# Food Recommendation Project

David Nguyen
University of Tennessee, Knoxville
Knoxville, Tennessee
dnguye18@vols.utk.edu

Catherine Fei
University of Tennessee, Knoxville
Knoxville, Tennesee
catfei1@vols.utk.edu

Spencer Harper
University of Tennessee, Knoxville
Knoxville, Tennessee
hjw848@vols.utk.edu

Hudson Doerr
University of Tennessee, Knoxville
Knoxville, Tennessee
hdoerr@vols.utk.edu

Isaac Sikkema
University of Tennessee, Knoxville
Knoxville, Tennessee
isikkema@vols.utk.edu

*Abstract—* **The growing data of food reviews across different platforms allows for the possibility of manipulating this data to recommend a food to a person. Our team will collect a variety of datasets from different food review spots (Yelp, Google, etc.) and categorize this information to be able to recommend different foods to people.**

*Keywords—food, data, recommend, people*

## I. Objective

The objective is to obtain the Yelp dataset and analyze it. The Yelp dataset provides over 200,000 businesses and their information. Using this, we are able to analyze where what types of restaurants were included and where they are most populated. We will break this information down and see the different types of statistics these restaurants provide for us.

## II. Motivation

For the motivation of this project, the idea came when I couldn't decide on what to eat for the day. I just needed to be shown a restaurant that has something I like. I just wished I was given a suggestion on where to eat with something I like. YouTube has an algorithm that suggests videos based on what you previously watched. I wanted to make something similar but less advance because YouTube's algorithm is a little too complex.

## III. Data And Models/Algorithms

For the data of this project, we used a dataset that was provided from Yelp. The dataset from Yelp gave us the information of roughly 200,000 restaurants. The information included for each restaurant was vast and included things like: rating, reviews, location, category, etc. We took this vast dataset and chopped off the least important bits. For example, since Yelp included businesses that were not restaurants, we removed them from the dataset. As for models/algorithms, no specific algorithms were used in processing this data. We did use some graphing libraries like seaborn and matplotlib to plot the data for visuals.

## IV. Responsiblities

Each member will be responsible for roughly the same things. Since some of us are more experienced in data scraping, they could help a little more with structuring the code. We all have very similar coding skills so that we can contribute a little to the program.

## Results

From our analysis, we were able to successfully see the different types of restaurants that were included in this dataset. We can see that the highest rated restaurants are in Las Vegas and include some very famous places. Not only that, but we were able to see a ton of restaurants in Ontario, Canada as well. We were able to perform some analysis like seeing what were the top 10 restaurants with the highest review counts. By seeing these restaurants, you could see all their information like rating, location, and reviews. If someone had unlimited money, they could use our list as a basis on where to eat in the world.

## Problems

A primary issue that occurred during our project was finding out the dataset had restaurants in Canada in it. We initially wanted to recommend restaurants in the United States but then pivoted to analyzing the different restaurants of the U.S. and Canada. Although we wanted to recommend restaurants using this dataset, we were able to pivot to something that still made effective use of this dataset.

### A. Responsibilities of Each Member

David Nguyen
1) Issue tracking on GitHub
2) Assigning Issues
3) Implementing Code
4) Paperwork
5) Adhering to timeline

Hudson Doerr
1) Implementing Code
2)Issue Tracking on GitHub
3) Paperwork
4) Adhering to timeline

Catherine Fei
1) Implementing Code
2) Issue Tracking on GitHub
3) Paperwork

4) Adhering to timeline

Isaac Sikkema
1) Implementing Code
2)Issue Tracking on GitHub
3) Paperwork
4) Adhering to timeline

Spencer Harper
1) Implementing Code
2) Issue Tracking on GitHub
3) Paperwork
4) Adhering to timeline

## B. Timeline

October 6th - Meet with team and figure out structure of code

October 8th - Test parsing of a dataset

October 13th - Make different food categories

October 15th - Test putting a dataset into the categories

October 20th - Store user's preferences

October 22nd - Parse the other datasets

October 27th - Parse the other datasets part 2

October 29th - Find the largest category and best restaurant in that category

November 3rd - Rest Day

November 5th - See if that category matches user preference

November 10th - Randomize best restaurants

November 12th - Finishing touches

November 17th - Finishing touches

November 19th - Finish

## C. Future Work

Since we found out that this dataset was very selective and did not include as many restaurants as we would've liked, we could find other dataset that include other parts of the U.S. and merge them with this one. By merging multiple datasets together, we should be able to form a deeper analysis of the United States and could possibily expand it to other countries.