

1. So far I have worked on task 1 of my project proposal which is the following:

*Task 1: Take user input, which is the user's desired location, and search for all the vegan friendly restaurants in the desired location (according to Yelp) in order to gather a list of all potential vegan friendly restaurants nearby.*

*Then for each restaurant in that list, I will scrape and review its Yelp reviews...*

To do this I built two scrapers:

The first one is in Python and it scrapes Yelp data given any keyword and location. My scraper extracts the businesses that show up on Yelp's search result given a particular keyword and location. (For example: keyword = "vegan", location= "chicago"). It extracts the high level details of each business in that search result such as the business name and Yelp URL. To use this scraper I would provide the keyword "vegan" and the location that the user provides as input to gather the list of businesses along with their Yelp URLs that show up on Yelp's search result for that particular keyword and location.

The second scraper is also in Python and it uses BeautifulSoup and the web scraping API ScrapingDog to scrape public reviews for each business in the list gathered from the first scraper. For each business on the list it makes a GET request to a target URL (which is a business's Yelp URL that I get from the first scraper) to get a business's raw HTML data. BeautifulSoup parses this and from here I run a for loop to get every review for that business.

2. The tasks pending are implementing a good search query, implementing a document selection algorithm, performing sentiment analysis, and performing evaluation.
3. As of now I am not facing any challenges.