Chinatowns Around the World: Are They All the Same?

Ayan Ashkenov

Coursera: IBM Applied Data Science Capstone

3 August 2020

Introduction:

The purpose of this analysis is to determine whether biggest Chinatowns around the world share any similarities. Chinatowns can be found in numerous major cities across the globe. Given their common status worldwide, it is worth measuring the extent to which they are similar or different from each other. Chinatowns do not only act as historical homes to ethnic Chinese communities, but also provide a cultural bridge to the host population and tourists from abroad.

Individuals or groups seeking to expand their business to other Chinatowns may find this analysis helpful in their business decision. Individuals or groups thinking of starting business in the Chinatown area of the world's biggest cities may also find great use in this evaluation.

Data:

In order to access location data through Foursquare API, such as nearby venues and their categories, it was first necessary to acquire coordinates of Chinatowns around the world. Initially, 17 Chinatown locations were chosen across the world: New York, Chicago, Los Angeles, San Francisco, Philadelphia, Houston, Honolulu, Portland, Boston, Melbourne, Toronto, Vancouver, Paris, London, Manchester, Amsterdam, and Lima. Further analysis discovered that Vancouver's Chinatown only generated 4 venues through FourSquare API, which resulted in its omission in the final analysis. For each location, its latitude and longitude were found using one of the following sources: Wikipedia, latlong.net, or latitude.to. All of this was then manually compiled into an Excel spreadsheet (Table 1). Analysis was performed using Python.

Location	Latitude	Longitude
New York, NY, USA	40.7176638	-74.00149999
Chicago, IL, USA	41.851215	-87.634422
Los Angeles, CA, USA	34.062888	-118.23789
San Francisco, CA, USA	37.79016351	-122.4043317
Philadelphia, PA, USA	39.9535	-75.1563
Houston, TX, USA	29.705	-95.5453
Honolulu, HI, USA	21.30749877	-157.8584966
Portland, OR, USA	45.52528	-122.67246
Boston, MA, USA	42.3501	-71.0624
Melbourne, Australia	-37.8118	144.9676
Toronto, Canada	43.6666144	-79.3472958
Paris, France	48.82222	2.36528
London, UK	51.50666464	-0.125499498
Amsterdam, Netherlands	52.3739	4.9001
Lima, Peru	-12.0524	-77.0255
Manchester, UK	53.4786	-2.2401

Table 1: Manually compiled Excel Spreadsheet containing Chinatown locations and their coordinates.

Methodology:

The analysis consisted of 3 stages. First, select which Chinatown locations to include in the analysis and acquire the necessary information for each one. Second, use FourSquare API to collect all venues within a certain radius of each Chinatown location and determine which are the 10 most common venue categories for each. Third, use the K-means algorithm to cluster the Chinatown locations to evaluate which ones are similar.

The second stage was the largest by volume. Using the coordinates of each Chinatown location, all venues on the FourSquare platform were compiled in the radius of 350 meters. In total, there were 829 venues and 189 unique venue categories for the 16 Chinatown locations. Ultimately, the venue distribution was not balanced as some Chinatown locations had more venues than others. For example, New York had 100 venues and Amsterdam had 76, while Honolulu and Toronto had 15 and 14, respectively. Each venue had its name, coordinates, and category recorded. This venue data was then transformed and manipulated to display the 10 most common venue categories for each Chinatown location.

Results:

Once the 10 most common venue categories were calculated for each Chinatown, it was time to employ the K-means algorithm to cluster the locations. Before that, however, it was first necessary to establish the optimal number of clusters for the algorithm. To accomplish this, it was decided to utilize the Elbow Method. The results pointed towards using 3 cluster groups.

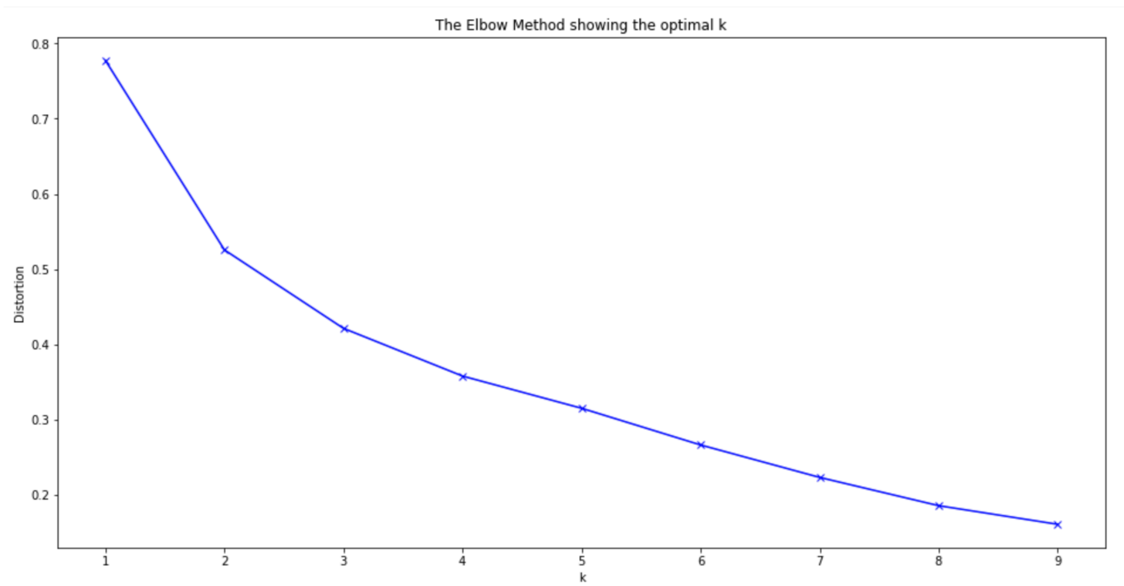

Figure 1: Distortion vs. K plot showing the optimal number of clusters.

The resulting 3 clusters showed surprising results. Since a large portion of Chinatown locations in the data was from the United States, it was logical to assume that they would potentially be grouped together. However, as the composition of common venue categories showed initially, this was far from the truth. For example, Los Angeles' Chinatown had more in common with the

Chinatown in Melbourne than in Chicago or Boston. On the other hand, both Chinatowns in the United Kingdom, those in London and Manchester, were clustered together. Overall, the American Chinatowns in the data were divided among all 3 clusters.

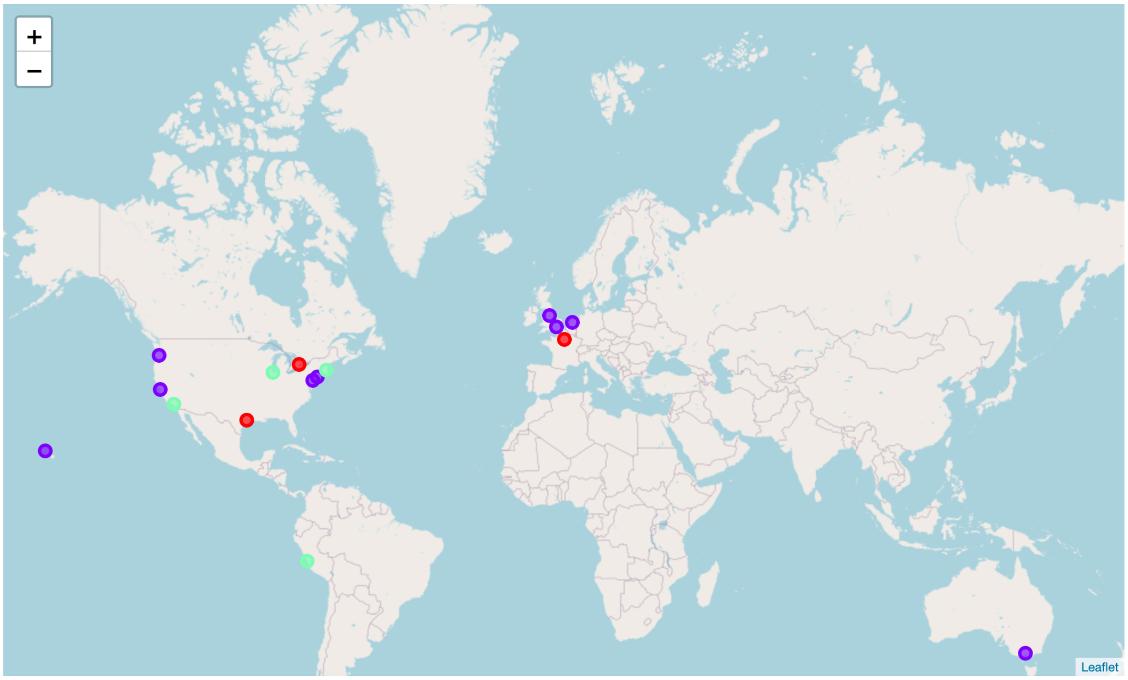

Figure 2: Map with Chinatown locations used in the analysis, with 3 colors representing 3 cluster groups.

The first cluster included Chinatowns in Houston, Toronto, and Paris. The second included New York, San Francisco, Philadelphia, Honolulu, Portland, Melbourne, London, Manchester, and Amsterdam. The third cluster included the rest, Chicago, Los Angeles, Boston, and Lima.

Judging from the cluster grouping and the locations' most common venue categories, it became clear that cluster one and three share more similarities with each other than with cluster two. Cluster one's most common venue categories were Vietnamese restaurant, Chinese restaurant, Asian restaurant, and Bubble tea shop. Cluster three's most common categories included Chinese restaurant, Dim sum restaurant, Vietnamese restaurant, Bakery, and Asian restaurant. It is very likely that if two cluster groups were chosen instead of three, cluster one and three would be merged together.

Cluster two was significantly different from the other clusters. The prevalence of Asian cuisine was drastically lower in comparison and many categories not present in cluster one and three were found in cluster two. Overall, it would be difficult to identify cluster 2 locations as Chinatowns given their common venue category distribution. Lastly, New York's Chinatown was something off an outlier in cluster two as its most common category was Chinese restaurant, making it more similar to cluster one and three Chinatowns in that regard. However, it is evident that cluster two's Chinatowns have very different venue compositions, featuring more coffee shops, bars, and hotels than the other two clusters.

Discussion:

Business owners seeking expansion to other Chinatown locations may find it most useful to look to Chinatowns in their location's cluster. For example, a Dim Sum restaurant owner in Amsterdam may find more success in branching out to New York, rather than Chicago or Paris. An owner of a Bubble Tea Shop in Chicago, on the other hand, does not necessarily have to only look at Lima or Boston, but potentially consider locations in cluster one, such as Houston or Toronto. Additionally, individuals or groups seeking to make business in any of the analyzed Chinatown locations will now have a better understanding of their preferred location's venue composition.

In the end, however, choosing a location for moving or expanding a business is not as simple as finding the neighborhood in the same country or abroad that most closely matches its original location in terms of composition. The same goes for establishing a business, since knowing what the composition and competition is like in a location is not a full stack solution. Many more things should go into such a business decision, including but not limited to area demographics, rent cost, labor cost, ingredient sourcing, ease of making business, cultural specifics, and competition environment. This analysis only tackles one area of concern for businesses or individuals seeking to expand, relocate, or establish themselves. The nature of this analysis allows it to act as a supplement in making a business decision. When used in conjunction with other research and findings, this analysis will be most effective.

Conclusion:

The purpose of this analysis was to determine how similar Chinatowns across the world are and to evaluate which ones share most similarities. Using FourSquare API, it was determined that out of the 16 Chinatowns analyzed, some are hotspots of Asian cuisine while others are more diverse in terms of venue categories. This report points to the idea that location and host culture play a very important role in shaping ethnic communes over time. Additionally, it clearly shows that not all Chinatowns are the same and that Chinatowns in one country are not often alike.

Further Notes:

- The code for this analysis and Table 1 Excel file can be found on the author's GitHub page (username: sha-naya; repository: IBM-Capstone). Reading of this report is best in conjunction with the actual code used to generate the analysis.
- The venues acquired through FourSquare API were of date version 01/01/2020. Given the world pandemic due to COVID-19, it is possible that many venues used in this

analysis ceased to exist as of today. The date was specifically chosen to represent the time before the pandemic began worldwide and to include as many venues as possible.

- When viewing the code on GitHub (a Jupyter notebook file), it is best to use NBViewer, which can be found on Jupyter's website.
- It is best to contact the author through LinkedIn if there are questions, concerns, or commentary on the report (Ayan Ashkenov).
- Python version used: 3.7.6
- Programming environment: Jupyter Notebook
- System: macOS Catalina 10.15.5; MacBook Pro (Retina, 13-inch, Mid 2014)