# COGS 108 - Final Project Proposal

## Names

- Shih Wen Ma
- Andrea Diaz
- Viet Nguyen
- Manda Tran

## Group Members IDs

- A15469802
- A15084975
- A13702483
- A14452168

## Research Question

Is there a "bandwagon" effect where customers are likely to give ratings close to a restaurant's current rating on Yelp? How can we quantify this effect? How would we correct this effect?

## Background and Prior Work

Notes:
- How corrupt is Yelp? (How much 'weight' should you put on a yelp review?)
    - What gets filtered out? (criteria)
    - Filter's effects of score (reference #1)
    - Paying to hide bad reviews and highlight good ones (reference #4)
        - Leads to extortion to maintain reputation (reference #3)
- Bandwagon
    - Does current rating affect future ratings?
        - Consider both highlighted reviews and average score as predictors
        - Highlighted extreme ratings affect overall ratings more
    - How does this affect small businesses vs. large businesses?
        - Large businesses get more ratings than small businesses, and may be more vulnerable to such an effect
    - How does demographic / area of interest affect ratings?
        - E.g., we expect San Diego to have a higher quality standard for Mexican food than New York. Chipotle in San Diego may be lower rated than Chipotle in New York.

- Do older reviews have less weight to them? How is Yelp's average review score calculated?

References (include links):

- 1) The Yelp Review Filter:
  https://vivial.net/blog/how-to-avoid-the-yelp-review-filter-and-get-more-positive-reviews/
- 2) The Yelp Effect
  https://rbj.rady.ucsd.edu/index.php/topics/social-impact/2014/01/the-yelp-effect/
- 3) Fake Yelp Reviews
  https://smallbusiness.yahoo.com/advisor/writing-fake-reviews-yelp-might-sued-173135209.html
  https://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9
- 4) The Yelp Factor
  https://hbswk.hbs.edu/item/the-yelp-factor-are-consumer-reviews-good-for-business

# Hypothesis
Yelp elicits the "bandwagon" effect and thus affects the overall reliance of the customer reviews.

Average yelp score encourages similarly rated future reviews. In terms of data, we predict a "funnel" effect where ratings become more condensed over time. Similarly, the ratings of highlighted reviews affect how users will review a restaurant. We predict that a highlighted positive review will move scores up, whereas a highlighted negative review will move scores down.

# Data
*Explain what the ideal dataset you would want to answer this question. (This should include: What variables? How many observations? Who/what/how would these data be collected? How would these data be stored/organized?)*

The ideal dataset would include multiple reviews from both local and chain restaurants from multiple similar regions / areas. This data ideally would be collected without emphasis on certain types of reviews (without the influence of yelp's advertisement services). In addition, the data will be collected by utilizing Yelp's API to get access to their data. Yelp also has a "Yelp Dataset Challenge" which provides a link to a downloadable JSON dataset. The data will ideally be organized based on region, size of business, and amount of reviews. We also need information on how reviews change over time. The data will be stored by utilizing the data wrangling method presented in lecture.

# Ethics & Privacy

- Corrected Yelp Scores could help or hurt some businesses
- Frequent reviewers could potentially be identified by the location of their reviews (since we are collecting temporal data)
  - Since the reviews are available to the public and the reviewers know what they write can be tracked to them, they are consenting to the information they put on their profile.
  - Yelp does not have an anonymous reviewer option, which means that people need to jeopardize their information (ex: location), which could lead to safety concerns, in exchange for writing a review
- Bandwagon and corruption relationship: Can the corruption be traced?

# Team Expectations

- Active communication and equal participation
- Consistently meeting up per week
- Fulfilling tasks assigned

# Project Timeline Proposal

| Meeting Date | Meeting Time | Completed Before Meeting | Discuss at Meeting |
|---|---|---|---|
| 1/27 | 6 PM | Brainstormed possible project ideas. | Finalize research question and project proposal. Find background research. |
| 2/3 | 6 PM | Find data / research for project | Data wrangling, come up with possible factors that affect data |
| 2/10 | 6 PM | Outline analytical approach. | Analyze data and visualizations |
| 2/17 | 6 PM | Start thinking of visualizations for our data | Finish check in, discuss and edit analysis and visualization |
| 2/24 | 6 PM | Continue analysis and visualisations. | Discuss/edit analysis and visuals |

| 3/9 | 6 PM | Complete a draft for the conclusion and discussion. | Revise the entire project thoroughly. |
|---|---|---|---|
| 3/19 | Before 11:59 PM | NA | Turn in Final Project & Group Project Surveys |