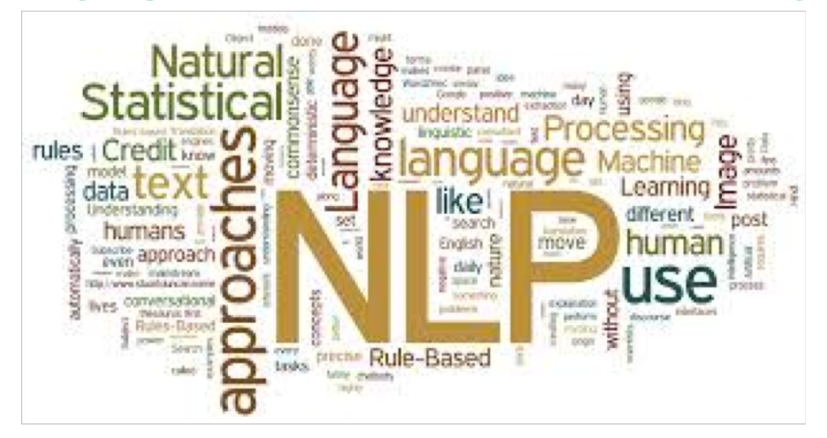
MAGICS

Yelp Improved: Aggregating Restaurant Reviews



Kunal Sonar , Paul Intrevado

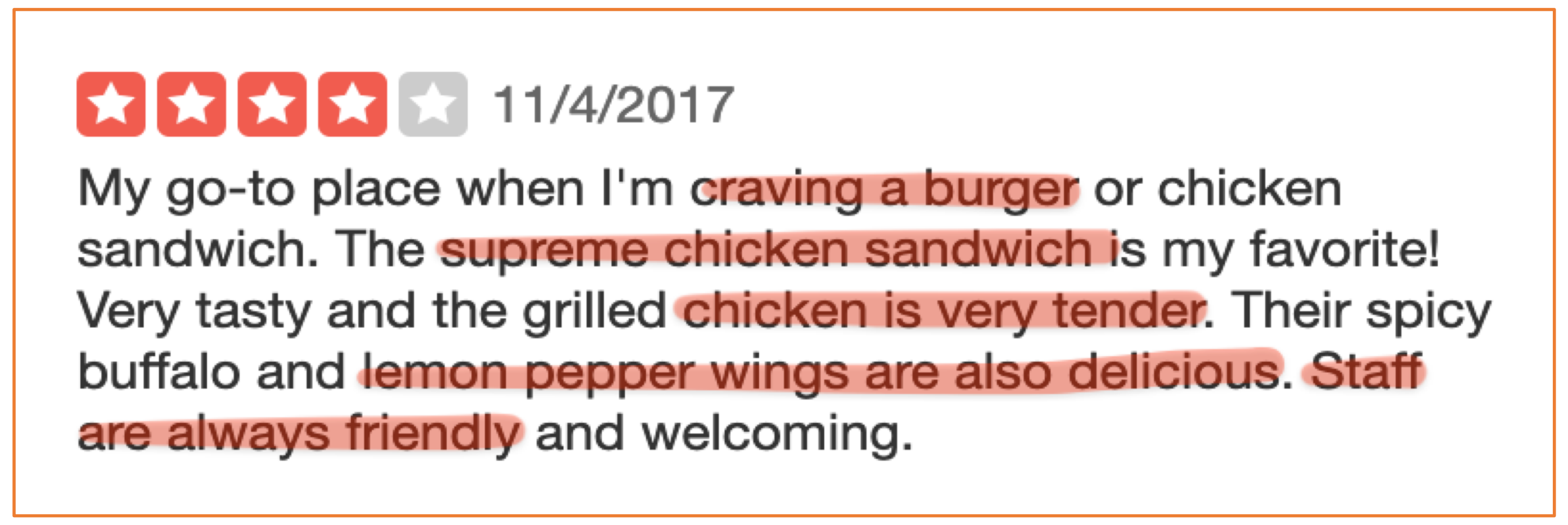
yelp

USF UNIVERSITY OF SAN FRANCISCO

Problem Statement

It is cumbersome to read many restaurant reviews before making a dining decision:

1. Employ Natural Language Processing (NLP) to extract insights
2. Aggregate insights to provide condensed information of all reviews
3. Customers can more easily make well-informed decisions

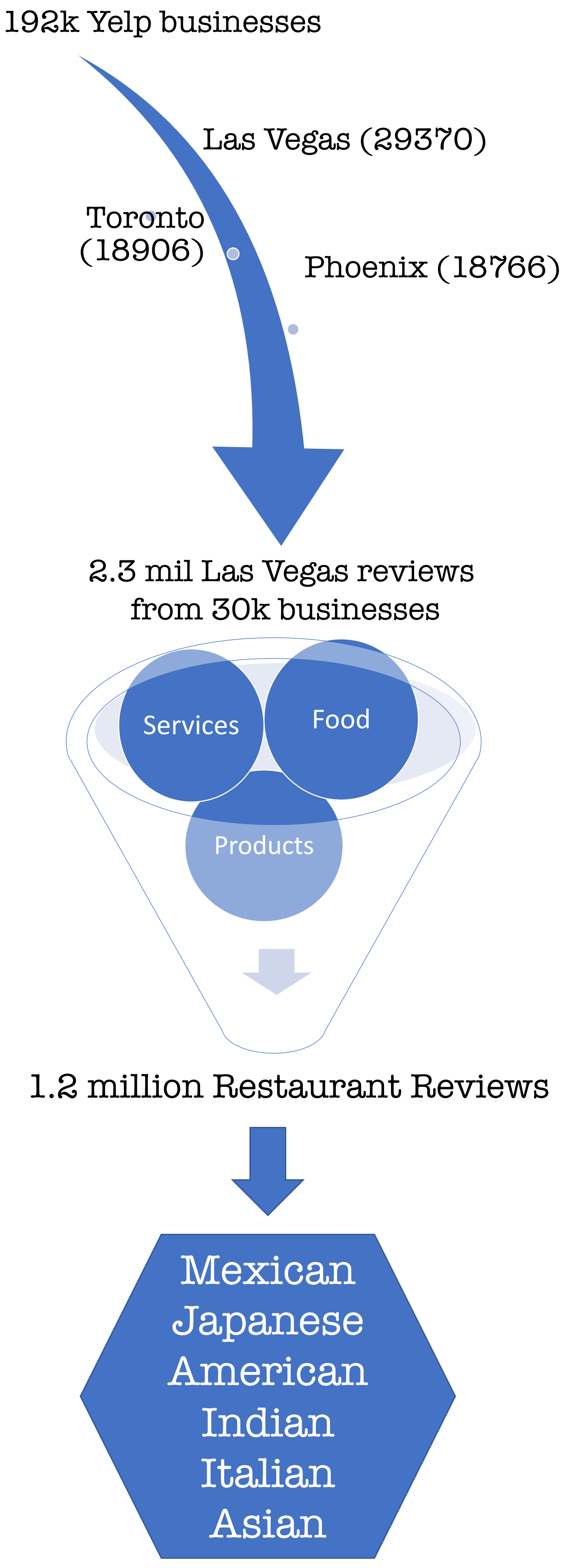


★★★★☆ 11/4/2017

My go-to place when I'm craving a burger or chicken sandwich. The supreme chicken sandwich is my favorite! Very tasty and the grilled chicken is very tender. Their spicy buffalo and lemon pepper wings are also delicious. Staff are always friendly and welcoming.

Customer rating : 4/5
How to extract helpful data from this review?

Methods



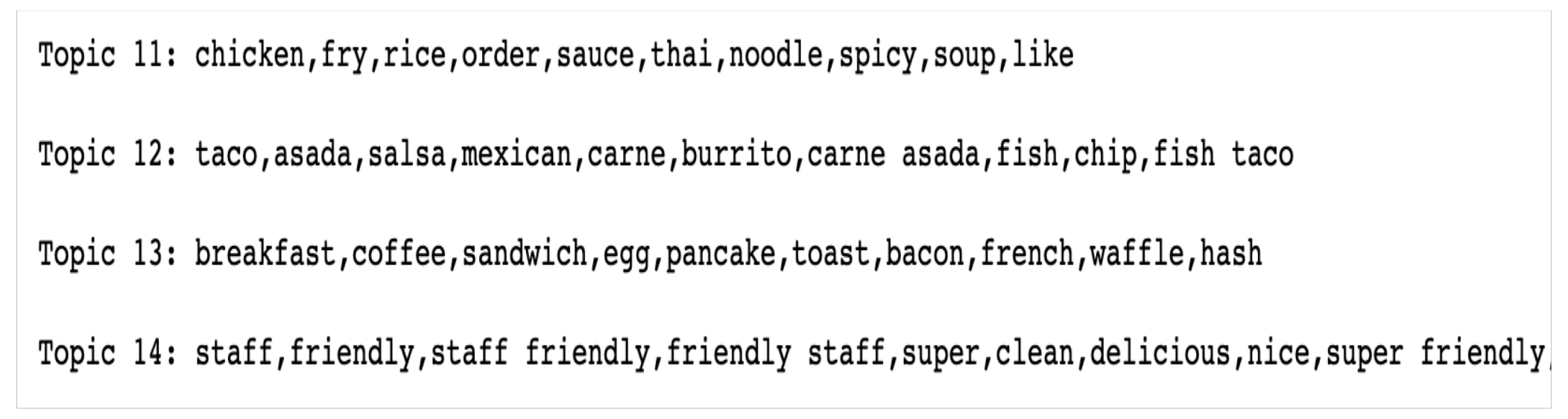
- **Topic Modeling** What are customers talking about? Non-Negative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA) classifies text into discrete topics. Pre-processing includes removal of stop words and lemmatization.
- **Chunking** Phrases are '*chunked*' out from reviews for analysis. SpaCy's Matcher is used on Part-of-Speech (PoS) tags.

The food was delicious and it has vegan options
['DET', 'NOUN', 'VERB', 'ADJ', 'CCONJ', 'PRON', 'VERB', 'NOUN', 'NOUN']

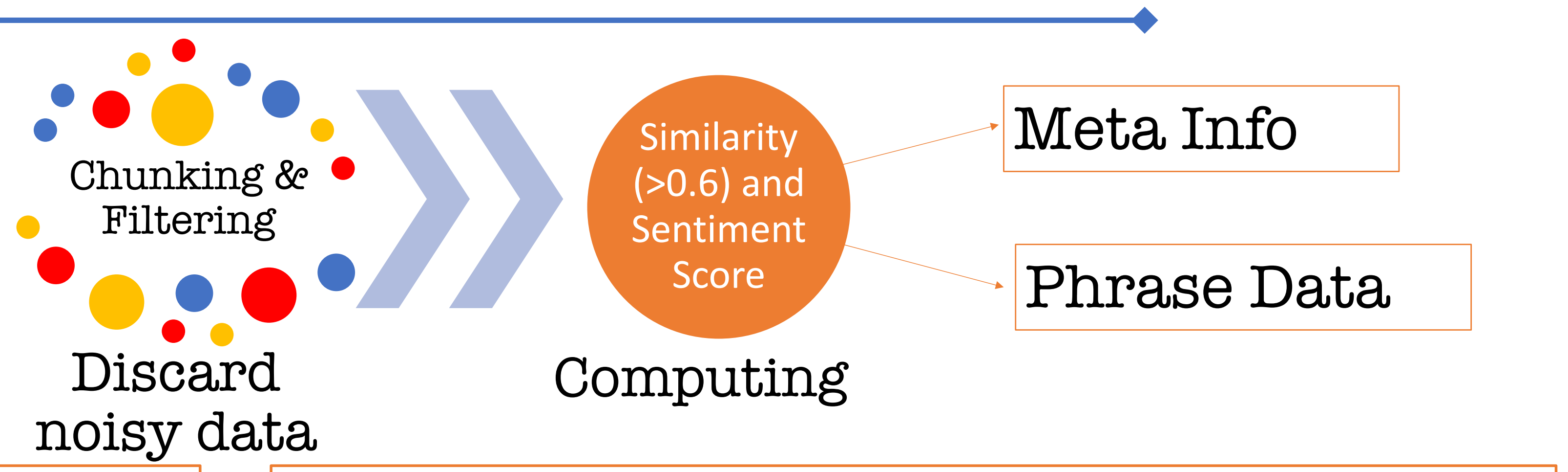
- **Filtering** Discard phrases based on custom algorithm and keep longest matching unique phrases.
- **Similarity Analysis** Map phrases to attributes like food, service, price, etc., based on SpaCy's similarity score.
- **Sentiment Analysis** Bucket phrases in each category with NLTKs Vader SentimentAnalyzer. Due to an incomplete lexicon set, we use a Naïve Bayes Classifier to move certain phrases (containing at least one <ADJ>) from neutral to either positive or negative.
- **Storing Data** Streamline computational pipeline for analysis and evaluation

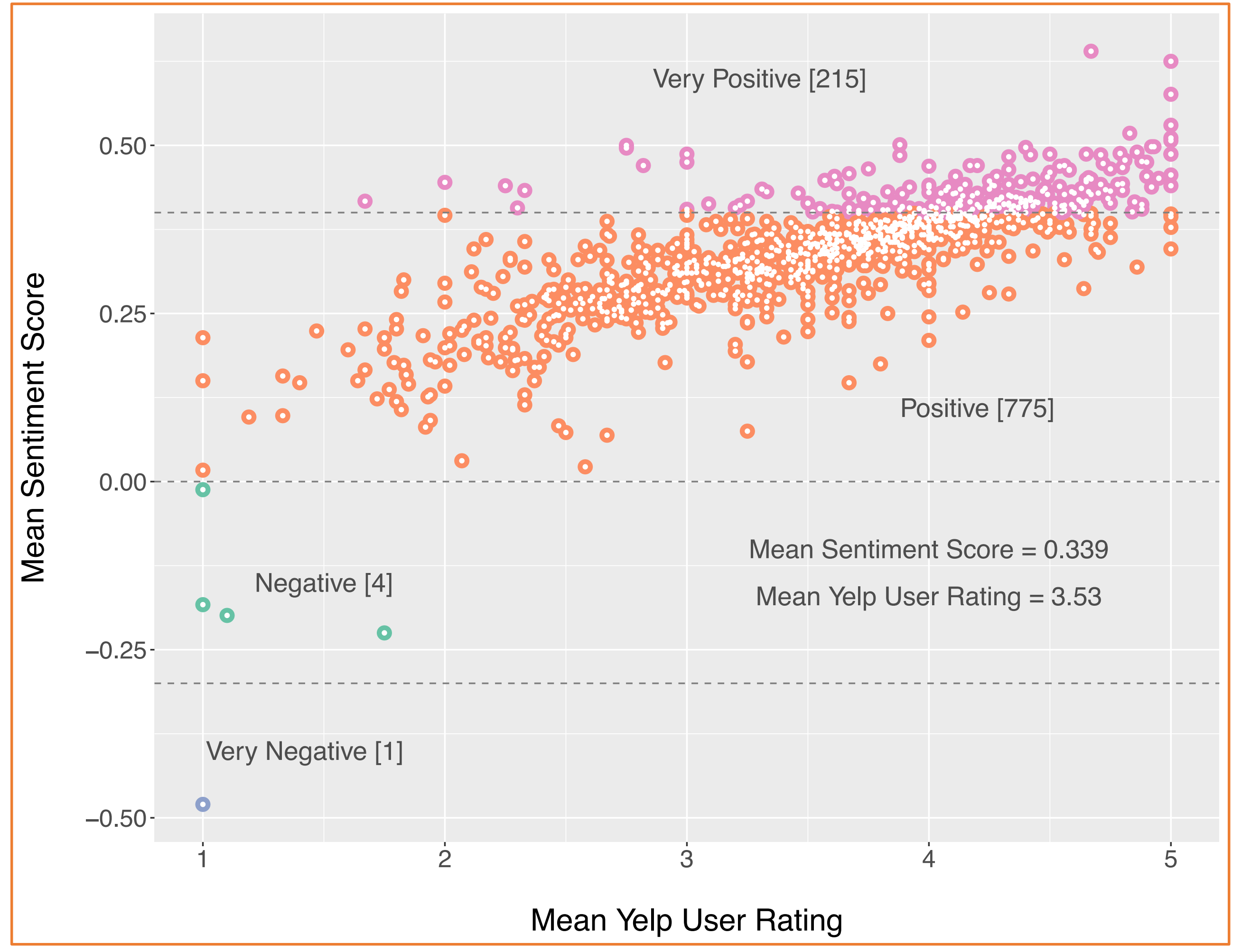
Results

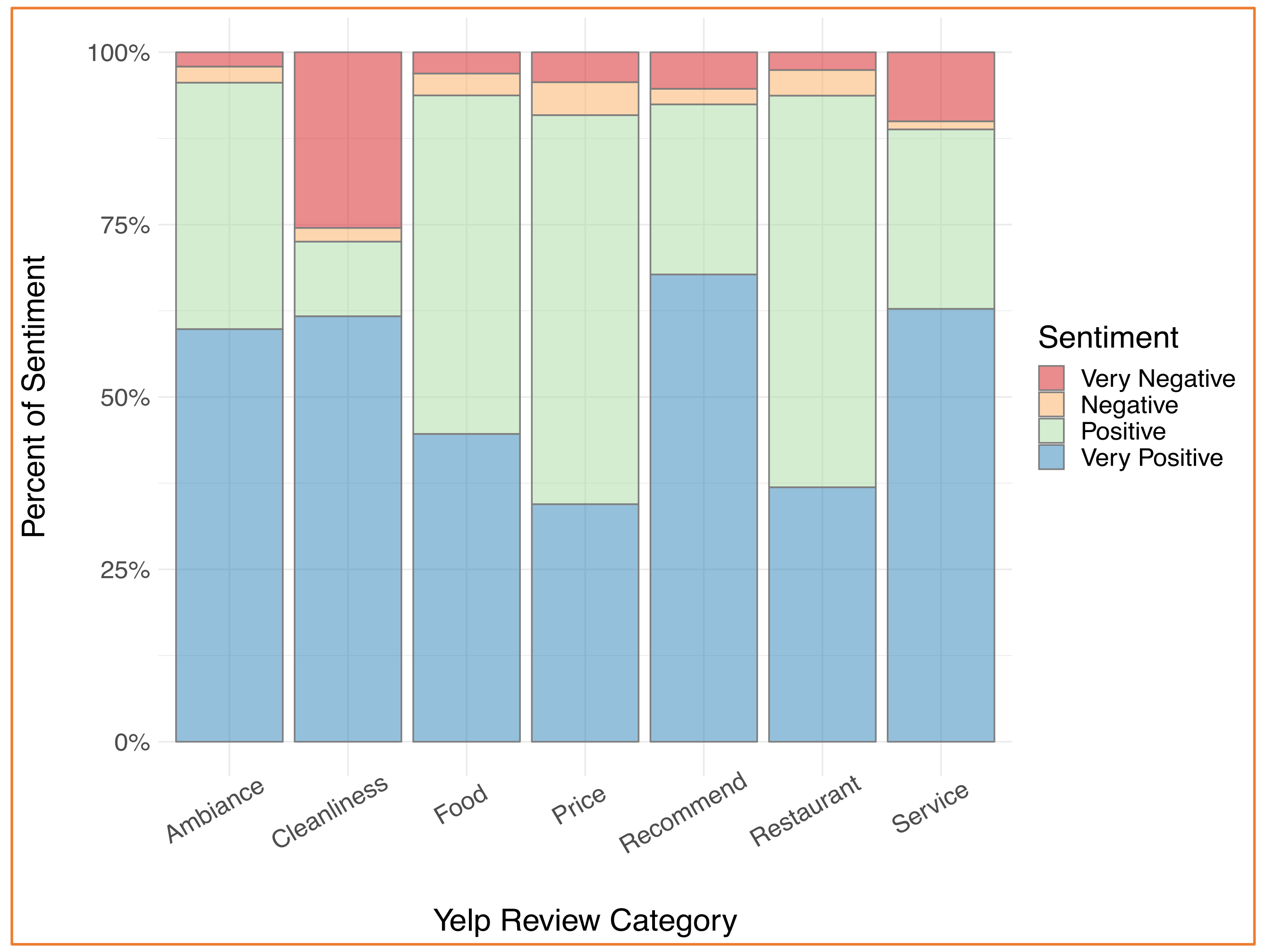
Topic Modeling (k=15)



```
Topic 11: chicken,fry,rice,order,sauce,thai,noodle,spicy,soup,like

Topic 12: taco,asada,salsa,mexican,carne,burrito,carne asada,fish,chip,fish taco

Topic 13: breakfast,coffee,sandwich,egg,pancake,toast,bacon,french,waffle,hash

Topic 14: staff,friendly,staff friendly,friendly staff,super,clean,delicious,nice,super friendly,
```

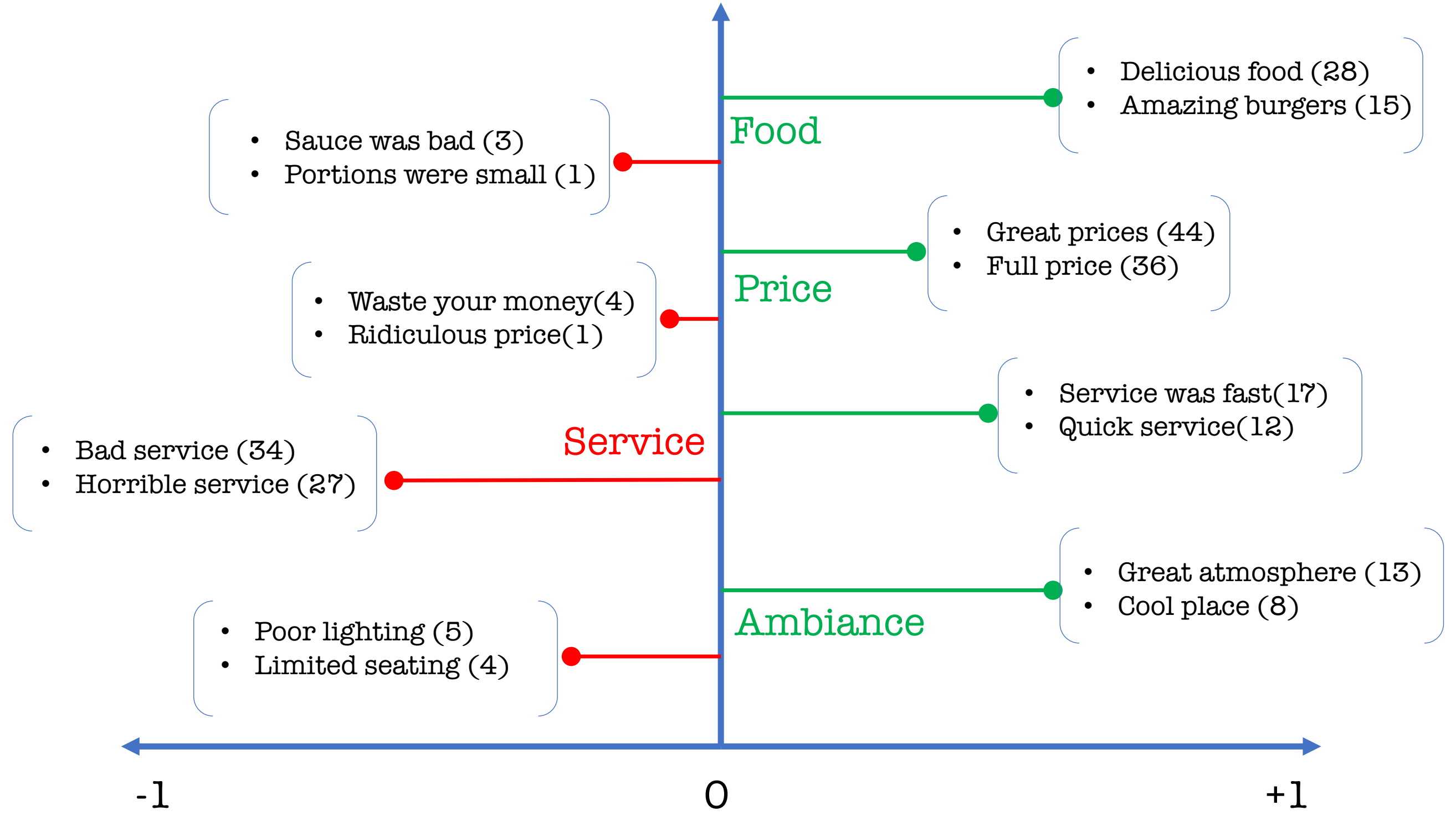






- ➢ Mean Yelp user ratings strongly correlate with calculated mean sentiment score (0.75)
- ➢ Majority of the customers comment positive reviews. Very few restaurants have an overall negative score by sentiment.
- ➢ Projecting sentiment and top words in each category helps a customer to quickly gauge a restaurant by what customers are saying.

Sentiment by Category



Future Work

- Enrich chunking by adding more well defined regexes and also better discarding logic.
- Map more phrases in similarity algorithm by including phrases from unseen categories. For example indoor/outdoor seating under ambiance.
- Custom algorithms for topics like delivery, hours open and cuisine specific information
- Smarter logic to move wrongly classified phrases during sentiment evaluation.
- Build a recommendation system