

Yelp Review Categorizer

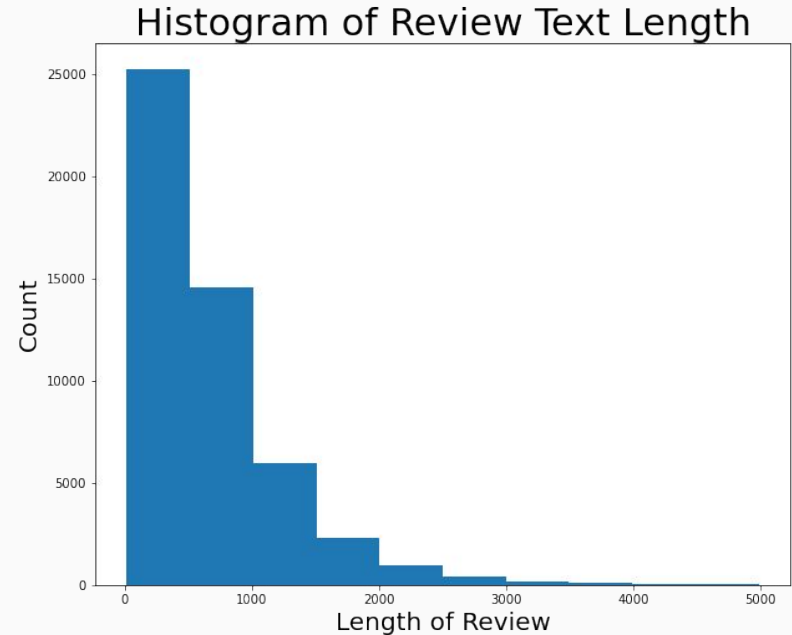
By Injung Ahn

Problem Statement



Data

- 50 restaurants
- 1000 reviews from each
- Average review length: 677 characters



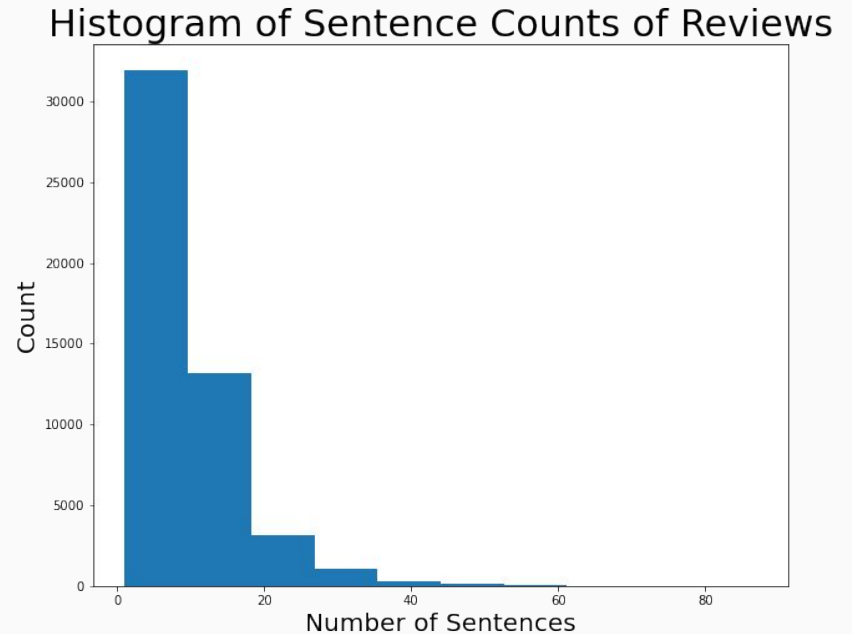
Data Cleaning

Filtered For

- HTML characters
- Hyper links
- Emojis
- Whitespaces

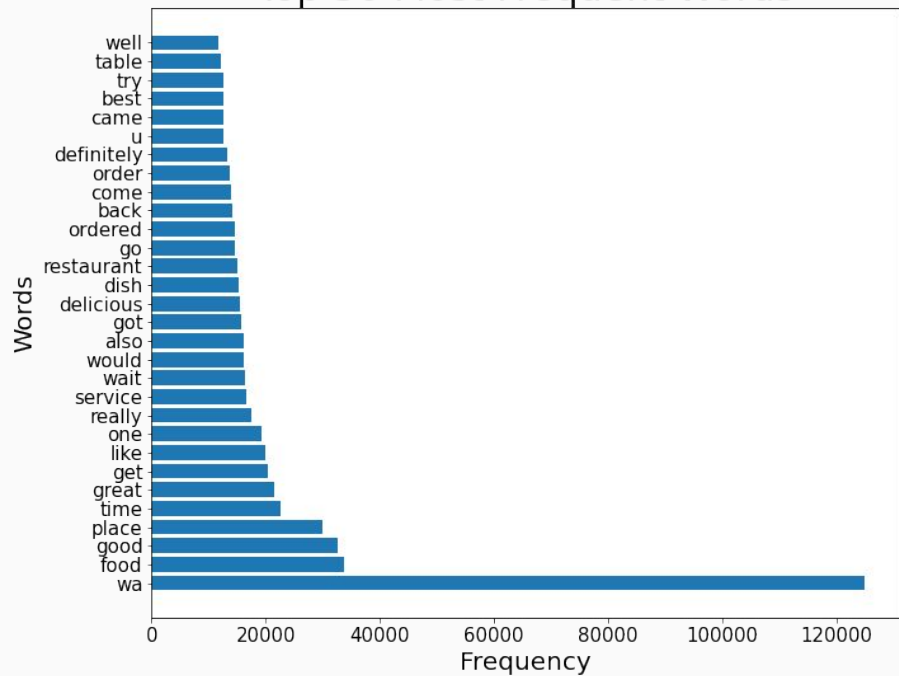
Data

- Average sentences per review: 9
- Total of 469_663 sentences

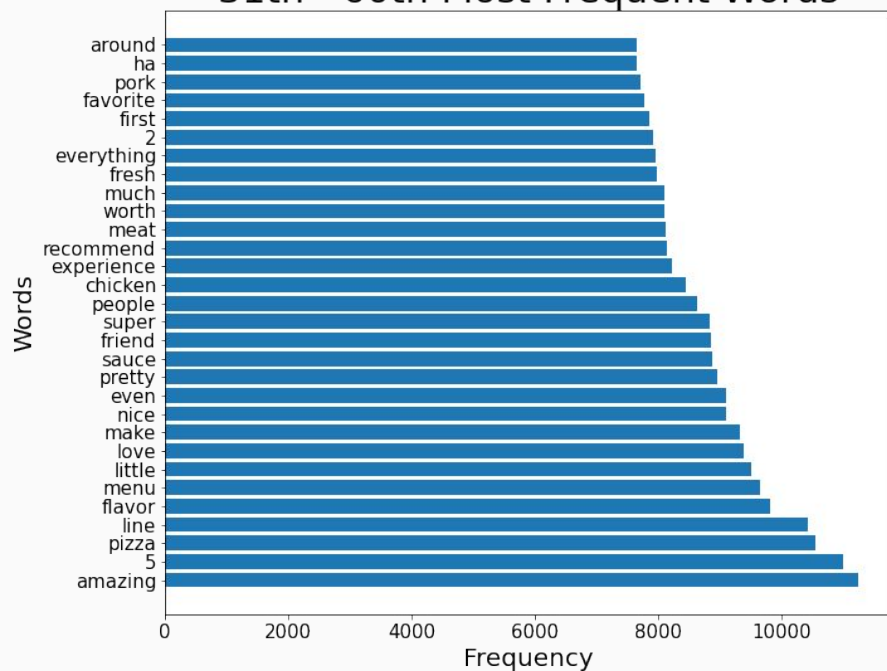


Lemmatized Terms

Top 30 Most Frequent Words



31th - 60th Most Frequent Words



Sentiment Analysis

Restaurant	Yelp Rating	Sentiment Analysis Rating
Brenda's French Soul Food	4.0	4.45
Tartine Bakery	4.0	4.26
House of Prime Rib	4.0	4.41
Hog Island Oyster	4.5	4.56
Gary Danko	4.5	4.65

Sentiment Analysis

1000 Most Recent Reviews

Restaurant	Yelp Rating	Food Rating	Service Rating	Time Rating
Brenda's French Soul Food	4.0	3.228686	3.335638	2.991169
House of Prime Rib	4.0	3.122370	3.290248	2.959223
Gary Danko	4.5	3.279305	3.420495	3.125987
Tartine Bakery	4.0	3.229126	2.938706	2.985375
Hog Island Oyster	4.5	3.350521	3.460033	3.202031

All Reviews

Restaurant	Yelp Rating	Food Rating	Service Rating	Time Rating
Brenda's French Soul Food	4.0	3.249160	3.413877	3.013338
House of Prime Rib	4.0	3.106466	3.369225	2.999174
Gary Danko	4.5	3.263361	3.340689	3.095244
Tartine Bakery	4.0	3.244576	3.033528	3.024083
Hog Island Oyster	4.5	3.334523	3.499479	3.083014

Conclusions

- Next steps for categorizing would be to use the categorized sentences and build a classifier instead of just using the stop words
- Getting a more wide variety of data. All of the restaurants had ratings of 4-5 and wasn't able to get all of the data due to computing restraints
- These numbers are subjective and the difference between a 3.2 versus a 3.4 is hard to judge