

# Moving from Rome to Toronto

*Using Machine Learning to recommend neighbourhoods for relocation*

# Rome and Toronto

- Since both are quite large and important cities, neighborhoods within those cities are quite different from each other.
- As an example, a person from Centocelle (Rome) would adjust better in Harbourfront East (Toronto) than in Summerhill East (Toronto), where he/she can enjoy a similar density of restaurants and cinemas; while Summerhill may be better for someone from Trastevere (Rome), where he/she can appreciate the green area.
- The purpose of this project is to use Foursquare location data to identify the types of venues within each neighborhood and calculate the "similarity" a Roman neighborhood will have to a Toronto neighborhood.
- At the end of the project, the person from Rome who needs to relocate will type his/her neighborhood and he/she will be provided with a ranking of the "less distant" Toronto neighborhoods from his/her own.

# Data acquisition and cleaning

- Canada postcode available at [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Canadian postcode latitude and longitude available at [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)
- Italian postcode latitude and longitude available at <http://download.geonames.org/export/zip/>
- Roman neighborhoods available at [http://www.roma-o-matic.com/zoom.php3?actions\[Z\]=x](http://www.roma-o-matic.com/zoom.php3?actions[Z]=x)
- Data have been grouped together to obtain one row per neighbourhood with latitude and longitude, for a total of 123 observations
- Foursquare location data

# The Kmeans algorithm

- The goal is to obtain a square matrix of euclidean distance across neighbourhoods.
- This can be obtained easily by using an initial number of clusters (K) equals to the number of neighbourhoods.
- During the execution, it became evident that few roman neighborhoods are so similar to each other that the algorithm can not separate them into different cluster (same distance with any other cluster).
- An example of the above consists of "Torre Angela" and "Borghesiana", two neighborhoods next to each other in Rome that can be reasonably grouped into a single neighbourhood.

- **Procedure for the choice of number of Cluster (K)**
- 1. K=123 (number of neighbourhoods) has been chosen as a starting K
- 2. If the Kmeans algorithm is not able to converge because of the existence of too similar neighbourhood, the k is decrease by 1
- 3. The loops stops when the algorithm is able to converge.
- Starting from 123, the final k has converged into 107 clusters and few Roman neighbourhoods have been grouped together

# Recommendation matrix with euclidean distances

Aurelio (suburbio)	Trastevere	Settecamini	Prima Porta	Lido di Ostia Levante,Tor di Valle,Casal Palocco,Maccarese Sud	Torricola	Casal Morena,Torre Maura,Capannelle,Appio Claudio	Trevi	...	St. James Town,Cabbagetown	Ryerson,Garden District	Sant'Angelo	Esquilino	Castro Pretorio	Berczy Park	Underground city,First Canadian Place	Commerce Court,Victoria Hotel	Neighbourhood	City
1.016415	0.618951	1.006529	1.006529	0.446331	1.016415	0.522590	0.268514	...	0.179072	0.157162	0.314780	0.313528	0.343074	0.167433	0.138924	0.136748	Harbourfront East,Toronto Islands,Union Station	Toronto
1.172604	0.866025	1.172604	1.172604	0.485913	1.172604	0.612372	0.630872	...	0.602166	0.616604	0.698181	0.654064	0.699285	0.633373	0.624179	0.634508	Torre Angela,Borghesiana	Rome
0.000000	1.172604	1.414214	1.414214	1.092906	1.414214	1.118034	1.040673	...	1.016395	1.012522	1.054731	1.054893	1.064894	1.012996	1.017153	1.018627	Aurelio (suburbio)	Rome
1.172604	0.000000	1.172604	1.172604	0.697217	1.172604	0.790569	0.676757	...	0.602166	0.628649	0.698181	0.683959	0.706399	0.619369	0.640000	0.634508	Trastevere	Rome
1.414214	1.172604	0.000000	1.414214	1.092906	1.414214	1.118034	1.040673	...	1.016395	1.012522	1.054731	1.054893	1.064894	1.012996	1.017153	1.018627	Settecamini	Rome

# Results

- Taking Tuscolano as an example

	Tuscolano	Neighbourhood	City
99	0.211368	Ryerson,Garden District	Toronto
37	0.217007	Adelaide,Richmond,King	Toronto
94	0.223483	St. James Town,Cabbagetown	Toronto
103	0.224318	St. James Town	Toronto
105	0.227741	Design Exchange,Toronto Dominion Centre	Toronto

- The closest place to Tuscolano would be Ryerson,Garden District.

- **Additional statistics:**
- ***Most similar couple of roman/Toronto neighbourhoods:*** Monti and Design Exchange,Toronto Dominion Centre with a distance of 0.20815036178207
- ***Most dissimilar couple of roman/Toronto neighbourhoods:*** Aurelio (suburbio) and Roselawn with a distance of 1.4142135623731
- ***The distance average*** between Toronto and Rome is 0.542667019235158

## Conclusion

- By using Foursquare location data, Rome and Toronto geo data, the kmeans cluster algorithm and the euclidean distance, it has been possible to create a recommendation matrix to help relocate people from Rome to Toronto (or the other way around) based on the neighbourhood of origin.
- This kind of data is really easy to obtain, meaning that this analysis can be replicated for any city