EXPLORING THE YELP DATASET

Mary Ann Sundermeyer

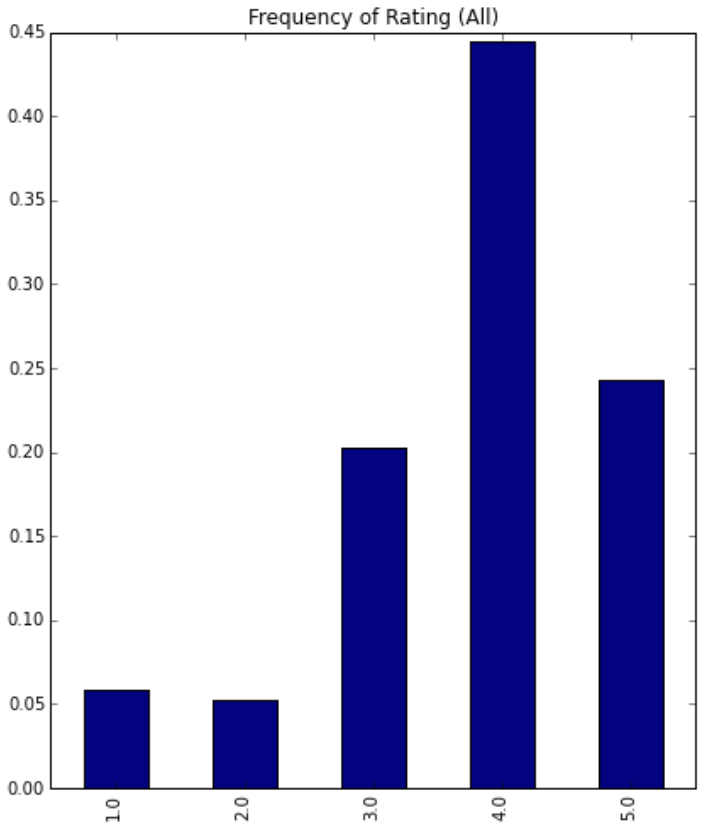
What is it?

- **1.6M reviews by 366K users for 61K businesses**
- **Check-ins over time for each business**
- **481K business attributes, e.g., hours, parking availability, ambience…**

Investigating the ‘One Review User’

- About ¼ of all users have only written one review.
- Is their rating behavior different from users who write multiple reviews?

- 5 stars ≈ 25%
- 4 stars ≈ 45%
- 1 star ≈ 6%

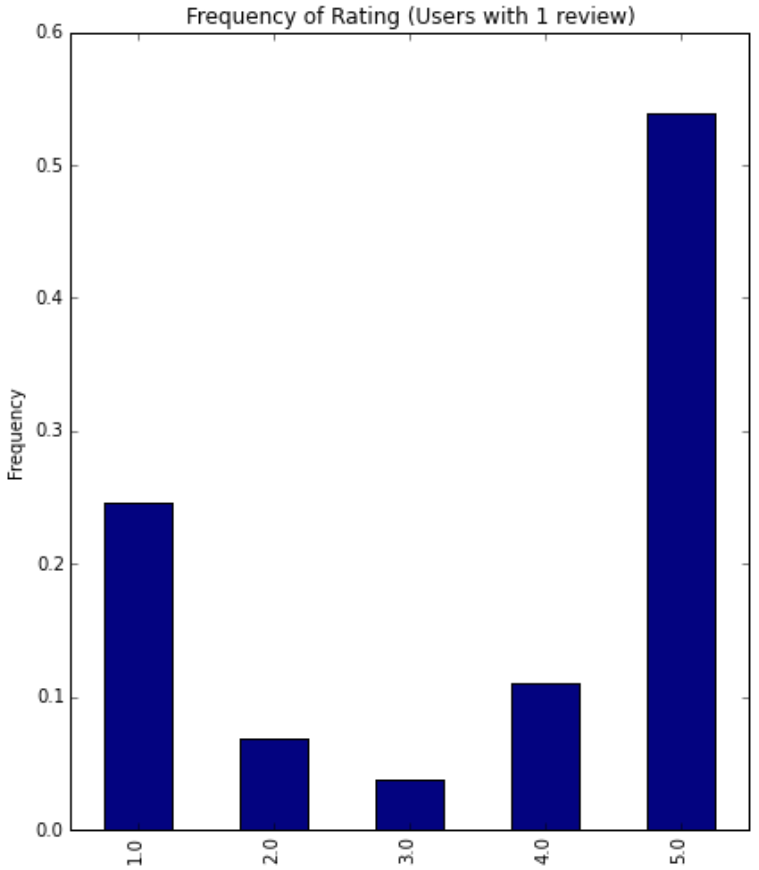


Investigating the ‘One Review User’

- About 80% of the time a user who is writing their first and only review will give an ‘extreme’ rating (1 star of 5 stars)

- People are motivated by polarizing experiences

- Restaurants writing reviews for themselves or for their competitors?



Predicting whether a user gives positive or negative reviews

- Each user has a value 1-5 representing the average rating that they give.
- Define a user as giving positive reviews if this average is between 3.5 and 5.
- Use logistic regression on a set of features to try to predict positive or negative

The results

- Accuracy: 72%

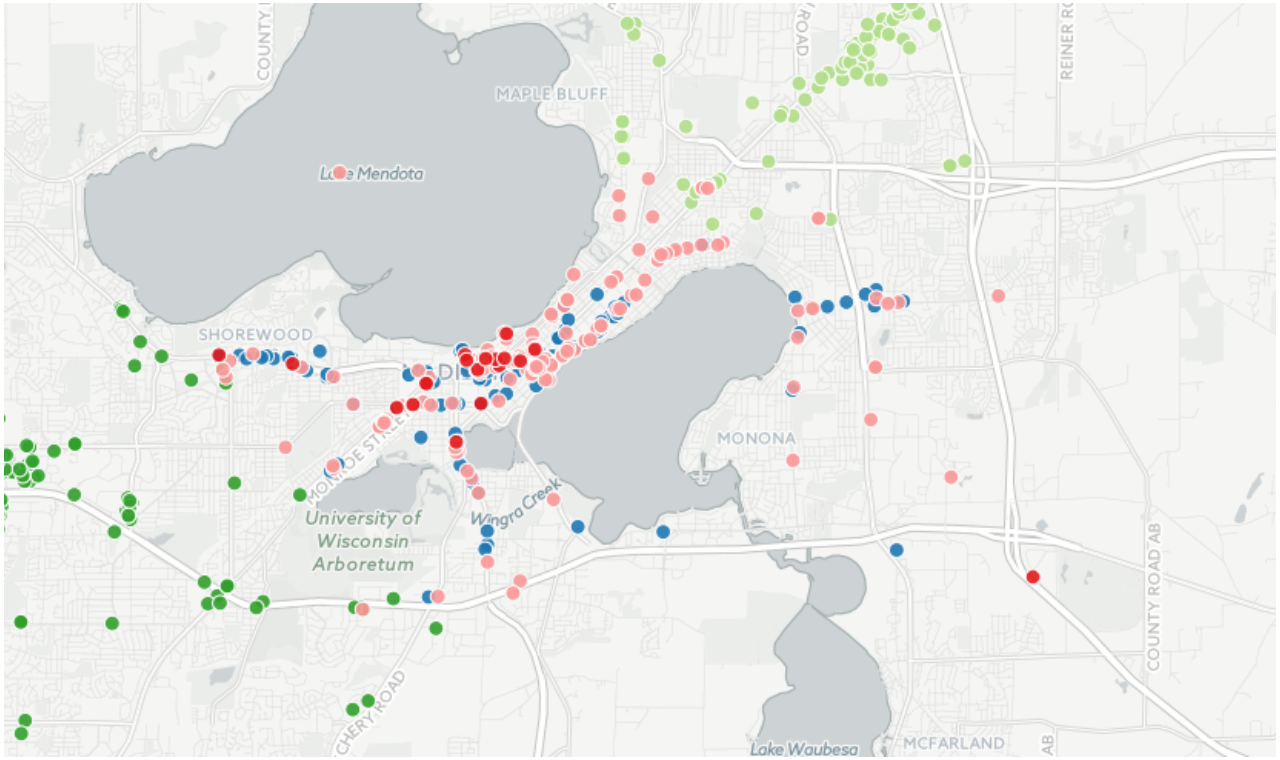
- **People find negative reviews more useful**
- Negative reviews also get more “funny” votes
- **Positive reviews get more “cool” votes**
- More friends → more likely to give a positive review
- More fans → much more likely to give a positive review
- People with a lot of “writer” compliments tend to give positive reviews.

Location, location, location?

- Are the geographic location and the stars of a restaurant related?
- Prediction: High star (4-5) restaurants will tend to be found in prime geographic locations (e.g downtown)

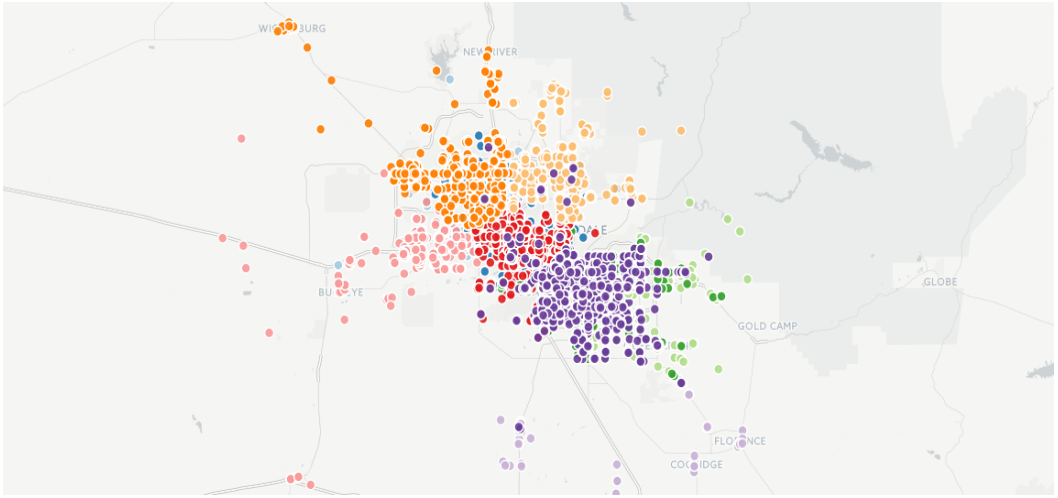
- K-Means clustering on latitude, longitude, and restaurant stars

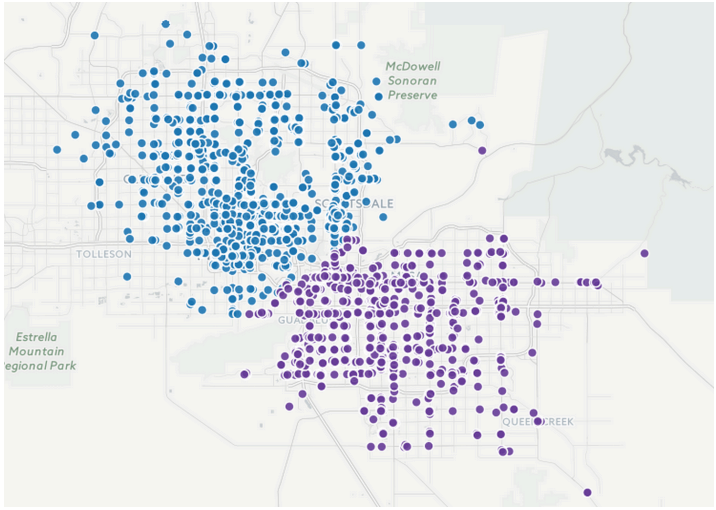
Results: Madison



Pink: 4.14 Stars
Red: 2.16 Stars

Results: Phoenix





How to stay in business

- Use logistic regression to try to predict whether a restaurant has been closed

- Factors: restaurant type, rating, parking options, Wi-Fi…
- Predictions: Restaurants that offer lot of garage parking will be more likely to stay open. Restaurants with a high rating a more likely to stay open. Restaurants that accept credit cards are more likely to stay open.

Results

- Accuracy: 80.3%

- Restaurants least likely to close: Fast food, Chinese, Bars
- Restaurants with Wi-Fi and a high rating are more likely to be open (Wi-Fi more than stars)
- Restaurants that accept credit cards are more likely to be open
- Restaurants serving alcohol are more likely to be closed
- Restaurants with a parking lot a more likely to be closed
- American restaurants and coffee shops are more likely to be closed.