CITIES COMPARATOR

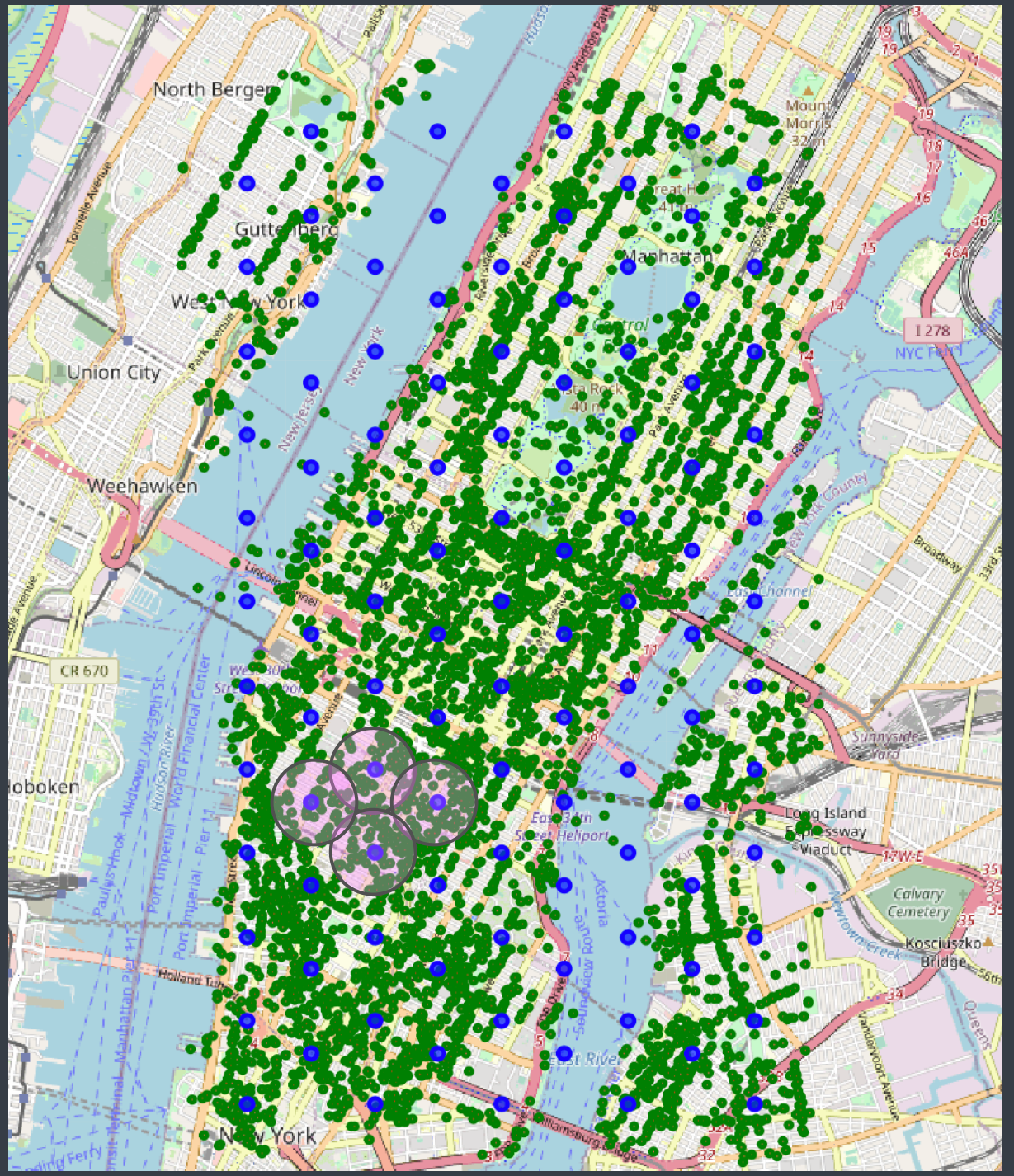
AN ANALYSIS ON CITIES SIMILARITIES AROUND THE WORLD

INTRODUCTION

With the current trend toward Globalization how cities around the world compare?
Is country characterization as strong as it was 50 years ago?
In this analysis we will compare a group of cities around the world to find if cities that are near each other are more similar than distant one or if nowadays country and culture have less importance and what matters are other factors and, if so, discover which are those factors.

HOW

- For each city in our study, we created a grid of zones.
- Each zone is a circle with a center (blue dot) and a radius (in pink in the image).
- For each zone, using Foursquare API, we fetched all the venues/center of interest present in the zone (the green dots).
- Each Venue belongs to a Category (French Restaurant, Museum, Pizza place, etc.)
- The number of zones inquired and the dimension of each zone are the same for all cities



ZONE CATEGORIZATION

Using the Categories of the venues in a zone we created a profile.
The profile is a list of the numbers that for each venue category tells how many venues of that category are present in the zone. Then using these profiles we have grouped the zones in 10 Zone Types using **Kmeans** clustering algorithm.

zone type: 0		zone type: 1		zone type: 2	
Coffee Shop	531.0	Mexican Restaurant	234.0	Italian Restaurant	541.0
Italian Restaurant	318.0	Taco Place	177.0	Pizza Place	282.0
Pizza Place	307.0	Bakery	47.0	Café	257.0
Bar	300.0	Coffee Shop	43.0	Hotel	231.0
Café	282.0	Seafood Restaurant	43.0	Ice Cream Shop	174.0
Bakery	257.0	Restaurant	41.0	Plaza	139.0
Hotel	256.0	Convenience Store	38.0	Restaurant	114.0
Gym / Fitness Center	211.0	Ice Cream Shop	35.0	Cocktail Bar	101.0
American Restaurant	198.0	Pizza Place	35.0	Japanese Restaurant	82.0
Sandwich Place	173.0	Bar	30.0	Bakery	75.0

CITY CATEGORIZATION

Using the number of each zone type within a city we clustered, using always kmeans , the cities in homogeneous groups.
Here is the result using 9 as number of clusters.
The columns with number from 0 to 9 indicates the zone types and the values in those columns the number of zone of that type present in the city. The column city_type contains the cluster in which the city has been placed. So Tokyo is alone in group 0, Barcelona and Madrid are in group 2 and so on

[33]:

city	0	1	2	3	4	5	6	7	8	9	city_type
tokyo	0	0	0	2	65	2	0	0	0	31	0
los_angeles	8	3	0	1	0	0	0	0	0	87	1
washington	28	0	0	0	0	2	1	0	0	72	1
boston	24	0	1	0	0	6	0	1	0	75	1
dallas	6	0	0	0	0	1	0	0	0	88	1
moscow	11	0	0	2	0	4	9	1	0	72	1
phoenix	3	0	0	0	0	2	0	0	0	94	1
philadelphia	16	0	0	1	0	1	1	0	0	81	1
barcelona	0	0	0	4	0	1	0	0	30	63	2
madrid	0	0	0	1	0	2	0	0	39	57	2
milan	0	0	42	1	0	4	0	0	0	57	3
rome	0	0	38	1	0	1	0	0	0	58	3
sao_paulo	26	0	0	2	1	17	4	0	0	54	4
london	25	0	0	15	0	6	6	0	1	51	4
new_york	48	0	0	1	0	12	0	7	0	40	4
munich	0	0	4	20	0	1	1	0	0	76	5
berlin	0	0	1	33	0	2	4	0	0	61	5
mexico_city	0	42	0	0	0	2	0	0	0	60	6
istanbul	0	0	0	0	0	3	41	13	0	47	7
paris	0	0	1	55	0	7	0	0	0	37	8

CONCLUSIONS

- We found that cities with same culture/ethnicity tends to be grouped together, i.e. are more similar using our comparison method.
- This is probably due also to the source of our data. Foursquare venue categories are very (too much?) detailed. Especially restaurants are categorized by cuisine type and the distribution of restaurants types depends largely on the nation the city is.