

Analysis the Chinatown in the U.S

Zetong Gu 04/24/2021

1 Introduction

Chinatown is an ethnic enclave of Chinese people located outside mainland China, Hong Kong, Macau or Taiwan, most often in an urban setting. Areas known as "Chinatown" exist throughout the world, including Europe, North America, South America, Asia, Africa and Australasia.

Chinatowns are cities within cities: adventurous, bustling, full of distinctive signage, street vendors selling unusual items, specialty shops, a noted lack of big chains, a variety of dialects being spoken, and multitudes of unique and exciting food choices.

The Chinese have been established in the United States since the mid-19th century, when laborers were needed for gold mining and railroad work, but the immigrant population also grew during the 1990s and 2000s; in fact, more than one-third of the Chinese immigrants now living in the U.S. arrived in 2000 or later. Currently, there are more than 3 million Chinese in America, according to the 2008 census report. Whether leaving China for issues ranging from poverty, famine or political reasons, across the decades, the Chinese have built strong communities that keep their ethnic heritage and shared identity; this maintained and rich culture is a defining reason that Chinatowns endure and why they're so appealing for residents and tourists alike.

Around the globe, there are Chinatowns in many major cities, from London (Europe's largest) to Vancouver (Canada's largest), Melbourne to Manila; and fortunately for us, there are many within the United States. Many of these districts share their community with other immigrant cultures, making the sights, sounds and eating choices that much more exotic. In a neighborhood where English is not the primary language, a visitor can feel as though they've left the U.S. altogether — and now they are the foreigner, a tourist in their own city.

So, what exactly makes a Chinatown great?

In order to compile my capstone, I will take a look at Top10 in America, and analysis the quality of authentic dining options, size, cultural experiences available, and whether a visitor will feel like they've left the United States as they explore the neighborhood.

2 Data

2.1 Get the City List

I use a list from USA today for my research items. They offer us a top10 list for Chinatowns in the United States. I will explore the all the Top 10 Chinatowns and compare them from all the aspects.

- San Francisco
- New York City
- Chicago
- Seattle
- Philadelphia
- Honolulu
- Boston
- Los Angeles
- Dallas
- Washington, D.C.

The list is from here: <https://www.usatoday.com/story/travel/destinations/2014/03/08/chinatown-chinese-asian-food/6173601/>

2.2 Get the City Chinatown Geo Data

I will use the Geocoder Python package to get the latitude and longitude values data of Chinatown in each city. Data from: <https://geocoder.readthedocs.io/index.html>.

Geocoder is a simple and consistent geocoding library written in Python. Dealing with multiple different geocoding provider such as Google, Bing, OSM & many more has never been easier.

2.3 Explore the neighborhoods using Foursquare API

Lastly, I will utilize the Foursquare API to explore the Chinatown neighborhoods and analysis them. I will get the top 100 venues that are in Chinatown within a radius of 1000 meters.

Foursquare is a service that lets users announce their presence at a venue ("check in"). They can build up a social graph on the service, or just announce their checkins on other services like Twitter or Facebook.

The following analysis will be performed by using the data we get:

- Analysis the quality of authentic dining options, size, cultural experiences
- Analysis the venues between Restaurant and other catalogs.
- Find the most similar Chinatowns among those ten cities.

3 Methodology

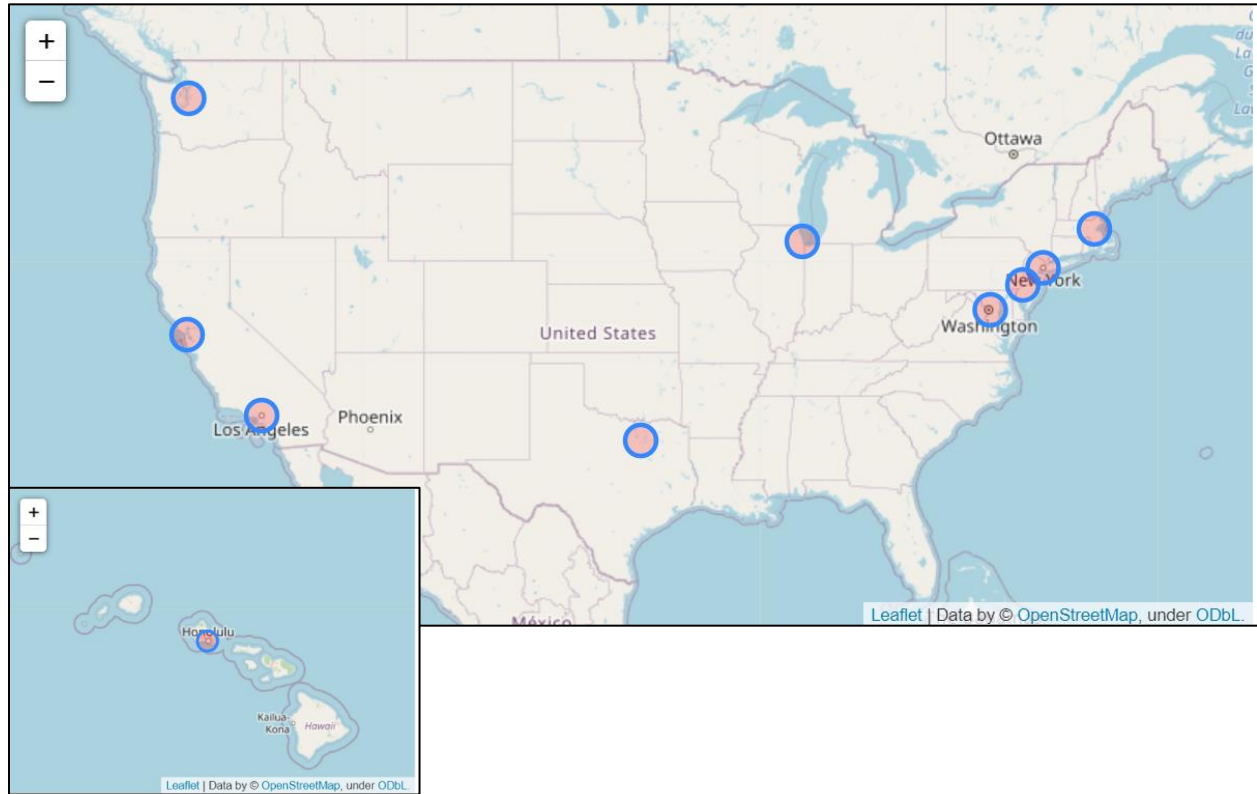
First, we load city list into pandas data frame

	CityName	Chinatown_address
0	San Francisco	Chinatown, San Francisco
1	New York City	Chinatown, New York City
2	Chicago	Chinatown, Chicago
3	Seattle	Chinatown, Seattle
4	Philadelphia	Chinatown, Philadelphia
5	Honolulu	Chinatown, Honolulu
6	Boston	Chinatown, Boston
7	Los Angeles	Chinatown, Los Angeles
8	Dallas	Chinatown, Dallas
9	Washington, D.C.	Chinatown, Washington, D.C.

Then, we use Geolocator to get Chinatown latitude and longitude in each city. Enter the Chinatown address in Geolocator, it will return the latitude and longitude value of Chinatown. After getting the latitude and longitude data, they will attach to the same data frame of previous.

	CityName	Chinatown_address	Lat	Lon
0	San Francisco	Chinatown, San Francisco	37.794301	-122.406376
1	New York City	Chinatown, New York City	40.716491	-73.996250
2	Chicago	Chinatown, Chicago	41.851658	-87.633138
3	Seattle	Chinatown, Seattle	47.599226	-122.323447
4	Philadelphia	Chinatown, Philadelphia	39.953446	-75.154622
5	Honolulu	Chinatown, Honolulu	21.312903	-157.862800
6	Boston	Chinatown, Boston	42.351329	-71.062623
7	Los Angeles	Chinatown, Los Angeles	34.063840	-118.235868
8	Dallas	Chinatown, Dallas	32.953118	-96.728050
9	Washington, D.C.	Chinatown, Washington, D.C.	38.900342	-77.021443

For next step, we use Folium package to display the Chinatown location. This help us to have a visual overview of distribution. Among those ten cities, nine of ten are locates in the US continent, and the other one is from Hawaii.



We use four squares API to explore each Chinatowns. For each point, I explored the 100 venues within 1000 meters range. Typically, Chinatown is not a big neighborhood, so 1000 meters should be good enough to cover the area.

Example for venues data

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Chinatown, San Francisco	37.794301	-122.406376	Blue Bottle Coffee	37.792771	-122.404833	Coffee Shop
1	Chinatown, San Francisco	37.794301	-122.406376	Red Blossom Tea Company	37.794643	-122.406379	Tea Room
2	Chinatown, San Francisco	37.794301	-122.406376	Hinodeya	37.794656	-122.404544	Ramen Restaurant
3	Chinatown, San Francisco	37.794301	-122.406376	Old Saint Mary's Cathedral (Old Cathedral of S...	37.792772	-122.405910	Church
4	Chinatown, San Francisco	37.794301	-122.406376	Yi Fang Taiwan Fruit Tea	37.795038	-122.407614	Bubble Tea Shop

It total returned 961 venues for Chinatowns in ten cities. Unique categories return from four square API.

After that, we use one-hot encoding to analysis the frequency of venue in each city Chinatown. This is not only analysis the frequency of venue, but also prepare for further clustering. One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction. One hot encoding creates new (binary) columns, indicating the presence of each possible value from the original data. In one hot encoding, 0 indicates non existent while 1 indicates existent.

Example for one-hot encoding

	Yoga Studio	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	Austrian Restaurant	Automotive Shop	...	Travel Lounge	Turkish Restaurant	Vegetarian / Vegan Restaurant	Video Game Store
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0

Then using the mean square and grouped function, we can get frequency of occurrence of each category.

----Chinatown, Boston---- venue freq 0 Chinese Restaurant 0.08 1 Bakery 0.06 2 Coffee Shop 0.05 3 Sandwich Place 0.04 4 Asian Restaurant 0.04	----Chinatown, Chicago---- venue freq 0 Chinese Restaurant 0.22 1 Pizza Place 0.05 2 Asian Restaurant 0.04 3 Mexican Restaurant 0.03 4 Korean Restaurant 0.03
----Chinatown, Dallas---- venue freq 0 Chinese Restaurant 0.11 1 Indian Restaurant 0.07 2 Burger Joint 0.05 3 Mediterranean Restaurant 0.05 4 Coffee Shop 0.05	----Chinatown, Honolulu---- venue freq 0 Bakery 0.07 1 Coffee Shop 0.06 2 American Restaurant 0.05 3 Chinese Restaurant 0.05 4 Pharmacy 0.04
----Chinatown, Los Angeles---- venue freq 0 Chinese Restaurant 0.11 1 Mexican Restaurant 0.08 2 Sandwich Place 0.05 3 Coffee Shop 0.05 4 Bakery 0.04	----Chinatown, New York City---- venue freq 0 Sandwich Place 0.05 1 Cocktail Bar 0.05 2 Bakery 0.05 3 Ice Cream Shop 0.04 4 Chinese Restaurant 0.04
----Chinatown, Philadelphia---- venue freq 0 Chinese Restaurant 0.06 1 Bakery 0.04 2 History Museum 0.03 3 Hotel 0.03 4 Sandwich Place 0.03	----Chinatown, San Francisco---- venue freq 0 Coffee Shop 0.07 1 Pizza Place 0.04 2 New American Restaurant 0.04 3 Cocktail Bar 0.04 4 Hotel 0.04
----Chinatown, Seattle----	----Chinatown, Washington, D.C.----

	venue	freq		venue	freq
0	Vietnamese Restaurant	0.11	0	American Restaurant	0.07
1	Chinese Restaurant	0.09	1	Hotel	0.06
2	Coffee Shop	0.05	2	Italian Restaurant	0.04
3	Cocktail Bar	0.04	3	Cocktail Bar	0.04
4	Café	0.03	4	Pizza Place	0.04

After using one-hot encoding, we use k -means to cluster the Chinatown into 3 clusters. The K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

To process the learning data, the K-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids

It halts creating and optimizing clusters when either:

- The centroids have stabilized — there is no change in their values because the clustering has been successful.
- The defined number of iterations has been achieved.

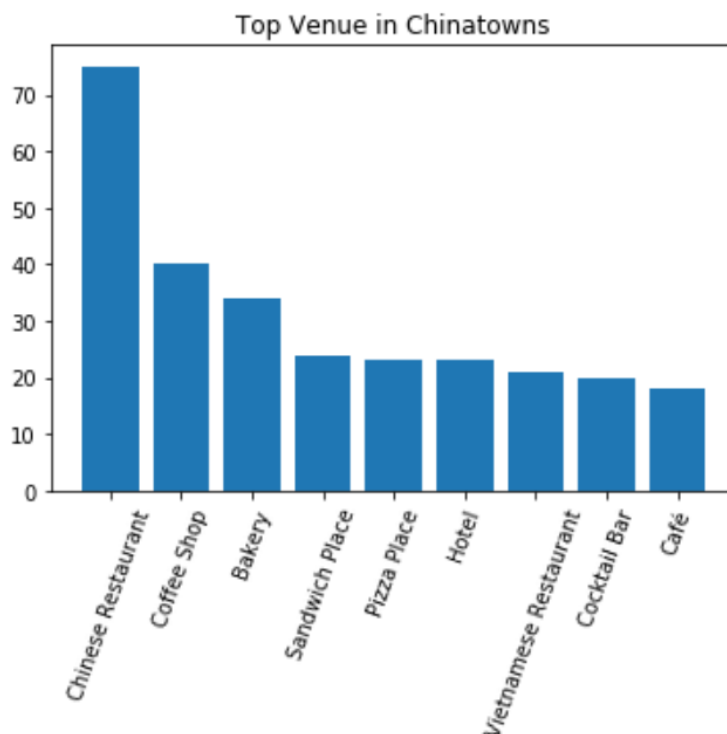
Here we use “sklearn” package for processing. The cluster results are saved at cluster label.

Cluster the Chinatown into 3 groups by K-means

	CityName	Chinatown_address	Lat	Lon	Cluster Label
0	San Francisco	Chinatown, San Francisco	37.794301	-122.406376	0
1	New York City	Chinatown, New York City	40.716491	-73.996250	2
2	Chicago	Chinatown, Chicago	41.851658	-87.633138	0
3	Seattle	Chinatown, Seattle	47.599226	-122.323447	0
4	Philadelphia	Chinatown, Philadelphia	39.953446	-75.154622	0
5	Honolulu	Chinatown, Honolulu	21.312903	-157.862800	1
6	Boston	Chinatown, Boston	42.351329	-71.062623	0
7	Los Angeles	Chinatown, Los Angeles	34.063840	-118.235868	1
8	Dallas	Chinatown, Dallas	32.953118	-96.728050	0
9	Washington, D.C.	Chinatown, Washington, D.C.	38.900342	-77.021443	1

4 Results

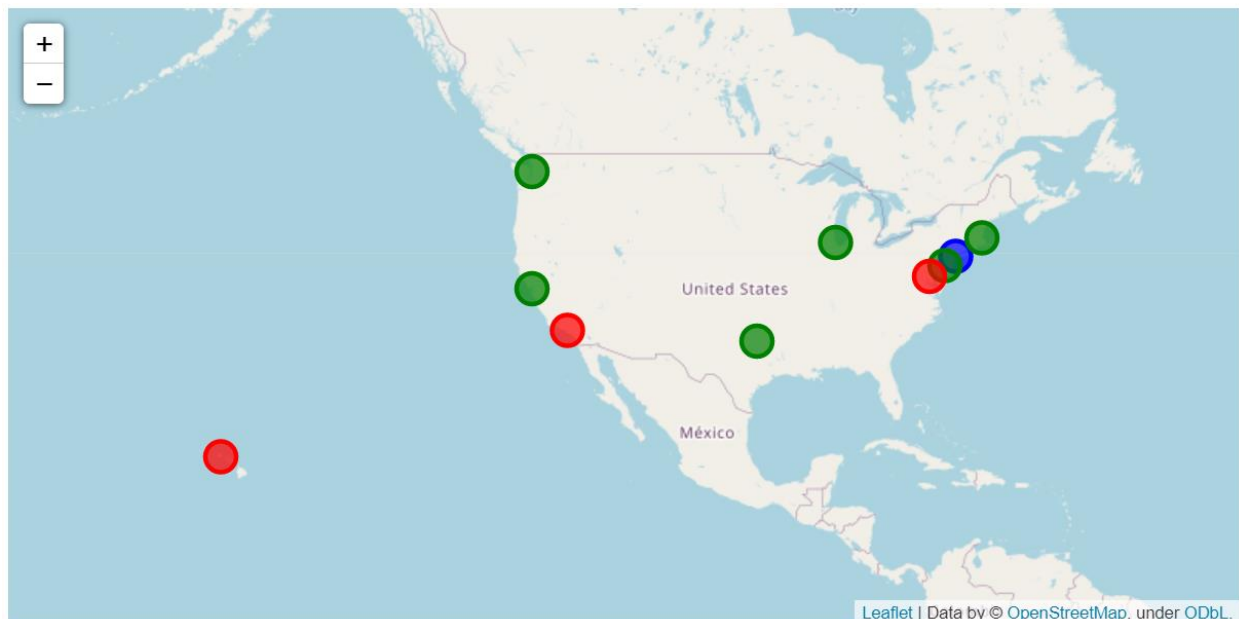
After getting the venues near Chinatown, we plot overall venue catalog among these Chinatowns. The figure demonstrate that the majority of venues are relate to EAT. "Food is the first necessity of the people" is a famous Chinese old saying, which reflects that Chinese have had paid much attention to food since the ancient times. The importance of food in people's daily life in China can be seen in the way they send greetings to each other: "Have you eaten?", which is far away from the way people greet in western countries and has been seen as a characteristic of Chinese culture. And obviously, Chinese Restaurant is the top 1 venue here following by coffee shop and bakery.



With the table below, we got the top 10 common venue in each Chinatown specifically. Besides the Chinese restaurant, we also observed other type of restaurant, like Korean restaurant, Japanese restaurant, Indian restaurant etc. The table shows the diversity in Chinatown.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Chinatown, Boston	Chinese Restaurant	Bakery	Coffee Shop	Asian Restaurant	Sandwich Place	Performing Arts Venue	Park	Theater	Hotel	Seafood Restaurant
1	Chinatown, Chicago	Chinese Restaurant	Pizza Place	Asian Restaurant	Mexican Restaurant	Korean Restaurant	Rental Car Location	Thai Restaurant	Bakery	Bubble Tea Shop	Storage Facility
2	Chinatown, Dallas	Chinese Restaurant	Indian Restaurant	Coffee Shop	Mediterranean Restaurant	Burger Joint	Café	Thai Restaurant	Noodle House	Bubble Tea Shop	Convenience Store
3	Chinatown, Honolulu	Bakery	Coffee Shop	American Restaurant	Chinese Restaurant	Pharmacy	Café	Dessert Shop	Japanese Restaurant	Pizza Place	Shopping Mall
4	Chinatown, Los Angeles	Chinese Restaurant	Mexican Restaurant	Coffee Shop	Sandwich Place	Bakery	Historic Site	Vietnamese Restaurant	Plaza	Art Gallery	Café
5	Chinatown, New York City	Bakery	Cocktail Bar	Sandwich Place	Chinese Restaurant	Ice Cream Shop	Wine Bar	Mexican Restaurant	Coffee Shop	Thai Restaurant	Pizza Place
6	Chinatown, Philadelphia	Chinese Restaurant	Bakery	Sandwich Place	History Museum	Coffee Shop	Hotel	Tea Room	Restaurant	Shanghai Restaurant	Brewery
7	Chinatown, San Francisco	Coffee Shop	Pizza Place	Cocktail Bar	Hotel	New American Restaurant	Bakery	Chinese Restaurant	Men's Store	Sushi Restaurant	Restaurant
8	Chinatown, Seattle	Vietnamese Restaurant	Chinese Restaurant	Coffee Shop	Cocktail Bar	Japanese Restaurant	Pizza Place	Café	Thrift / Vintage Store	Bubble Tea Shop	Tea Room
9	Chinatown, Washington, D.C.	American Restaurant	Hotel	Pizza Place	Cocktail Bar	Italian Restaurant	Theater	Mediterranean Restaurant	Salad Place	Indian Restaurant	Plaza

Last but not least, we cluster all the Chinatown into 3 clusters by using the k-means algorithm. We cluster the majority of Chinatowns into group1. Los Angeles , Honolulu, and Washington DC are in group2. Finally, Chinatown in New York city is the only one candidate in group3. The New York Chinatown is unique among all of them. So if you would like to go to a special Chinatown, don't miss the New York city one!



5 Discussion

From previous map, we had the Chinatown map inside the US. Among those ten cities, nine of ten are located in the US continent, and the other one is from Hawaii. For these nine Chinatowns inside the United States continent, four of them are located at the east coast, two of them are located at central US, and the remaining three are at the west coast. By looking at the city, it pretty much aligns with the city with the most population in the US. Chinatowns in the United States have historically been located in the "big cities" and existed initially as enclaves that ease transition into the American culture. The earliest Chinatowns tended to be on the west coast while the newer ones are being built in lesser profile cities as opportunities shift. As the migration trends toward returning to China, many Chinatowns, especially smaller ones like the one in Washington DC, begin to lose their initial mission. In Honolulu, since it is near China geographically, so it is not surprising that there is a nice Chinatown there.

From the clustering results, we found the New York (Manhattan) Chinatown is the most special one. The New York (Manhattan) Chinatown is one of nine Chinatown neighborhoods in New York City, as well as one of twelve in the New York metropolitan area, which contains the largest ethnic Chinese population outside of Asia, comprising an estimated 893,697 individuals as of 2017. Walking its busy, narrow streets reveals surprise after surprise.

6 Conclusion

In this report, we use python to explore ten Chinatowns in the United States. Using the foursquare API to get the venues among those places. We discovered the most common venues in each Chinatown as well as overall venues study. Lastly, we clustered the Chinatown into 3 groups, and found the most unique Chinatown: New York City.