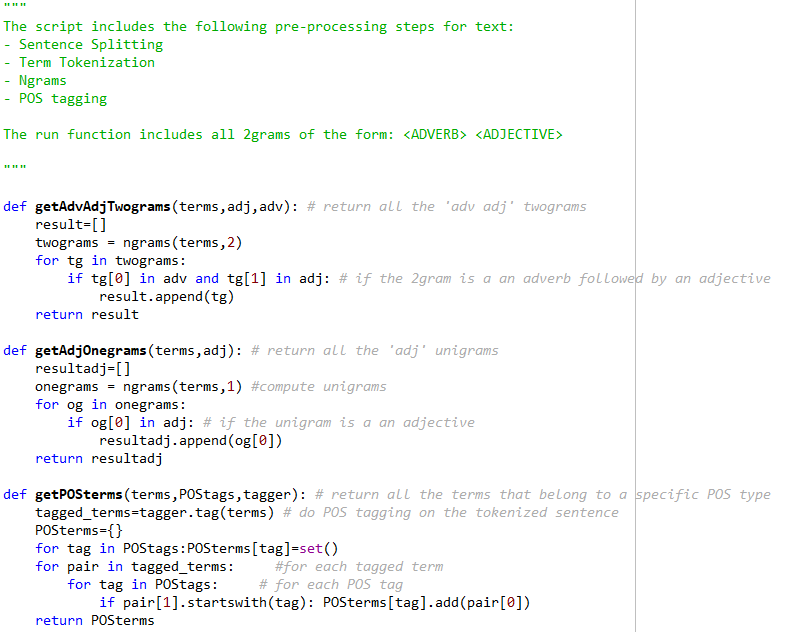
Step 2: Scrape review highlights



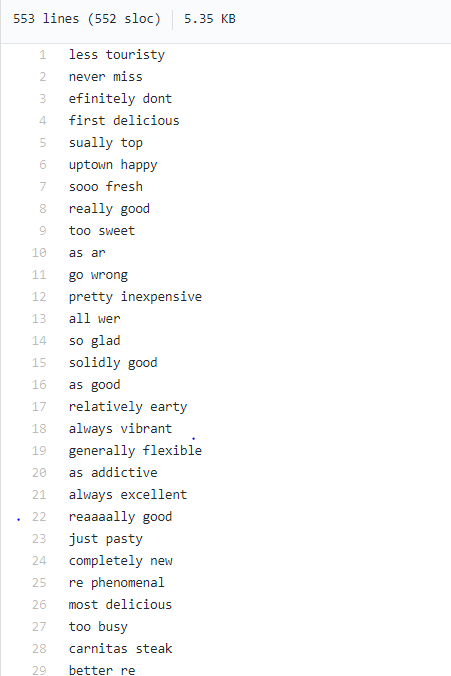
Output:



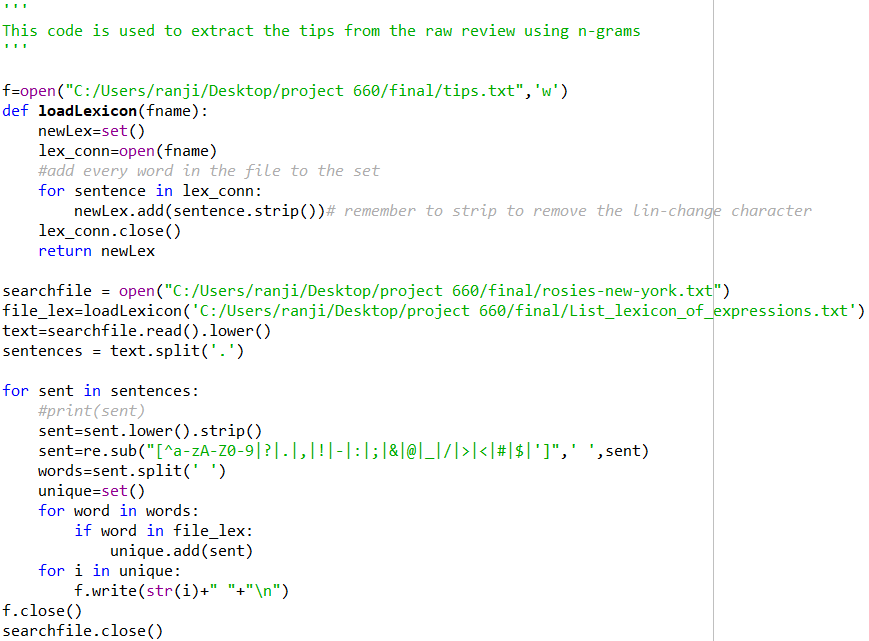
Step 3: Prepare training lexicon



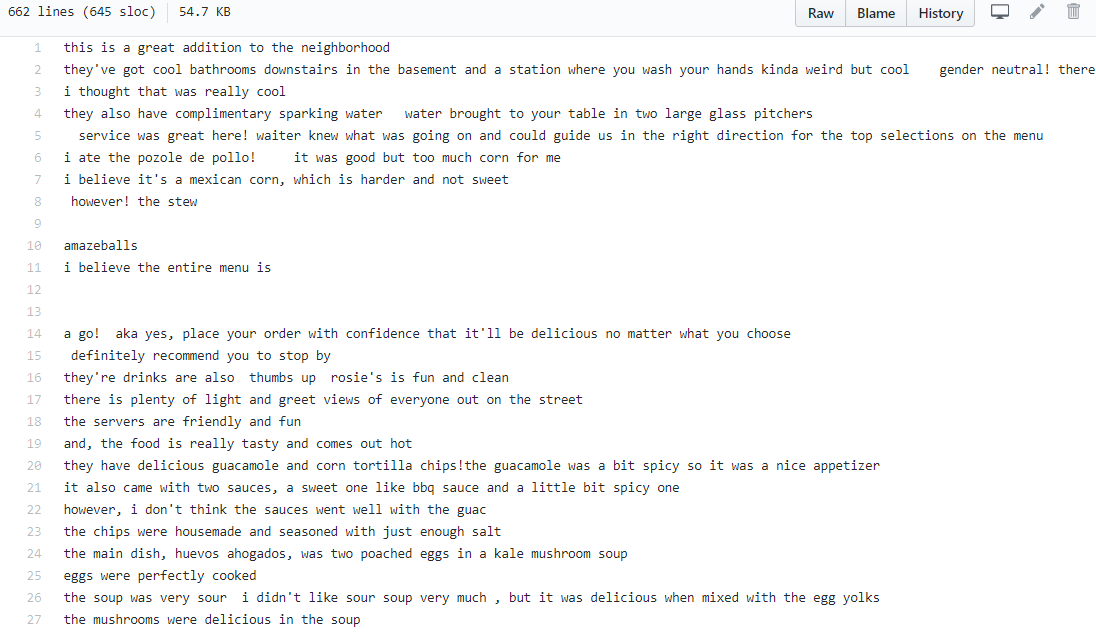
Output:



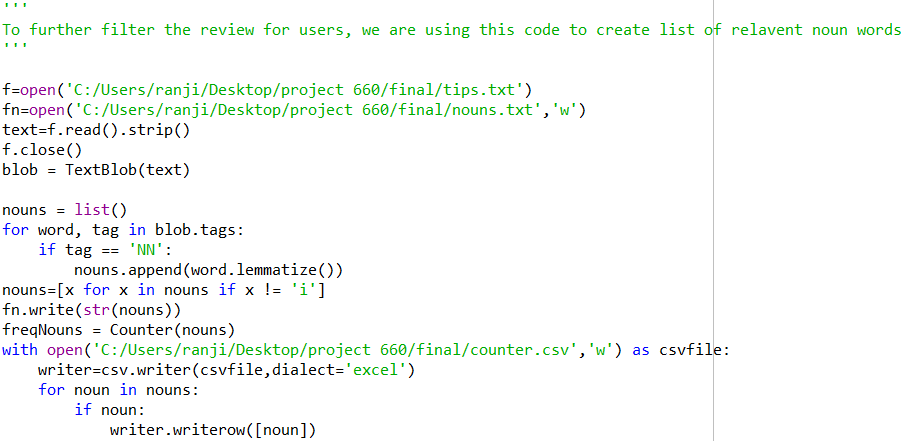
Step 4: Use Lexicon to extract relevant reviews (interim tips)



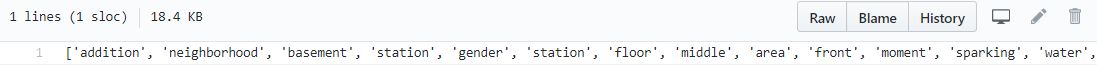
Output:



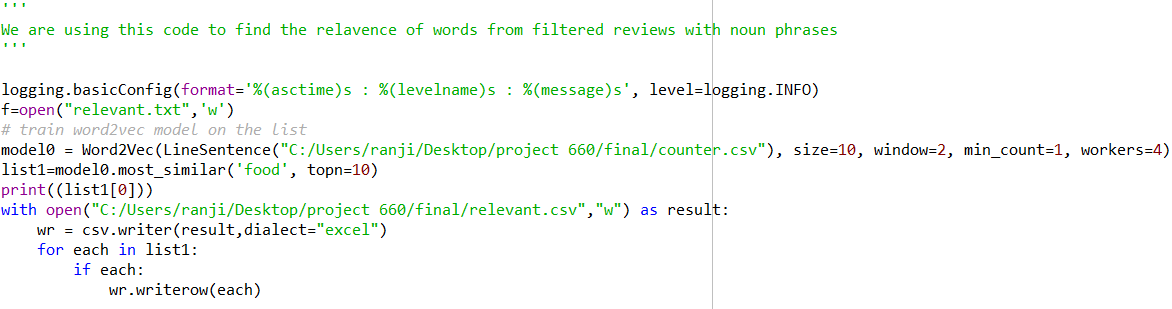
Step 5: Using textblob to generate list of nouns from the interim tips



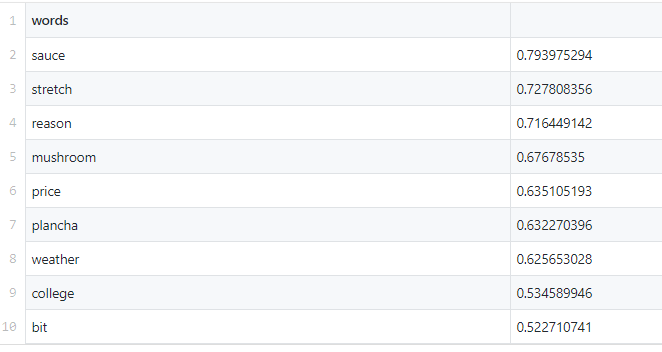
Output:



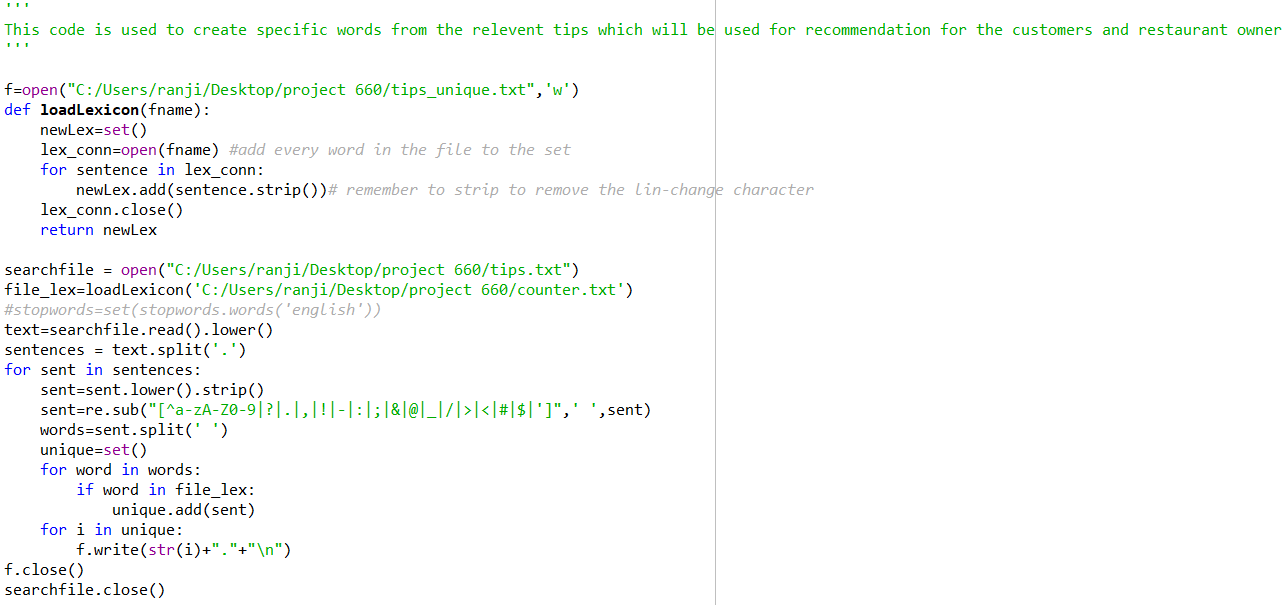
Step 6: Using Word2Vec model, find similar words with the words of interest (e.g. food, restaurant, Mexican etc.)



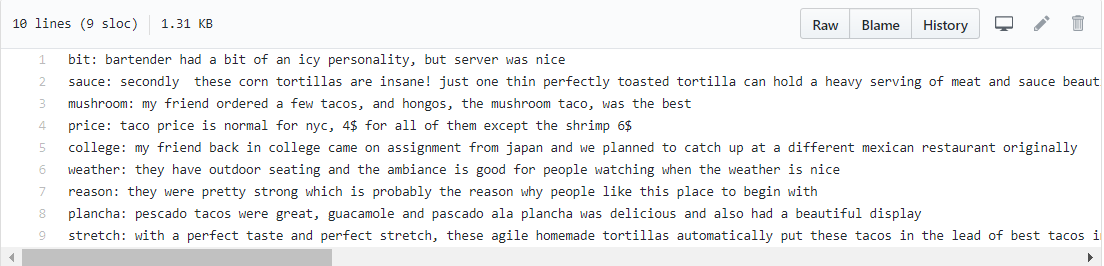
Output:



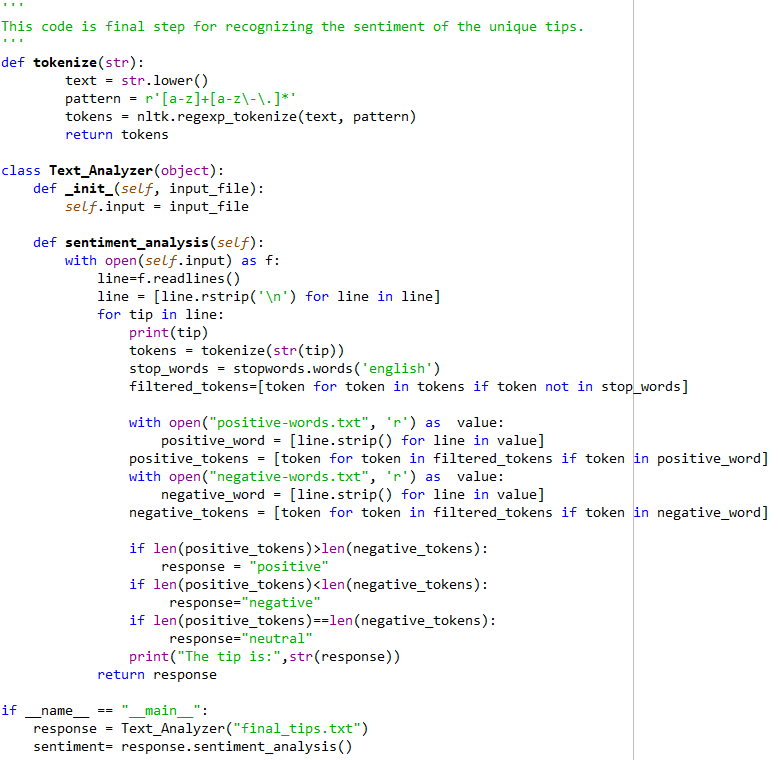
Step 7: Extract the relevant tips associated with the relevant words generated above



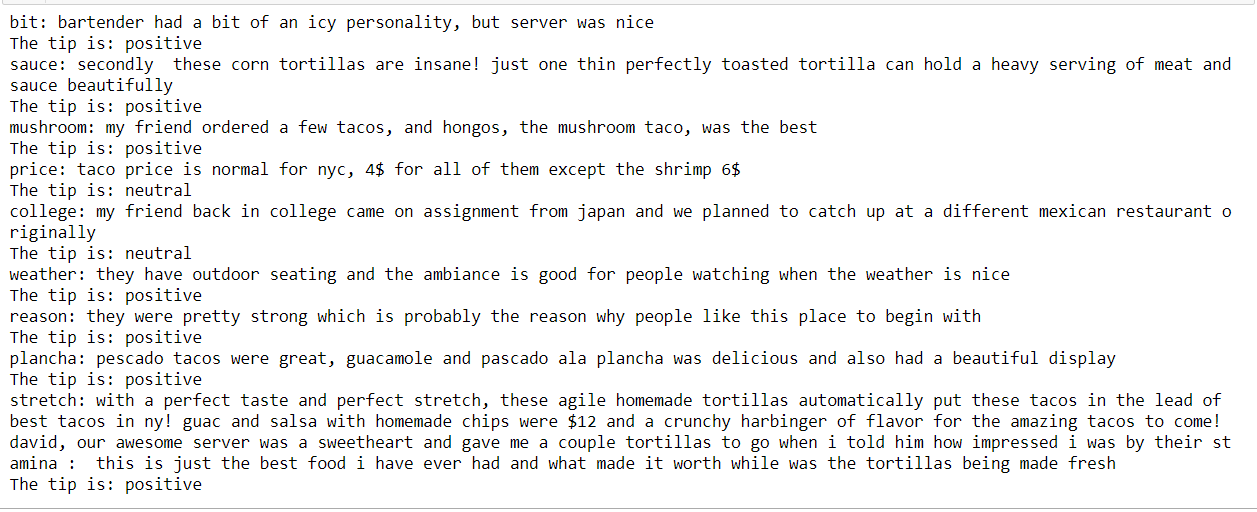
Output:



Step 8: Perform Sentiment Analysis of the relevant tips generated above



Output:



**Conclusion**

This model would efficiently assist restaurants in keeping up with good work they are doing, and improve upon the negatives. The restaurant owners would specifically know what to improve upon rather than just vague positive or negative reviews. The customers too would be highly benefitted by the positive tip feed on the website, thus saving the time to go through the entire reviews and looking for what is good or bad at a restaurant. This would in turn help the restaurants to promote their business through the websites and attract potential new customers with their key food items.