THE BATTLES OF NEIGHBORHOODS

Potential Neighborhood for a New Restaurant, Toronto

Introduction

Investors need to:

- Decide where to open new restaurants in Toronto, Canada.
- Know who are the competitors.
- Know each neighborhood in Toronto to decide the perfect place.
- And to make profit!

Evaluation Criteria:

It has to be a good recommendation in terms of profit and investors expectations.

Data

Data used in this project:

- ❑ Neighborhood Information (i.e. name, coordinates, populations, income).
- ❑ Restaurants information (i.e. name, category, coordinates)

Data Sources:

- www.foursquare.com
- www.toronto.ca

Data Processing

✓ Data cleaning

✓ Data merging between the source codes

✓ Features Extraction

	Population	Income	Longitude	Latitude
Islington-City Centre West	43965	52787	-79.543317	43.633463
Malvern	43794	29573	-79.222517	43.803658
Dovercourt-Wallace Emerson-Junction	36625	39740	-79.438541	43.665677
Downsview-Roding-CFB	35052	34168	-79.490497	43.733292
Parkwoods-Donalda	34805	42516	-79.330180	43.755033
Mimico	33964	54438	-79.500137	43.615924
West Humber-Clairville	33312	31771	-79.596356	43.716180
Mount Olive-Silverstone-Jamestown	32954	26548	-79.587259	43.746868
Church-Yonge Corridor	31340	53583	-79.379017	43.659649
Niagara	31180	70623	-79.412420	43.636681

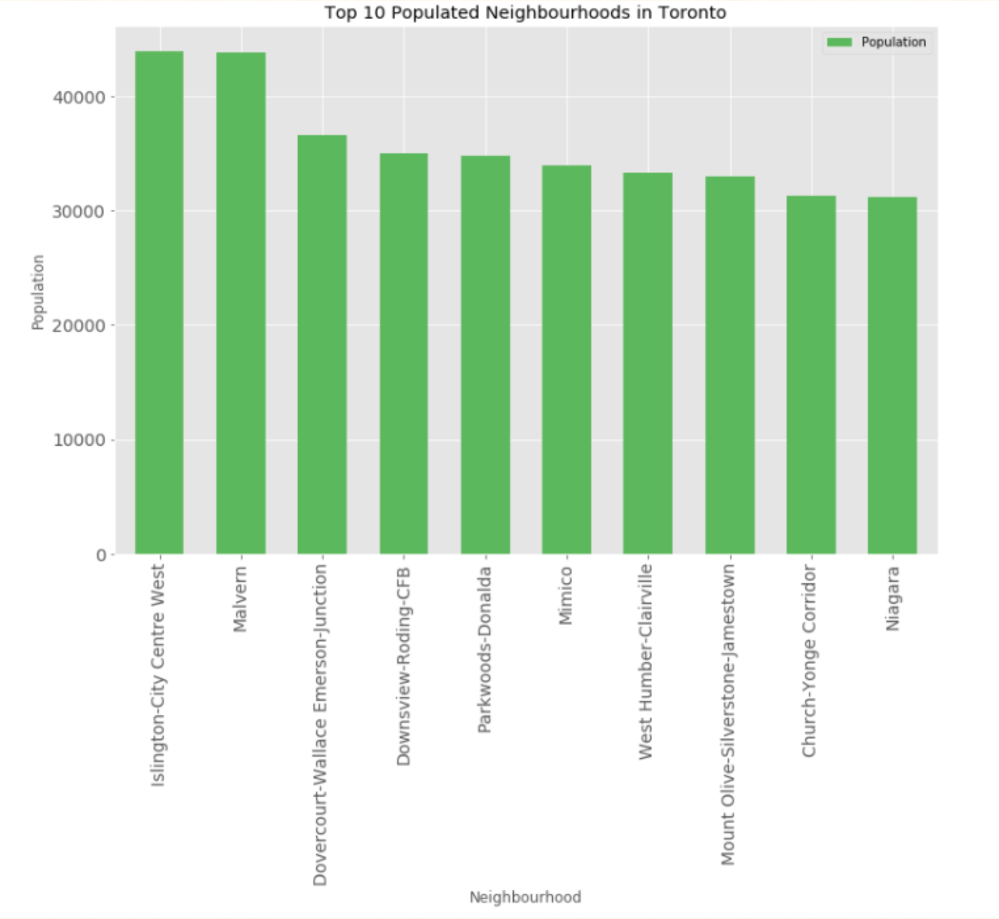
Methodology

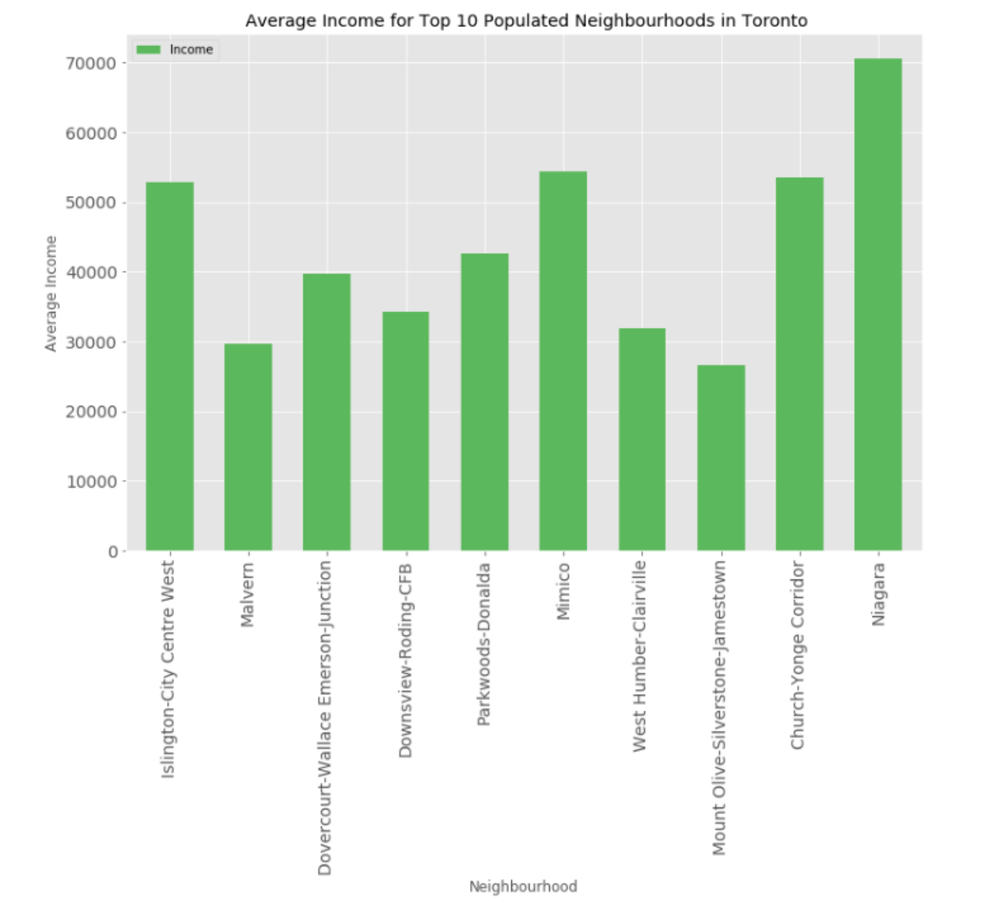
1. Extracting Neighborhoods' data from Toronto.ca
2. Retrieving restaurants information per neighborhood from FourSquare API.
3. Clustering similar neighborhoods based on total scoring using K-Means Clustering

4. Total scoring per neighborhood based on population, income, and number of restaurants.

Result (1)

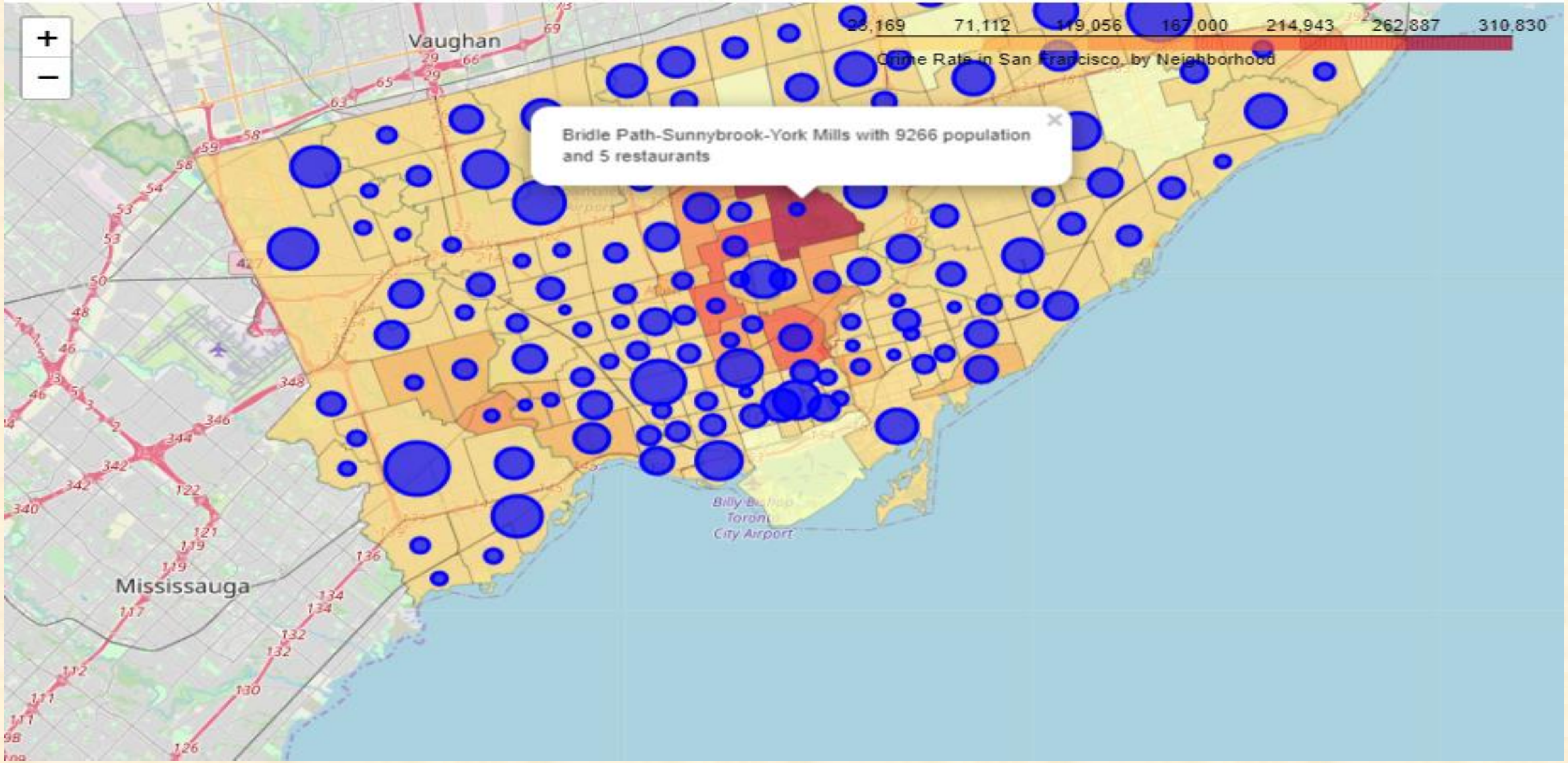
Getting top neighborhoods based on population and comparing their average income.





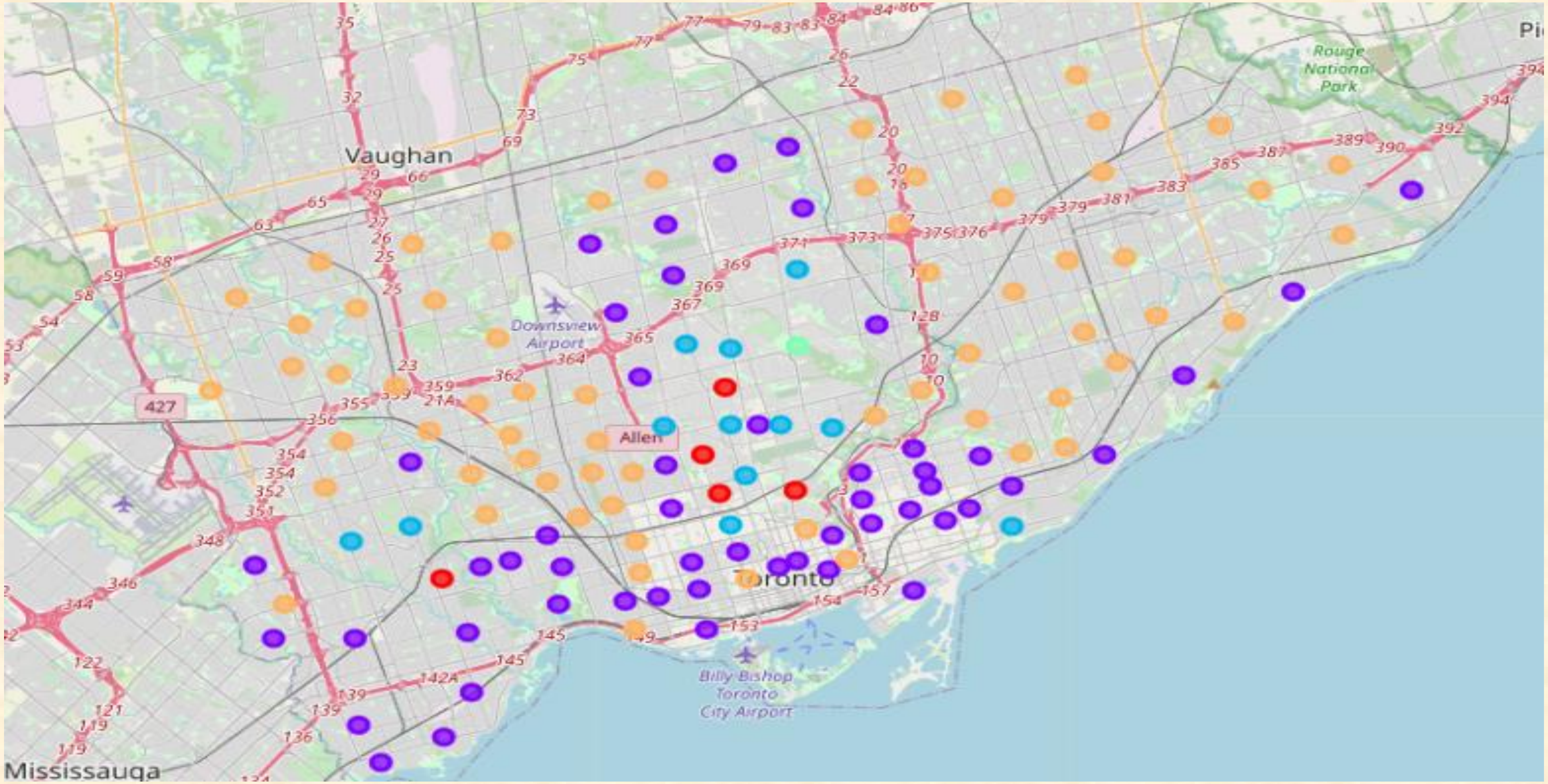
Result (2)

Visualizing population and income for the neighborhoods to be able to compare visually.



Result (3)

Clustering similar Neighborhoods using the scoring



Result (3) – cont.

```
# Cluster 1
toronto_df.loc[toronto_df['Cluster Label'] == 0].sort_values('Population Score', ascending=False).head(4)
```

	Population	Income	Longitude	Latitude	Population Score	Income Score	counts	Restaurants Score	Total Score	Cluster Label
Rosedale-Moore Park	20923	207903	-79.379669	43.682820	0.846657	2.765917	1.0	0.080841	1.231138	0
Lawrence Park South	15179	169203	-79.406039	43.717212	0.614224	2.251057	4.0	0.323363	1.062881	0
Casa Loma	10968	165047	-79.408007	43.681852	0.443824	2.195766	4.0	0.323363	0.987651	0
Forest Hill South	10732	204521	-79.414318	43.694526	0.434274	2.720923	2.0	0.161681	1.105626	0

```
# Cluster 2
toronto_df.loc[toronto_df['Cluster Label'] == 1].sort_values('Population Score', ascending=False).head(4)
```

	Population	Income	Longitude	Latitude	Population Score	Income Score	counts	Restaurants Score	Total Score	Cluster Label
Islington-City Centre West	43965	52787	-79.543317	43.633463	1.779059	0.702272	13.0	1.050930	1.177420	1
Mimico	33964	54438	-79.500137	43.615924	1.374365	0.724237	5.0	0.404204	0.834269	1
Church-Yonge Corridor	31340	53583	-79.379017	43.659649	1.268184	0.712862	50.0	4.042037	2.007694	1
Niagara	31180	70623	-79.412420	43.636681	1.261710	0.939560	15.0	1.212611	1.137960	1

```
# Cluster 3
toronto_df.loc[toronto_df['Cluster Label'] == 2].sort_values('Population Score', ascending=False).head(4)
```

	Population	Income	Longitude	Latitude	Population Score	Income Score	counts	Restaurants Score	Total Score	Cluster Label
Annex	30526	112766	-79.404001	43.671585	1.235245	1.500226	17.0	1.374293	1.369921	2
Bedford Park-Nortown	23236	123077	-79.420227	43.731486	0.940253	1.637402	21.0	1.697656	1.425103	2
The Beaches	21567	92580	-79.299601	43.671050	0.872716	1.231673	30.0	2.425222	1.509871	2
St.Andrew-Windfields	17812	100516	-79.379037	43.756246	0.720769	1.337253	0.0	0.000000	0.686007	2

```
# Cluster 4
toronto_df.loc[toronto_df['Cluster Label'] == 3].sort_values('Population Score', ascending=False).head(4)
```

	Population	Income	Longitude	Latitude	Population Score	Income Score	counts	Restaurants Score	Total Score	Cluster Label
Bridle Path-Sunnybrook-York Mills	9266	308010	-79.378904	43.731013	0.374952	4.097729	5.0	0.404204	1.625628	3

```
# Cluster 5
toronto_df.loc[toronto_df['Cluster Label'] == 4].sort_values('Population Score', ascending=False).head(4)
```

	Population	Income	Longitude	Latitude	Population Score	Income Score	counts	Restaurants Score	Total Score	Cluster Label
Malvern	43794	29573	-79.222517	43.803658	1.772140	0.393436	4.0	0.323363	0.829646	4
Dovercourt-Wallace Emerson-Junction	36625	39740	-79.438541	43.665677	1.482044	0.528696	6.0	0.485044	0.831928	4

Discussion (1)

Based on the map of population and income:

the neighborhoods in the middle of Toronto were good choices for the new restaurants as they have high income average and good population.

Discussion (2)

Based on clustering:

- The first and fifth clusters (0 and 4) had the best potential places as the first one has very low supply (low restaurants number) with average demand (population) and the fifth one has very high demand (high population score) with low restaurants number (low supply).
- As we have the average income, it is recommended that the new restaurants focus on high prices menu for the first cluster and low prices menu for the fifth one.

Conclusion

We have concluded that the best potential new restaurant place in Toronto will be either first or fifth cluster as there is average to high demand with low supply (low competitors).