CS410 Final Project Progress Report
Xin Jin (xinj3)
Zhicong Fan (zhicong2)

- **Progress made thus far**

  Tried different ways to scrape data from websites. Firstly, we tried scraping data from the bartender website: tipsy bartender. After getting familiar with the html layout, Zhicong began to create a script to scrape the recipes of the drinks. Firstly, he tried using the python package scrapy. By parsing the Xpath and writing the selector, the script was able to scrap the data from a single html. However, the script cannot be running throughout the websites. After investigation, Zhicong found that 99% of the web pages on tipsy bartender have this robots.txt to prevent being scraped from scripts. Then, we tried to find another website to crawl. We then found another website called yummly which contains thousands of recipes for drinks. What's more, we also found that it has an API for us to search on the website. However, the website's API was shut down in 2019 and no new developer could register for using the API. We then found the third website, called all recipes and successfully crawled 500 different drinks from the website and created a csv file containing the image of the drink, the link of the drink, the ingredients of the drinks and the instructions on making the drinks. After that, the framework of the frontend was created. Some local images and text could be put on the website. In the middle,we would display several pictures of the drinks in the future. Along with the picture, ingredients of the drinks would be listed. We also wrote the basic algorithm for calculating the ranking on the drinks' for the backend.

- **Remaining tasks**
  - Improving the performance of the ranking system that compares the user input information with the rest of the items in the dataset
  - Finalizing the recommendation system based on the result of the ranking system and return the result to the frontend
  - Building the communication path between frontend and backend for sending user inputs to backend recommendation system and sending results from the recommendation system to the front end for display

- **Any challenges/issues being faced**
  - Failed to find a website that is easy to crawl and contains fruitful thorough recipes of drinks.
  - Busy with other courses, graduate research work and finding jobs. Hard to manage time among all the tasks