Digging In: Text Mining of Yelp Reviews for Subtopic Identification and Sentiment

Identifying Review Sub Categories and Rankings

BY SAM GUTENTAG
MARCH 3, 2018

Introduction

Yelp is the gold standard in the field of restaurant and business reviews, tips, and ratings. These ratings are boiled down to just a single star ranking between one and five. But as with may user generated ranking systems, locations tend to be heavily favored or frequently despised. Many users create accounts just to give poor reviews aft er a particularly bad experience, or on the other hand sign up to sing the praises of a local favorite.

But how does a user know to compare a four star sushi to a 3 star sushi restaurant to a 5 star roofing contractor? It may not be likely that a person would be looking at all of these just to find a good spot for date night, but it shows the downside of a single number rating system.

This project aims to user Reviews and Tips for restaurants in Arizona to build a system that will identify key positive and negative terms for each location. From here we will generate sub scores in the fields of Food Quality, Value, Cleanliness, Service, and possibly more!

This will help Yelp users make better, more informed decisions as well as provide Restaurant owners and managers helpful insight into areas in which their business is leading or tailing customer expectations and make improvements to drive sales.

The Client

This project will serve both Yelp users and restaurant owners and managers. Having a mid step between a single overall star ranking and reading through possibly hundreds of reviews is a good middle ground for the time pressed but still scrutinous user or owner.

The Yelp Open Dataset

The Yelp Open Dataset is provided directly by Yelp and consists of tips, reviews, users, businesses and check-ins datas sets. Available for download in SQL or JSON format the dataset consists of more than 5.25 million reviews and over 1 million user tips from more than 1.1 million users. Included also is a collection of more than 156,000 business profiles.

Data Wrangling

The first big hurdle for analysis of the datasets is that the data file are large, and stored in JSON format. The first step was to make decisions on how best to prune down the total data set to only include bits of data that we will be using. The first step is loading in business data only for businesses located in Arizona. And from there filtering down to only Restaurants within Arizona. This will provide us a still substantially large data set while limiting the domain of user reviews and tips to be more focused on restaurant related topics.

Business Dataset

The 'business.json' file consists of data for each business included in the datasets. Each business is assigned a unique 'business_id' attribute that can be used to cross reference reviews, tips, and checkins reported in the other data files. Business data contains information on the name, location, neighborhood, the total review count, the business' star rating, several attributes such as parking, validation, byob, and several other factors. These are largely ignored in the wrangling as they are all user reported and by default are set to nan, only updating to True or False when enough users report an update.

Business data also includes an array of cuisine categories for each business. This categories array and location are the largest filters used to limit the data set to only locations that are considered a 'restaurant' and are in the state of Arizona. A list of restaurant cuisines was manually collected from all possible reporting categories and a list of 221 possible options was retained. This list is used to cross reference each business, removing any location that included a category value not in this list of restaurant categories. Finally, a secondary data frame of only locations not classifying as 'Fast Food' were created with a similar process.

This reduced the initial set of 174,567 businesses to just 7066 total restaurants, 5561 non fast food restaurants and 1505 Fast Food restaurants.

Reviews Dataset

The original 'review.json' contains over 5.25 million restaurant reviews, after we have pruned down our selection of businesses to include only Arizona Restaurants, we can filter these reviews to include only those reviews given to restaurants in this list. We end up with a set of Arizona Restaurant Reviews consisting of 501,250 reviews split into 40,349 Fast Food Reviews and 460901 Non Fast Food Reviews.

Next up, we clean up reviews by stripping out reviews with a text length less than 80 characters. The 80 character limit was reached by looking at the 5% Quantile of the review lengths year to year. Review lengths at this metric only dip below 80 characters in early 2005 so it is a safe line to draw in the sand.

Tips Dataset

The original 'tip.json' file consists of 1,098,324 tips which we quickly reduce in the same method as the review data described above. We prune down to 1119,274 Arizona restaurant tips, with 12,052 Fast Food and 107,222 Non Fast Food Tips. Next, we prune out tips less than 10 characters with the same method as reviews again. Finally, tips are scrubbed of commas, for the benefit of writing out to csv files.

Initial Findings - Restaurant Scores

Initial findings of the data set are interesting, and corroborate early believes that users are often creating accounts to write a few negative reviews or positive ones. Rarely is a review neutral (3 stars in this case). A wide disparity exists in the Business Star ratings of Fast Food restaurants and Non Fast Food restaurants.
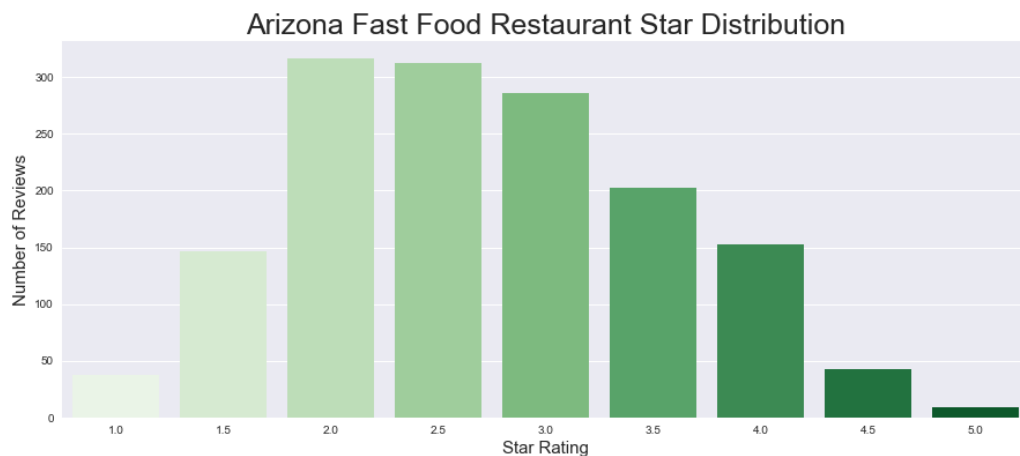
Not quite perfect mirrors of each other, but it is immediately apparent that users are much more critical of fast food locations than they are of non fast food locations.

Initial Findings - Review Texts

After initial wrangling and prep of the review data, we tokenize the review text by removing stop words, removing punctuation, and lemmatizing (stemming) all reviews to clean tokens. Initial exploration of review text yields interesting results.

For each analysis, the tokens from all reviews are counted by frequency, and only tokens occurring more frequently than the 10,000th most frequent term are kept. Example if the 10,000th most frequent token appears 4 times in the corpus, only terms appearing 5 or more times are kept.

Given the difference between Fast Food and Non Fast Food review trends, I completed initial analysis of several subsets of the total reviews data set. Latent Dirichlet Allocation models are generated from 5 groups of reviews.

Several adjustments were made when running LDA models on the data set, on the next page, each run of the model and parameters I show each set o topics generated. Highlighted in Red are the the most easily summarized topic sets.

Model 1 Fast Food Review Categories:
4:Speed, 5:Drive Through, **12:Sanitary Status**, 20:Order Accuracy, **24:Seating**, **27:Price Value**, **35:Customer Service**, 41:Payment Options, **43:Order Wait Time**, 46:Deals and Specials **48:Atmosphere**

Model 2 Non Fast Food Review Categories:
1:Seating, **5:Loyalty**, 8:Service, **15:Value**, 16:Speed, **18:Customer Service**

Model 3a All Restaurant Review Categories:
5:Loyalty, 8:Service, **11:Customer Service**, 16:Bar Scene, **17:Value**

Model 3b All Restaurant Review Categories:
1:Atmosphere, 8:Menu Option, 16:Wait Time, **17:Sanitary Status**, 27:Service Time, **30:Customer Service**, **33:Customer Service**, 40:Bar Scene, **42:Food Quality**, 44:Home Town Feel, 46:Menu Variety

Model 1 - Fast Food Reviews

Review Subset: Reviews Given to Fast Food businesses, most frequent tokens only
Number of Topics: 50
Number of Terms: 10
Number of Passes: 50
Results: 4:Speed, 5:Drive Through, 12:Sanitary Status, 20:Order Accuracy, 24:Seating,
27:Price Value, 35:Customer Service, 41:Payment Options, 43:Order Wait Time, 46:Deals and Specials
48:Atmosphere

```
0    ['chipotle', 'bowl', 'burrito', 'rice', 'bean', 'meat', 'chicken', 'steak', 'guacamole', 'get']
1    ['taco', 'bell', 'burrito', 'mexican', 'salsa', 'carne', 'asada', 'good', 'food', 'tortilla']
2    ['counter', 'one', 'behind', 'people', 'employee', 'order', 'line', 'guy', 'working', 'front']
3    ['cold', 'fry', 'food', 'old', 'never', 'hot', 'warm', 'ever', 'even', 'fresh']
4    ['lunch', 'quick', 'rush', 'dinner', 'work', 'stopped', 'get', 'today', 'busy', 'grab']
5    ['drive', 'thru', 'window', 'car', 'minute', 'get', 'order', 'food', 'inside', 'wait']
6    ['panda', 'express', 'job', 'team', 'professional', 'thanks', 'thank', 'polite', 'guy', 'done']
7    ['asked', 'said', 'told', 'back', 'ordered', 'wanted', 'got', 'went', 'would', 'could']
8    ['box', 'jack', 'server', 'surprised', 'waitress', 'rocket', 'johnny', 'mall', 'pleasantly', 'la']
9    ['chicken', 'wei', 'pei', 'rice', 'food', 'dish', 'roll', 'orange', 'chinese', 'teriyaki']
10   ['place', 'try', 'back', 'best', 'great', 'good', 'go', 'time', 'owner', 'definitely']
11   ['location', 'one', 'time', 'staff', 'new', 'visit', 'particular', 'store', 'employee', 'need']
12   ['dirty', 'clean', 'place', 'food', 'dining', 'area', 'kid', 'room', 'bathroom', 'floor']
13   ['fish', 'chip', 'bar', 'priced', 'outstanding', 'beer', 'game', 'reasonably', 'menu', 'selection']
14   ['lot', 'parking', 'across', 'street', 'wall', 'hole', 'space', 'sooo', 'spot', 'basic']
15   ['manager', 'called', 'back', 'call', 'order', 'store', 'phone', 'would', 'number', 'could']
16   ['place', 'would', 'like', 'good', 'really', 'better', 'much', 'get', 'little', 'pretty']
17   ['hubby', 'lemon', 'sugar', 'malt', 'mild', 'fruit', 'cinnamon', 'btw', 'gf', 'content']
18   ['review', 'star', 'yelp', 'write', 'read', 'based', 'negative', 'writing', 'reading', 'http']
19   ['drink', 'soda', 'ice', 'cup', 'machine', 'tea', 'water', 'coke', 'cream', 'iced']
20   ['order', 'time', 'get', 'wrong', 'go', 'every', 'right', 'location', 'never', 'correct']
21   ['food', 'fast', 'good', 'service', 'place', 'pretty', 'really', 'friendly', 'restaurant', 'like']
22   ['breakfast', 'mcdonald', 'egg', 'morning', 'coffee', 'mcdonalds', 'sausage', 'bacon', 'whataburger', 'brown']
23   ['sub', 'subway', 'firehouse', 'meatball', 'shop', 'best', 'meat', 'bread', 'way', 'veggie']
24   ['table', 'inside', 'restaurant', 'outside', 'area', 'seating', 'sit', 'clean', 'station', 'nice']
25   ['sandwich', 'bread', 'cheese', 'meat', 'good', 'subway', 'turkey', 'like', 'mayo', 'tomato']
26   ['pizza', 'topping', 'crust', 'place', 'pie', 'good', 'mod', 'great', 'like', 'sauce']
27   ['price', 'small', 'size', 'special', 'portion', 'large', 'nothing', 'money', 'full', 'average']
28   ['like', 'place', 'get', 'know', 'food', 'go', 'want', 'people', 'say', 'one']
29   ['chicken', 'wing', 'piece', 'sauce', 'kfc', 'got', 'spicy', 'meal', 'ordered', 'tender']
30   ['potato', 'wendy', 'gravy', 'mashed', 'popeye', 'sweet', 'airport', 'whopper', 'cheese', 'quarter']
31   ['u', 'man', 'young', 'lady', 'hair', 'took', 'came', 'gave', 'went', 'guy']
32   ['time', 'first', 'ordered', 'got', 'back', 'like', 'one', 'went', 'tasted', 'would']
33   ['salad', 'fresh', 'mike', 'healthy', 'jersey', 'dressing', 'option', 'love', 'pita', 'ingredient']
34   ['school', 'phoenix', 'carl', 'high', 'jr', 'bk', 'north', 'downtown', 'mcd', 'american']
35   ['customer', 'service', 'employee', 'rude', 'manager', 'ever', 'worst', 'food', 'bad', 'attitude']
36   ['night', 'open', 'late', 'close', 'hour', 'closed', 'door', 'sign', 'friday', 'closing']
37   ['year', 'last', 'week', 'time', 'ago', 'food', 'eat', 'day', 'month', 'went']
38   ['container', 'grill', 'heat', 'fork', 'plastic', 'put', 'throw', 'spoon', 'set', 'standard']
39   ['sauce', 'flavor', 'green', 'red', 'hot', 'good', 'fresh', 'menu', 'delicious', 'also']
40   ['yes', 'mac', 'cheese', 'big', 'drop', 'yeah', 'everyday', 'heart', 'blue', 'denny']
41   ['card', 'extra', 'charge', 'pay', 'cash', 'cent', 'use', 'credit', 'money', 'dollar']
42   ['always', 'love', 'get', 'location', 'time', 'go', 'friendly', 'never', 'staff', 'one']
43   ['order', 'minute', 'wait', 'food', 'time', 'line', 'ready', 'long', 'waiting', 'waited']
44   ['shake', 'burger', 'style', 'n', 'fry', 'animal', 'double', 'menu', 'love', 'east']
45   ['sonic', 'drink', 'tot', 'happy', 'hour', 'frosty', 'slush', 'cherry', 'w', 'car']
46   ['free', 'coupon', 'lettuce', 'wrap', 'buy', 'use', 'deal', 'birthday', 'one', 'learned']
47   ['dog', 'hot', 'beef', 'chicago', 'chili', 'italian', 'fry', 'cheese', 'good', 'portillo']
48   ['great', 'food', 'friendly', 'service', 'staff', 'place', 'delicious', 'fresh', 'clean', 'awesome']
49   ['burger', 'fry', 'onion', 'cheese', 'guy', 'five', 'good', 'cheeseburger', 'bun', 'bacon']
```

Model 2 - Non Fast Food Reviews

Review Subset: Reviews Given to Non Fast Food businesses, most frequent tokens only
Number of Topics: 20
Number of Terms: 15
Number of Passes: 50
Results: 1:Seating, 5:Loyalty, 8:Service, 15:Value, 16:Speed, 18:Customer Service

```
 0  ['place', 'back', 'try', 'got', 'time', 'good', 'really', 'went', 'first', 'ordered']
 1  ['great', 'bar', 'highly', 'patio', 'beer', 'atmosphere', 'place', 'nice', 'outside', 'area']
 2  ['salad', 'dessert', 'dish', 'bread', 'wine', 'pasta', 'meal', 'sauce', 'delicious', 'cheese']
 3  ['like', 'food', 'place', 'would', 'could', 'one', 'taste', 'bad', 'even', 'really']
 4  ['steak', 'shrimp', 'fish', 'fried', 'cooked', 'seafood', 'ordered', 'crab', 'potato', 'good']
 5  ['place', 'always', 'love', 'best', 'food', 'get', 'go', 'great', 'time', 'favorite']
 6  ['chicken', 'rice', 'dish', 'thai', 'soup', 'chinese', 'food', 'noodle', 'pho', 'sauce']
 7  ['breakfast', 'egg', 'coffee', 'pancake', 'bacon', 'morning', 'toast', 'waffle', 'potato', 'brunch']
 8  ['great', 'food', 'service', 'friendly', 'good', 'staff', 'place', 'back', 'delicious', 'definitely']
 9  ['bbq', 'pork', 'rib', 'meat', 'sauce', 'side', 'brisket', 'mac', 'cheese', 'pulled']
10  ['burger', 'sandwich', 'fry', 'cheese', 'onion', 'bread', 'good', 'vegan', 'bun', 'sub']
11  ['sushi', 'roll', 'hour', 'happy', 'tuna', 'fish', 'fresh', 'chef', 'spicy', 'salmon']
12  ['time', 'location', 'year', 'last', 'restaurant', 'visit', 'since', 'food', 'first', 'one']
13  ['restaurant', 'make', 'menu', 'know', 'business', 'owner', 'review', 'one', 'staff', 'experience']
14  ['pizza', 'wing', 'crust', 'cheese', 'good', 'sauce', 'italian', 'slice', 'topping', 'thin']
15  ['good', 'food', 'place', 'price', 'pretty', 'really', 'get', 'like', 'would', 'little']
16  ['table', 'u', 'server', 'minute', 'drink', 'food', 'came', 'wait', 'time', 'service']
17  ['chicken', 'salad', 'pita', 'gyro', 'dog', 'greek', 'hummus', 'meat', 'tea', 'lunch']
18  ['order', 'food', 'time', 'said', 'back', 'would', 'service', 'asked', 'get', 'customer']
19  ['taco', 'mexican', 'salsa', 'burrito', 'chip', 'bean', 'good', 'cheese', 'green', 'tortilla']
```

Model 3a - All Restaurant Reviews, Nouns Only

Review Subset: All reviews, most frequent noun tokens only
Number of Topics: 20
Number of Terms: 15
Number of Passes: 50
Results: 5:Loyalty, 8:Service, 11:Customer Service, 16:Bar Scene, 17:Value

```
 0  ['sushi', 'fish', 'shrimp', 'crab', 'place', 'seafood', 'tuna', 'chef', 'salmon', 'time']
 1  ['breakfast', 'egg', 'coffee', 'bacon', 'waffle', 'morning', 'place', 'toast', 'potato', 'gravy']
 2  ['dessert', 'cream', 'dish', 'menu', 'meal', 'dinner', 'restaurant', 'cake', 'appetizer', 'chocolate']
 3  ['chicken', 'sauce', 'meat', 'bbq', 'pork', 'rib', 'side', 'flavor', 'potato', 'tender']
 4  ['taco', 'street', 'fish', 'bell', 'tuesday', 'shrimp', 'shop', 'nacho', 'pastor', 'shell']
 5  ['love', 'place', 'food', 'family', 'time', 'service', 'kid', 'staff', 'year', 'owner']
 6  ['location', 'restaurant', 'phoenix', 'place', 'area', 'parking', 'lot', 'wall', 'spot', 'menu']
 7  ['steak', 'pita', 'gyro', 'chicken', 'salad', 'hummus', 'meat', 'wrap', 'plate', 'lamb']
 8  ['table', 'food', 'server', 'service', 'restaurant', 'time', 'drink', 'minute', 'waiter', 'meal']
 9  ['salsa', 'chip', 'bean', 'food', 'cheese', 'rice', 'tortilla', 'margarita', 'sauce', 'chicken']
10  ['place', 'review', 'food', 'star', 'thing', 'people', 'make', 'way', 'something', 'day']
11  ['order', 'time', 'food', 'customer', 'service', 'minute', 'manager', 'location', 'employee', 'people']
12  ['burger', 'fry', 'dog', 'onion', 'cheese', 'place', 'time', 'order', 'soda', 'shake']
13  ['chicken', 'airport', 'jungle', 'masala', 'phx', 'sample', 'jerk', 'tikka', 'flight', 'island']
14  ['dish', 'chicken', 'rice', 'soup', 'spicy', 'noodle', 'food', 'sauce', 'beef', 'curry']
15  ['pizza', 'wing', 'place', 'cheese', 'sauce', 'slice', 'time', 'pie', 'try', 'order']
16  ['bar', 'hour', 'night', 'drink', 'beer', 'place', 'music', 'friday', 'atmosphere', 'fun']
17  ['food', 'place', 'service', 'price', 'restaurant', 'lunch', 'time', 'quality', 'menu', 'staff']
18  ['salad', 'sandwich', 'bread', 'lunch', 'cheese', 'chicken', 'pasta', 'sub', 'tomato', 'meat']
19  ['bowl', 'pho', 'tea', 'spring', 'water', 'ice', 'cup', 'broth', 'pot', 'express']
```

Model 3b - All Restaurant Reviews, Nouns Only

Review Subset: All reviews, most frequent noun tokens only

Number of Topics: 50

Number of Terms: 10

Number of Passes: 30

Results: 1:Atmosphere, 8:Menu Option, 16:Wait Time, 17:Sanitary Status, 27:Service Time, 30:Customer Service, 33:Customer Service, 40:Bar Scene, 42:Food Quality, 44:Home Town Feel, 46:Menu Variety

```
0   ['food', 'service', 'price', 'place', 'quality', 'restaurant', 'portion', 'time', 'atmosphere', 'size']
1   ['kid', 'coupon', 'tuesday', 'deal', 'dad', 'adult', 'regret', 'meal', 'def', 'b']
2   ['burger', 'fry', 'onion', 'place', 'cheese', 'shake', 'ring', 'bacon', 'well', 'time']
3   ['star', 'review', 'yelp', 'reason', 'read', 'rating', 'rate', 'update', 'living', 'state']
4   ['lunch', 'day', 'today', 'time', 'work', 'week', 'dinner', 'spot', 'service', 'yesterday']
5   ['spicy', 'curry', 'food', 'spice', 'dish', 'chicken', 'pad', 'rice', 'flavor', 'place']
6   ['bowl', 'chicken', 'rice', 'noodle', 'egg', 'food', 'soup', 'ramen', 'sauce', 'place']
7   ['chicken', 'strip', 'mall', 'waffle', 'piece', 'breast', 'tender', 'crispy', 'shawarma', 'meal']
8   ['dish', 'restaurant', 'menu', 'course', 'meal', 'chef', 'well', 'cuisine', 'flavor', 'try']
9   ['sauce', 'onion', 'pepper', 'tomato', 'flavor', 'mushroom', 'oil', 'cheese', 'taste', 'lemon']
10  ['box', 'bomb', 'chop', 'stew', 'claim', 'alot', 'vacation', 'bento', 'earth', 'freshness']
11  ['hand', 'daughter', 'bathroom', 'child', 'hair', 'parent', 'restroom', 'paper', 'book', 'cover']
12  ['plate', 'lamb', 'vegetable', 'platter', 'rice', 'leaf', 'kabob', 'meat', 'sampler', 'combination']
13  ['order', 'time', 'line', 'minute', 'food', 'drive', 'counter', 'guy', 'girl', 'store']
14  ['salad', 'pita', 'gyro', 'hummus', 'lettuce', 'wrap', 'side', 'meat', 'bread', 'jungle']
15  ['place', 'food', 'thing', 'time', 'make', 'something', 'way', 'taste', 'try', 'cook']
16  ['table', 'minute', 'waiter', 'food', 'water', 'restaurant', 'waiting', 'place', 'service', 'people']
17  ['picture', 'groupon', 'step', 'baby', 'crap', 'floor', 'foot', 'photo', 'fly', 'hell']
18  ['night', 'dinner', 'friday', 'date', 'saturday', 'evening', 'friend', 'girlfriend', 'thursday', 'movie']
19  ['shrimp', 'crab', 'seafood', 'lobster', 'fish', 'scallop', 'puff', 'leg', 'oyster', 'sea']
20  ['area', 'phoenix', 'trip', 'restaurant', 'city', 'arizona', 'hotel', 'finding', 'town', 'summer']
21  ['soup', 'pho', 'noodle', 'spring', 'broth', 'place', 'beef', 'bowl', 'restaurant', 'service']
22  ['pie', 'joe', 'apple', 'market', 'goat', 'lasagna', 'farm', 'partner', 'bruschetta', 'pecan']
23  ['country', 'fruit', 'refreshing', 'lemonade', 'sugar', 'cinnamon', 'juice', 'pineapple', 'air', 'temperature']
24  ['cream', 'dog', 'ice', 'chipotle', 'chocolate', 'cookie', 'cheese', 'cooky', 'scoop', 'vanilla']
25  ['chip', 'salsa', 'fish', 'margarita', 'food', 'guacamole', 'place', 'restaurant', 'service', 'dip']
26  ['tea', 'drink', 'soda', 'cup', 'bag', 'tip', 'charge', 'coke', 'machine', 'container']
27  ['server', 'drink', 'food', 'hour', 'time', 'service', 'order', 'min', 'half', 'friend']
28  ['location', 'tempe', 'downtown', 'chandler', 'chain', 'mesa', 'drive', 'school', 'mile', 'year']
29  ['group', 'parking', 'lot', 'party', 'people', 'door', 'reservation', 'birthday', 'sign', 'everyone']
30  ['staff', 'food', 'family', 'year', 'time', 'visit', 'place', 'restaurant', 'service', 'son']
31  ['breakfast', 'egg', 'coffee', 'bacon', 'place', 'morning', 'toast', 'potato', 'gravy', 'brunch']
32  ['style', 'mouth', 'tasteless', 'word', 'suck', 'walking', 'bf', 'express', 'edamame', 'calzone']
33  ['customer', 'service', 'manager', 'time', 'food', 'employee', 'business', 'owner', 'care', 'restaurant']
34  ['pasta', 'bread', 'sausage', 'spaghetti', 'garden', 'olive', 'catfish', 'alfredo', 'lil', 'square']
35  ['beef', 'bean', 'rice', 'combo', 'cheese', 'chile', 'corn', 'tamale', 'side', 'ground']
36  ['taco', 'tortilla', 'street', 'place', 'salsa', 'quesadilla', 'bell', 'meat', 'nacho', 'food']
37  ['love', 'place', 'food', 'owner', 'time', 'service', 'try', 'everything', 'make', 'everyone']
38  ['pork', 'meat', 'bbq', 'rib', 'wing', 'sauce', 'side', 'place', 'flavor', 'potato']
39  ['pizza', 'place', 'cheese', 'slice', 'wing', 'time', 'try', 'pie', 'pepperoni', 'sauce']
40  ['bar', 'dining', 'drink', 'area', 'sit', 'room', 'seating', 'bartender', 'hour', 'beer']
41  ['wife', 'valley', 'gem', 'west', 'sister', 'disappoint', 'crepe', 'sooo', 'coast', 'scratch']
42  ['bland', 'money', 'salt', 'greasy', 'smell', 'waste', 'hate', 'disappointment', 'save', 'soggy']
43  ['wall', 'boyfriend', 'dollar', 'hole', 'mark', 'crave', 'college', 'father', 'spending', 'christmas']
44  ['beer', 'fun', 'music', 'atmosphere', 'game', 'watch', 'place', 'tv', 'folk', 'vibe']
45  ['option', 'delivery', 'mom', 'vegan', 'gluten', 'pop', 'driver', 'app', 'eater', 'jack']
46  ['menu', 'item', 'selection', 'buffet', 'variety', 'try', 'choice', 'everything', 'thing', 'lot']
47  ['sandwich', 'cheese', 'bread', 'sub', 'mac', 'meat', 'place', 'shop', 'subway', 'try']
48  ['steak', 'dessert', 'meal', 'dinner', 'cake', 'potato', 'entree', 'appetizer', 'bread', 'salad']
49  ['sushi', 'place', 'tuna', 'chef', 'salmon', 'fish', 'hour', 'california', 'service', 'price']
```

Model 4 - All Restaurant Reviews, Nouns and Verbs Only

Review Subset: All reviews, most frequent noun and verb tokens only
Number of Topics: 30
Number of Terms: 15
Number of Passes: 50
Results: 3:Service, 10:Customer Experience, 13:Wait Time, 18:Problem Resolution,
19:Atmosphere, 24:Value, 27:Menu Variety

```
0    ['shrimp', 'fish', 'crab', 'seafood', 'lobster', 'fried', 'leg', 'sauce', 'clam', 'oyster']
1    ['parking', 'door', 'lot', 'closed', 'drive', 'car', 'sign', 'street', 'find', 'mile']
2    ['food', 'eat', 'place', 'could', 'got', 'tasted', 'would', 'go', 'eating', 'looked']
3    ['u', 'server', 'food', 'drink', 'service', 'came', 'time', 'waiter', 'table', 'ordered']
4    ['flavor', 'sauce', 'dish', 'spicy', 'taste', 'bland', 'spice', 'bit', 'ordered', 'would']
5    ['ice', 'cream', 'bowl', 'soda', 'cup', 'drink', 'water', 'coke', 'machine', 'fountain']
6    ['taco', 'mexican', 'salsa', 'burrito', 'chip', 'bean', 'food', 'cheese', 'tortilla', 'margarita']
7    ['lunch', 'salad', 'tea', 'curry', 'today', 'pad', 'soup', 'dinner', 'lettuce', 'went']
8    ['phoenix', 'year', 'arizona', 'know', 'lived', 'find', 'life', 'city', 'moved', 'mouth']
9    ['star', 'place', 'review', 'give', 'would', 'try', 'say', 'yelp', 'think', 'go']
10   ['customer', 'service', 'food', 'employee', 'owner', 'people', 'make', 'need', 'guy', 'staff']
11   ['salad', 'dessert', 'pasta', 'wine', 'bread', 'cheese', 'chocolate', 'dish', 'cake', 'appetizer']
12   ['place', 'get', 'time', 'go', 'love', 'food', 'ca', 'order', 'wait', 'come']
13   ['table', 'minute', 'wait', 'u', 'food', 'order', 'waiting', 'seated', 'waited', 'people']
14   ['order', 'said', 'asked', 'told', 'time', 'would', 'ordered', 'called', 'manager', 'got']
15   ['pizza', 'wing', 'crust', 'cheese', 'place', 'sauce', 'topping', 'slice', 'pie', 'salad']
16   ['burger', 'fry', 'dog', 'onion', 'cheese', 'place', 'shake', 'got', 'ring', 'bacon']
17   ['sandwich', 'bread', 'sub', 'cheese', 'meat', 'beef', 'shop', 'subway', 'store', 'got']
18   ['night', 'dinner', 'went', 'u', 'time', 'husband', 'friend', 'meal', 'food', 'restaurant']
19   ['food', 'place', 'service', 'restaurant', 'love', 'staff', 'price', 'recommend', 'family', 'atmosphere']
20   ['bbq', 'pork', 'rib', 'meat', 'side', 'mac', 'sauce', 'brisket', 'cheese', 'pulled']
21   ['location', 'tempe', 'strip', 'mall', 'chandler', 'downtown', 'west', 'mesa', 'valley', 'corner']
22   ['steak', 'cooked', 'potato', 'well', 'medium', 'perfection', 'side', 'fried', 'filet', 'mashed']
23   ['chicken', 'pita', 'gyro', 'salad', 'hummus', 'meat', 'rice', 'lamb', 'plate', 'ordered']
24   ['money', 'tip', 'price', 'pay', 'card', 'dollar', 'charge', 'buck', 'coupon', 'cost']
25   ['chicken', 'rice', 'soup', 'fried', 'noodle', 'pho', 'beef', 'roll', 'food', 'dish']
26   ['breakfast', 'egg', 'coffee', 'bacon', 'pancake', 'waffle', 'morning', 'toast', 'brunch', 'biscuit']
27   ['menu', 'option', 'food', 'item', 'price', 'restaurant', 'buffet', 'portion', 'choice', 'variety']
28   ['bar', 'beer', 'place', 'drink', 'hour', 'music', 'atmosphere', 'fun', 'night', 'area']
29   ['sushi', 'roll', 'hour', 'place', 'tuna', 'ramen', 'chef', 'salmon', 'fish', 'spicy']
```

Next Steps

With a set of subtopics identified, I will next move to assign sub categories to reviews, and then accumulate these sub category ratings to each restaurant. The first model we generated with 50 Topics, 10 Terms, and 50 Passes. I will be using these six topics are common in all reviews and are not menu specific:

Topic 12:	Sanitary Status
Topic 24:	Seating
Topic 27:	Price Value
Topic 35:	Customer Service
Topic 43:	Order Wait Time
Topic 48:	Atmosphere

It looks like our first Model, trained on only Fast Food Restaurants, actually does the best job generalizing to other non fast food restaurants as well as provides enough topics of interest as well as topics of menu and cuisine collections that can be safely ignored.

Each review will be tokenized by sentence and each sentence assigned a most likely category by comparing the number of terms in each Topic found in the sentence. Each Review can be tagged with multiple Topics, thus each review can contribute to a single or potentially all of these Topics.

Code for these Analysis can be found at on my GitHub Page