



Katarina Keating, Aanya Jhaveri, Shyamsunder Sriram, and Samantha Stagg

## Summary:

Our group wants to look at Yelp data to examine the circumstances behind how reviews are written. We will start by attempting to answer the question of whether or not weather conditions during a restaurant visit impact reviews, and then go into seeing how location and income information affect reviews. We would also like to examine the text-content of reviews to see if we can predict from the words whether or not the review is positive or not. Additionally, we want to examine how the online delivery market has impacted the frequency and quality of reviews.

## Analyses:

1. Relationship between weather data + reviews:
  - a. Can we see any correlation between weather and average review? (5 or 6)
    - i. Good weather + bad reviews
    - ii. Good weather + good reviews
    - iii. Bad weather + bad reviews
    - iv. Bad weather + good reviews
2. Average Income/QOL information + reviews
  - a. Are higher-reviewed restaurants generally located in higher-income areas?
  - b. Disproportionate representation in reviewers; are they generally from higher-income areas?
3. Sentiment analysis
  - a. Analyzing text keywords; deciding which are positive or negative, and how they're related to ratings (can use mapreduce) (BONUS)
4. Population bias; are restaurants with a lot of reviews better than restaurants with a few reviews, or is review amount just due to population density?
5. See if certain people are just 'worse' reviewers
  - a. What's the average (numerical) review for an average reviewer?
6. Are people more inclined to leave very good or very bad reviews rather than "in between" reviews?
  - a. Classify people based on this?
7. Online delivery market vs Yelp Reviews (ubereats launch, grubhub launch, etc dates have # yelp reviews increased, look for key words) (5 or 6)
  - a. GrubHub launch (1st online food delivery service I think) vs # of yelp reviews

- b. Ubereats launch vs # of yelp reviews
  - c. Postmates launch vs # of yelp reviews
- 8. Increase of general use of internet vs frequency of Yelp reviews/ number of users
- 9. Yelp/Yelp-equivalents in other countries
  - a. Do non-Americans tend to review differently? Leave better/worse reviews?
- 10. Restaurants going out of business (if we can see that in the database)
  - a. Do they tend to have worse reviews (as one might expect) before closing?
- 11. Creating our own holistic restaurant score from our analyses
- 12. Analyze a particular cuisine or even food
  - a. The BEST pizza EVER??

## **Datasets:**

- 1. Yelp dataset <https://www.yelp.com/dataset>
  - a. #lines/entries: **6,685,900 reviews, 192,609 businesses**
- 2. Weather
  - a. <https://catalog.data.gov/dataset/u-s-hourly-precipitation-data>  
<https://catalog.data.gov/dataset/daily-weather-records>