

Greater Toronto Area



The Battles of Neighborhoods

Potential Neighborhood for a New Restaurant, Toronto

1. Introduction:

The goal is to help investors to decide where to open a potential new restaurant in Toronto. The data provided will be the income and population per neighborhood. Also, the surrounded restaurants that are considered competitors will be shown. The final project will help the investors to check where to open a profitable new restaurant.

The investors needs to:

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

As success, we should provide a good recommendation in terms of profit and investors' expectations.

2. Data:

The income and population per neighborhood in Toronto will be taken from Toronto's Neighborhood Profiles 2016. This data will be combined with Foursquare API to collect restaurants that are competitors in the same neighborhoods.

Toronto's Neighborhood Profiles data is publicly available at this website: <https://open.toronto.ca/dataset/neighbourhood-profiles>

From Toronto website, we got the neighborhood information including name, coordinates, populations, and income.

	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

From Foursquare, we got restaurants information per neighborhood including name, category, and coordinates.

	Neighborhoods	Neighborhood Latitude	Neighborhood Longitude	Restaurant	Restaurant Latitude	Restaurant Longitude	Restaurant Category
0	Islington-City Centre West	43.633463	-79.543317	Taste of Thailand Cuisine	43.635928	-79.540785	Thai Restaurant
1	Islington-City Centre West	43.633463	-79.543317	Mekong River Restaurant	43.632643	-79.544184	Vietnamese Restaurant
2	Islington-City Centre West	43.633463	-79.543317	Kebab 49	43.635943	-79.540422	Turkish Restaurant
3	Islington-City Centre West	43.633463	-79.543317	Swiss Chalet	43.632057	-79.544213	Restaurant
4	Islington-City Centre West	43.633463	-79.543317	Harvey's	43.632549	-79.543757	Fast Food Restaurant

3. Methodology

The neighborhoods' data was extracted from Toronto.ca. Also the restaurants information per neighborhood were retrieved from FourSquare API. Geocoder API was used to get the GPS coordinates of the neighborhoods in Toronto.

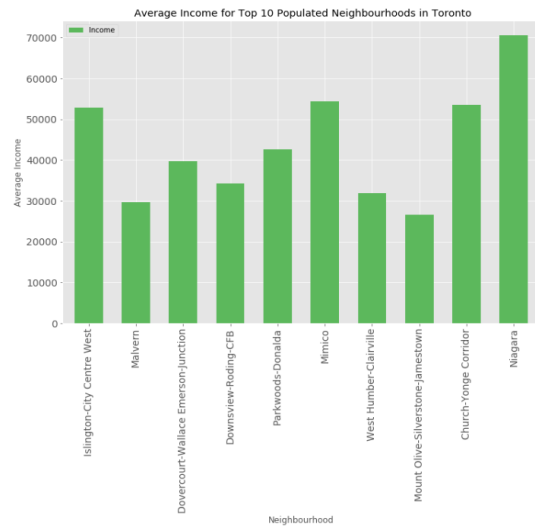
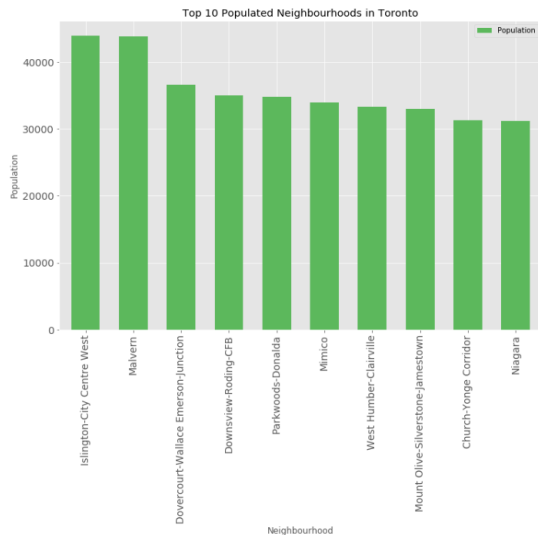
The population, average income, and the number of restaurants per neighborhood were used as features for the K-Means clustering. K-Means clustering was used to group the similar neighborhoods together. We added extra features to the method which are population score, income score, restaurants availability score, and total score. The first three scores are calculated by dividing each number by the total. Total scoring per neighborhood was calculated based on population, average income, and number of restaurants with same weight for each.

	Population	Income	Longitude	Latitude	Population Score	Income Score	counts	Restaurants Score	Total Score
Church-Yonge Corridor	31340	53583	-79.379017	43.659649	1.268184	0.712882	50.0	4.042037	2.007694
Mount Pleasant West	29658	57039	-79.393360	43.704435	1.200121	0.758840	50.0	4.042037	2.000333
Bay Street Corridor	25797	56526	-79.385721	43.657511	1.043885	0.752015	50.0	4.042037	1.945979
Yonge-St.Clair	12528	114174	-79.397871	43.687859	0.506950	1.518957	44.0	3.556993	1.860967
Kensington-Chinatown	17945	37422	-79.397240	43.653554	0.726151	0.497858	47.0	3.799515	1.674508

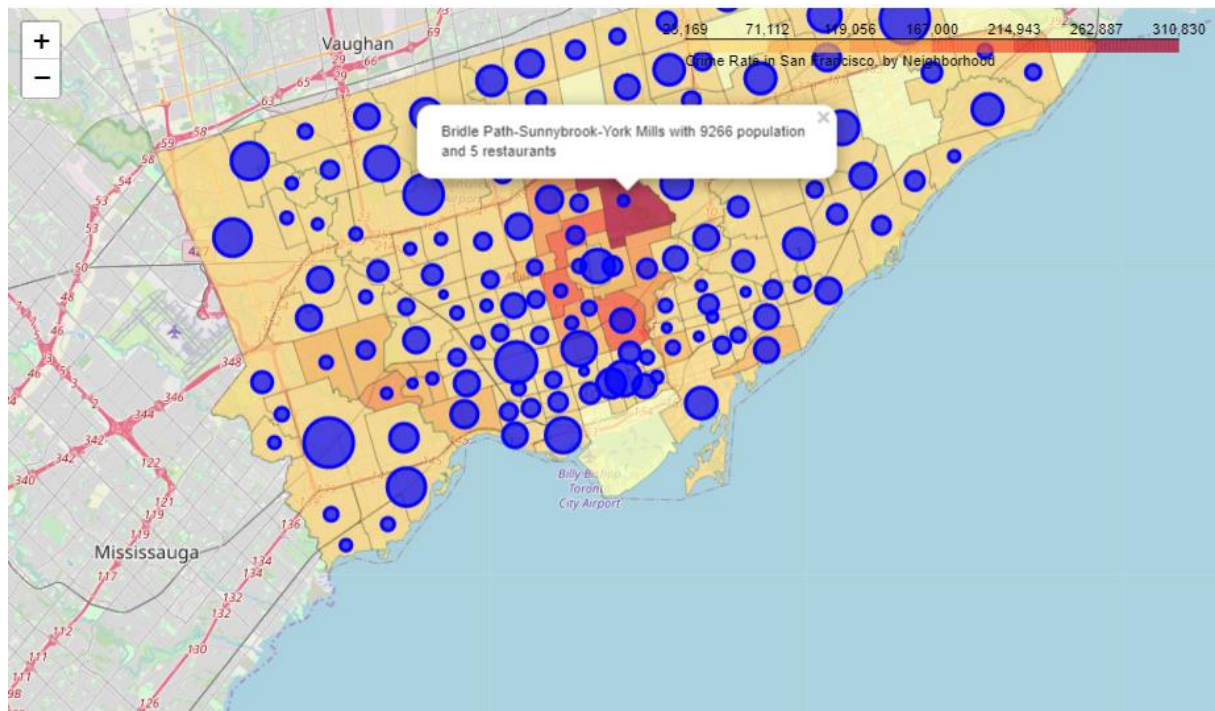
With the above, it will be possible to recommend the potential new restaurants place.

4. Result

We got the top 10 neighborhoods based on population. As first look, the potential new restaurants place will be in the neighborhood but the average income isn't increasing the same.

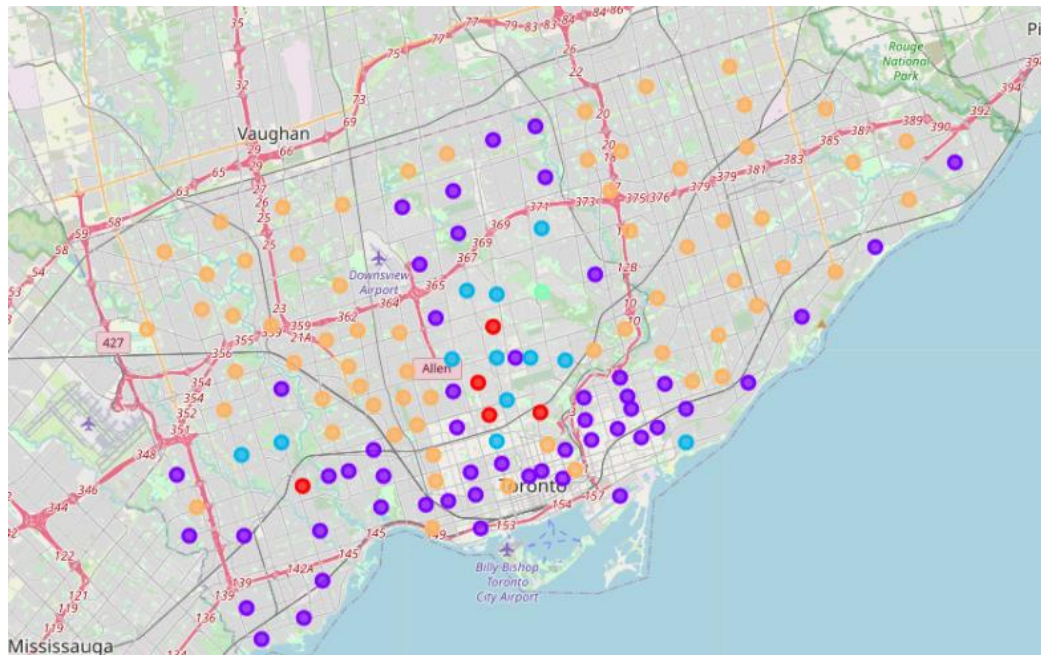


We also visualized the population and average income for the neighborhoods to be able to compare them visually.



The above map made it easy to decide where to have new restaurant based on demand (population and income) and supply (available restaurants). So for example, Bridle Path York Mills has very high average income but the population is very low so we can say having average demand. The supply in that neighborhood is also very low as it has only 5 restaurants so maybe it can be a potential new restaurant place with average to high prices menu.

By K-Means technique, we got 5 clusters of similar neighborhoods. The neighborhoods were grouped together according to the population, average income, restaurants, and total score.



We also check each cluster as K-Means clustering evaluation.

```
# 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	-79.408039	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	43.633483	1.779059	0.702272	13.0	1.050930	1.177420	1
Mimico	33964	54438	-79.500137	43.615924	1.374385	0.724237	5.0	0.404204	0.834269	1
Church-Yonge Corridor	31340	53583	-79.379017	43.659649	1.268184	0.712882	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	-79.299801	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

5. Discussion

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.

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.

6. 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).