

# COURSERA CAPSTONE PROJECT (WEEK 5) – FULL REPORT

Indranil Chatterjee

## Contents

Introduction/ Business Problem .....	2
Discussion on Data .....	2
Methodology.....	2
Detailed Analysis.....	3
Results.....	3
Discussion.....	5
Conclusion.....	6

## Introduction/ Business Problem

The owner of a Chinese restaurant chain is interested to open their first restaurant in Manhattan, New York. For this purpose, he is seeking advice on locations of existing similar restaurants. The owner mentioned that he would like to know the top 10 location/ neighborhood where there are many Chinese (or similar) restaurant already existing. This will help him to further strategize on the location aspect.

Hence here,

### **Target audience**

The owner of the Chinese restaurant chain

### **Problem statement**

To advise on current locations of Chinese (or similar) restaurants in Manhattan

## Discussion on Data

First we'll utilize the New York location data as available on the web (also in labs as part of the Capstone course). It contains the borough and neighborhood information of NY. This data should be first extracted into suitable DataFrames - containing Borough, Neighborhood, Latitude, Longitude.

After that we'll focus on the data for Manhattan borough and neighborhoods within this borough. The relevant latitude and longitude data will help us to utilize Foursquare APIs. Foursquare API will provide us top N number of venues and their categories, latitude, longitude etc. Then we'll focus on categories and try to find out number of Chinese (or similar) restaurants in each neighborhoods. Additionally, we'll map the top 10 neighborhoods on Manhattan map for easy visualization.

## Methodology

The following steps describes the methodology followed:

1. The NY data set is available in the internet. This should be utilized to get the following data: Borough, Neighborhood, Latitude, Longitude.
2. From the overall NY data, the data for Manhattan should be extracted. These data points are required subsequently to get venue details across the neighborhoods.

3. Foursquare API will be utilized to get 100 venues for each of the neighborhoods. The following data will be extracted: venue name, latitude, longitude, venue category
4. From the available venue categories, relevant categories should be explored. As we are looking for Chinese/ similar restaurants, the venue categories will be analyzed for these and for each neighborhood, the number of such restaurants will be calculated
5. The top 10 neighborhoods along with the number of restaurants should be identified
6. These data will be visualized on top of the map of Manhattan. For this Python Folium package will be used.

## Detailed Analysis

This is the link to the ipython notebook containing detailed analysis:

[https://github.com/einchat/Coursera\\_Capstone/blob/master/Course9\\_Capstone\\_Week5-V01.ipynb](https://github.com/einchat/Coursera_Capstone/blob/master/Course9_Capstone_Week5-V01.ipynb)

The below files are some important result obtained from the analysis:



manhattan\_grouped\_2-excel.xlsx



manhattan\_venues-excel.xlsx

Please refer to the ipython notebook for more details on analysis

### Note

- Here we have considered the following categories under 'Chinese/ similar'
  - Chinese Restaurant, Dim Sum Restaurant, Dumpling Restaurant, Shanghai Restaurant, Szechuan Restaurant
  - If this classification is changed, results may vary

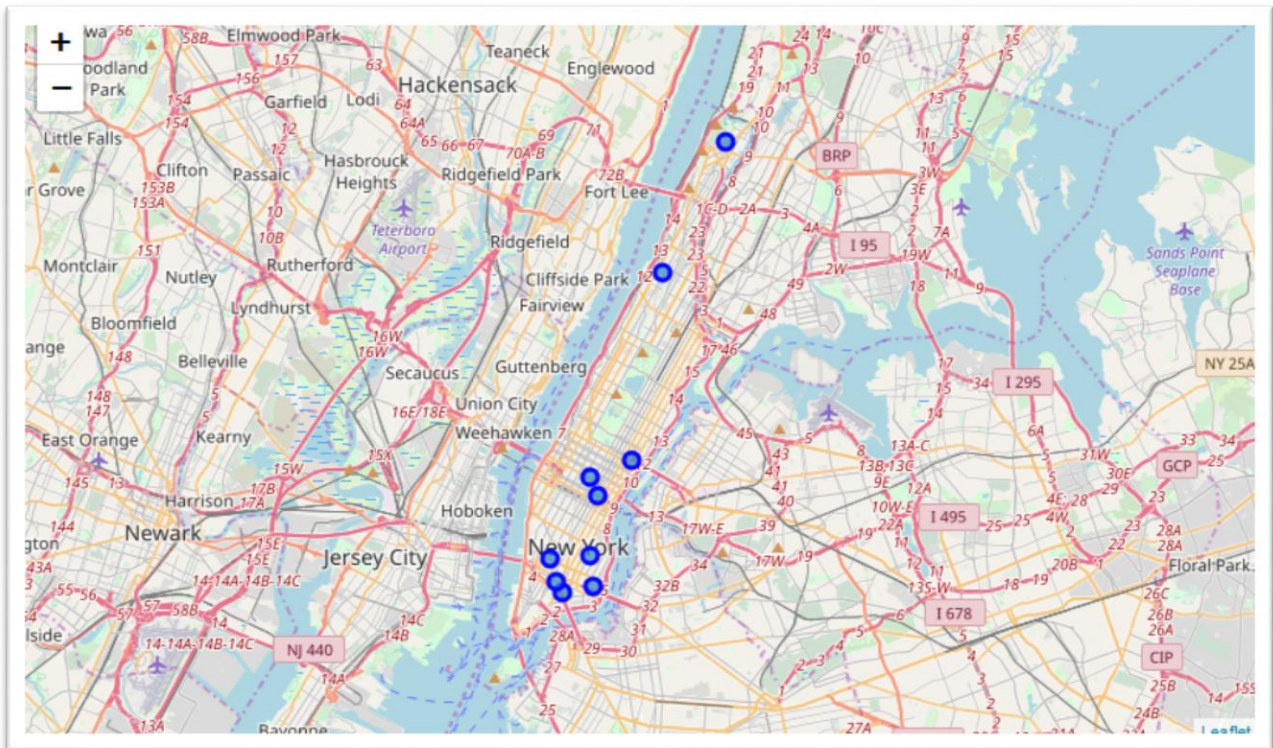
## Results

1. The following table shows Top 10 neighborhoods with highest number of Chinese/ similar restaurants:

Neighborhood	Chinese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Shanghai Restaurant	Szechuan Restaurant	Total
<b>Chinatown</b>	9	4	3	1	0	17.0
<b>Little Italy</b>	4	0	0	0	0	4.0

<b>East Village</b>	3	0	1	0	0	4.0
<b>Lower East Side</b>	3	0	0	0	0	3.0
<b>Midtown</b>	2	0	0	0	1	3.0
<b>Greenwich Village</b>	3	0	0	0	0	3.0
<b>Sutton Place</b>	1	0	0	0	1	2.0
<b>Inwood</b>	2	0	0	0	0	2.0
<b>Murray Hill</b>	1	0	0	1	0	2.0
<b>Hamilton Heights</b>	2	0	0	0	0	2.0

2. The map below shows the top 10 neighborhood as described in previous point



3. Also let us identify neighborhoods with 0 number Chinese/ similar restaurants:

Neighborhood	Chinese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Shanghai Restaurant	Szechuan Restaurant	Total
<b>Gramercy</b>	0	0	0	0	0	0
<b>Chelsea</b>	0	0	0	0	0	0
<b>Civic Center</b>	0	0	0	0	0	0
<b>Upper East Side</b>	0	0	0	0	0	0
<b>Turtle Bay</b>	0	0	0	0	0	0

<b>Stuyvesant Town</b>	0	0	0	0	0	0
<b>Marble Hill</b>	0	0	0	0	0	0
<b>Soho</b>	0	0	0	0	0	0
<b>Roosevelt Island</b>	0	0	0	0	0	0
<b>Noho</b>	0	0	0	0	0	0
<b>Hudson Yards</b>	0	0	0	0	0	0
<b>Lenox Hill</b>	0	0	0	0	0	0
<b>Midtown South</b>	0	0	0	0	0	0
<b>Yorkville</b>	0	0	0	0	0	0

4. Neighborhoods with only 1 Chinese/ similar restaurants:

<b>Neighborhood</b>	<b>Chinese Restaurant</b>	<b>Dim Sum Restaurant</b>	<b>Dumpling Restaurant</b>	<b>Shanghai Restaurant</b>	<b>Szechuan Restaurant</b>	<b>Total</b>
<b>Tudor City</b>	0	0	0	1	0	1
<b>Upper West Side</b>	1	0	0	0	0	1
<b>Morningside Heights</b>	1	0	0	0	0	1
<b>Battery Park City</b>	1	0	0	0	0	1
<b>Carnegie Hill</b>	1	0	0	0	0	1
<b>Lincoln Square</b>	1	0	0	0	0	1
<b>East Harlem</b>	1	0	0	0	0	1
<b>Financial District</b>	1	0	0	0	0	1
<b>Flatiron</b>	1	0	0	0	0	1

## Discussion

- It can be observed that out of total Manhattan, Chinatown has significantly large number of Chinese restaurants
- For rest of the top 10 locations, there are 2-4 restaurants in the neighborhoods
- It can be seen from the analysis that 14 neighborhoods have 0 number of such restaurants
- 9 neighborhoods have only 1 of such restaurants
- So depending on the business strategy, the organization can focus either in areas where there are quite a few such restaurants (like Chinatown, Little Italy etc.) OR in neighborhoods where there are only 0 or 1 such restaurant
- Some limitation of the model:
  - Here we have not done any sizing of the potential customer base, i.e. number of people per neighborhood who may be interested in Chinese food. This can subsequently be done as further refinement

## Conclusion

This concludes the report on the (hypothetical) problem statement of identifying number of Chinese/similar restaurants in Manhattan, NY neighborhoods. It utilizes the methodologies taught in the Coursera course for a real world problem.