Food Finder

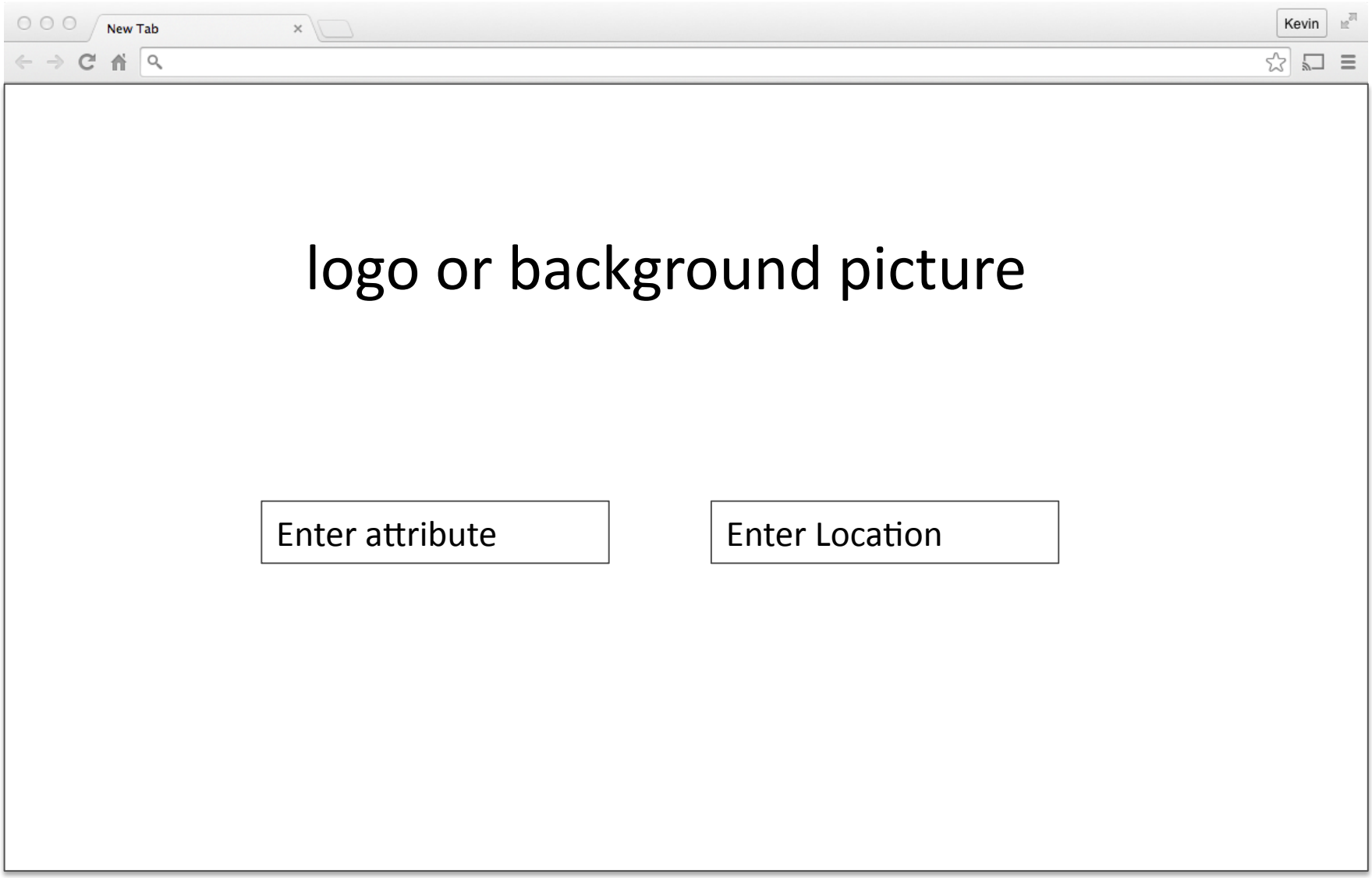
Finding you the best burrito

Kevin Wecht

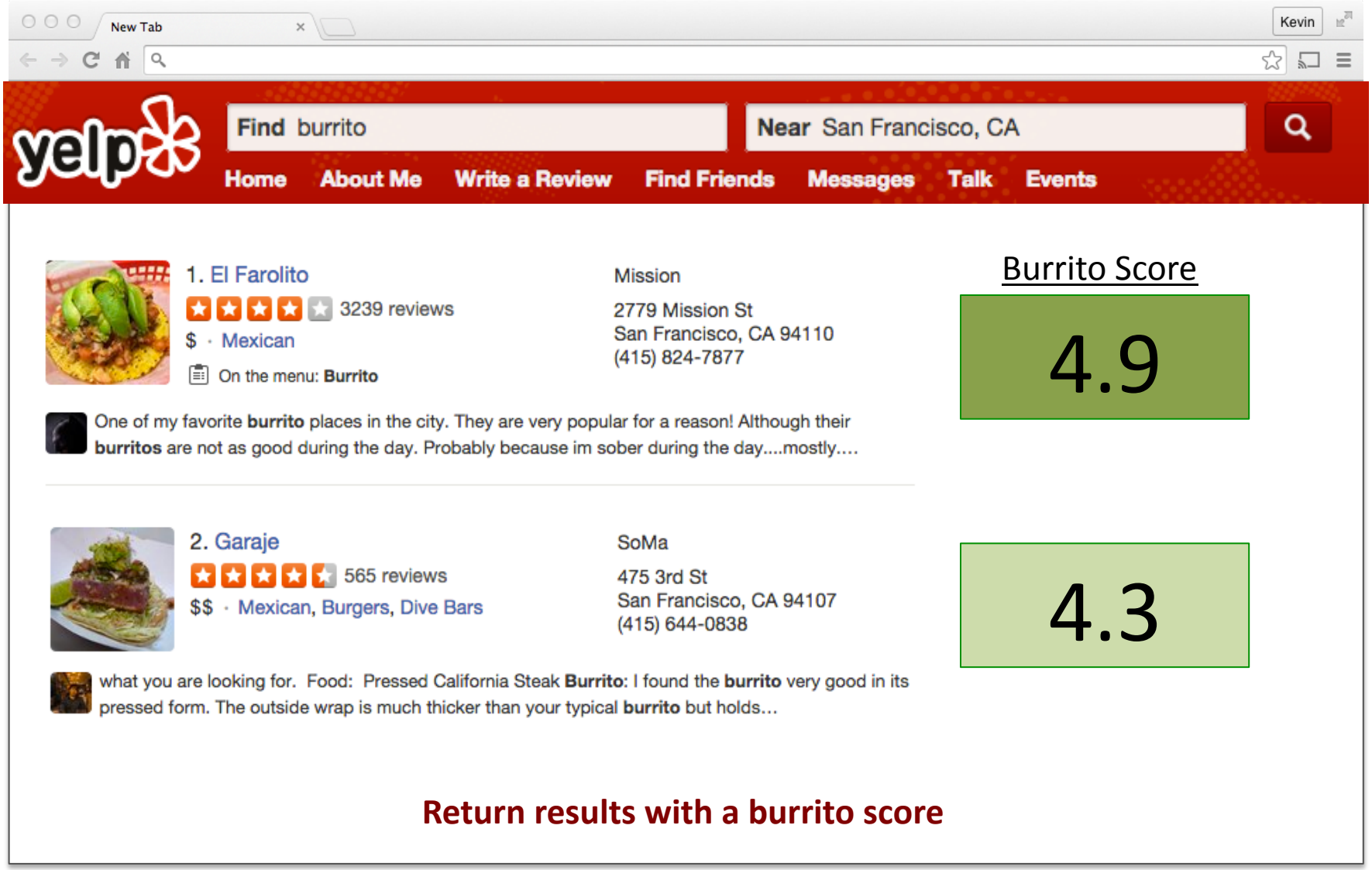
The Problem:

There is no way to aggregate reviews of the best restaurant “dishes” in your area.

Mock-up

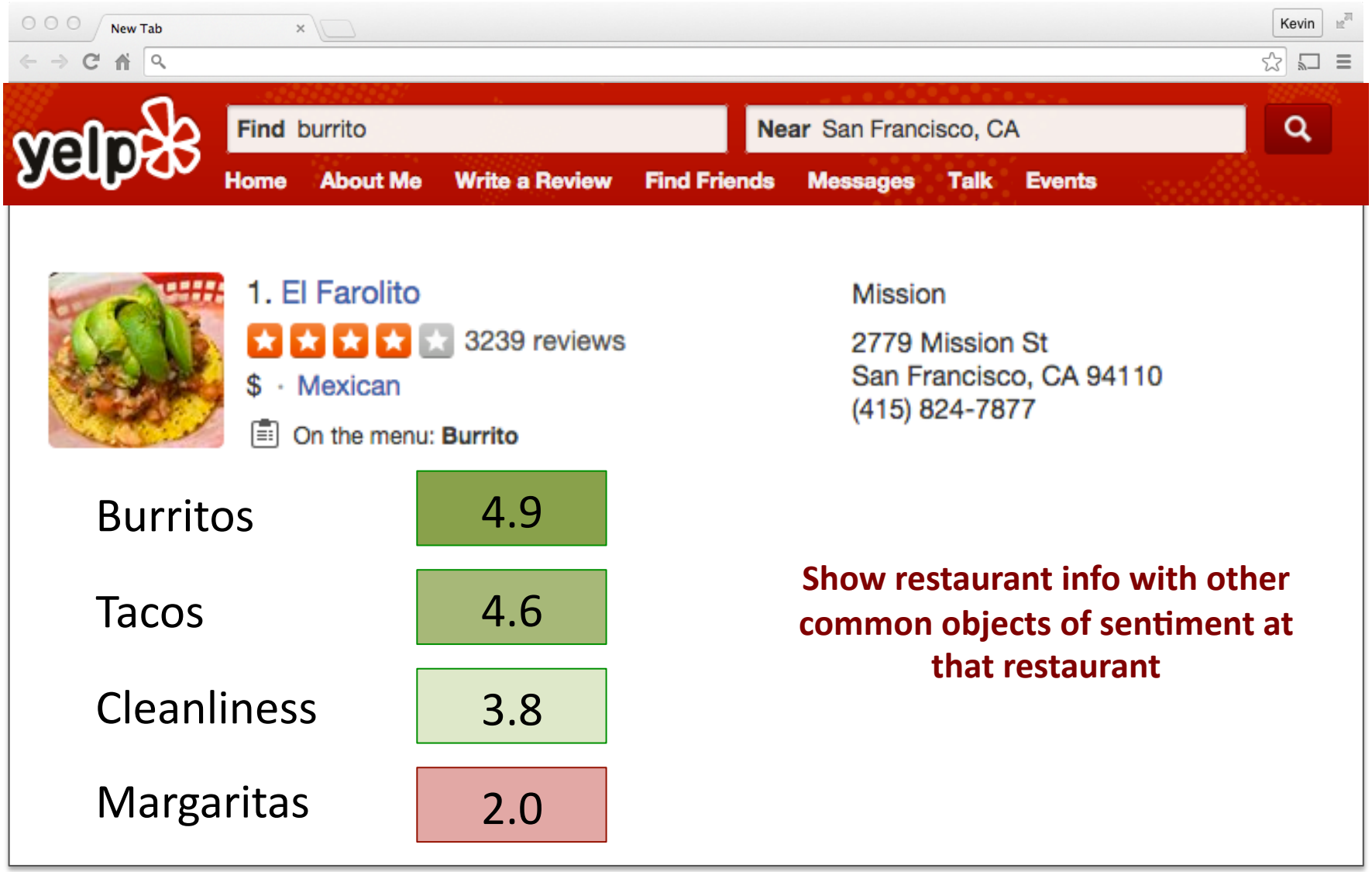

New Tab
Kevin
logo or background picture
Enter attribute
Enter Location

Mock-up



Return results with a burrito score

Mock-up


New Tab
Kevin
yelp
Find burrito
Near San Francisco, CA
Home
About Me
Write a Review
Find Friends
Messages
Talk
Events
1. El Farolito
3239 reviews
$ · Mexican
On the menu: Burrito
Mission
2779 Mission St
San Francisco, CA 94110
(415) 824-7877
Burritos
4.9
Tacos
4.6
Cleanliness
3.8
Margaritas
2.0
Show restaurant info with other common objects of sentiment at that restaurant

Data and Algorithms

- Data
  - ~300K Yelp restaurant reviews from Phoenix, AZ (hopefully SF for demo).
  - Business name and category
  - Rating: [1-5] Stars
  - Text of the review

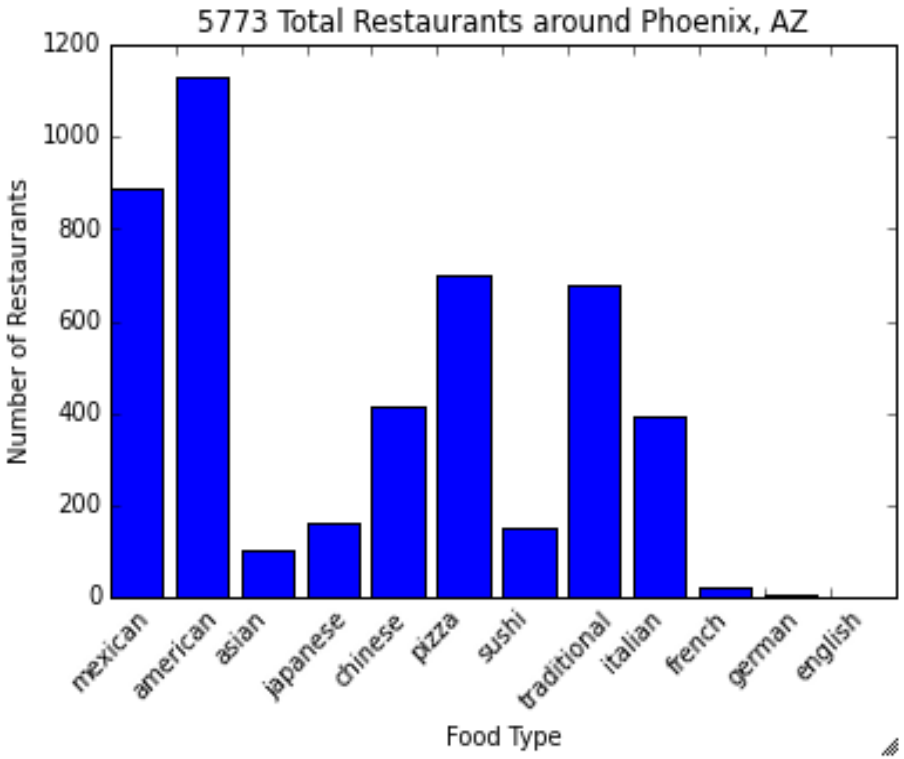
- Algorithms
  - 1. determine objects of sentiment and features that encapsulate sentiment
    - Discover domain dependent objects and features
  - 2. assign sentiment values to unknown features
  - 3. classify (calculate) the sentiment felt toward objects using full set of features

  Will involve NLP for sentiment analysis, clustering, classification?

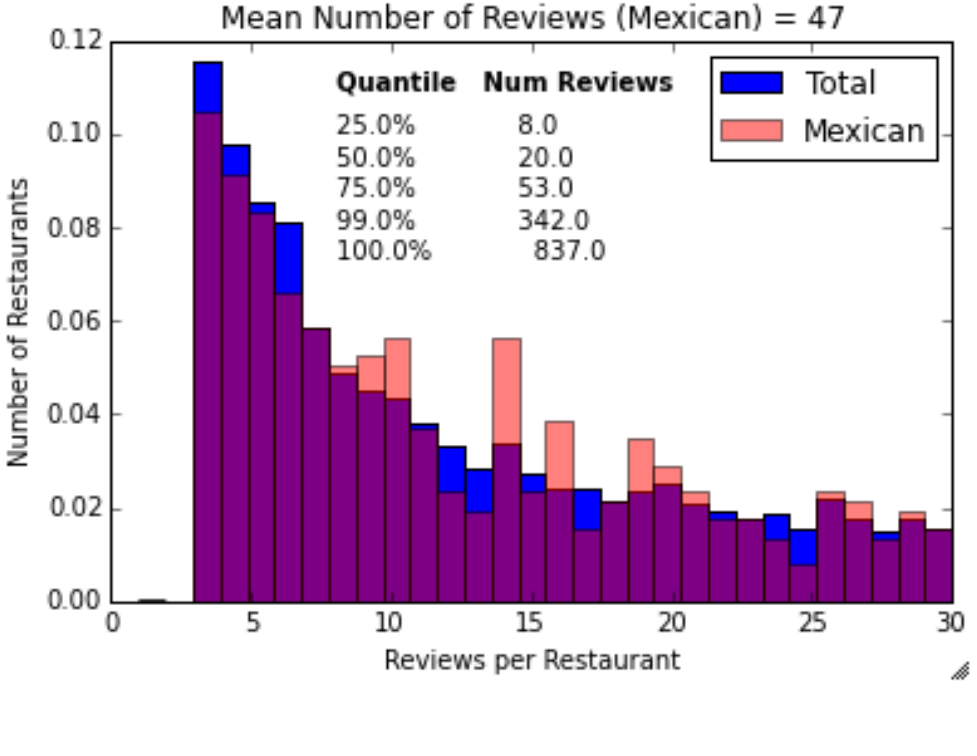
Visualizing the Data

- Yelp reviews from Phoenix, AZ (hopefully SF).
  - How many reviews, how many restaurants, what does the data look like?

How many restaurants?



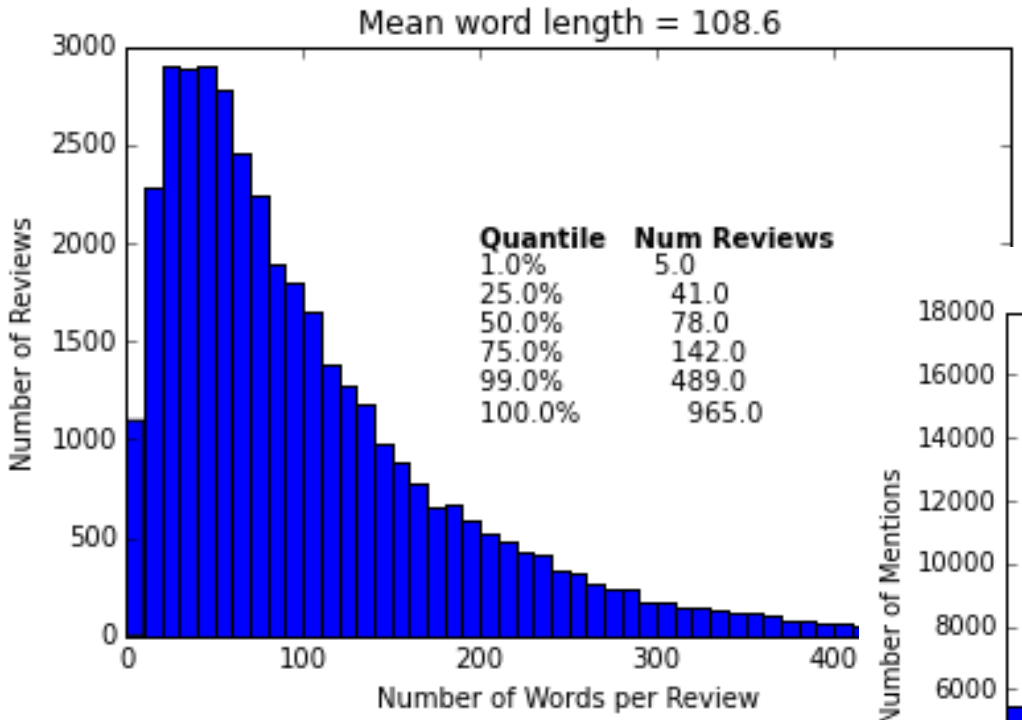
How many reviews?



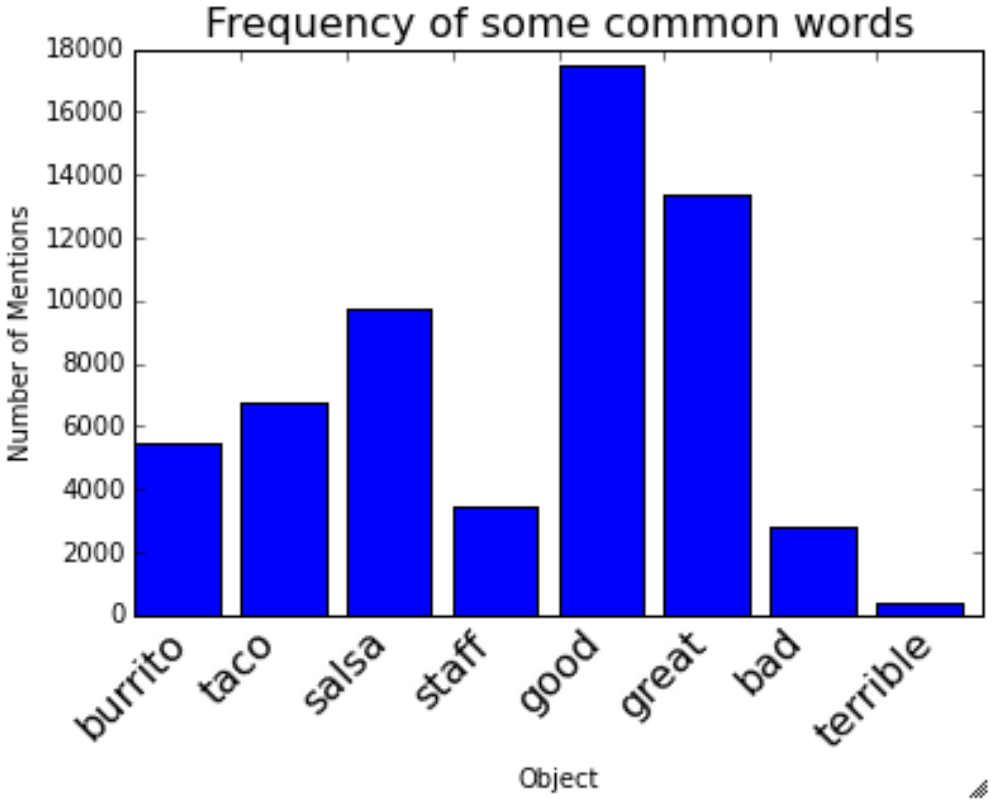
Focus on Mexican restaurants for minimum viable product

Visualizing the Data

How long are the reviews?

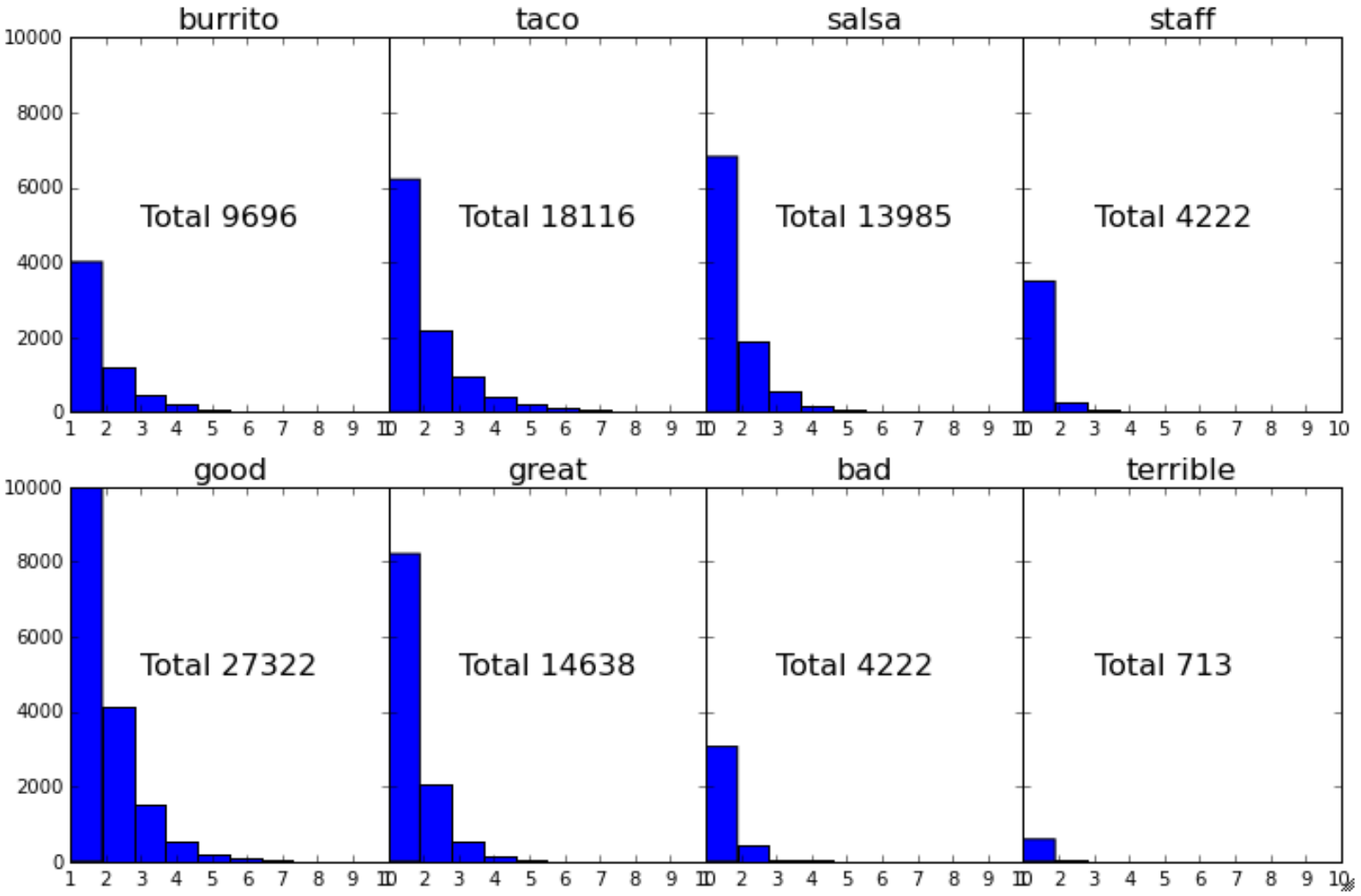


How many reviews mention common words?



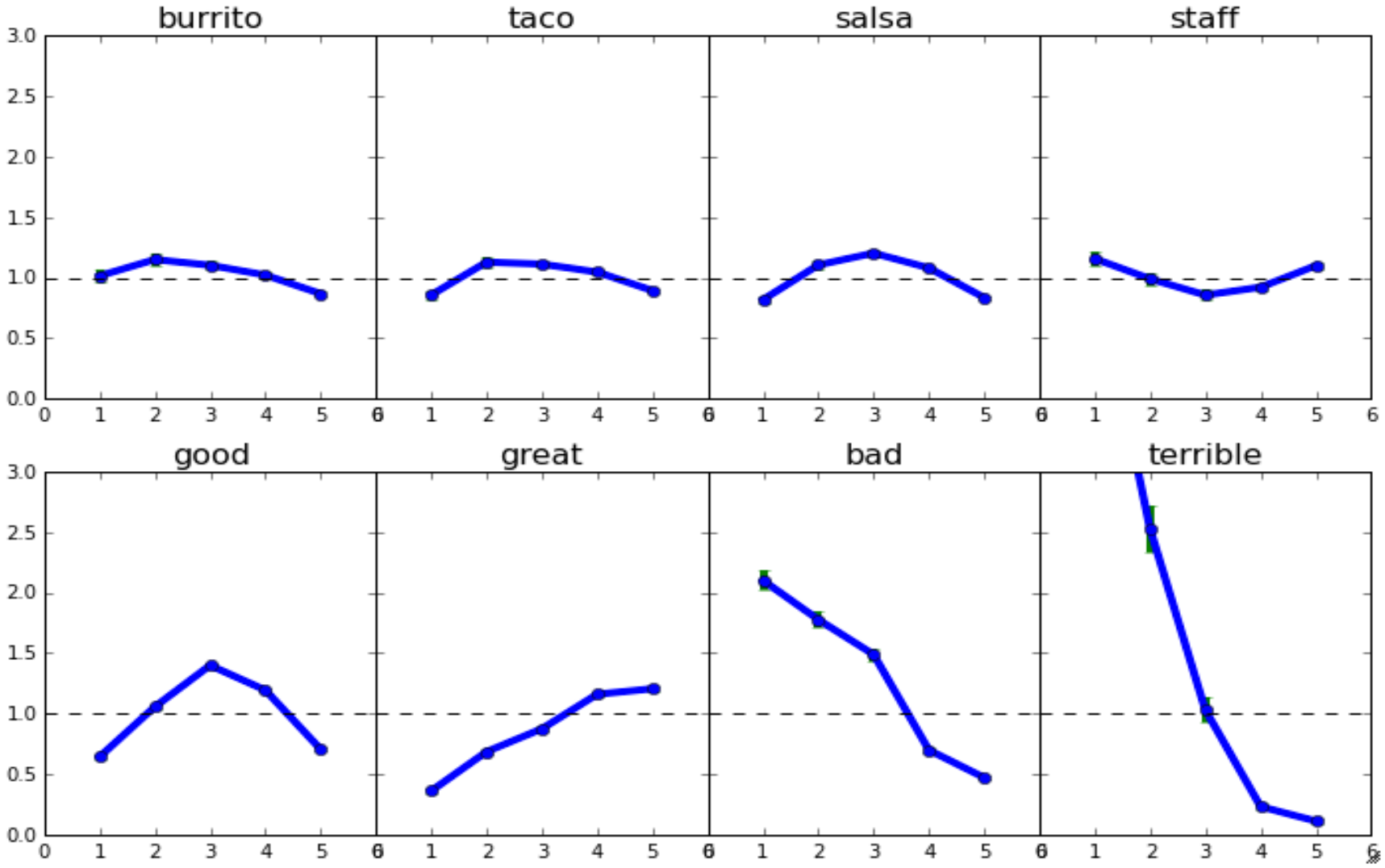
Visualizing the Data

How frequently are words repeated in a review?



Visualizing the Data

How frequently are words used based on their overall sentiment?



Validating Results

How do I know that my results are informative?

- Anecdotal
  - Select a few reviews to read by hand, see if my algorithm does what's expected.
- Sanity Check
  - Results should correlate with overall restaurant ranking.
- Evaluate against baseline performance
  - Baseline algorithm: avg. rating of reviews that mention each object for each restaurant.
- Independent Data
  - 538.com ratings of burritos (perform more highly in my rankings than would be expected given restaurant overall rating or even the baseline above?)
  - Health inspection records (testing sentiment toward cleanliness)
  - Other? Open to brainstorms!
- While in use
  - Do users use the product? Do returning users use it?
  - Do A/B tests swapping the default ordering of results or the sorting options

Expanding Analysis

How can I expand this beyond burritos?

- How to change for other cuisine types?
- How to change for other locations?
- How to change for other businesses types?

- I cannot answer these Qs. The goal entering is to use a framework that can easily be scaled (isn't sensitive to domain-dependent information). In two weeks, the limitations of my actual product will inform the answers to these questions.