The Battle of the Neighborhoods Clustering of cities

Maxime Menu

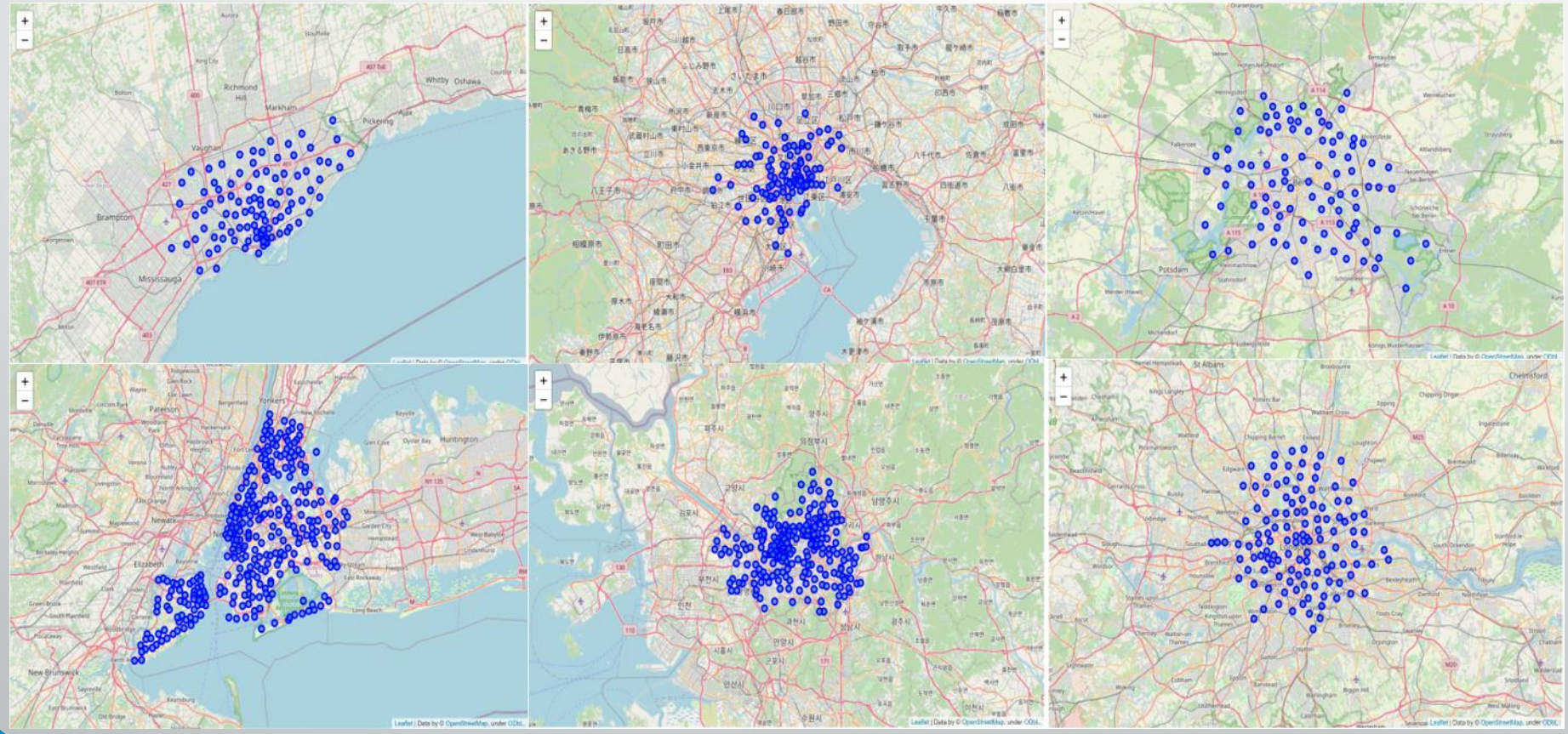
Finding similarities between cities is valuable to the travel industry

- Cheap air travel and globalization has made overseas travel easier
- Cultures all over the world have been cross-influencing each other
- A lot of countries and capital-like cities share many kinds of venues
- Knowing which cities are most similar could help travel industries recommending cities to potential customers
- Which city is most similar to New York City

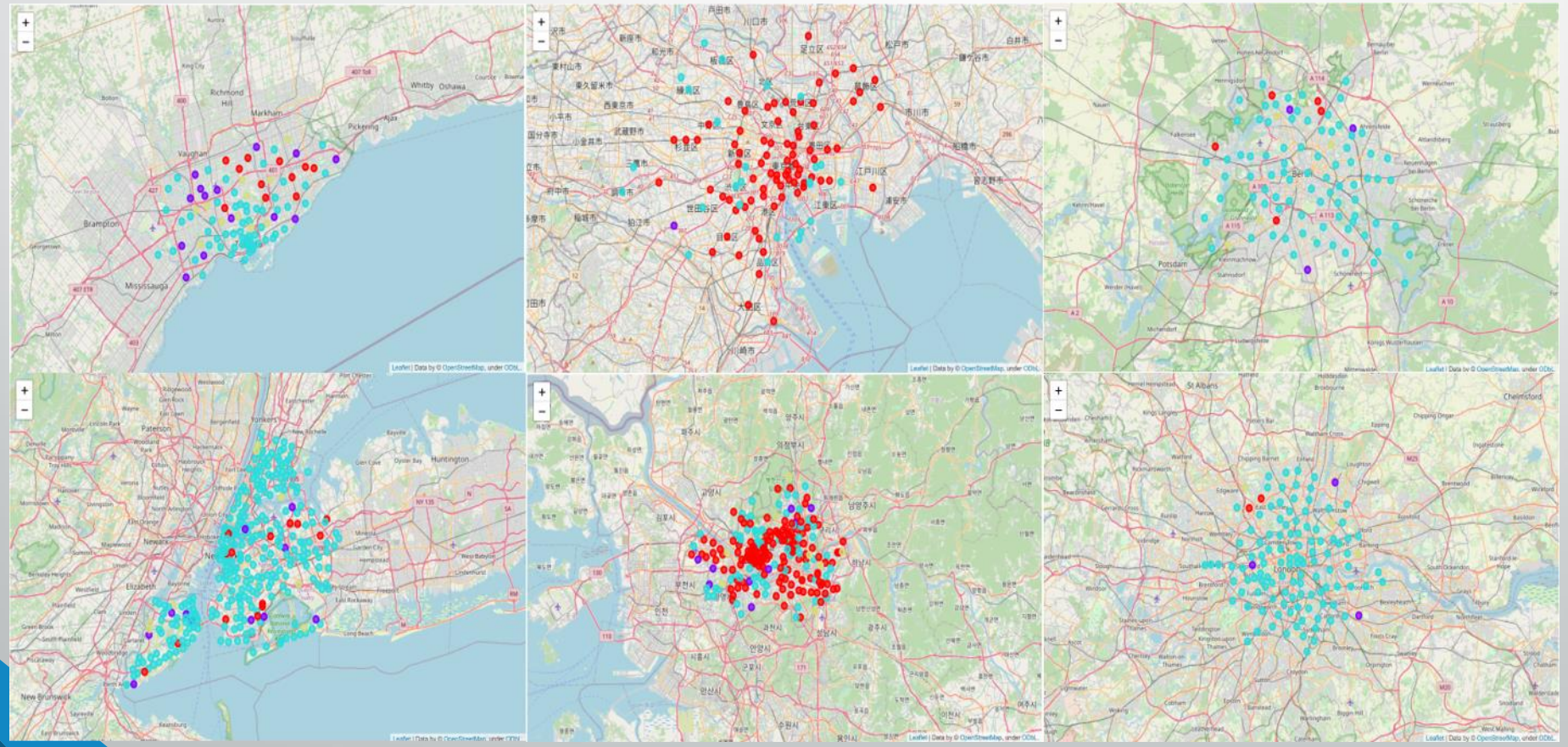
Data acquisition

- 6 Cities were selected: New York, Toronto, Tokyo, Seoul, Berlin and London
- For these 6 cities: borough and neighborhood names and coordinates were gathered through various data sources such as Wikipedia and geocoders
- Using the Foursquare API Venue data was gathered for each neighborhood and the venue categories were set at a semi-high level of specification (Not too broad e.g. Food, not too specific e.g. Korean restaurant

Neighborhoods



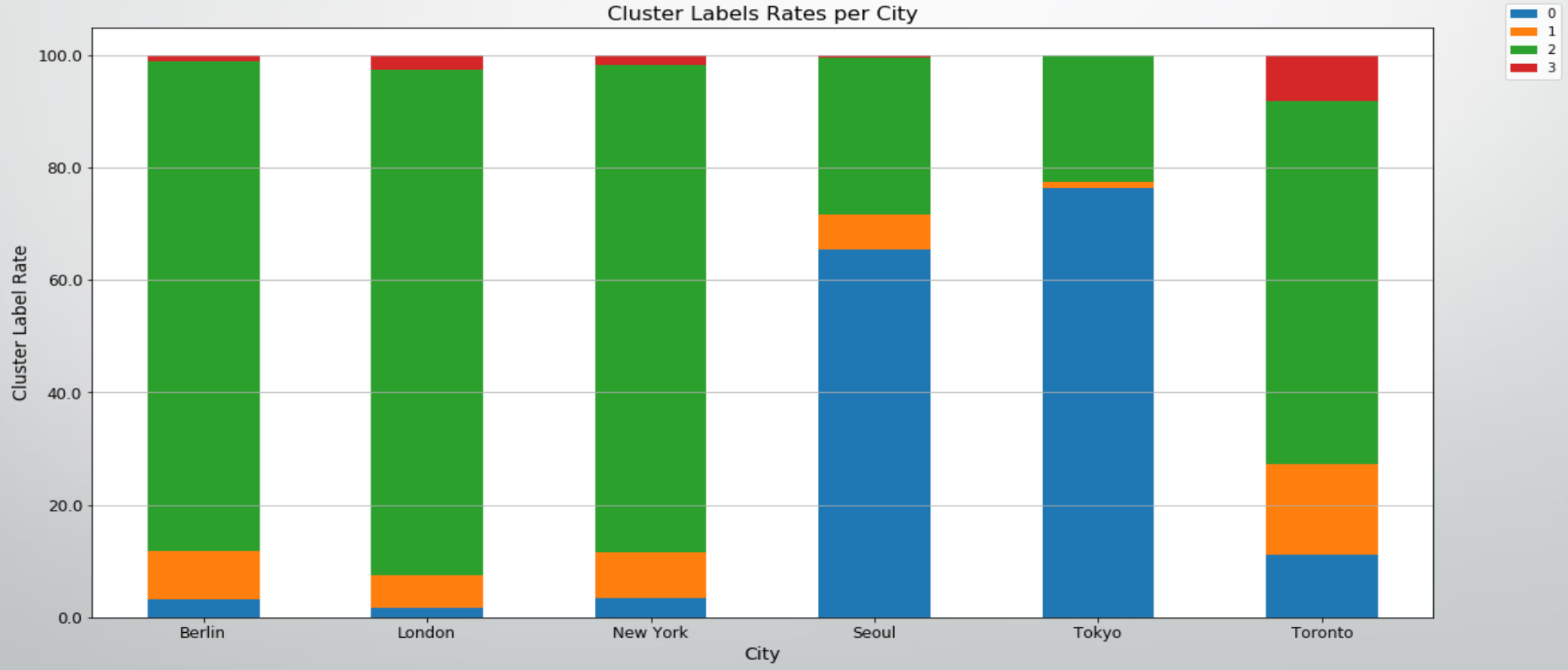
Clustering these neighborhoods based on most common venues leads to 4 distinct kinds of neighborhoods



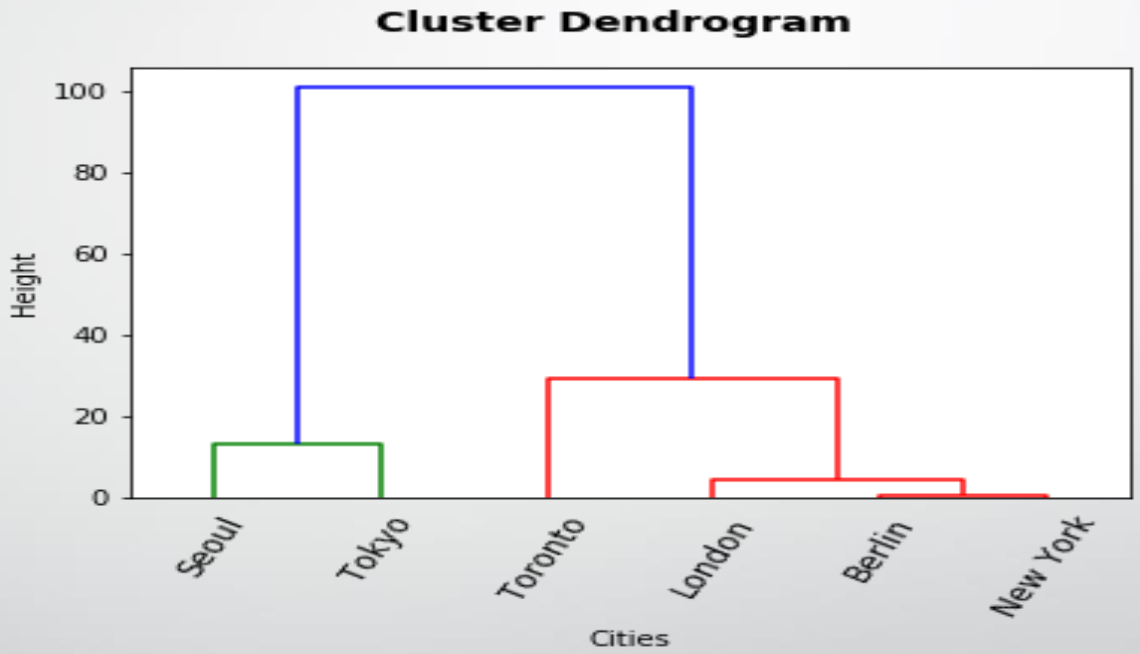
These clusters show specific types of venues

Cluster Label	Most Common Venues	Count
Cluster 3 Parks and Trails	Park	22
	Trail	4
Cluster 2 Bars and Food & Drink shops	Bar	77
	Food & Drink Shop	77
Cluster 1 Athletics & Sports and Food & Drink shops	Athletics & Sports	14
	Food & Drink Shop	4
Cluster 0 Asian Restaurants	Asian Restaurant	223
	Coffee Shop	10

Proportions of cluster labels per city show clear trends. A lot of Asian restaurants in Seoul and Tokyo would be expected.



Hierarchical clustering of cities shows expected and unexpected results



- Tokyo and Seoul are very similar in events
- The most similar city to New York is Berlin ?!
- Toronto stands out of the group of western cities

Conclusion

- Cities can be clustered based on the most common events in their neighborhoods
- Berlin would be the recommended travel spot for people who want a similar experience to New York
- This analysis could be extended to more cities to create recommendation engines