Project 1 - Proposal

Team Members:

Christopher Yang
Sameer Zubairi
Alan Deng
Jacob Evans

Project Description/Outline:
Analyzing Yelp business and user data via the Yelp API, and cross-referencing it if required to geolocation data via the Geoapify API to analyze correlation of Yelp ratings by city, proximity to busier areas, types of restaurant, hours of operation and user age on Yelp. Can either focus on one city (i.e. Toronto) or two different cities (i.e. Toronto vs Vancouver)

Research Questions to Answer(~4 Questions):

Yelp + Geoapify

- Which cities have the most numbers of highest rated restaurants (percentage wise and just pure total number) (yelp + geoapify)
- Average age/ age groups reviewing popular restaurants in certain areas
- Which cities have the most variety of restaurant categories? (Yelp dataset only)
    - Does having more variety reduce the overall ratings of all restaurants in a city?

Datasets to Be Used:
Yelp: https://www.yelp.com/dataset
Geoapify: https://apidocs.geoapify.com/

Rough Breakdown of Tasks:

1. Get access to the Yelp and Geoapify APIs
2. Initial analysis and cleanup of raw data
    a. Create tables/data frames from JSON files
    b. Combine them to come up with insights to answer question
3. Create visualizations for the data collected
4. Create presentation with the visualizations

Appendix

(Other questions)

- Does restaurant rating change with category type (total number of category type → which is most popularity, and then category type vs. average rating) (yelp dataset only)
- Does restaurant rating change for restaurants that stay open longer (24 hour takeout vs dinner service only restaurants) (Yelp dataset only)

- Restaurant distance from city center vs rating (Yelp + Geoapify)
- Expansion of Module 6 class 3 assignment 4 → cross reference Geoapify “nearest ethnic restaurant” with Yelp star rating review and sort from highest to lowest (Yelp + Geoapify)
- Restaurants that have permanently closed over COVID (Yelp dataset only)
- Amount of time user has been on yelp vs. reviews written (Yelp dataset only)
    - Statistical breakdown of how many reviews write per year (i.e. only 20% would write more than 10 reviews in their first year of use or something)
    - Does user average review scores go up or down the longer they’re on yelp
    - Does average user star rating correspond to the type of compliment they receive (funny, normal, cool, funny, etc.)
- Which cities have the lowest number of reviews of restaurants? Which cities have the most review counts?
    - We can also look at the relationship between the number of reviews to average rating