CITY COMPARISON

A TOOL TO COMPARE DIFFERENT CITIES BASED ON THEIR MOST COMMON VENUES

INTRODUCTION

Three cities have been considered:

- Toronto
- New York
- Rome

Are these cities similar? Are there any evident difference? How do their neighborhoods look like based on their most trending venues?

DATA

2 main sources of data:

- For each city, consider the city center or downtown neighborhoods. List them and retrieve their geo coordinates
- For each Neighborhood, explore the venues nearby using the Foursquare API

	City	Neighborhood	Abruzzo Restaurant	Accessories Store	Adult Boutique	Afghan Restaurant	Airport Lounge	Airport Service	American Restaurant	Animal Shelter	Antique Shop	Arepa Restaurant	Arge Rest
0	New York	Battery Park City	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.032258	0.00	0.00	0.00	0.(
1	New York	Carnegie Hill	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00	0.(
2	New York	Central Harlem	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.047619	0.00	0.00	0.00	0.(
3	New York	Chelsea	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00	0.(
4	New York	Chinatown	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.020408	0.00	0.00	0.00	0.(
5	New York	Civic Center	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00	0.(

METHODOLOGY I

- All the results obtained by the foursquare API are put together in a unique dataframe
- Thanks to get dummies, we are able to see which venue is present in which neighborhood and their frequency

	Neighborhood	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	Battery Park City	Park	Department Store	Cupcake Shop	Food Court	Sandwich Place	Pharmacy	Coffee Shop	Smoke Shop	Gastropub	Salad Place
1	Carnegie Hill	Café	Gym / Fitness Center	Italian Restaurant	Gym	Spa	Dance Studio	Deli / Bodega	Korean Restaurant	Sports Bar	Pizza Place
2	Central Harlem	Cosmetics Shop	Breakfast Spot	Bagel Shop	Cycle Studio	Café	French Restaurant	Fried Chicken Joint	Music Venue	Lounge	Caribbean Restaurant
3	Chelsea	Mexican Restaurant	Nightclub	Café	Hotel	Speakeasy	Cupcake Shop	Asian Restaurant	Liquor Store	Coffee Shop	Event Space
4	Chinatown	Bubble Tea Shop	Chinese Restaurant	Japanese Restaurant	Hotpot Restaurant	Sandwich Place	Vietnamese Restaurant	Korean Restaurant	Spa	Noodle House	Record Shop

METHODOLOGY 2

- We can also find the most common venues for each neighborhood:

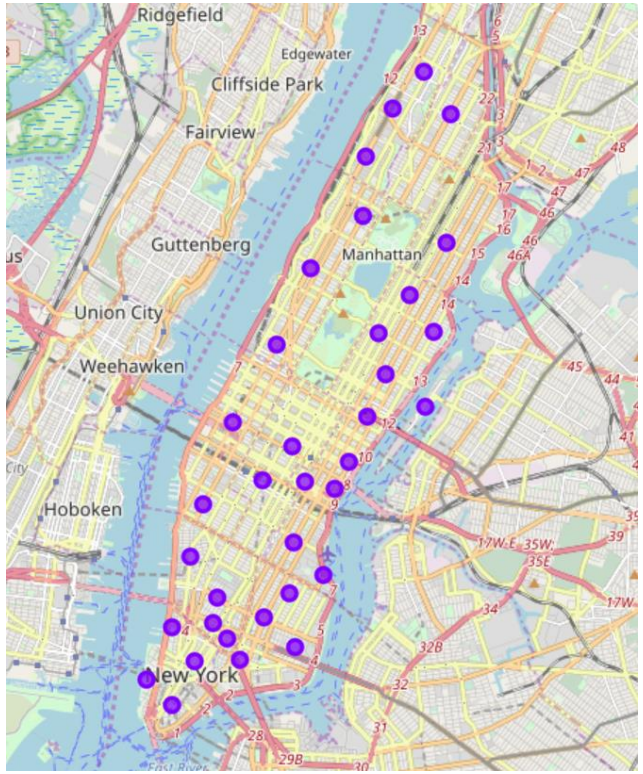
CLUSTERING

- K mean is used to cluster all the neighborhoods

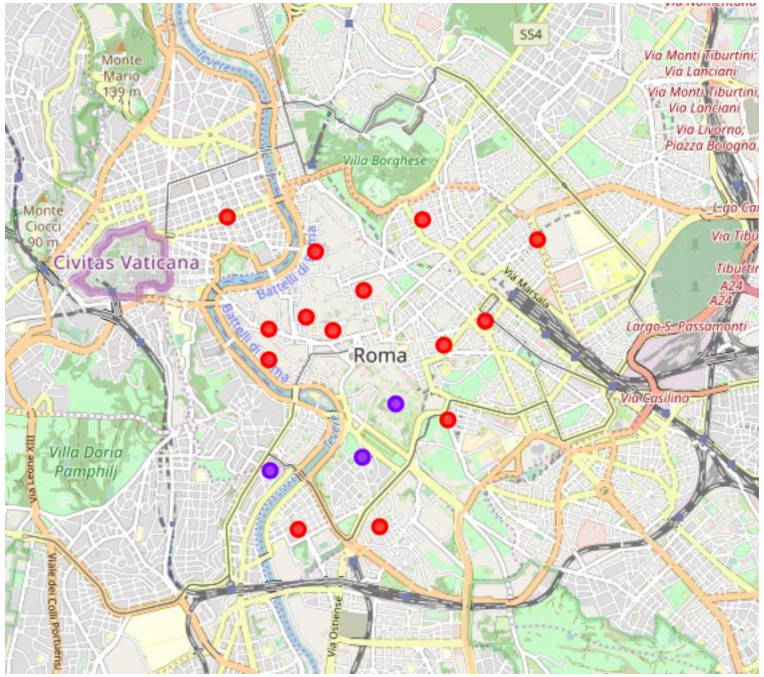
	city	Neighbourhood	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	New York	Marble Hill	40.876551	-73.910660	0.0	Sandwich Place	Pizza Place	Deli / Bodega	Pharmacy	Supermarket
1	New York	Chinatown	40.715618	-73.994279	1.0	Bubble Tea Shop	Chinese Restaurant	Spa	Korean Restaurant	Hotpot Restaurant
2	New York	Washington Heights	40.851903	-73.936900	1.0	Café	Chinese Restaurant	Mobile Phone Shop	Deli / Bodega	Park
3	New York	Inwood	40.867684	-73.921210	1.0	Café	Pizza Place	Restaurant	Frozen Yogurt Shop	Bakery
4	New York	Hamilton Heights	40.823604	-73.949688	1.0	Mexican Restaurant	Yoga Studio	Cocktail Bar	Deli / Bodega	Coffee Shop
5	New York	Manhattanville	40.816934	-73.957385	1.0	Coffee Shop	Bank	Bus Station	Climbing Gym	Gas Station
6	New York	Central Harlem	40.815976	-73.943211	1.0	Cosmetics Shop	Juice Bar	Deli / Bodega	Dessert Shop	Ethiopian Restaurant

RESULTS

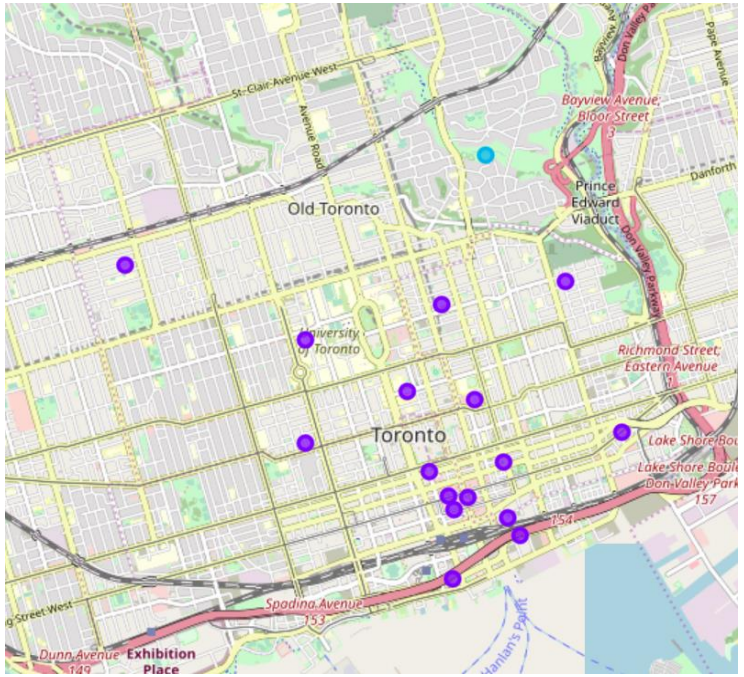
- We can now assign a different color for each cluster (in our case 5) and display them in the city map



New York



Rome



Toronto

CONCLUSION

- As we could easily expect, New York and Toronto are very similar, while Rome appears to be much different from the other two
- K means gave trustworthy results in identifying the differences between the cities

- Future works can focus on including many more cities in order to see the differences and similarities, and identify for example which neighbourhood has a more European vibe or a more Asian vibe or rather international.
- Another further application could include a Machine Learning algorithm to classify a new neighbourhood give the data of many other as starting point database