**Group Members:** Marissa Stanvick and Maxwell Lloyd

**Github Repository:** https://github.com/maxwelllloyd/cs171-finalproject.git

1. **Background and Motivation.** Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

Going out to eat is a common occurrence in many households. When going out to eat, many people base their decision on

* 1. Proximity of Restaurant
  2. Cost of Restaurant
  3. Type of Cuisine
  4. Restaurant Reputation

Yelp is a website which tracks a lot of the above data. They allow users to rate restaurants and write reviews. This allows others to gain more information about the restaurant before visiting.

Currently, Yelp has a map functionality. However, you can only view 10 restaurants at a time and cannot easily view all of the restaurants in a certain area. Currently, you also can’t compare restaurants. The comparison of restaurants is critical in selecting where to dine, and the aggregation of reviews is useful in deciding if the restaurant is right for you. Perhaps you are looking for a pizza joint, but the reviews specify that the restaurant actually specializes in subs. There is also no aggregation of the text in the reviews. Our Yelp visualization intends to build on these areas.

1. **Project Objectives.** Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

The primary goal of this visualization is to be able to choose a place to dine based on Yelp reviews. From this visualization, you will learn:

* 1. What restaurants are in the area
  2. What restaurants have the best rating
  3. Where specific cuisine restaurants are located
  4. Which restaurants have the highest number of reviews
  5. What the major topics in the reviews are for each restaurant
  6. What the specifics of each restaurant are

1. **Data.** From where and how are you collecting your data? If appropriate, provide a link to your data sources.

The data for this visualization will be from the Yelp data set challenge (<https://www.yelp.com/dataset_challenge/dataset>). This data is a collection of JSON files that include information about each business, the reviews, and the reviewers.

1. **Data Processing.** Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

Because we are focusing on our design on restaurants, we will have to clean-up the Yelp data to include only business that serve food and beverages. Currently, the data set includes other businesses such as hospitals or nail salons. By cleaning up the data, our data set will be smaller and help with the load time for our website.

From the data set we will be deriving the following quantities:

* 1. Average # of Stars
  2. Total # of Reviews
  3. # of Times Specific Words are Mentioned in Reviews for a Restaurant

1. **Visualization.** How will you display your data? Provide some general ideas that you have for the visualization design. Include sketches of your design.

The sketches for the design are located in the PDF file “Yelp Project Proposal Sketch”. Because our design is trying to answer the question of “where should I eat”, it will primarily be map based. The user will have the opportunity to interact with the map to select a certain area to eat. They will be able to filter the restaurants based on type of cuisine or other important features such as delivery. They will then be able to look at a small subset of restaurants to compare the number of reviews and what words are mentioned most often in the reviews.

1. **Must-Have Features.** These are features without which you would consider your project to be a failure.

Our design must be map based. The user needs to be able to first select the general section of the country they are eating. They then must be able to brush a smaller area.

The user must be able to filter based on the type of cuisine. This will be crucial to limiting the amount of restaurants the user sees. They could also filter on delivery or cost.

The nodes on the area map (visualization 2) need to be colored by rating so that the user can have a general idea of how a restaurant is regarded.

The user must be able to compare restaurants based on the number of ratings because the number of ratings is an important marker of how popular a restaurant is.

We must be able to aggregate the top words from each review. For example, if a review mentions pizza a lot but the user wanted a sub, that would be useful in selecting the restaurant they want to eat at. At minimum, we should be able to provide this information in a bar graph or force layout.

1. **Optional Features.** Those features which you consider would be nice to have, but not critical.

One optional feature would be to display the review information in a word cloud. A word cloud is a neat visual way to quickly pick out which words are mentioned most often in the reviews. Ideally, we would color these words based on the average number of stars given in the reviews which include the word. For example, if 100 reviews mentioned pizza but the average rating of these reviews was 1 star, the pizza at the restaurant may not be good.

An additional optional feature would be for the user to be able to zoom in and out of the selected area similar to Google Maps functionality. The required feature is the ability to brush a certain area.

An additional optional feature would be to have a rating and # of reviews filter on the area map visualization. The user can then select if they only want to look at restaurants with a X number of reviews and an average rating of greater than Y.

1. **Project Schedule.** Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

**April 3rd:** Project Proposal Due

**April 5th – April 11th:**

* Data: (Max)
  + Gather data source
  + Format data source
* Map: (Marissa)
  + Learn D3 map functionality
* Visualization Pieces: (Marissa)
  + SVG 1: Map
  + SVG 2: Selections / Options
    - * Cuisine
      * Other Attributes
  + SVG 3: Detailed Visualization
    - Bar Chart for # of Reviews Comparison
    - Word Cloud for Review Text
    - Overview Box with Specific Info

**April 12th – April 18th:**

* Map
  + Plot restaurants on map
  + Select map area
  + Plot restaurants on specific map area
  + Color points based on review

**April 17th:** Milestone 1

**April 19th – April 25th:**

* Map
  + Filter restaurants based on cuisine and attributes
  + Brush map to select a small number of restaurants to compare
* Comparison Visualization
  + Bar chart to compare the number of reviews
* Detailed Info Visualization
  + Word cloud to determine most popular words
  + Text box with detailed restaurant information

**April 26th – May 5th :**

* Final details!

**May 5th:** Final Project Due