



Best Neighborhoods to open an Italian Restaurant in Toronto

David Mateo Cangrejo



Business Problem

- The problem consist in identifying **the optimum place to open** an Italian restaurant in the city of Toronto, Canada.
- The target audience of this project are entrepreneurs of Italian food or owners of Italian restaurants.
- More specifically, stakeholders who are interested in identifying the best Neighborhood candidates in Toronto to open a new Italian restaurant, so with the best group of candidates decide finally where to open it.



Background



Some **key** aspects to take into account for this analysis are:

- Parking.
- Visibility.
- Number of people who can pass near the restaurant.
- Income of the Neighborhood.
- Presence of near similar restaurants.
- Presence of other business that can attract people to the restaurant:
Stadiums, parks, theaters, medical majors.



Data Description

For this analysis the following data is going to be used:

- List of Neighborhoods in Toronto: from [The City of Toronto's Open Data Portal](#).
- Latitude and longitude coordinates of neighborhoods: from [The City of Toronto's Open Data Portal](#).
- Venues near every neighborhood: from Foursquare Api.
- Population and income of each neighborhood: from [The City of Toronto's Open Data Portal](#)

The City of Toronto's Open Data Portal is an open source delivery tool to bring people and data together.



Data Usage

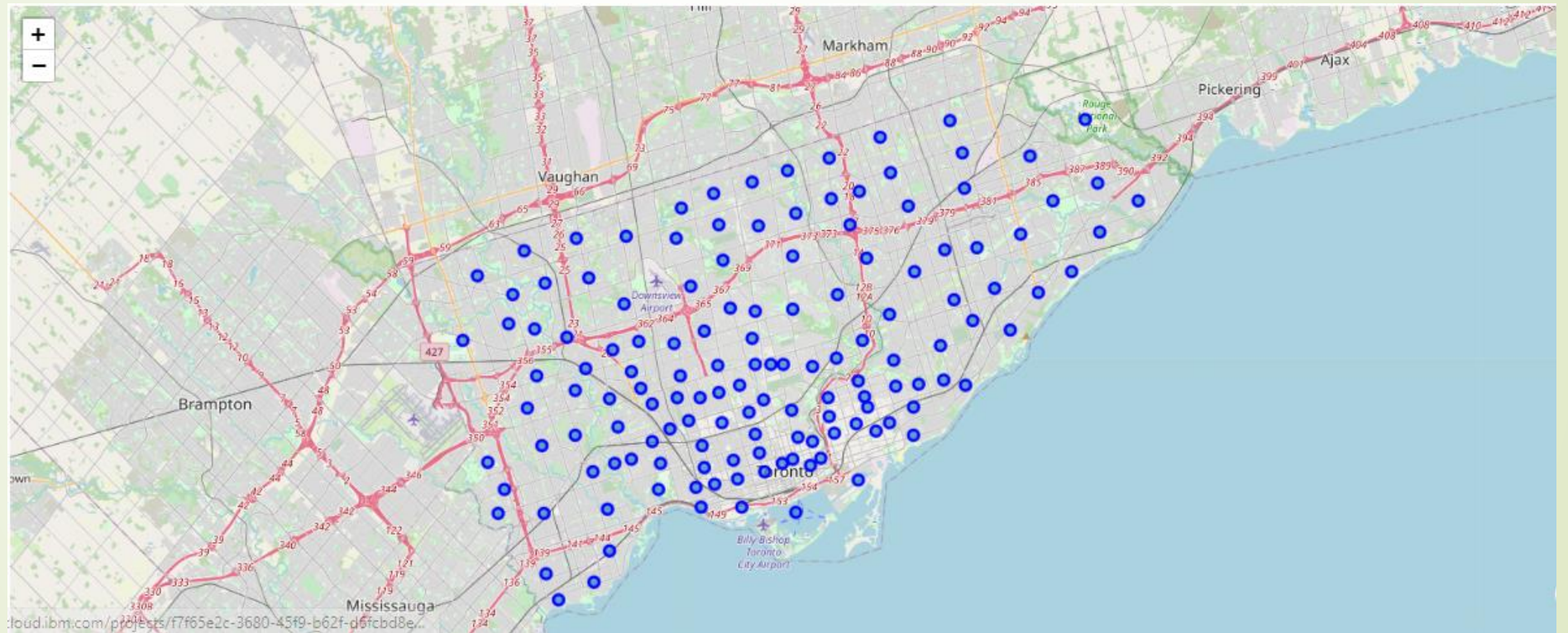


The data is going to be used in the following way to solve the problem:

- The list of Neighborhoods and its coordinate's data are going to be merged to identify the location of each neighborhood.
- The population and income data of each neighborhood is going to be also merged with the previous data.
- The venues data of every neighborhood is going to be classified in the following way: Parking and presence of other business that can attract people to the restaurant (like stadiums, theaters, medical majors) are going to be count as a "Collaborator Index" and similar restaurants (Italian food) are going to be count as a "Competitor index".
- Finally, K-mean machine learning will be used to cluster the neighborhoods with this features: **Population**, **Income**, **Collaborator index** and **Competitor index**.

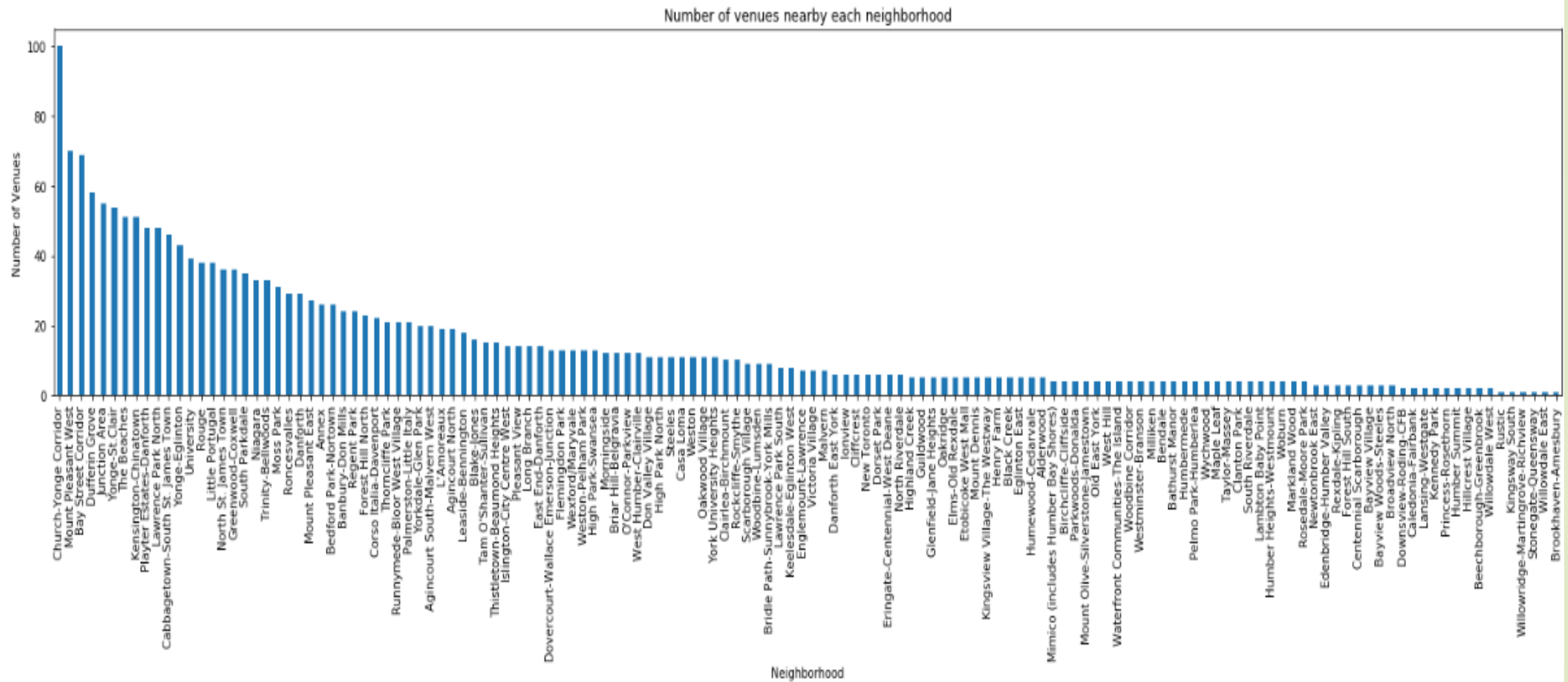
Toronto Neighborhoods map

With location dataset



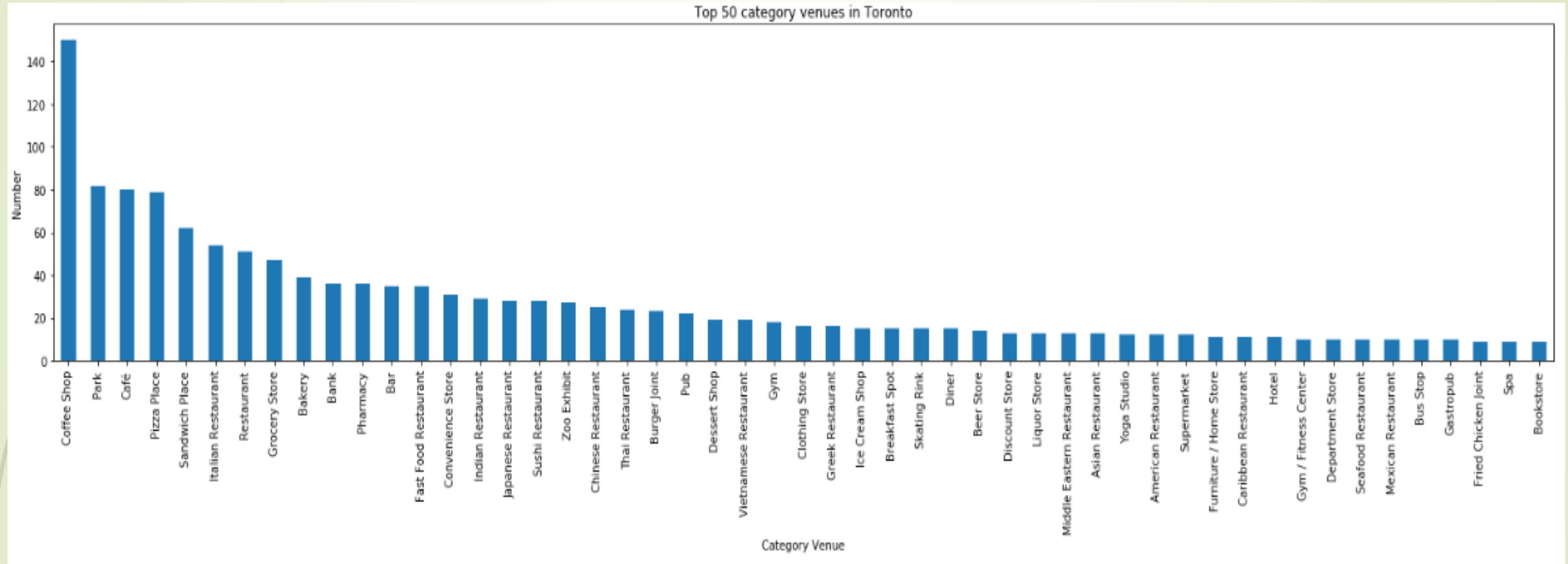
Exploratory analysis

Number of venues nearby each neighborhood



Exploratory analysis

Top 50 Category venues in Toronto



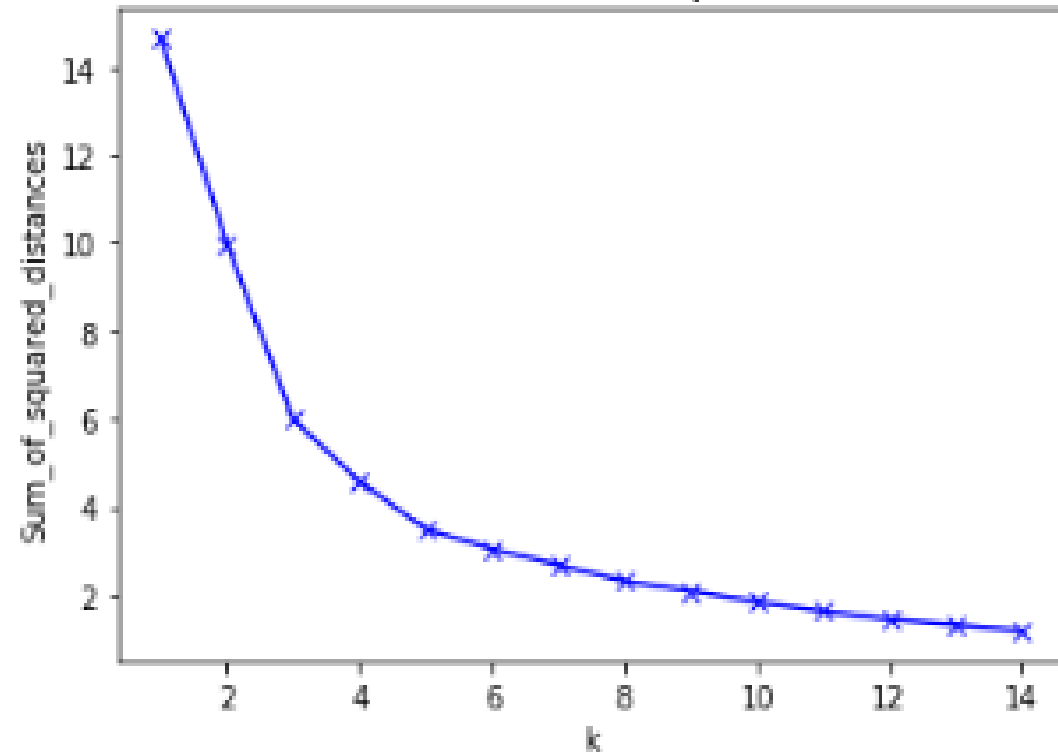


Competitor's and collaborator's indices

- Competitor's index: if it is a restaurant +1 point; and if it is an Italian Restaurant +10 points. We try to penalize a lot higher if it is an Italian Restaurant. As far as this index increase, the sector is very competitive for an Italian Restaurant.
- Collaborator's index: Every venues which is not a Restaurant we give +1 point. As far as this index increase, the sector have more venues to attract more people, which is good for the restaurant.

Clustering

Elbow Method For Optimal k

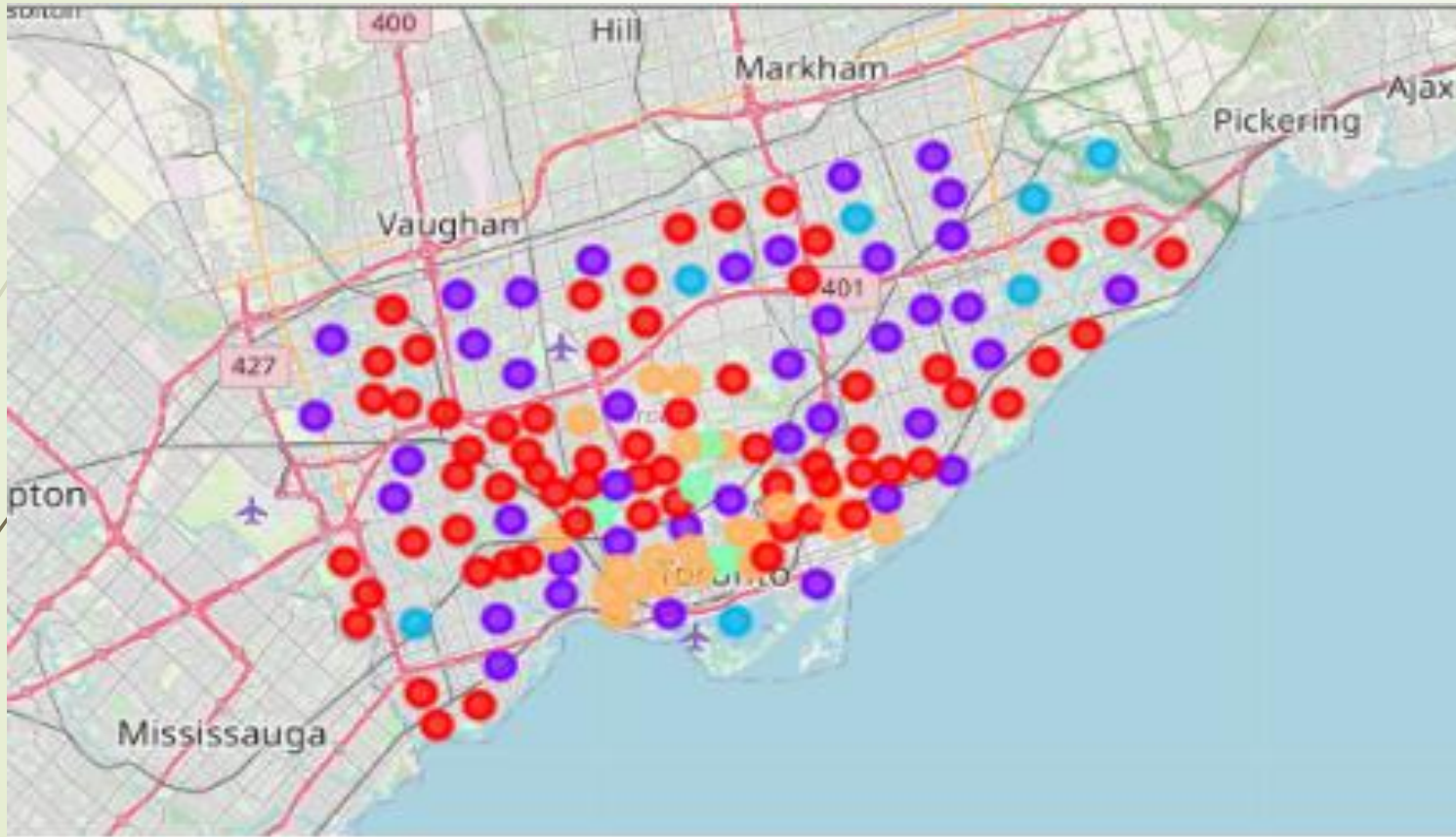


Best k=5

Clustering Dataset (Head)

	Competitors	Collaborators	Population	Income
Neighborhood				
Agincourt North	5	14	29113	25005
Agincourt South-Malvern West	15	5	23757	20400
Alderwood	0	5	12054	10265
Annex	4	22	30526	26295
Banbury-Don Mills	12	22	27695	23410

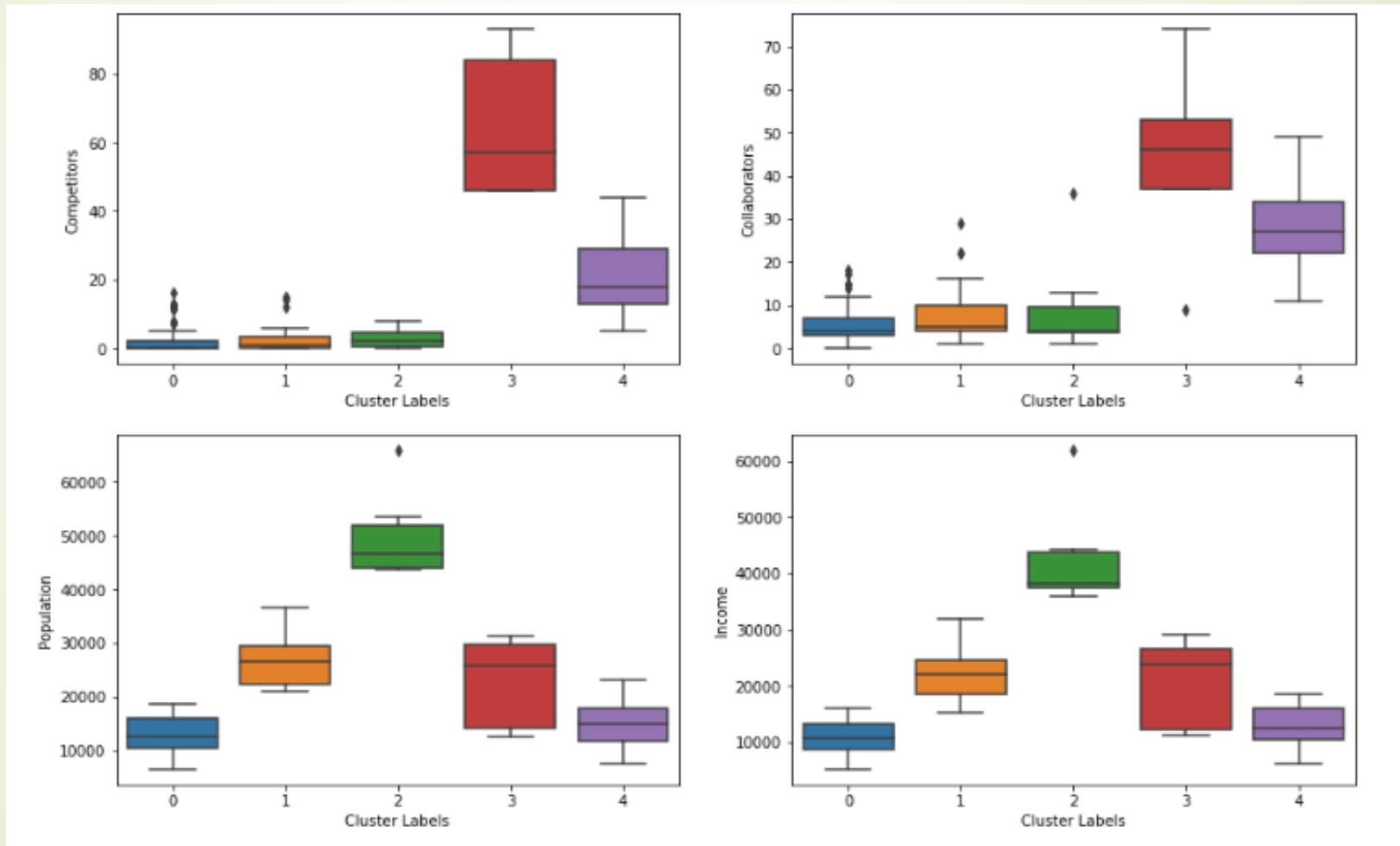
Cluster Toronto's Map



*Cluster identification: 0-Red / 1-Purple / 2-blue
/ 3-Green / 4-Orange*

Results

Cluster Examination





Discussion

Cluster Analysis

- ✓ Cluster 0: Low Competitors, low collaborators, low population and low income.
- ✓ Cluster 1: Low Competitors, low collaborators, medium population and medium income.
- ✓ Cluster 2: Low Competitors, low collaborators, high population and high income.
- ✓ Cluster 3: High Competitors, high collaborators, medium population and medium income.
- ✓ Cluster 4: Medium Competitors, medium collaborators, low population and low income.



Conclusion

As a summary we obtain the following groups:

- Good neighborhoods for renowned restaurants.
- Good neighborhoods for new restaurants.
- Good neighborhoods for experienced stakeholders.
- Bad places for a restaurant (for 2 clusters).

The model may be improved in the following aspects:

- ❑ Optimizing the functions of Collaborator's and Competitor's index: like giving a higher score to a stadium and a lower to a coffee shop, because the first one attracts more people.
- ❑ Adding more features: like customers scores of the existing restaurants and visibility of the places.