

# CS 410 Course Project Progress Report:

## Team Mango

**Project Coordinator:** Saket Vissapragada

**Team Members:**

Ameya Gharpure ([ameyapg2@illinois.edu](mailto:ameyapg2@illinois.edu))

Amrith Balachander ([amrithb2@illinois.edu](mailto:amrithb2@illinois.edu))

Sruthi Kilari ([skilari2@illinois.edu](mailto:skilari2@illinois.edu))

Saket Vissapragada ([saketv2@illinois.edu](mailto:saketv2@illinois.edu)),

Shreya Sharma ([ssharm90@illinois.edu](mailto:ssharm90@illinois.edu))

### 1.) Progress made thus far:

- Project Planning and Task Assignment
  - i.) 10 hours total = 2 hours/person \* 5 group members
- Web Scraper for obtaining text from Google reviews
  - i.) 25 hours total = 5 hours/person \* 5 group members

After developing our implementation plan for our project, we created organized schedules for our team and created a scrum board to organize the tasks that we want to complete (since we have a bigger team, this will help us stay on track).

### 2.) Remaining tasks:

- Apply algorithms covered in class to generate sentiment analysis scores for each review based on commonly used words
  - i.) 30 hours total = 6 hours/person \* 5 group members
- Rank these scores against each other to determine standardized star ratings and compute an average standardized score for restaurants.
  - i.) 25 hours total = 5 hours/person \* 5 group members
- User testing to determine accuracy of standardized scores
  - i.) 10 hours total = 2 hours/person \* 5 group members

As a team, we have gone through a high level approach to completing these tasks and read through online resources to get a better understanding of how we will develop our project. We hope to leave more time aside at the end for the purposes of bug fixes and code refactoring.

### 3.) Any challenges/issues being faced:

- With the webscraper, we were having issues with accurately gathering the text from the Google reviews. We attempted to use a Yelp API to remedy this issue, but then we reverted back to a web scraping library called BeautifulSoup.
- Now that we have gathered the raw text from the web pages, we are facing challenges with organizing the data. Once we work through this challenge, we will be able to apply our algorithms as well as standardize star ratings