



Grouping major world cities

Using venue data to group major world cities

Grouping major world cities

- Tourists going on city trips
- Improving suggestion algorithms
- Hotel booking or airplane ticket websites
- Grouping on basis of venue data of all major world cities



Data acquisition

- Limitations of chosen data
- All world cities with over 500.000 inhabitants (<https://worldpopulationreview.com/world-cities/>)
- Geographic location data from geopy package in Python
- Venue data from the Foursquare API



Data processing

- Major world cities:

	Rank	Name	Country	2020 Population	2019 Population	Change
0	1	Tokyo	Japan	37393129	37435191	-0.11%
1	2	Delhi	India	30290936	29399141	3.03%
2	3	Shanghai	China	27058479	26317104	2.82%
3	4	Sao Paulo	Brazil	22043028	21846507	0.90%
4	5	Mexico City	Mexico	21782378	21671908	0.51%

(55642, 7)

- Geographic location data:

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Tokyo	35.682839	139.759455	Palace Hotel Tokyo (パレスホテル東京)	35.684644	139.761302	Hotel
1	Tokyo	35.682839	139.759455	Kokyo Gaien (皇居外苑)	35.679928	139.758562	Garden
2	Tokyo	35.682839	139.759455	Wolfgang's Steakhouse	35.679185	139.762134	Steakhouse
3	Tokyo	35.682839	139.759455	Aman Tokyo (アマン東京)	35.685236	139.765401	Hotel
4	Tokyo	35.682839	139.759455	Mitsubishi Ichigokan Museum (三菱一号館美術館)	35.678420	139.763260	Art Museum

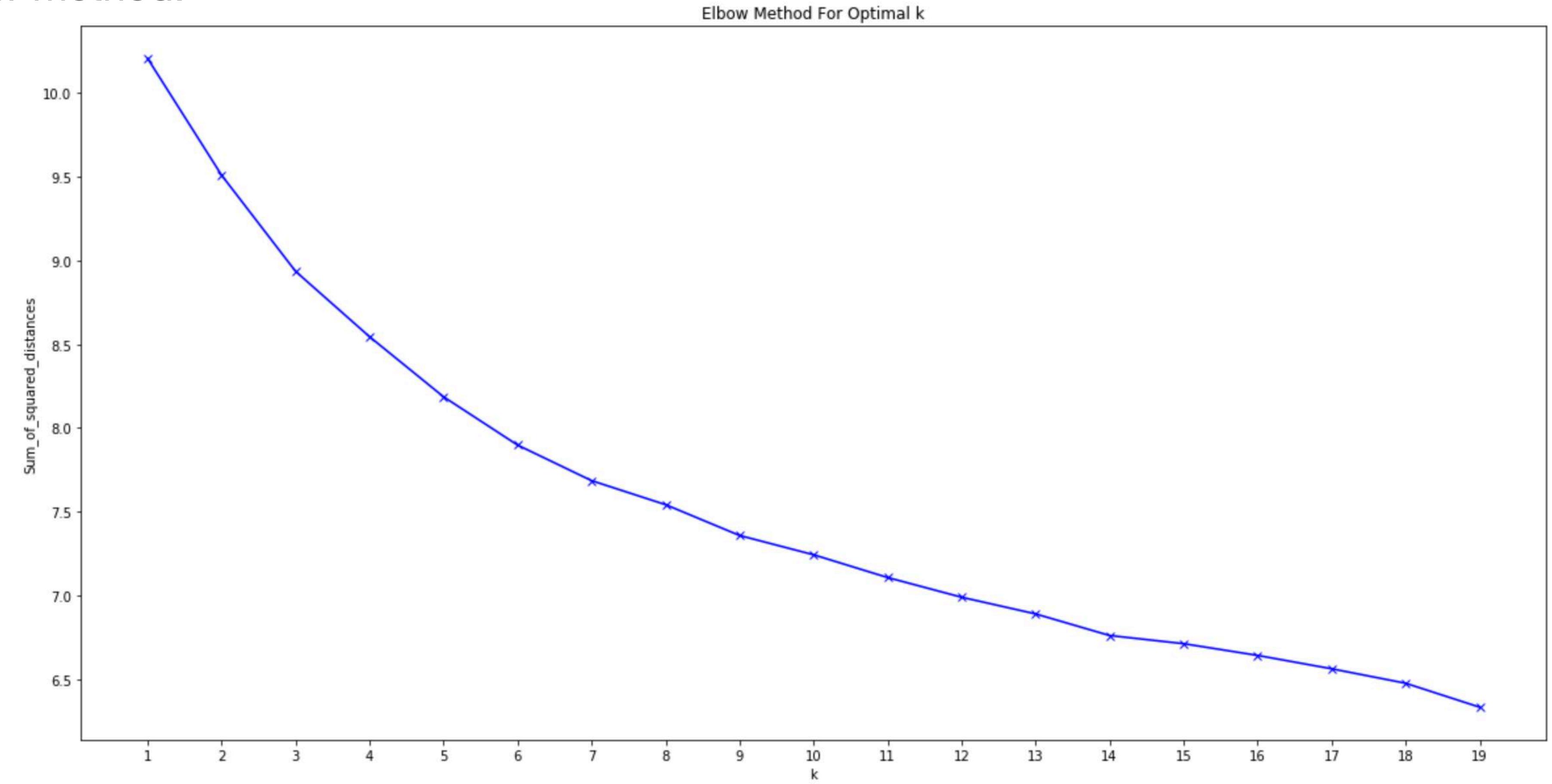
- Venue data:

	City	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	Abu Dhabi	Café	Hotel	Middle Eastern Restaurant	Coffee Shop	Chinese Restaurant	Restaurant	Park	Dessert Shop	Asian Restaurant	Beach
1	Abuja	Hotel	Restaurant	Lounge	Fast Food Restaurant	Arcade	Nightclub	Café	Pizza Place	BBQ Joint	Movie Theater
2	Acapulco De Juarez	Seafood Restaurant	Mexican Restaurant	Taco Place	Restaurant	Beach	Coffee Shop	Food Truck	Italian Restaurant	Burger Joint	Ice Cream Shop
3	Accra	Hotel	African Restaurant	Fast Food Restaurant	Bar	Cocktail Bar	Shopping Mall	Bakery	Pizza Place	Nightclub	Restaurant
4	Ad Dammam	Coffee Shop	Bakery	Juice Bar	Ice Cream Shop	Donut Shop	Café	Supermarket	Dessert Shop	Restaurant	Middle Eastern Restaurant



Data analysis

- K-means cluster analysis
- Elbow method:



Data analysis

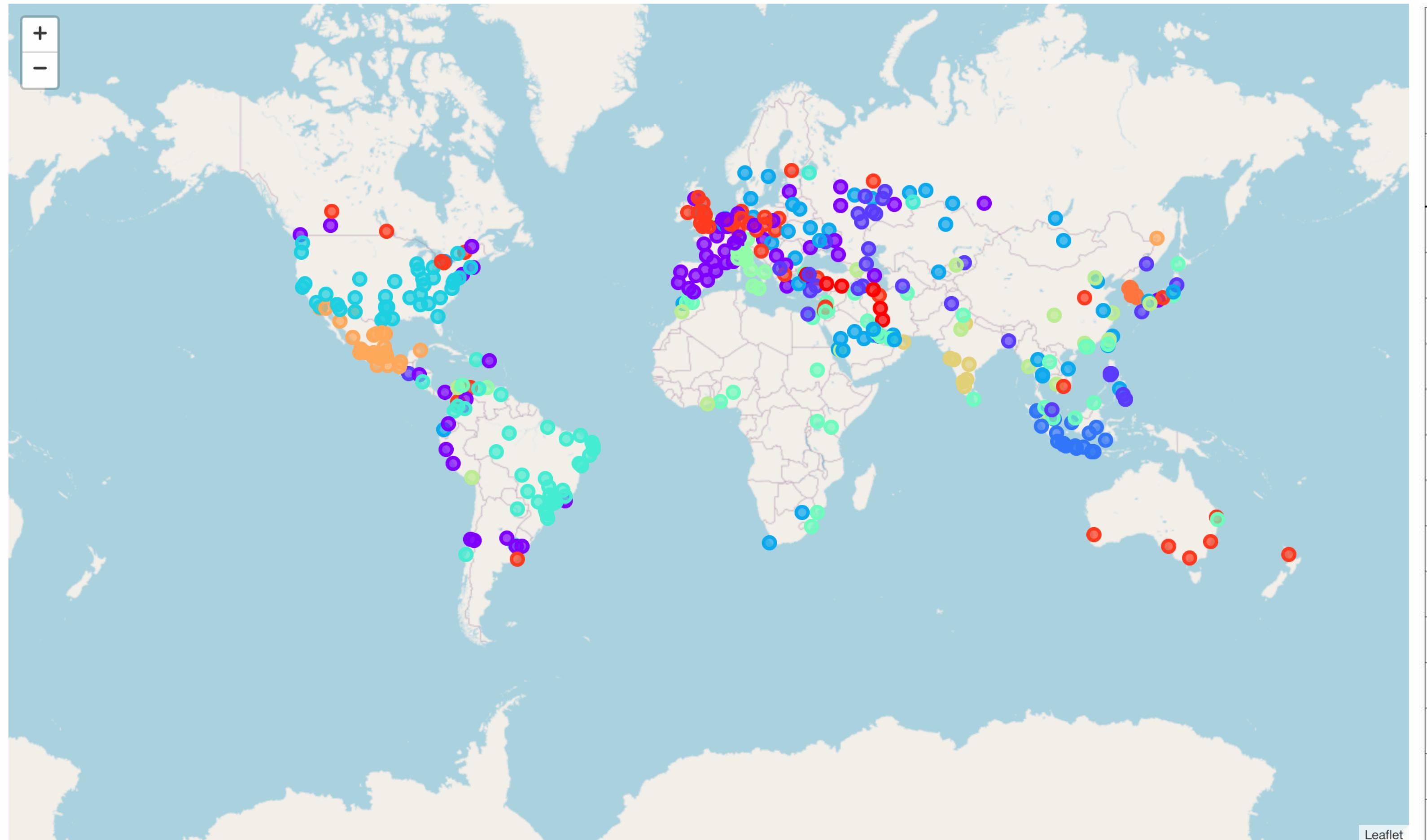
- Final clustering:

	City	Country	Population	latitude	longitude	Cluster Labels	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
0	Tokyo	Japan	37393129	35.682839	139.759455	7.0	Hotel	Coffee Shop	Japanese Restaurant	Soba Restaurant	Café	Ramen Restaurant	Japanese Curry Restaurant	Hobby Shop
1	Delhi	India	30290936	28.651718	77.221939	10.0	Indian Restaurant	Hotel	Café	Lounge	Snack Place	Bakery	History Museum	Breakfast Spot
2	Shanghai	China	27058479	31.232276	121.469207	9.0	Hotel	Coffee Shop	French Restaurant	Spa	Shopping Mall	Dumpling Restaurant	Italian Restaurant	Lounge
3	Sao Paulo	Brazil	22043028	-23.550651	-46.633382	6.0	Italian Restaurant	Ice Cream Shop	Pizza Place	Theater	Bakery	Brazilian Restaurant	Cosmetics Shop	Art Museum
4	Mexico City	Mexico	21782378	19.432630	-99.133178	11.0	Mexican Restaurant	Bakery	Art Museum	Ice Cream Shop	Plaza	History Museum	Taco Place	Deli / Bodega



Results

- Cultural grouping of cities

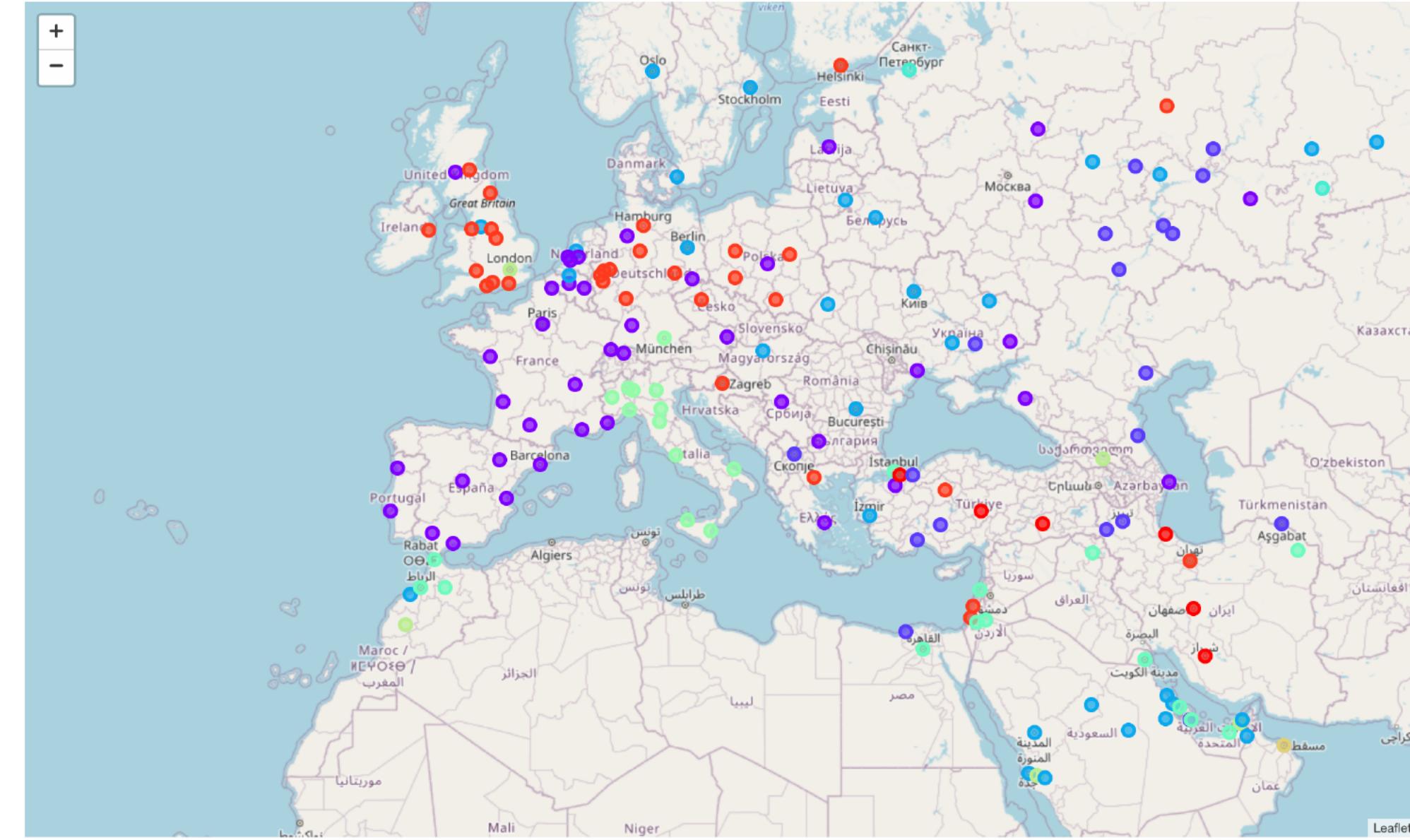
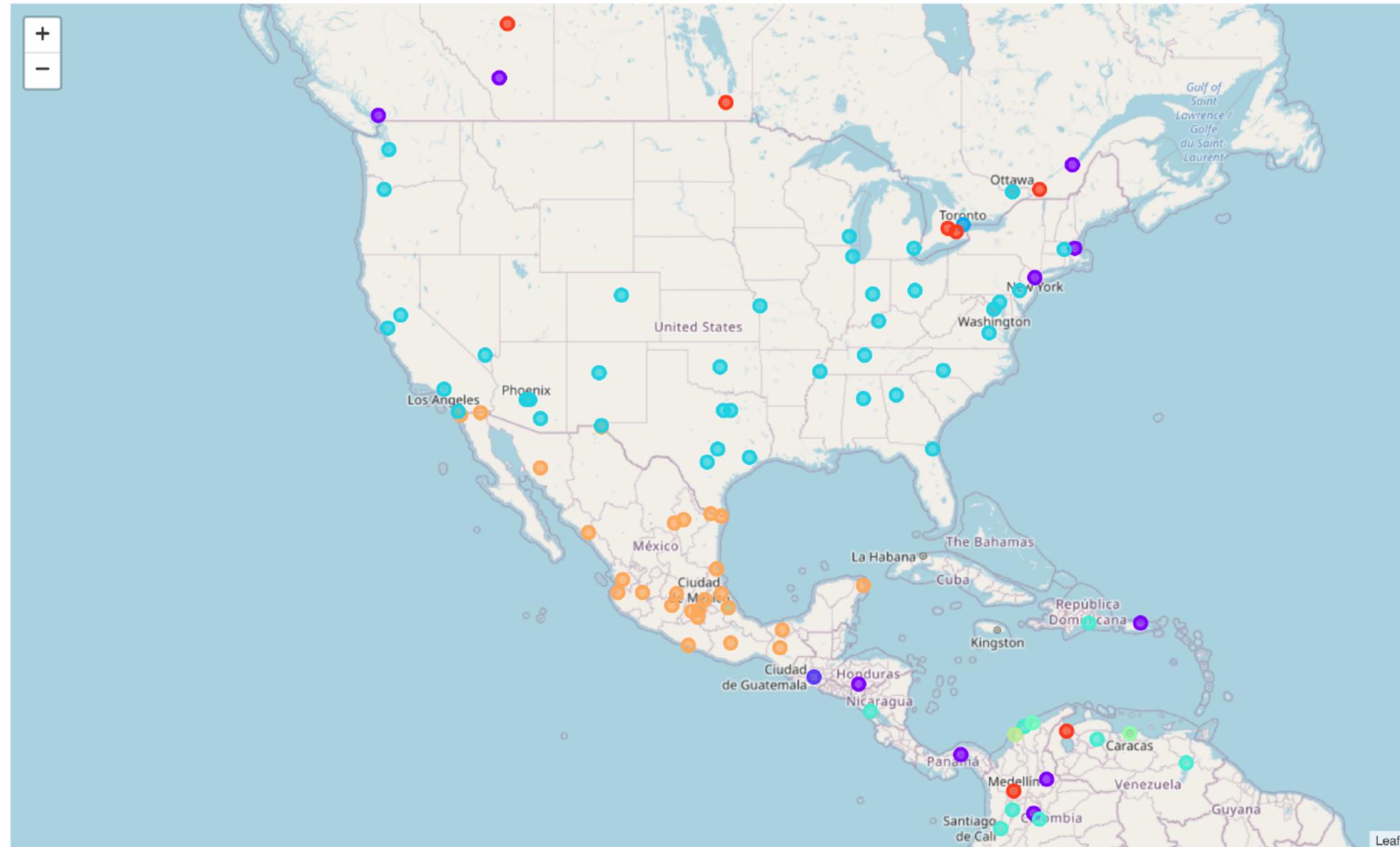


Cluster Labels	City
13	49
12	8
11	27
10	8
9	21
8	13
7	38
6	41
5	38
4	51
3	21
2	33
1	63
0	6



Results

- For example USA, Italy or Spain/France



Discussion

- Effectiveness of city grouping
- Model limitations
- Future improvements
 - *also cities with less than 500.000 inhabitants*
 - commercial Foursquare API account
 - *more computing power*



Conclusion

- Successful model
- Use as exploratory model to see the possibilities
- Room for improvement

