Hongyu Dai

Unique name: hongyud

SI 507

Final Project Proposal

For the final project, I plan to use two Web API that I have not used before that requires API key. The first one is Google Places API in Google Maps Platform, with the following website: https://developers.google.com/maps/documentation/places/web-service/overview, and the second one is Yelp Fusion API, with the following website: https://www.yelp.com/developers/documentation/v3/get_started . Two of them in combination will contribute 8 challenge scores.

In addition, I would like to scape a new single page, which is the QSR50 2020: https://www.qsrmagazine.com/reports/2020-qsr-50 . This will contribute 4 challenge scores. In sum, we will have 12 challenge scores.

In practice, I plan to create an interactive interface for users to examine the quality of America's 50 biggest fast-food restaurant chains in different states of the country based on the ratings from Google and Yelp API. As I imagine, I will first let the user to enter the name of the brand that he/she want to examine, and use web scraping to check if the brand that has been entered is in our QSR50 list. If it is one of the 50 brands, then the user can continue; if not, I will tell the user that the name he/she enter does not belongs to one of the 50 biggest brands.

Next, I will let the user to enter the alpha code of the state that he/she chooses (for example, enter CA if he wants to search the restaurants in California). I may drop this step and instead only examine the quality in a specific state (like Michigan) because this step might be difficult for me to design, and the result that web API returns might also be problematic for different states.

Then, for the brand and the state that the user has directed, I will find all the restaurants of that brand in that states, and scrape their rating in google and yelp API. Suppose Macdonald has 500 restaurants in Michigan, after this process I will get each of the 500 restaurants' google rating and yelp rating. We will average the rating and make the difference of the two ratings and get 2 additional datasets.

Finally, I will plot 2 histograms for average rating and difference of the two ratings respectively, and give the mean, variance, skewedness, and kurtosis of the 2 datasets to help users examine the quality. The first histogram is mainly used in examining the overall quality and quality control of the brand. For example, if our histogram is heavy tailed, we may think that the quality control of MacDonald in Michigan is not that good. The second histogram is mainly used in examining the difference of preferences between yelp users and google users.

I have not made any attempts to access the two APIs. But I have checked several articles and all of them said it is not that difficult to access the Google and Yelp APIs. If I fail to access one of the two APIs , then I will just use another API's rating as my source instead of averaging the two, and drop the difference of the two ratings part. This will still ensure that I have 8 challenging scores.