Restaurant Recommendation

Group Members:
Anjani Sruti Doradla
Prajwal Khot
Priyanka Raghavendra
Samruddhi Kalsekar
Sharanya Manohar

Contents

Introduction

Our hypothesis for this project was to develop a paradigm that recommends restaurants to users based on their location and popularity of the restaurants in amalgamation with providing a solution to business aspirants who intend to set up a business in a particular area by using the stats of existing restaurants. We used the Yelp API to procure the data and performed analysis to obtain meaningful insights from it. Yelp thrives on business data and the various reviews given by users for many restaurants. It is significant to analyze this data and derive useful information to build various models.Lastly, we implemented models to aid the recommendation process for both common users and business aspirants and also performed sentiment analysis on the data to obtain the information of whether a review was positive or negative for a particular restaurant.

Data section

The dataset used for the analysis and development of our project is the yelp dataset procured via the yelp API. The files downloaded were in the JSON format.

Our dataset consists of 5 files.

- *Business Dataset* - Consists of details of the business, i.e, the restaurant related information.
- *Review Dataset* - Provides insight on the user reviews and ratings for particular restaurants.
- *Users Dataset* - Comprises user information associated with a particular user id.
- *Check-in Dataset* - Consists of user check-in date and time for particular businesses.
- *Tips Dataset* - Includes user opinions for a particular business and the dates these tips were written.

The dataset most widely used in our project is Business, Review and Users.

Methods section

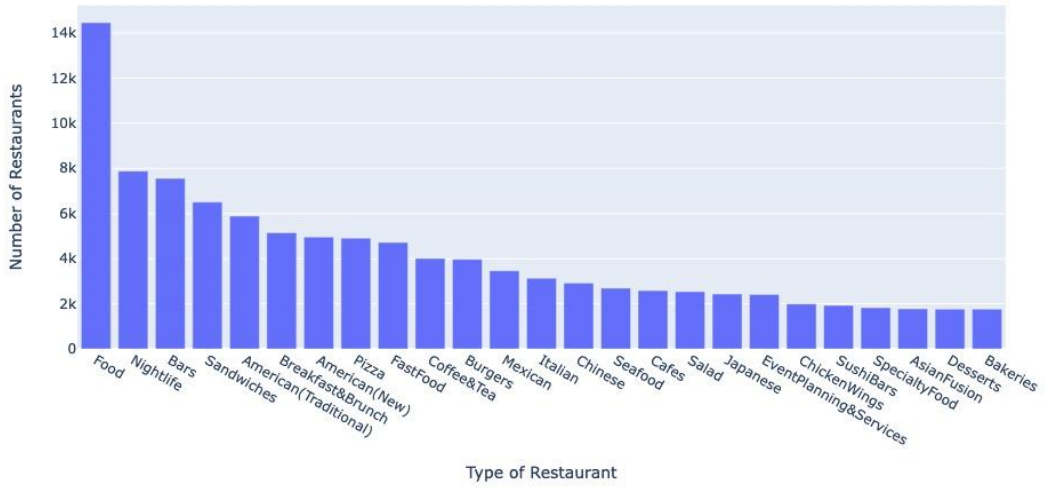
Data Collection :
Yelp provides us with an API to obtain data. Once a profile is created, a private key is generated. We then used this key to fetch the data using Postman.

Data Cleaning :

- Removed the duplicate, irrelevant data and filtered the outliers which was not necessary for analysis.
- Handling missing values and inconsistent data by replacing it with null values.
- Columns which were insignificant individually, were merged with other attributes to form a more efficient dataset for analysis.

Data Analysis :

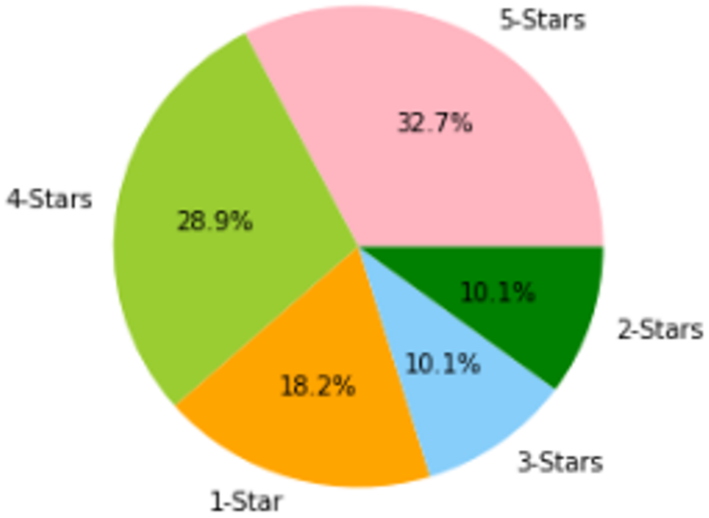
1) <u>Distribution of restaurant type against restaurant count</u>
   The Yelp business data consists of the location (state, city, neighborhood, longitude, latitude, zip code, etc.), number of reviews, average star rating,working time and category of cuisine offered for various restaurants.We analyzed this data to obtain and understand the types of restaurants and the count of them. This gave us an edge in understanding that the highest count of restaurants was for the type “Food” and the least count was of “Bakeries”.



2) Data visualisation for star rating distribution

The Review data consists of the star rating of the restaurant that ranges from 1 to 5, review texts provided by the user for every restaurant, time at which the review was written, and whether or not the review was noted as cool, funny or useful by other Yelp users.

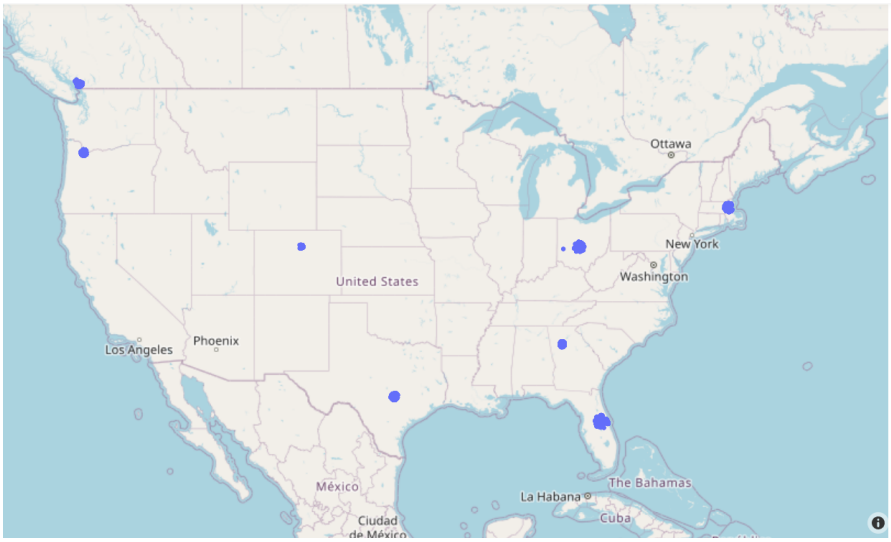
Out of all the reviews provided, we analyzed the composition of the star ratings and the stats obtained were such that the 32.7% were given a rating of 5 stars, 28.9% were given a rating of 4 stars, 10.1% were given a rating of 3 stars, 10.1% were given a rating of 2 stars and 18.2% were given a rating of 1 star.



3) Top 10 states with highest number of restaurants

The clusters on the map represent restaurants in that particular state. Hovering over one of the points on the clusters will result in displaying the name of the restaurant present at the point.

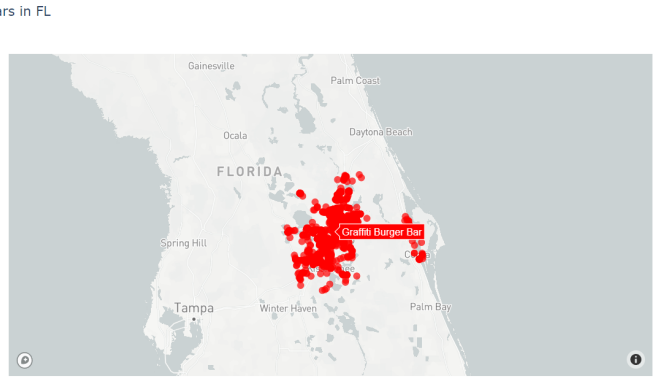
```
{'MA': 8676,
 'FL': 6601,
 'OR': 6319,
 'BC': 6092,
 'GA': 5161,
 'TX': 4629,
 'OH': 3774,
 'CO': 730,
 'WA': 657}
```



4) <u>Preferred cuisine in preferred state</u>

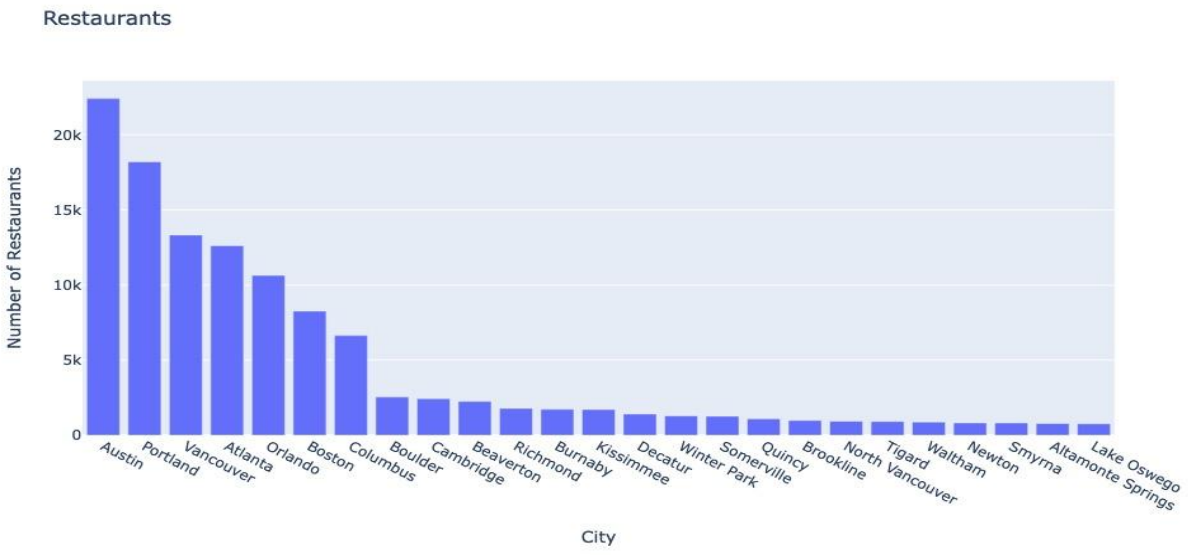
The user can enter the state and the type of cuisine they are interested in. The Map representation shows the clusters of the desired cuisine for the state. One such example of bars in the state of Florida is shown below.

```
Pick a state: FL
Pick Restaurant type: Bars
```



5) <u>Top 25 cities with number of restaurants</u>

It is seen that Austin is the city with the most number of restaurants around 23K in the business category.



Analysis section

The analysis is based on providing two kinds of products to the user:

1) Nearest restaurant recommendation based on location entered by the user
2) Efficient solution for business aspirants with respect to the best sub category of restaurants based on state entered.
   Additionally, we have implemented analysis on the review dataset.
3) Sentiment Analysis to obtain the expression of the reviews.

1.Restaurant recommendation

In this section ,we obtain the results of the recommendation by considering the location of the user. The user is asked to input their current address and their preferred restaurant type.Then, the latitude and longitude of that location is obtained by making use of Nominatim class from geopy and the restaurants that are closer to the user and the ones that match their preference are suggested. This is then displayed on a map implemented using graph_objects provided by plotly.

Step 1: User is asked to input their name, location and choice of cuisine

```
Enter your name:Prajwal
Enter your address:Boston Uni
Enter cuisine:Bar
Boston University, Manchester Road, Coolidge Corner, Boston, Norfolk County, Massachusetts, 0
(42.35031725, -71.10122725124097)
```

Step 2: The resulting latitude and longitude of the user's location is generated and stored in a dataframe format.

	Username	lat	lon
0	Sharanya	43.534248	-70.990948
1	Anjani	48.660160	-122.586133
2	Priyanka	46.622195	-122.578420
3	Samruddhi	42.827666	-73.924628

2. Business perspective

Here, we are implementing a model that provides insight about the success of a business based on the popularity of a particular type of restaurant in that area. We ask the user to input the type of restaurant that they would like to invest in and the location. Based on the user's input, one of the following 3 results are obtained.

i) A business has fairly good potential if it is amongst the top 10 popular types in a given area.

ii) A business is highly successful if there is potential for that particular type of restaurant to be put up in that location and if it is one of the top 5 popular restaurants.

iii) A business fails if that particular type of restaurant is not popular in the given area.

In the third case, we also suggest those types of restaurants that might be successful in that particular area.

Additionally, we have also analyzed the data to check for top 10 types of restaurants when the location is provided.

3. Sentiment Analysis

In addition to the above implementations, we also performed sentiment analysis on the review dataset. We have implemented the same by omitting the stopwords that have no base meaning to them. The remaining set of words were then processed. The NLTK library provides us with the PorterStemmer() method which aids in the stemming process to obtain the base form of a word by removing affixes from them. We used the n-grams approach instead of the bag-of-words approach to count the frequency of pairs/triplets of words that occur simultaneously. For example, French fries. Additionally, we implemented the Word2Vec model to do the same. Continuing , we used VADER analyzer from the NLTK package which provides us with a predefined list of words with respective sentiment scores. We are then provided with 4 scores. I.e, Negative, Neutral, Positive and Compound. In our model, we have used the compound type of score. If the value was greater than 0, we classified the review as Positive(1) and if the value was lesser than 0 the review was classified as Negative(0).

[Representations of all analysis can be found in the 'Results' section.]

Results

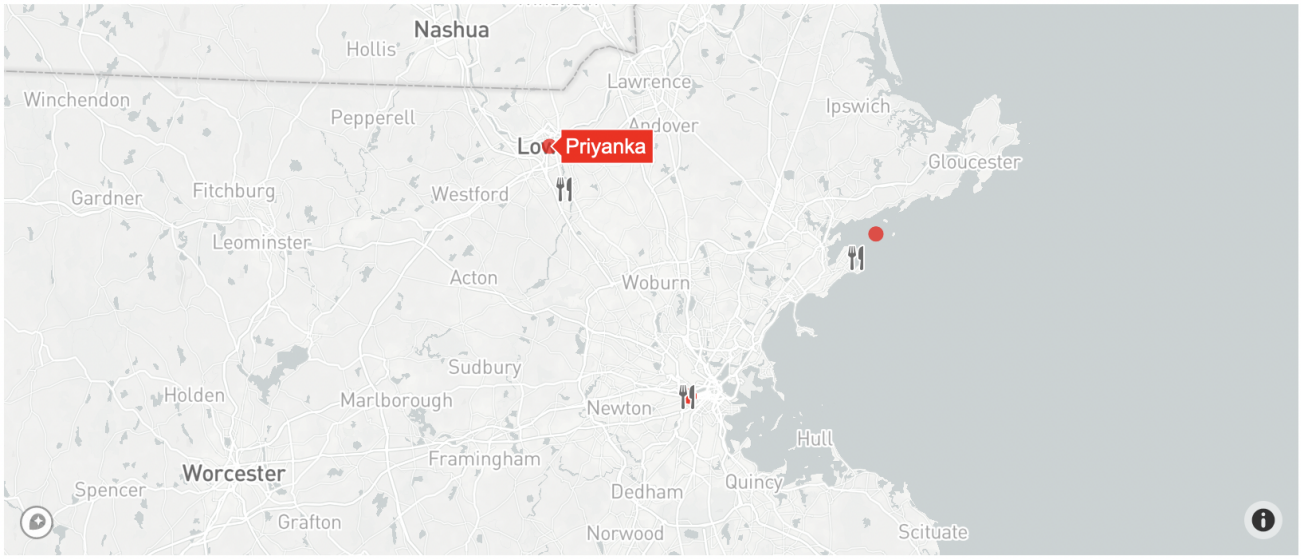
Restaurant Recommendation:

The following data frame representation shows the name and coordinate details of the restaurant closest to the user:

	Username	m_lat	m_lon	name	h_lat	h_lon	stars
0	Sharanya	43.534248	-70.990948	Iron Chef	42.599823	-70.959027	3.5
1	Sharanya	43.534248	-70.990948	Iron Chef	39.923154	-82.783890	3.5
2	Anjani	48.660160	-122.586133	Sushi Topia	49.118241	-122.888557	3.5
3	Priyanka	46.622195	-122.578420	Mahoney's Public House	45.698985	-122.703598	4.5
4	Samruddhi	42.827666	-73.924628	Saltbox Kitchen	42.457227	-71.395106	3.5
5	Prajwal	42.350317	-71.101227	Basho Express	42.349341	-71.103982	3.0

Each red point on the map represents the location of each user:

Hovering over each point results in the displaying the name of the user and the restaurant closest to them. An example of the same is shown below.



Business Solution:

The representations for the various cases discussed under the analysis of the business perspective are shown below.

```
Enter the type Restaurant: Mexican
State: FL
User would like to open Mexican in FL
There is a good business potential for Mexican.
```

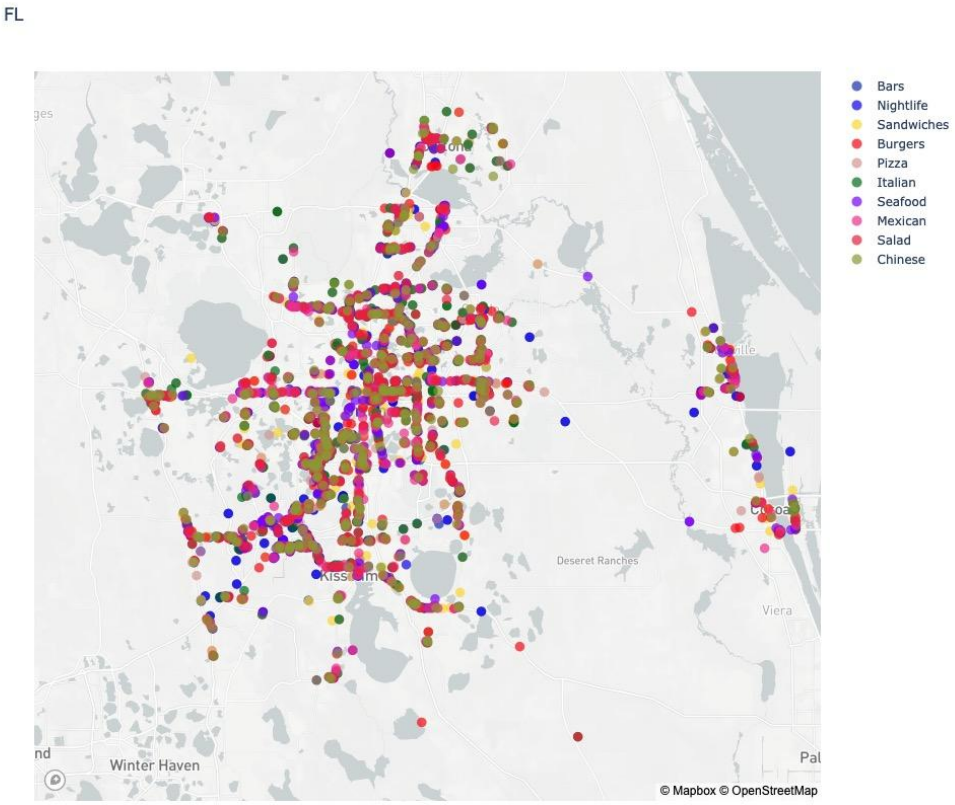
Case (i)

```
Enter the type Restaurant: Bars
State: FL
User would like to open Bars in FL
This location is a great choice for Bars, but you will have quite a lot of competition.
```

Case (ii)

```
Enter the type Restaurant: Indian
State: FL
User would like to open Indian in FL
We would suggest running some other type of restaurant.
Maybe you can consider the following:
        * Bars
        * Nightlife
        * Sandwiches
        * Burgers
        * Pizza
        * Italian
        * Seafood
        * Mexican
        * Salad
        * Chinese
```

Case (iii)



The above representation shows the top 10 cuisines in the state of Florida:

Sentiment Analysis:

The result for the positive and negative analysis is as shown below. Here, label ‘0’ represents negative review and label ‘1’ represents positive review.

	review	label
0	Thank goodness for their lettuce-wrapped burge...	1
1	This place is wonderful!! Also they have restr...	1
2	Great, tasty burgers that look and taste fresh...	1
3	We were traveling by this Novato restaurant an...	1
4	I've been here 3 times. The first, it was per...	0
...	...	...
10656	Yes, someone finally called because they fucke...	0
10657	Super Duper makes it's way to Oracle Park, and...	1
10658	So I heard that they opened a Super Duper at t...	0
10659	Love this place! Highly recommend the vegetari...	1
10660	Wanted to try a Super Duper Burger before the ...	0

Conclusion

1. To summarize, the model helps analyze various factors that influence the current cuisine trends taking the location into consideration.
2. The project is beneficial to users for achieving customized recommendations based on the required cuisine/food.
3. In order to provide better performance, we performed collaborative filtering methods for recommending a restaurant, based on ratings and reviews given by other users.
4. The proposed system also helps businesses to target profitable locations for setting up restaurants as well.