Roadmap of files on Github RePo for Foodies – Project 1

Group Members: Carly Russell, Loba Quasem, Phillip Choi, Natalie Stanislov

Please note: all files are on the master, unless specified that they are on another branch

1. **Project Planning**

Cities.csv in the Resources folder, was finding the cities in the bay area, and using those to generate YELP API queries

Cities\_by\_Zip, file of the bay area zips, cities and counties

Counties\_by\_zip was a different version of the file, only has the zips and the counties

1. **Yelp API Calls**

calls\_by\_city.ipynb – first Yelp API calls were performed by city

restaurant\_data.csv in the master resources file, is an output file of the calls\_by\_city.ipynb Yelp queries

calls\_by\_city\_ratings.ipynb – Yelp API calls performed by city, sorted by ratings since each query call could only give 1000 results back, thought we had data we were going to use, yet we had the cities we called for not localized to the bay area

restaurant\_data.csv in the master resources file, is an output file of the calls\_by\_city\_ratings.ipynb Yelp queries

calls\_by\_zip.ipynb – Yelp search by zip code

coordinates\_grid.ipynb is a jupyter notebook creating lat & lng coordinates through out the bay area. The first jupyter notebook had the cooridates spaces every 5 miles, the second every 3 miles.

grid\_coordinates.csv located in the project-1 master resources folder is an output file showing the bay area lat and lng coordinates spaced 5 miles apart

calls\_by\_coord\_ratings\_ipynb – calls by coordinates 5 miles apart with 5 mile search radius, query return is a sorted results from YELP

**restaurant\_data\_coords.csv, located in github master resources folder is an output file of lt&lng spaced 5 miles apart**

Hypotenuse theory calls\_by\_coordinate\_radius\_4023\_testing.ipynb – A test file with Yelp API to see if reduced graphical coordinates and shorter search radius would be better. If on thinks about the geographical coordinates spaced out every 5 miles, then the search radius selected was 2.5 miles and we are getting 78% of bay area coverage when search for the restaurants.

Hypotenuse theory calls\_by\_coordinate\_radius\_5690\_testing.ipynb – A test file with Yelp API to see if reduced graphical coordinates and shorter search radius would be better. If on thinks about the geographical coordinates spaced out every 3 miles, then the search radius

grid\_coordinates2.csv located in the project-1 master resources folder is an output file showing the bay area lat and lng coordinates spaced 3 miles apart

calls\_by\_coord.ipynb - calls by coordinates 3 miles apart with a smaller search radius

calls\_for\_2\_coords\_rating.ipynb – running code to get back restaurants sorted by ratings to 2 coordinates that were over 1K

calls\_by\_coord\_ver2.ipynb – calls by coordinates 3 miles apart with a reduced radius with updated code for category parameters

calls for\_2\_cords\_rating\_ver2.ipynb – running code to get back restaurants sorted by ratings to 2 coordinates that were over 1K

1. **Data Cleaning**

Merged\_zip.csv file, was the output from the first merge in coordinates 2 data set

Cleaning\_Yelp\_Data\_On\_Coodinates\_2.ipynb First Pass at Cleaning with Coordinates 2 Data set, which was supposed to be prep for cleaning Coordinate 3 Data Set, the final Yelp API.

Zip\_of\_9\_counties.csv – starting file of bay area counties, and zip codes to use in the Cleaning\_Yelp\_Data\_on\_Coordinates3\_Calls.ipynb in the resources folder in github master

Cleaning\_Yelp\_Data\_On\_Coordinates3\_Calls.ipynb cleaning notebook

Cleaned\_zip.csv – Final Clean Bay Area Zip Codes to merge with the data set to get the bay area counties on the data table by matching with their zip codes, an output from Cleaning\_Yelp\_Data\_On\_Coordinates3\_Calls.ipyn

Cleaning\_Yelp\_Data\_On\_Coordinates3\_Calls.ipynb. Final version of cleaning for Coordinates 3 Data Set (final data set)

Cleaning\_2.csv, located in the github project-1 master resources file was an output file used in Cleaning\_Yelp\_Data\_On\_Coordinates3\_Calls.ipynb to merge with counties\_by\_zip.csv file

Restaurants\_cleaning\_2.csv – file to find categories of cuisines that our group approved of

Original Categories.list.xlsx – first pass of categorizing cuisine types

Category Lumped.csv -file helped categorize cuisine types

Category Lumped2.xlsx – a pass at categorizing cuisine types

Categories.ipynb – file of initial/draft category cleaning, in the Resources folder on master github

Preliminary\_Category\_Sorting\_Lumping\_by\_Counties.ipynb – pass at cleaning the categories that are unwanted and then creating a list of categories by county by restaurant ratings on final clean data set.

Final\_Clean\_Restaurants.csv – in the Resources folder, is the first final cleaned data set

1. **Data Analysis & Graphing**

Plot\_coord.ipynb Creates scatter plot using lat/lng and total restaurants at each coordinate pair

Pricing\_&\_Rating.ipynb file for analysis on the Final Clean Data Set

Restaurantsbyprice.png analysis graph over view of the bay area on the Final Clean Data Set

**Bayarea.mostpop10.png – analysis was performed on Final Clean Restaurants.csv in the Resources git hub file**

**Bayarea\_piechart.mostpop10.png – Analysis performed on Final Clean Restaurants.csv**

Percentages of the 3.5 to 5 Star Ratings by county, on Final Clean Data.xlsx – An analysis to look at the distribution of the ratings by county.

1. **Project Presentation & Conclusions**
   1. All\_food\_in\_the\_hood.pptx – final presentation