Find a Foodie Friend:

Team members:

Ravneet Singh Sidhu

Harsha Vardhan Reddy Gorla

Data processing:

Filtered the users, business and review data sets using the APACHE HUE. We have installed the Cloudera virtual box which has HUE and Hadoop pre installed in it. We imported the JSON data files provided by the yelp into virtual box ran queries to get the required data. (Queries are provided in the Huescripts file)

Read the users data JSON file from the yelp data sets and uploaded the userid, used yelping since as the user password, friends of the user.

Query\_result.txt, query\_result3.txt contains the processed HUE data which are categories of the users who have rated the reviews as high(>3 star rating) and low (<3 star rating). Next step is to take the userid, category and find the number of food categories (e.g. Indian, Chinese, Italian etc) per user. This will help us estimate which type of food user likes (>=3 star rating) and which type of food user dislikes (<3 star rating). The resulting files after processing in Hadoop are stored in resulthigh.json and resultlow.json. usercategory\_mapper.py and usercategory\_reducer.py are used to achieve this task.

Now the mongodb database is updated with the user likeness and dislikeness information.

The sample user record looks like this

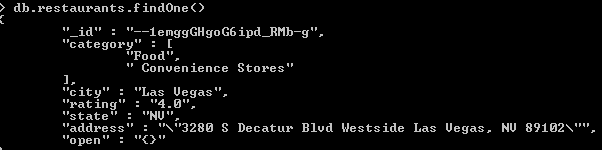


We have given the default passwords to all the users which is “yelpdata”

When user signs in his location will be updated from where is logging in based on his public IP address. Program updates his city, state, latitude and longitude location.

This will help to fetch all the users who are in the same city as he is in to get some suggestions from the fellow yelpers which restaurant or which category of food to choose, especially for the users who are new to the location.

The business data in the MongoDB looks like follows.



Open in the above record means open hours. For few records data for open hours is not present as shown in the above screen shot. So web site displays Open hours: info not available.

AWS EC2 server:

Started a new AWS EC2 instance to host the MongoDB data and also the website.

Website can be seen by using the following address:

<http://ec2-52-10-114-225.us-west-2.compute.amazonaws.com:5000/login>

Python Flask Frame work:

We have used the python flask and jinja2 templates for building our website.

It has static (that contains javascript and css files), templates (that contain html webpages) and index.py file that contains the coding.

Once user logs into the website using his user Id and the default password – yelpdata,

His location will be updated into the users collections. Then we find all the users who are in the same city as he is in.

We loop through all the resulting users records and get the high rating categories for all the users and find the Manhattan distance for the high rating categories. This helps in finding users who has the same taste as the logged in user. Website will display only the top matched users.

Next task is to find the restaurants which has the category (type of food) that the logged in user and the matched user has in common.

For example logged in user has Steakhouses and one of the matched user has the Steakhouses in their likeness. Website will fetch all the restaurants in that city which are Steakhouses and display their address and the open hours.

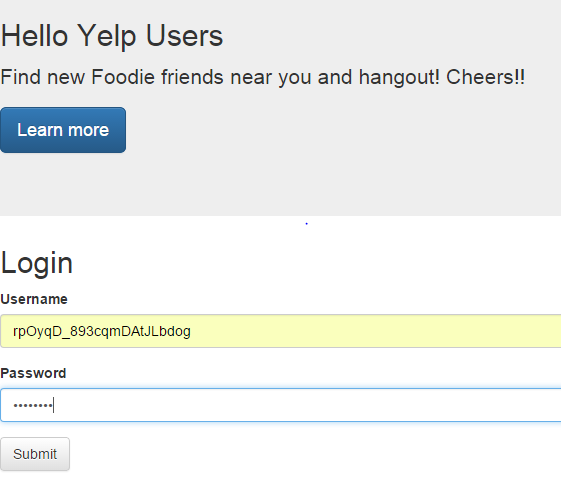
User may want to know which category is famous in the city. For example in a particular city or region, Chinese food is more famous but user might not this. So we fetch all the categories in the city and the count of number of categories and display the data in the pie chat.

For example, a city might have Italian: 7(this number is from the high rating categories of the user), Indian:4, Brazilian:1 then displaying this info on pie chart. This will help user to understand that the Italian category food is rating high in this city so he may want to try it.

Second pie chart displays number restaurants per category in the city. For example how many Italian restaurants are there in the city, how many Indian or Chinese restaurants are present in the city.

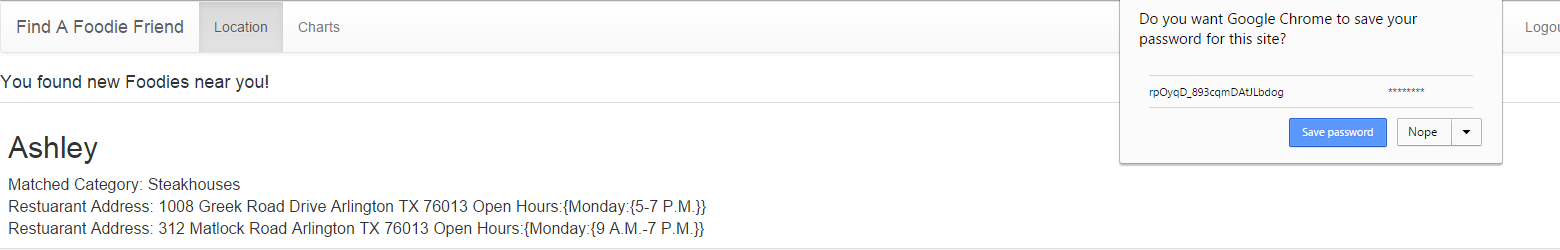
Test:

Give user id and password –yelpdata (default for all users)



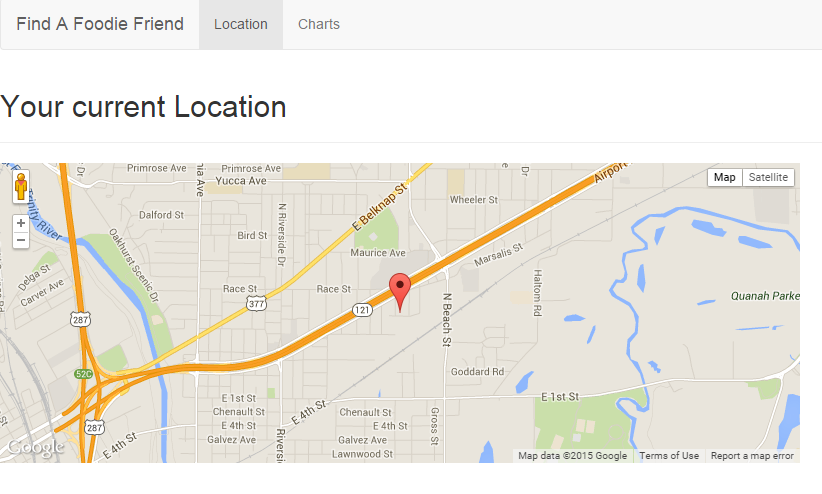
Learn more tab above will take you to GitHub account to download the files

Click on submit will take you to below screen

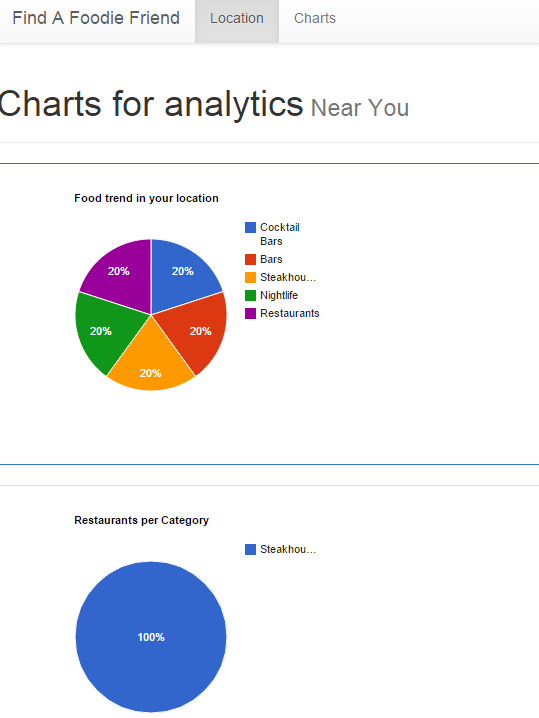


Next page will display the matched users along with the categories and the open hours of the restaurant.

Location tab will display your location on the map as shown below.



Chats will display information on the categories and the restaurants.



We currently inserted only one restaurant record for testing while preparing this document so it shows only Steakhouse in user location.

Future work:

We can give a facility for the user to enter a city name and then find the famous restaurants and the food types people prefer in that location.