

# Restaurant Data Map Project Report

---

Chang Liu(10373152), Zeyu Wang(10376853), Chengzhao Li(10374848)

Instructor:Professor Mukund Iyengar

May 4, 2014

## ABSTRACT

This project collected different restaurant informations around New York City,including restaurant name, rating value, address, then demonstrated them in forms of different data maps. Through the visualized data, a more intuitive and completed result can be generated from one single search compared to result from ordinary website. Provided a macroeconomic view of information such as the locations, rating values and distribution density of restaurant in New York area. In addition, a graphic user interface was implemented for better interactive experience.

## 1 INTRODUCTION

We are living in an interconnected world. Large amount of information could be easily access from the Internet. For example, you can get as many as 47947 results when browsing Yelp.com for restaurants in New York City area. Traditional applications bring people lots of convenience when finding the restaurant they want. But when a user wants to make a choice with comprehensive information, traditional applications' functionality became limited. For instance, when a user wants to find an area that has three specify kinds of restaurant, he has to search three times and compare the data manually. Or, if a traveler wants to find a place with the highest density of one cuisine, he has to zoom in and zoom out several times and do the comparison by himself. In our research, we develop software that plotting different kinds of restaurants from one search result in different colors on a geography map. Users could distinguish restaurant type easily from the color and find the location that has all restaurants

he need without any other operation. Also, users could know the density of a certain kind of restaurant from our application.

Our system is separated into three modules: data collecting, data visualizing and graphic user control panel. Data collecting module is response for grabbing restaurant information from the Internet and save the information into several files. Data visualizing module utilizes the data collected from the Internet, process it and create the data map. Graphic user control panel creates a graphic panel for users to control what kinds of data will be plotted on the map.

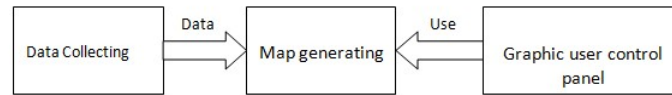


Fig 1.1 System Architecture

## 2 DATA COLLECTION

Restaurant information are needed from Yelp, including name, ratings, number of reviews, price range and addresses.

### 2.1 WEB SCRAPING

- URL Analysis
  - A Uniform Resource Locator(URL), as a specific character string, constitutes a reference to a resource.
  - The source from Yelp has a common format. According to its URL, restaurant information are listed as a regular pattern.
  - Here is an example of Indian food in Jersey City that we searched from Yelp, and some of the data from its URL are listed below.

```

<div class="search-result natural-search-result biz-listing-large" data-key="1">



<span class="indexed-biz-name">1. <a class="biz-name">
href="/biz/sapthagiri-taste-of-india-jersey-city#query:indian%20restaurant"
data-hovercard-id="gQw3yAXJMI3ApbtcoeiHg">Sapthagiri Taste Of India</a>
</span>

<i class="star-img stars_4" title="4.0 star rating">


<span class="review-count rating-qualifier">
148 reviews
</span>

<span class="business-attribute price-range">$</span>

<div class="secondary-attributes">
<address>
804 Newark Ave<br>Jersey City, NJ 07306
</address>
key = str(MediaBox.get('data-key'))

```

Fig 2.1 URL of Indian food in JC from Yelp

- From the data we receive through the URL from Yelp, the information we need are high lighted above.
- Browser Simulation
  - In this paper, we use Python urllib2 module to open the URL of Yelp.
- Data extraction
  - A Python library called BeautifulSoup are adopted for scraping. Some of the code are listed below, so that we can extract the information that needed.

```

Name = str(MediaBox.find(attrs = {'class':"photo-box-img"}).get('alt').encode('ascii',
'ignore'))
AddressList = str(MediaBox.find(attrs =
{'class':"secondary-attributes"}).address.replace('<br/>','\n').split('\n')
Rating = str(MediaBox.find('i').get('title'))

ReviewCount = str(MediaBox.find(attrs={"class":"review-count
rating-qualifier"}).get_text().strip())
try:
    PriceRange = str(MediaBox.find(attrs={'class':"business-attribute
price-range"}).get_text().strip())

```

Fig 2.2 Beautiful Soup codes for collecting data

## 2.2 ADDRESS RECOGNITION

- As we have collected all the restaurant addresses, it is necessary to locate all the addresses in the map.
- We convert all the addresses into geographic coordinates (latitude and longitude) using Google Geocoding API.

```

lcz@lcz-VirtualBox: ~/Downloads/proj
501 Shiro of Japan 3.5 star rating 5 reviews None 89 E
42nd St, New York, NY 10017 40.752471, -73.977295
502 Okami Fusion Sushi 3.0 star rating 20 reviews $$
63 Reade St, New York, NY 10007 40.714781, -74.006661
503 A Canaan 2.5 star rating 81 reviews $$ 154
W 29th St, New York, NY 10001 40.747348, -73.992311
504 Kumo 3.0 star rating 32 reviews $$ 37-18 Ditmar
s Blvd, Astoria, NY 11105 40.773526, -73.907190
505 Sushi Express 3.5 star rating 9 reviews $$ 230
Park Ave, New York, NY 10169 40.754033, -73.975785
506 Jc Hibachi 3.0 star rating 5 reviews $ 198-
A Orchard St, Manhattan, NY 10002 40.722326, -73.987776
507 Wei West 3.0 star rating 64 reviews $$ 235
Murray St, New York, NY 10282 40.715394, -74.014755
508 Zen 6 4.0 star rating 147 reviews $$ 328 E 6th St
, New York, NY 10003 40.726682, -73.987265
271 Smith St, Brooklyn, NY 11231 <No Results>
Traceback (most recent call last):
  File "Multigenre.py", line 99, in <module>
    saveGenre(genre,OutFolder)
  File "Multigenre.py", line 69, in saveGenre
    Coordinates = "%.6f, %.6f" %(CoorList[0],CoorList[1])
TypeError: float argument required, not NoneType
lcz@lcz-VirtualBox:~/Downloads/proj$

```

Fig 2.3 Screen shot when running codes

- However, in order to prevent abuse of the service, a user can submit only 2500 requests per 24 hour period from Google Geocoding API. As the screen shot above showed, no results is returned for the last restaurant address, which means requests have reached their limit.
- Finally, we save different cuisine of restaurant information extracted in different textfile documents.

3	maai marke indian cuisine	4.0 star rating	230 reviews	\$5	31W E 6th St, New York, NY 10003	40.720834, -73.987670
4	The Masalakala	4.5 star rating	279 reviews	\$5	179 Essex St, New York, NY 10002	40.721923, -73.986668
5	Dhaba Indian Cuisine	4.0 star rating	783 reviews	\$5	108 Lexington Ave, New York, NY 10016	40.742408, -73.982928
6	Bombay Duck Co.	4.0 star rating	109 reviews	\$5	190 Bleecker St, New York, NY 10012	40.729876, -74.001530
7	Seva Indian Cuisine	4.0 star rating	546 reviews	\$5	30-07 34th St, Astoria, NY 11103	40.765432, -73.918979
8	Chote Nawab	4.0 star rating	186 reviews	\$5	115 Lexington Ave, New York, NY 10016	40.742448, -73.982304
9	Tiffin Wallah	4.0 star rating	462 reviews	\$5	127 E 28th St, New York, NY 10016	40.742976, -73.982740
10	Junoon	4.0 star rating	514 reviews	\$5	27 W 24th St, New York, NY 10010	40.743193, -73.998838
11	Masala Times	4.0 star rating	239 reviews	\$	194 Bleecker St, New York, NY 10012	40.729153, -74.001681
12	Bombay Masala	4.0 star rating	274 reviews	\$5	678 Franklin Ave, Brooklyn, NY 11238	40.675341, -73.956518
13	Chola Eclectic Indian Cuisine	4.0 star rating	340 reviews	\$5	232 E 58th St, New York, NY 10022	40.760159, -73.965823
14	Heart of India	4.0 star rating	148 reviews	\$5	79 2nd Ave, New York, NY 10003	40.726657, -73.989439
15	Benares	4.0 star rating	85 reviews	\$5	45 Murray St, New York, NY 10007	40.714274, -74.009103
16	Thelewala	4.0 star rating	189 reviews	\$	112 Macdougal St, New York, NY 10012	40.729559, -74.000543
17	Brick Lane Curry House	3.5 star rating	595 reviews	\$5	306-308 E 6th St, New York, NY 10003	40.727083, -73.988024
18	Nirvana	4.0 star rating	212 reviews	\$5	346 Lexington Ave, New York, NY 10016	40.750019, -73.977412
19	Spice Symphony	4.0 star rating	155 reviews	\$5	182 Lexington Ave, New York, NY 10016	40.744722, -73.981202
20	Moti Mahal Deluxe	4.0 star rating	129 reviews	\$5	1149 1st Ave, New York, NY 10005	40.765629, -73.957915
21	Darbar Grill	4.0 star rating	219 reviews	\$5	157 E 55th St, New York, NY 10022	40.759341, -73.968839
22	Om Restaurant	4.0 star rating	217 reviews	\$5	1593 2nd Ave, New York, NY 10028	40.775834, -73.953629
23	Bukhara Grill	4.0 star rating	271 reviews	\$5	217 E 49th St, New York, NY 10017	40.755819, -73.976270
24	India Place	4.0 star rating	137 reviews	\$5	655 Vanderbilt Ave, Brooklyn, NY 11238	40.677153, -73.968646
25	Taj Mahal	4.0 star rating	139 reviews	\$5	7315 3rd Ave, Brooklyn, NY 11209	40.632329, -74.027081
26	Kati Roll Company	4.0 star rating	490 reviews	\$	49 W 39th St, New York, NY 10018	40.752803, -73.984577
27	Joy Indian Restaurant	4.0 star rating	126 reviews	\$5	301 Flatbush Ave, Brooklyn, NY 11217	40.678852, -73.973266
28	Indikitch	4.0 star rating	73 reviews	\$5	25 W 23rd St, New York, NY 10010	40.742221, -73.990445
29	Masala Indian Cuisine	4.0 star rating	163 reviews	\$5	39-26 61st St, Woodside, NY 11377	40.746544, -73.903200
30	Indus Valley	4.0 star rating	193 reviews	\$5	2636 Broadway, New York, NY 10025	40.796989, -73.969707
31	Milon	3.5 star rating	266 reviews	\$	93 1st Ave, New York, NY 10003	40.726316, -73.986644
32	Ashoka Indian Grill	4.5 star rating	44 reviews	\$5	1436 Flatbush Ave, Brooklyn, NY 11210	40.635562, -73.951028
33	Swagat	4.0 star rating	188 reviews	\$5	411 Amsterdam Ave, New York, NY 10024	40.783486, -73.977780
34	Bay Leaf	4.0 star rating	76 reviews	\$5	1358 N 5th St, Brooklyn, NY 11211	40.717826, -73.959179
35	Tamarind	4.0 star rating	152 reviews	\$5	99 Hudson St, New York, NY 10013	40.719017, -74.006975
36	Desi Grill	4.0 star rating	58 reviews	\$5	46 Wyckoff Ave, Brooklyn, NY 11237	40.705651, -73.921829
37	Ayurveda Cafe	4.0 star rating	144 reviews	\$5	786 Amsterdam Ave, New York, NY 10025	40.793109, -73.971486
38	Nanaste	4.0 star rating	280 reviews	\$5	31-15 38th Ave, Astoria, NY 11102	40.766755, -73.920704
39	Delhi Heights	3.5 star rating	179 reviews	\$5	37-66 74th St, Jackson Heights, NY 11372	40.747098, -73.891763
40	Doaba Deli	4.5 star rating	84 reviews	\$	945 Columbus Ave, New York, NY 10025	40.799405, -73.962364

Fig 2.4 saved data in textfile

### 3 DATA VISUALIZATION

Four kinds of maps were implemented to visualize all the gathered data: The first one shows the location of restaurants drawing on the basic map(Location Map); The second one shows the density distribution of restaurants drawing on the basic map(Density Map) ; The third one shows the rating of restaurants on their locations drawing on the basic map(Rating Map A) and the last one shows the rating of restaurants on their locations drawing on the google map background(Rating Map B).

The examples of each kind of maps are shown in fig.1 to fig.4.

Chinese Restaurant Locations, New York



Fig.1 location map

Chinese Restaurant Density, New York

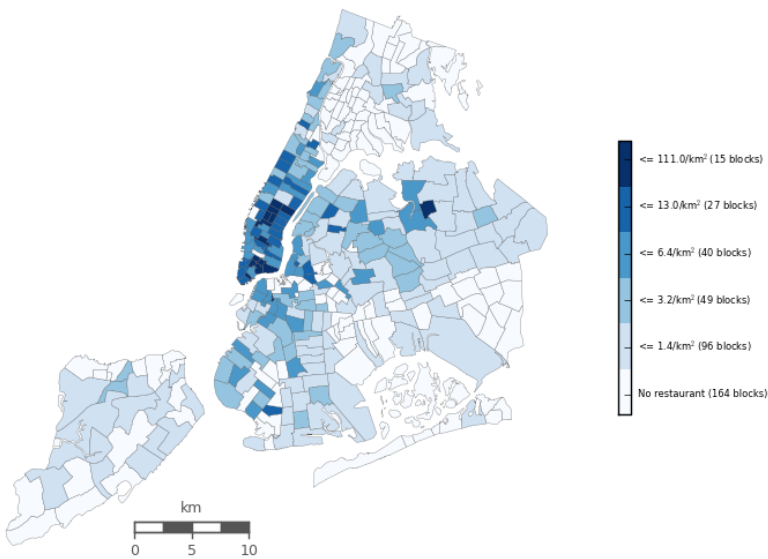


Fig.2 density map

Korea Restaurant Locations, New York

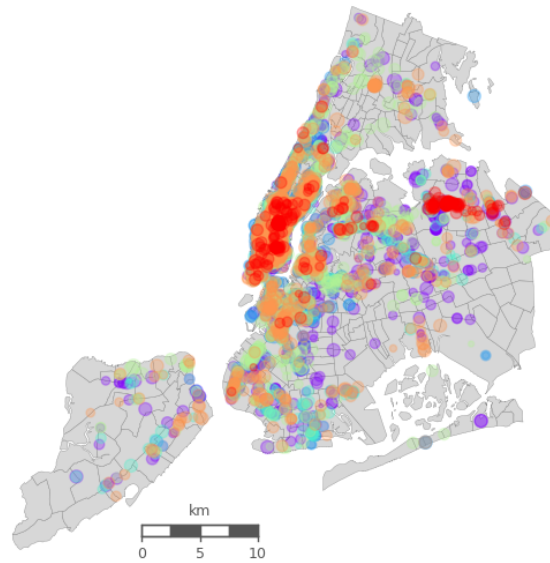


Fig.3 rating map A

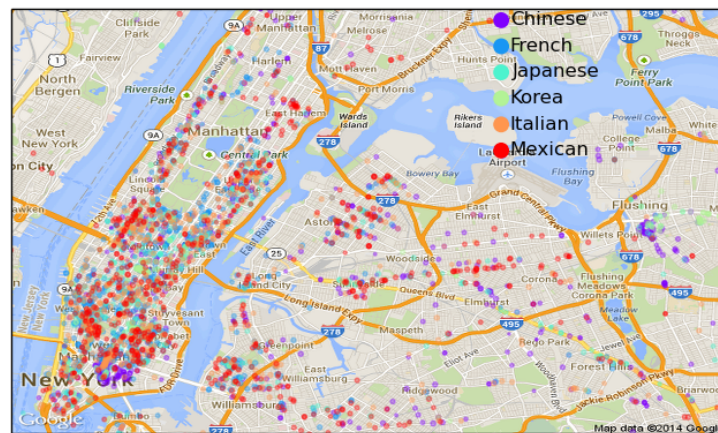


Fig.4 rating map B

Note that, in fig.1, one point represents one restaurant; In fig.2 and fig.3, one point represent one restaurant, the area of one point represents the rating value of the correlated restaurant,



different colors represents different kinds of restaurants; In fig.4, basic map are divided into many blocks, the color of one block represents the density of restaurants of the block, to be specific, blocks of the darkest color has the highest density of restaurants, blocks of the lightest color represents the lowest density of restaurants.

### 3.1 ENVIRONMENT REQUIREMENTS

- **Package**

Four independent modules were coded for plotting four kinds of map, respectively. For running each of these modules, the following non-stdlib are required :

- **Pandas**: Provides high-performance, easy-to-use data structures and data analysis tools for python
- **Numpy**: The fundamental package for scientific computing with python
- **Matplotlib**: A python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.
- **Basemap**: A Matplotlib toolkit, library for plotting 2D data on maps in Python
- **Shapely**: A package for manipulation and analysis of planar geometric objects.
- **Fiona**: Provides uncomplicated Python interfaces to functions in OGR
- **Descartes**: Use geometric objects as matplotlib paths and patches
- **PySAL**: a spatial analysis library

- **Shapefiles**

A New York City shapefile is also required for plotting the graph. The shapefile contains data such as map boundaries and polygon datas that divide the map into different blocks.

- **Restaurants' rating, longitude and latitude**

Rating Values and Coordinates of restaurants are required for plotting points.

### 3.2 VISUALIZATION PROCESS OUTLINE

The Ouelines for plotting each map.

#### 3.2.1 PLOTTING A LOCATION MAP

1. Extract boundary info using fiona from New York City shapefile:
2. Create a Basemap instance which works like a canvas that we can draw the map on,
3. Extract Polygon objects from the shapefile: each Polygon objects describes a block on the map.
4. Extract coordinates of each restaurant, Transform them into Point objects.



5. Filter each Point object out if it doesn't fall into the New York City map.
6. Transform each Polygon objects to Patch object.
7. Draw Patch objects on the Basemap instance.
8. Draw Points objects on the Basemap instance

### 3.2.2 PLOTTING A DENSITY MAP

1. Extract boundary info using fiona from New York City shapefile:
2. Create a Basemap instance which works like a canvas that we can draw the map on,
3. Extract Polygon objects from the shapefile: each Polygon objects describes a block on the map.
4. Extract coordinates of each restaurant, Transform them into Point objects.
5. Count Point objects that fall into each Polygon object.
6. Normalize count values for each Polygong object.
7. Transform each Polygon objects to Patch object, correlated count value determines the color of the Patch object.
8. Draw Patch objects on the Basemap instance.

### 3.2.3 PLOTTING A RATING MAP A

1. Extract boundary info using fiona from New York City shapefile:
2. Create a Basemap instance which works like a canvas that we can draw the map on,
3. Extract Polygon objects from the shapefile: each Polygon objects describes a block on the map.
4. For each genre of restaurant
  - a) Extract Both coordinates and rating value of each restaurant, Transform them into Point objects.
  - b) Filter each Point object out if it doesn't fall into the New York City map.
  - c) Draw Points objects on the Basemap instance. for each Point object, genre determines the color of the point, rating value determines the diameter of the point.
5. Transform each Polygon objects to Patch object.
6. Draw Patch objects on the Basemap instance .

### 3.2.4 PLOTTING A RATING MAP B

1. Request a map image in png format from google.com
2. Create a Basemap instance which works like a canvas that we can draw the map on, For each genre of food
3. Extract Both coordinates and rating value of each restaurant, Transform them into Point objects.
4. Filter each Point object out if it doesn't fall into the New York City map.
5. Draw Points objects on the Basemap instance. for each Point object, genre determines the color of the point, rating value determines the diameter of the point.
6. Show the png image as background on the Basemap instance.

## 4 GRAPHIC USER PANEL

After finish plotting restaurants information on geography map, the next step for us is designing a graphic user control panel.

### 4.1 GUI LIBRARY SELECTION

There are certain kinds of Python GUI libraries in the market include

- PyQt
- Pyside
- PyGTK
- wxPython
- TkInter

PySide became our choice for its functionality and maintainability. PySide is an open source software project providing Python bindings for the Qt framework. Qt is a cross-platform application and UI framework, allowing the developers to write applications once and deploy them across many operating systems without rewriting the source code, while Python is a modern, dynamic programming language with a vivid developer community. Combining the power of Qt and Python, PySide provides the wealth of Qt framework for developers writing software in Python and presents a first-class rapid application development platform available on all major operating systems(<http://qt-project.org/wiki/About-PySide>).

## 4.2 USER PANEL DESIGN

We include four buttons and six checkboxes in our user panel which is shown in Fig. 4.1. The buttons controls the type of map to be generated; the checkboxes indicates restaurants's nationality.



Fig 4.1

1. 'Show Basic' and 'show Google' allows users to select multiply restaurant type at the same time. To be specific, checking 'Show Google' button with the selection of Chinese and Italian would bring the user a restaurant information map with all Chinese restaurants and Italian restaurants in Google style (See Fig. 4.2).



Fig 4.2

2. 'Show Location' and 'Show Density' requires users select only one restaurant type each time. If a user selects more than one type for these two functions, he will get an warning of 'Only Accept One Input' at the bottom of the panel. For instance, if I select both Korea and French and check 'Show Density Map', no map would be display and a warning message would be shown (See Fig. 3).



Fig 4.3

## 5 EVALUATION

(1) In Data Collection part, we gathered all the data by extracting from Yelp.com. Because the usage limitation of Google Geocoding API, we can only gather no more than 2500 useful data per day, which limited our research region to only New York City area and only 6 genres of restaurant. If there is a better data source we can apply the research result on wider area.

(2) In Data Visualization part, if we can implement a interactive data map, the result can be more useful.

## 6 CONCLUSION

In our research, we offer a static solution for showing all restaurant information of a certain area in one single map. We use urllib2 and BeautifulSoup to collect data from Yelp website; use basemap toolkit to generate the map together with visualized restaurant data; use PySide lib to create user control panel. For the next step, we want to (1) expand the scale to nationwide, (2) use database instead of text files to organize data and (3) implement an interactive data map with more information showed for the restaurant.