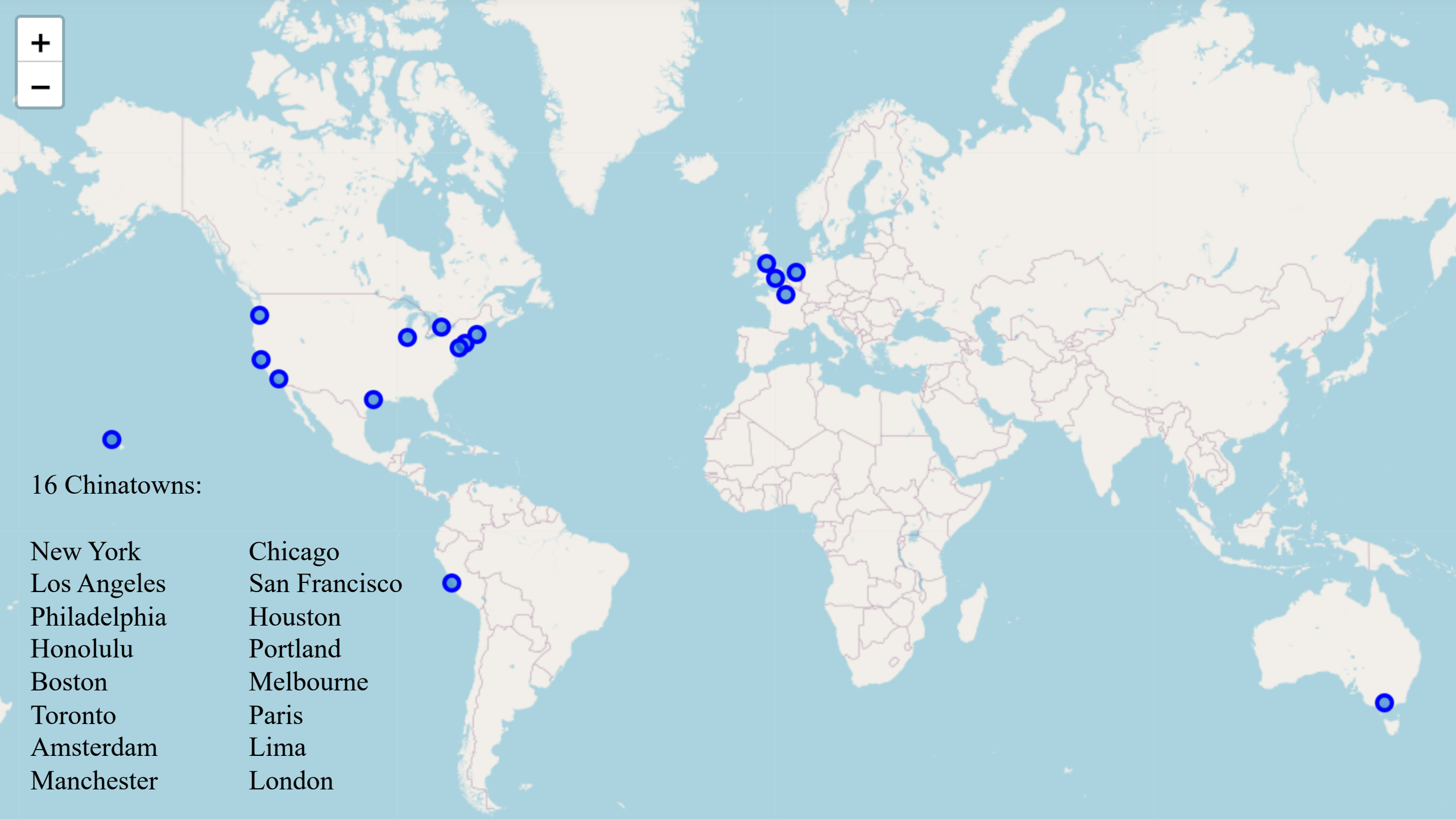
Grouping Chinatowns Around the World

Ayan Ashkenov

Coursera: IBM Applied Data Science Capstone

3 August 2020


+
−
16 Chinatowns:
New York
Chicago
Los Angeles
San Francisco
Philadelphia
Houston
Honolulu
Portland
Boston
Melbourne
Toronto
Paris
Amsterdam
Lima
Manchester
London

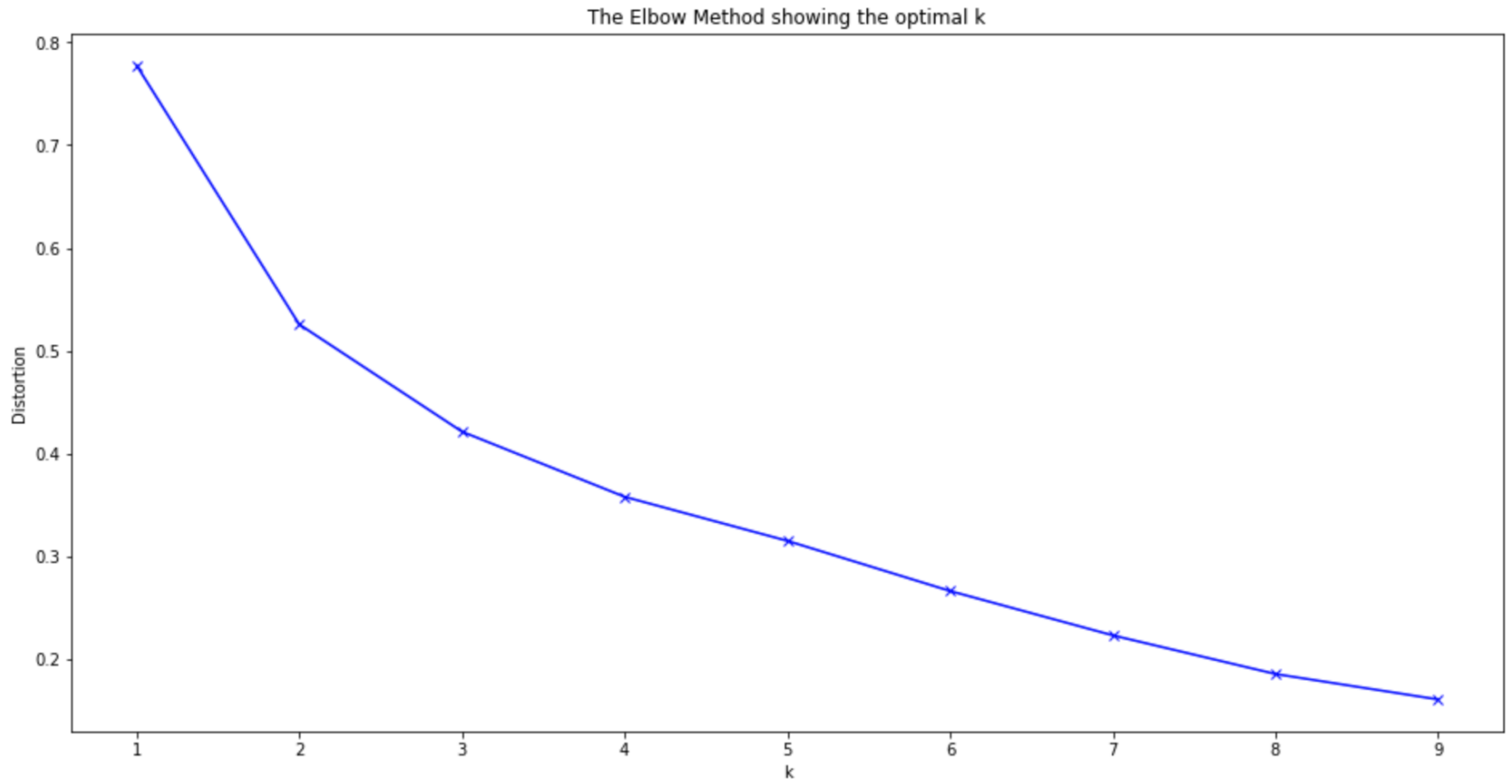
Data

- Coordinates of Chinatowns using Wikipedia, latlong.net, or latitude.to
- Foursquare API location data (venues, categories, coordinates)

Methodology

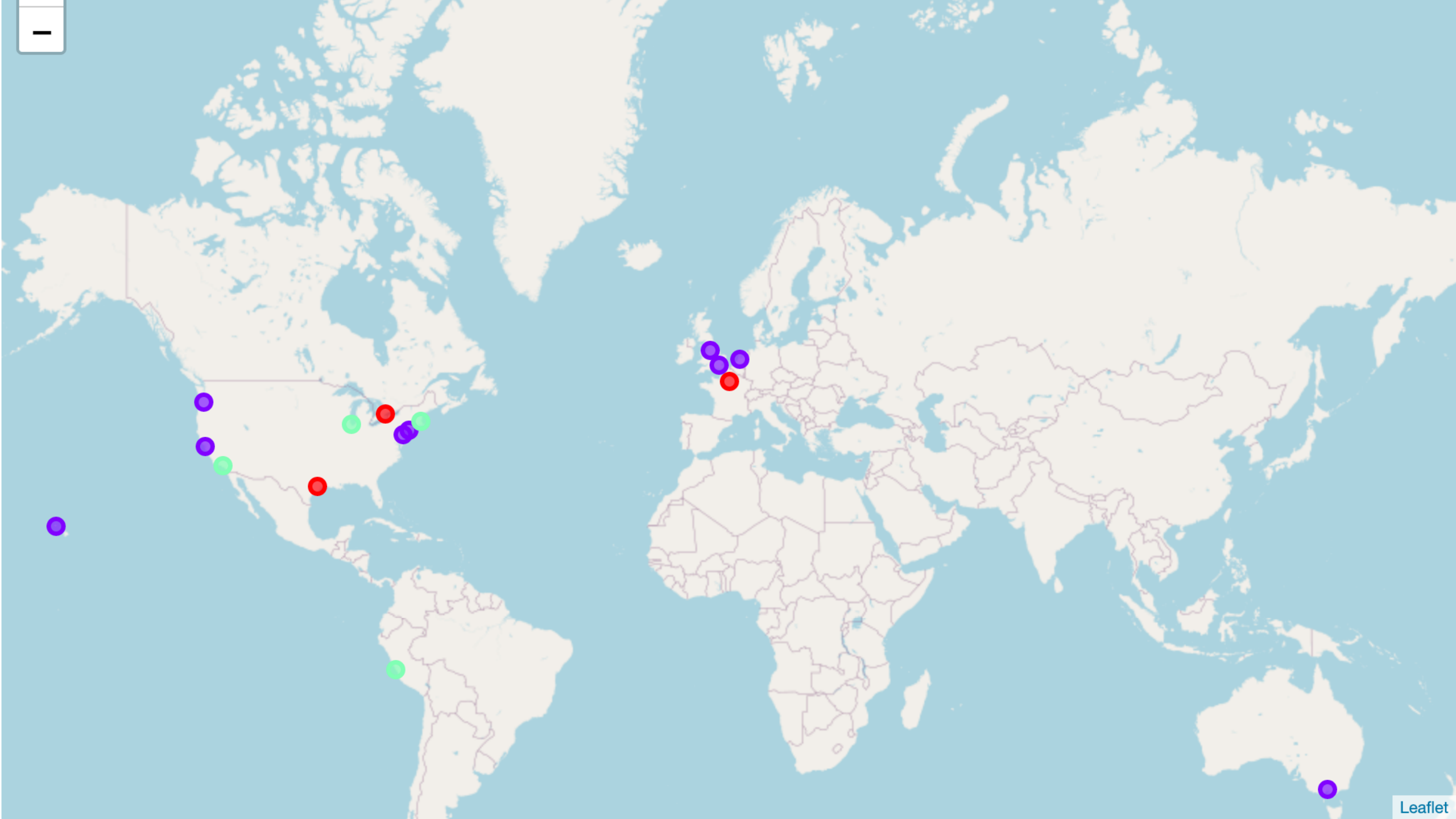
1. Create an Excel spreadsheet containing the 16 Chinatown locations and their coordinates.
2. Use Python and Foursquare API to get all venues in the radius of 350 meters for each Chinatown location. Then, calculate the 10 most common venue categories for each Chinatown.
3. Use K-means algorithm to cluster Chinatowns to determine which ones are similar.

The Elbow Method showed that 3 clusters is the optimal number.



3 Clusters

1. Houston, Toronto, and Paris. Red
2. New York, San Francisco, Philadelphia, Honolulu, Portland, Melbourne, London, Manchester, and Amsterdam. Purple
3. Chicago, Los Angeles, Boston, and Lima. Green


Leaflet

Results and Conclusion

- Chinatown business owners seeking to relocate or expand may find it most effective to do so in their current location's cluster.
- This analysis maps the composition of 16 Chinatowns across the world, acting as a useful tool for individuals or groups looking to establish a business in one of those locations.
- Chinatowns in America were divided between all 3 clusters. Chinatowns in the UK were part of the same cluster.
- Clusters 1 and 3 share some similarities. The biggest difference between them and cluster 2 seems to be the prevalence of Asian cuisine and absence of some venue categories found in cluster 2.