DS 4200 Final Project: Yelp Visualization

Elise Chan
Northeastern University
Bouve College of Health Sciences

Ryan Chapados
Northeastern University
Khoury College of Computer Science

Jinwoo Jung
Northeastern University
Khoury College of Computer Science

Xining Xu
Northeastern University
Khoury College of Computer Science

ABSTRACT

A solution to late night hunger - we are creating visualizations that allow users to assess restaurant quality, popularity, and distance. While many map softwares already exist, they lack high level differentiating features that can deter those who are not "foodies" from searching for new restaurants. We bring a new approach that emphasizes simplicity by evaluating quality and popularity against each other head to head and incorporating better design choices. With this combination of features, we provide solutions to the following domain tasks: 1) identify food options that are geographically accessible, 2) allow users to prioritize restaurant popularity or quality, and 3) quickly assess what recent customer sentiment for a restaurant has been to overall 4) make the most informed decision to purchase takeout in the most time efficient way.

1 INTRODUCTION

Food is a constant that often has great cultural significance. Especially in recent times, food has become a prominent element of pop culture that has become a popular trend. On the flip side, this has only exacerbated the indecision that comes with choosing what to eat. This can be distressing when you are already hungry and unfamiliar with the geographical location you're currently in. With our visualization tool, we want to address the average user who often just needs to find a restaurant that is relatively good quality, close by, and not too overloaded with orders. By preventing information overload and streamlining the visualizations to include these main points, users can make informed food decisions more quickly.

Thus, our domain tasks are to:

1. identify food options that are geographically accessible
2. allow users to prioritize restaurant popularity or quality
3. quickly assess what recent customer sentiment for a restaurant has been
4. make the most informed decision to purchase takeout in the most time efficient way

Please refer to Appendix section 9.2 for further details about domain task abstraction.

The goal end user is someone like a college student who lives in a dorm. This user is under immense stress, but does not have the means to cook for themselves and likely unfamiliar with their surroundings. To create the visualization tool for these end users, we will be cleaning data from a subset of the Yelp dataset to fit project parameters.

2 RELATED WORK

- The research conducted in 2021 3rd International Symposium on Smart and Healthy Cities (ISHC) [2] analyzes restaurants in the area using sentiment analysis of online restaurant review. Similar to our idea of using Yelp, the study uses online platforms Meituan and Weibo that are widespread in China. Although the ultimate goal of their project is to provide help for restaurants, we realized that most steps in the process align with our goal of the project, which is to utilize and analyze online reviews of restaurants. They had an interesting visualization, where they used a scatter plot format to plot restaurants in the area using color saturation to rate restaurants, providing us with a solid base work for our own visualizations. For instance, Figure 1 shows one of the graphs that they created plotted restaurants in the area using color saturation to show the ratings of 'taste' criteria. We are planning to extend this data visualization by adding interactions to each dot, so that when we hover over each point, the website will provide more specific information like the name of the restaurant, avg. rating, and even distance from a user's current location.

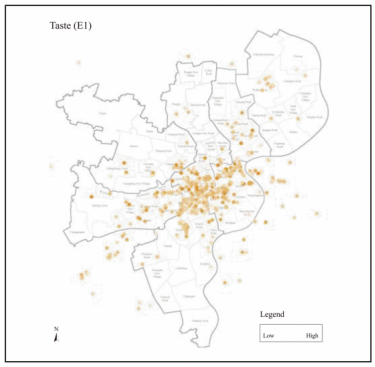

Figure 1: A visualization of restaurants in the area while using 'taste' as a criteria. The image is from [2] and is in the public domain.

- A research in 2020 0f 'Open AccessArticle Customer Restaurant Choice: An Empirical Analysis of Restaurant Types and Eating-Out Occasions' [1] shows that an investigation of key factors driving customers' restaurant choice in eating-out decision making not only can help restaurateurs understand restaurant customer perception of key factors when selecting a restaurant, but also form appropriate marketing strategies to attract existing and potential customers and outperform competitors. This article extends the body of knowledge on the relative importance of restaurant selection criteria. According to their research, they found that menu price is a really important function, which is a really important factor we are going to take a look at in our analysis. Also, eating- out is the key determinant of restaurant selection criteria. In order to solve this problem, our group created a fake map which can show the location and the price of the restaurant.

3 Use Case

Say you've just moved to a UC Santa Barbara dorm and don't have any means to cook. You look up "take out restaurants" on popular search tools like Google or Yelp, but you are tired and overwhelmed. Using our tool, the user enters their current location, current time and preferred mile radius. All Santa Barbara restaurants within that range that are currently open are displayed on a mini map, along with the current location. Each restaurant also lists the name, rating, number of reviews, and the distance to the restaurant. When a restaurant is clicked, a bar chart shows on the side showing the distribution of all reviews by star rating, with each category labelled with the date of the most recent review. The user can either use the map to browse what is located next to each other, or use the bar chart to identify recent sentiments about that restaurant. The two visualizations allow the user to choose whether they prioritize distance, rating, or popularity without having to step foot into the restaurant.

4 Data

Our data set is taken from Yelp, and is a compilation of businesses and reviews that Yelp put out for public use. A link to the data set can be found https://www.yelp.com/dataset. This data was not what we originally expected, as it did not contain data from all cities but only a small sample of cities. However, in the cities it did include, a wide variety of restaurants and other businesses were selected. For our purposes, we chose only locations in Santa Barbara, and we filtered out all values that were not restaurants which left us with around 700 lines of data. While the restaurants were not selected in any particular way, we will also be using reviews in our visualization. Reviews are naturally biased, as they are a reflection of an individuals experience, but the hope is that with a high enough quantity of reviews we will be able to get a good idea of the actual quality of the restaurant. However, for restaurants with very few reviews associated with them, we may not be able to accurately assess that restaurants quality. Please see Appendix 9.1 for further details about data abstraction.

5 Design Process

5.1 Sketches

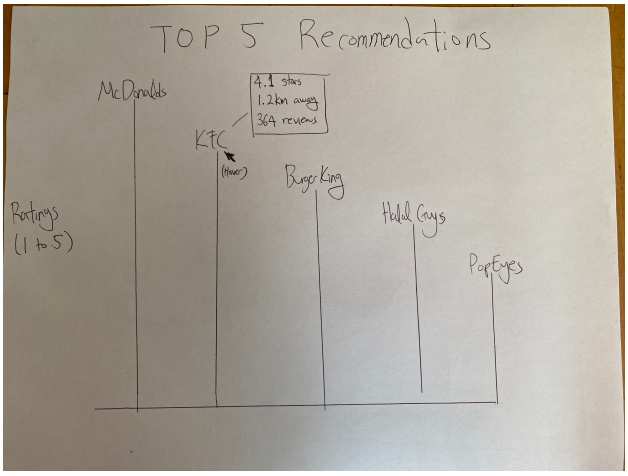

Figure 2: Rough Sketch 1 - show top 5 restaurants against each other showing star rating as height, with tool tip showing average star rating, distance from user, and number of reviews

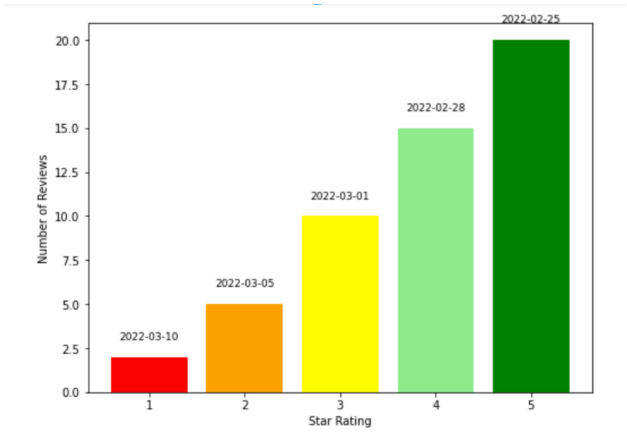

Figure 3: Rough Sketch 2 - show the review distribution of one restaurant by star rating, as well as the most recent date of each category's review

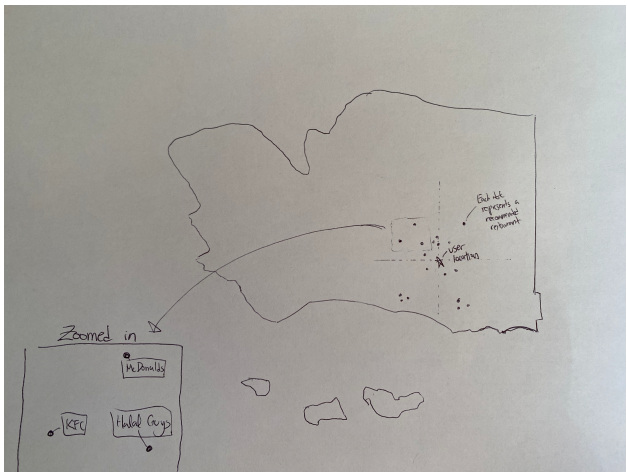

Figure 4: Rough Sketch 3 - plot user location and restaurants based on x, y coordinate plane scatter plot, can map over an actual geographical area to help user visualize

5.2 Final Sketch

In the final sketch, we incorporates components from sketches 2 and 3, but made minor adjustments over what we thought may be more visually appealling.

For Visualization 1, or the map portion, the Marks were the points indicating the different locations of interest, such as restaurants or the user's location. Each location is plotted along an x, y coordinate plane with the user automatically centered at (0, 0) and the restaurants accurate converted to a point on the system that reflects its real life distance away. Channels included position of each point to represent relative geographical locations, as well as color to differentiate between restaurants and users. Although color was not implemented in the sketch since it was done in black and white, it will be implemented in the final implementation. There is also a tool tip that appears when hovering over a point. This includes the restaurant's name, average star rating, number of reviews, and distance to user.

For Visualization 2, or the bar chart portion, the Marks were "lines", or the thick bars. When the user clicks on a restaurant, the visualization will show the distribution of reviews by star rating for that restaurant. There were numerous Channels, the first of which is horizontal and vertical position. The position of the bar was associated with the quantity of reviews for that star rating. The second Channel was color, to more easily differentiate between the increasing star ratings. The bar colors between 1 star and 5 star ratings were arranged along a continuum, with red associated with 1 star reviews which have a "bad" connotation and bright green associated with 5 star reviews which have a "good" connotation. Again, although color was not implemented in the sketch since it was done in black and white, it will be implemented in the final implementation.

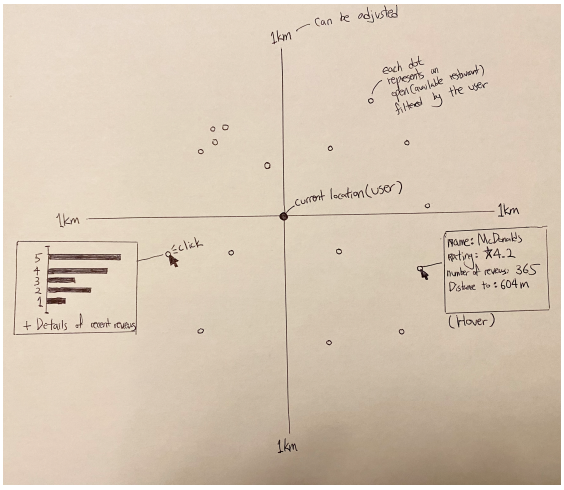

Figure 5: Final Sketch

5.3 Usability Testing

We found the usability testing in class to be a good trial run in our process, especially as 2nd to 5th year college students, we fit the target end user of our tool quite well. We made sure to include classmates who are well versed in sites like Yelp, as well as those who don't use online tools to find places to eat at all.

Since an important component of our tool is to improve ease and efficiency in decision making. Our first test was comparing time to evaluate which restaurant had a better review distribution. We did this by giving our testers the same graph of star review distributions with a vertical and a horizontal bar chart. We found that users unanimously preferred and were more efficient using the vertical bar chart. When we asked for comment, they indicated that they were more used to a vertical format, and that it was more intuitive that the star ratings increased when reading bars from left to right, rather than increasing from top to bottom. We also had a more observational study regarding the original "fake map" on a scatterplot in our final sketch. Although our users understood it, it required explanation and was harder for users to evaluate which choice was best based on distance. Due to those results, we decided to look into changing Vis 1 to something closer to an actual map, since we realized it would significantly impact user satisfaction and decrease decision time by precious minutes, just what we need to have an edge over other tools. We also decided to change the Vis 2 bar chart to vertical bars and overall keep that visualization simple, for much the same reasons.

6 Final Design

The final design deviated from our final sketch, but we decided to implement these changes for an improved user experience following the results from usability testing. Rather than translate a map onto an x, y coordinate plane with the user centered at (0, 0), we were able to instead simply use the Google Maps API to include a direct map widget. We instead mapped Yelp restaurants directly onto the map using their actual given latitude and longitude coordinates, as well as allowed the user to input their latitude and longitude to be plotted onto the map alongside the restaurants. The trade-off for using a real map meant that we decided to only include the restaurant name in the hover data for each restaurant pin. Since setting up the map was more resource intensive and was easier for the user to visualize the restaurants in space, we decided to only focus on presenting the review distribution in the linked bar chart and allow the user to make their own conclusions using the bar chart, as well as judge the distance for themselves on the Google Map visualization. The final bar chart was also rotated to be a vertical bar chart rather than the sketched horizontal bars since vertical bars were more intuitive for the user to interpret and what they were more likely to be used to.

To use HungryMap (Figure 6 - see following page), the user firsts types in their latitude and longitude into the designated input fields. After hitting submit, the user's location will be plotted onto the map showing all Santa Barbara restaurants plotted. The map will also center around the user. The user can then zoom in and out or drag the map around to browse various locations. The user can hover over each pin to see the name of each business. If the user sees one of interest, they can click on the pin, and a linked bar chart will appear on the right. This shows the distribution of Yelp star reviews. The user can also hover over each bar to reveal a corresponding tool tip. Each bar's tool tip will reveal the date of the most recent review with this rating and will decrease the opacity of the bar to show that it is selected. With these features in mind, users can choose restaurants within a certain distance of their current location and assess if the recent quality is good enough to warrant ordering food from there.

7 Discussion

The final visualization effectively accomplished our goal domain tasks and exceeded our initial design goals by improving upon our final sketch. However, we did want to note that it we were able to accomplish this by adapting to and working with the limitations of the Yelp API dataset. While this dataset contains many points with interesting information, it lacks several attributes that are more suitable to the average consumer, as well as notable datapoints of interest. We originally wanted to create a tool that serves the Northeastern community. However, we found that the dataset lacked datapoints from the entire Northeast United States, so we pivoted to Santa Barbara to work with a major city that had enough datapoints. A consumer would also like to know more about the average price of dishes at a restaurant and may be interested in seeing images of dishes. Neither of these attributes was available, but is something regular users of Yelp spend a significant portion of time clicking

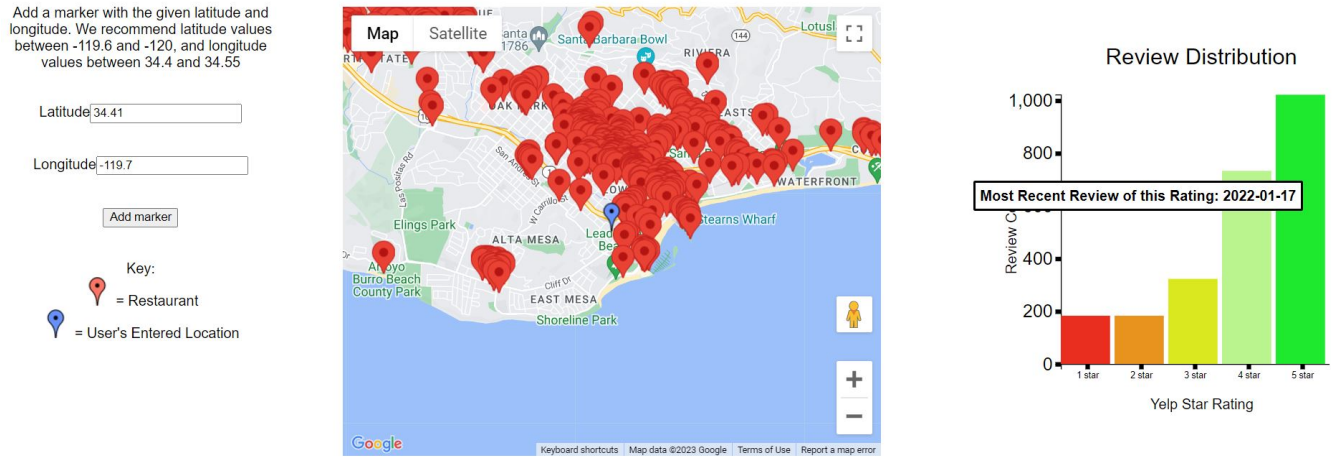

Figure 6: Final Visualization Screenshot

through the site to find. If we were to iterate through this process again to perfect our tool, we would look into a way to collect this information and incorporate it either in a tooltip display on the map, or perhaps in separate linked charts, similar to the review bar chart. In addition to this relative difficult improvement, some smaller scale improvements include aesthetic choices. Although our design went through usability testing and was well received, we could add different formatting and fun color choices on our site so that it was more appealing than a simple and utilitarian site. While this does not contribute to function, this would potentially drive more traffic to our tool as users tend to recommend sites that have outstanding UI design.

8 Conclusion

HungryMap is a tool developed to serve new transplants to the city, college students, or the generally indecisive eater. Where existing tools, like Google Maps and Yelp, overwhelm users with a wealth of hard to decipher information that can require many clicks to uncover, HungryMap presents key attributes involved in meal purchase decision-making in a more streamlined manner, allowing users to make split second decisions and reducing time standing around hungry and unsure. We hope that users will use this tool to evaluate new restaurants in their neighborhood by popularity and quality, prioritize personal preferences, and ultimately have something tasty to eat in their hands in less time than it takes to browse other tools.

9 Appendices

9.1 Data Abstraction

Business Dataset
address (string/Categorical) - The current address of the business
take_out (bool/Categorical) - True if the business has takeout, false otherwise
business_id (string/Categorical) - A unique identifier for the business
latitude (float/Diverging) - The latitude of the business
longitude (float/Diverging) - The longitude of the business
name (string/Categorical) - The name of the business
review_count (int/Sequential) - The number of reviews the business has received
stars (float/Sequential) - The average star rating of the business

Review Dataset
business_id (string/Categorical) - A unique identifier for the business
review_id (string/Categorical) - A unique identifier for the review
stars (int/Sequential) - The star rating given by the review

9.2 Task Abstraction

1. identify food options that are geographically accessible
   (a) Target: Distribution
   (b) High Level: Discover
   (c) Medium Level: Look up
   (d) Low Level: Identify
2. allow users to prioritize restaurant popularity or quality
   (a) Target: Features
   (b) High Level: Present
   (c) Medium Level: Browse
   (d) Low Level: Compare
3. quickly assess what recent customer sentiment for a restaurant has been
   (a) Target: Trends
   (b) High Level: Present
   (c) Medium Level: Locate
   (d) Low Level: Identify
4. make the most informed decision to purchase takeout in the most time efficient way
   (a) Target: Similarity
   (b) High Level: Present
   (c) Medium Level: Locate
   (d) Low Level: Compare

References

[1] B.-L. Chua, S. Karim, S. Lee, and H. Han. Customer restaurant choice: An empirical analysis of restaurant types and eating-out occasions. *International Journal of Environmental Research and Public Health*, 17(17):6276, Aug 2020. doi: 10.3390/ijerph17176276

[2] C. Liu, L. Li, C. Shan, X. Hu, Z. Diao, and M.-e. He. Exploring neighborhood service and development strategies by multi-dimensional sentiment analysis of online restaurant review. In *2021 3rd International Symposium on Smart and Healthy Cities (ISHC)*, pp. 120–125, 2021. doi: 10.1109/ISHC54333.2021.00031