Identifying optimum locations for starting a Restaurant in Toronto, Ontario in Canada

Mohab O. Hassan

Coursera

Applied Data Science Capstone

Problem

Identifying the optimum location(s) for starting a new restaurant in Toronto, Ontario in Canada.

Proposed Solution

Finding:

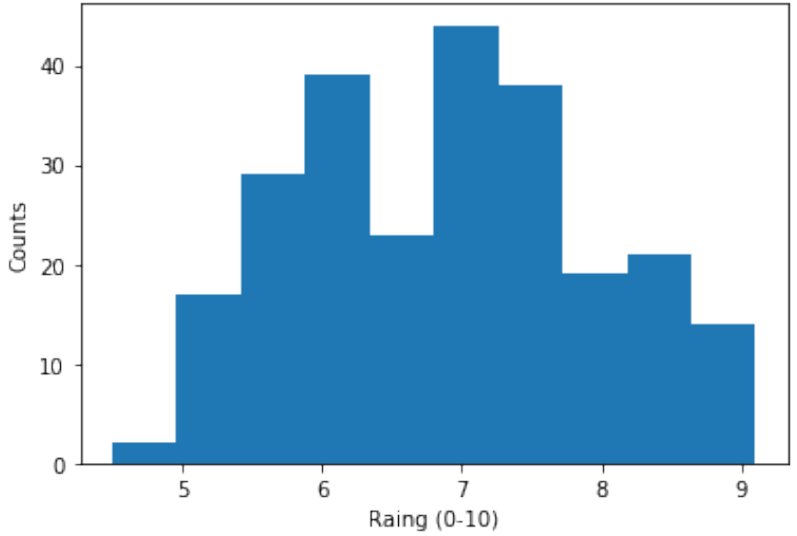
- A location with a minimum competition means a location where the number of restaurants is low with respect to the number of people or the current population living there.
- A location with a weak competition means a location where there are some restaurants, but their quality or rating is low.

Data Sources

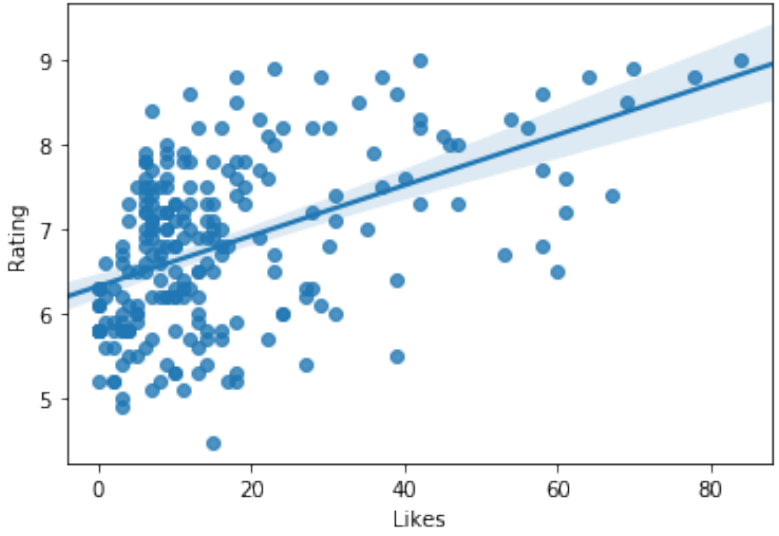
Data type	Source
Postal Codes, Boroughs and Neighbourhoods in Canada	Wikipedia
Postal Codes, and geographical locations in Toronto, Ontario	Coursera (CognitiveClass.ai)
The population in each Neighbourhood in Canada in 2016	Canada Stats
The geographical boundaries for each Neighbourhood in Toronto, Ontario	Canada Stats
Venues' IDs, names, locations, distances from the corresponding Neighbourhood centre location, postal codes and categories type	Foursquare API
Restaurants' ratings, likes, checkins, price tier and tips	Foursquare API

Data Cleaning

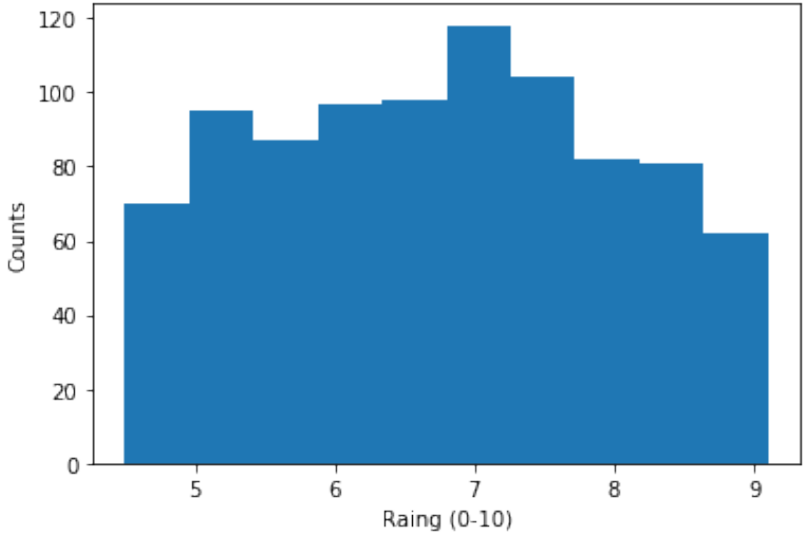
- More than 50% of the rating values of the restaurants are missing!
- Trying to get the missing rating values using the Like Counts in SLR.
- Weak correlation (~0.5) between Rating and Like Counts.
- Assuming uniform random distribution for missing values



Distribution of the available Ratings in the restaurants



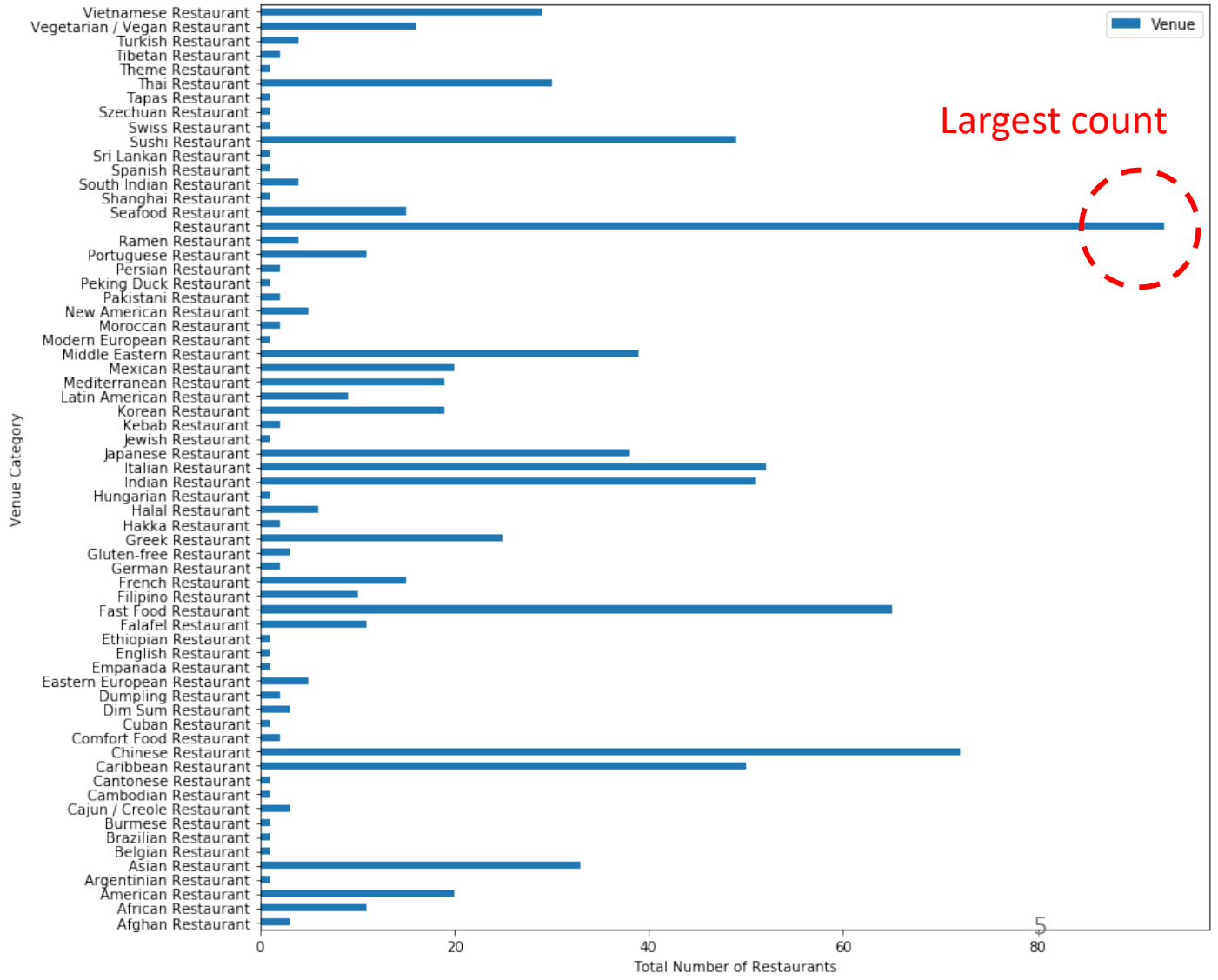
Single Linear Regression between the Like Count and Rating



Distribution of Rating after using a uniform random distribution.

Exploratory Data Analysis

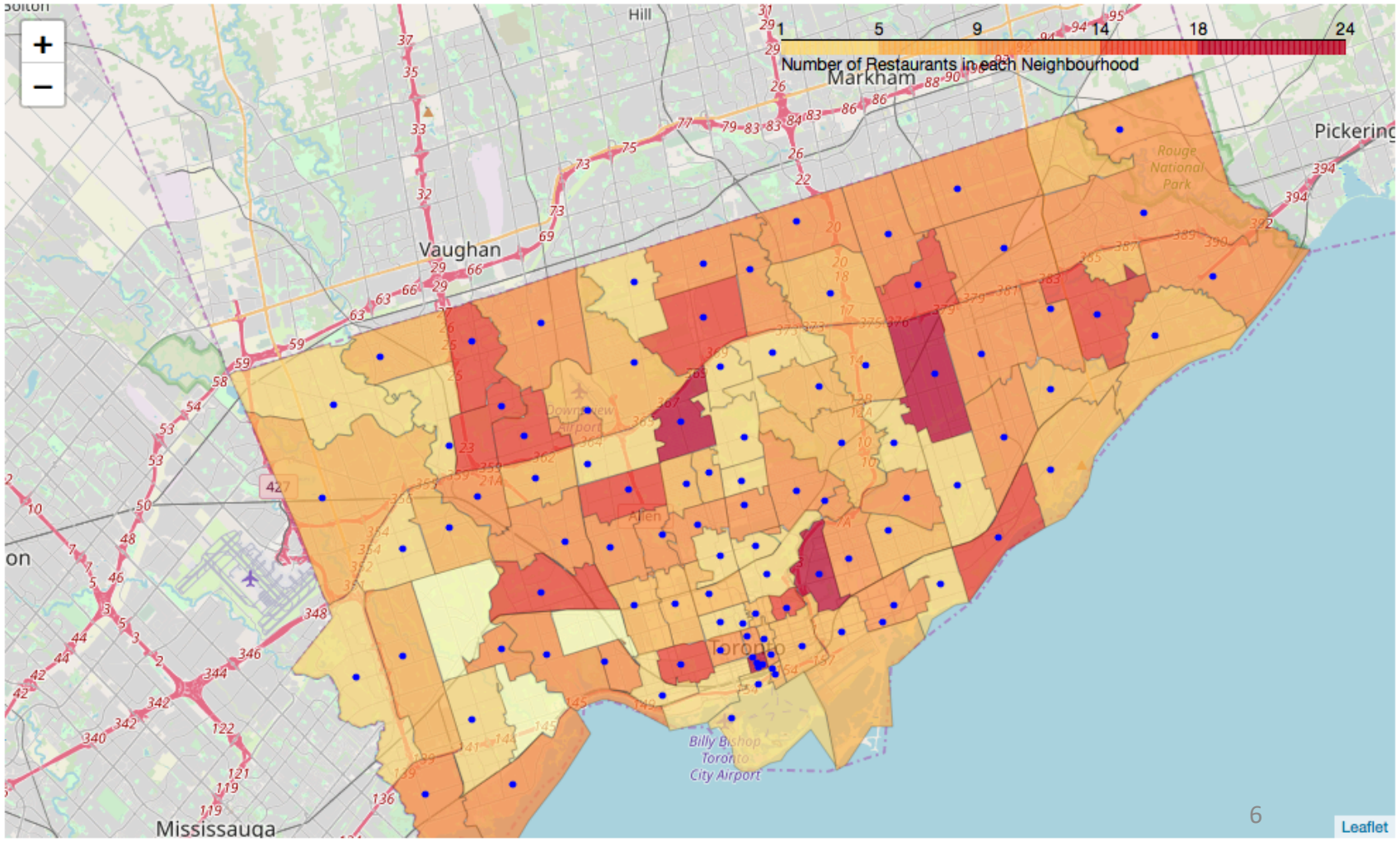
- Total number of restaurants for each cuisine.
- The largest count belongs to generic restaurants (without a specific cuisine).
- Not possible to do analysis based on cuisine.



Exploratory Data Analysis

Number of restaurants in each neighbourhood

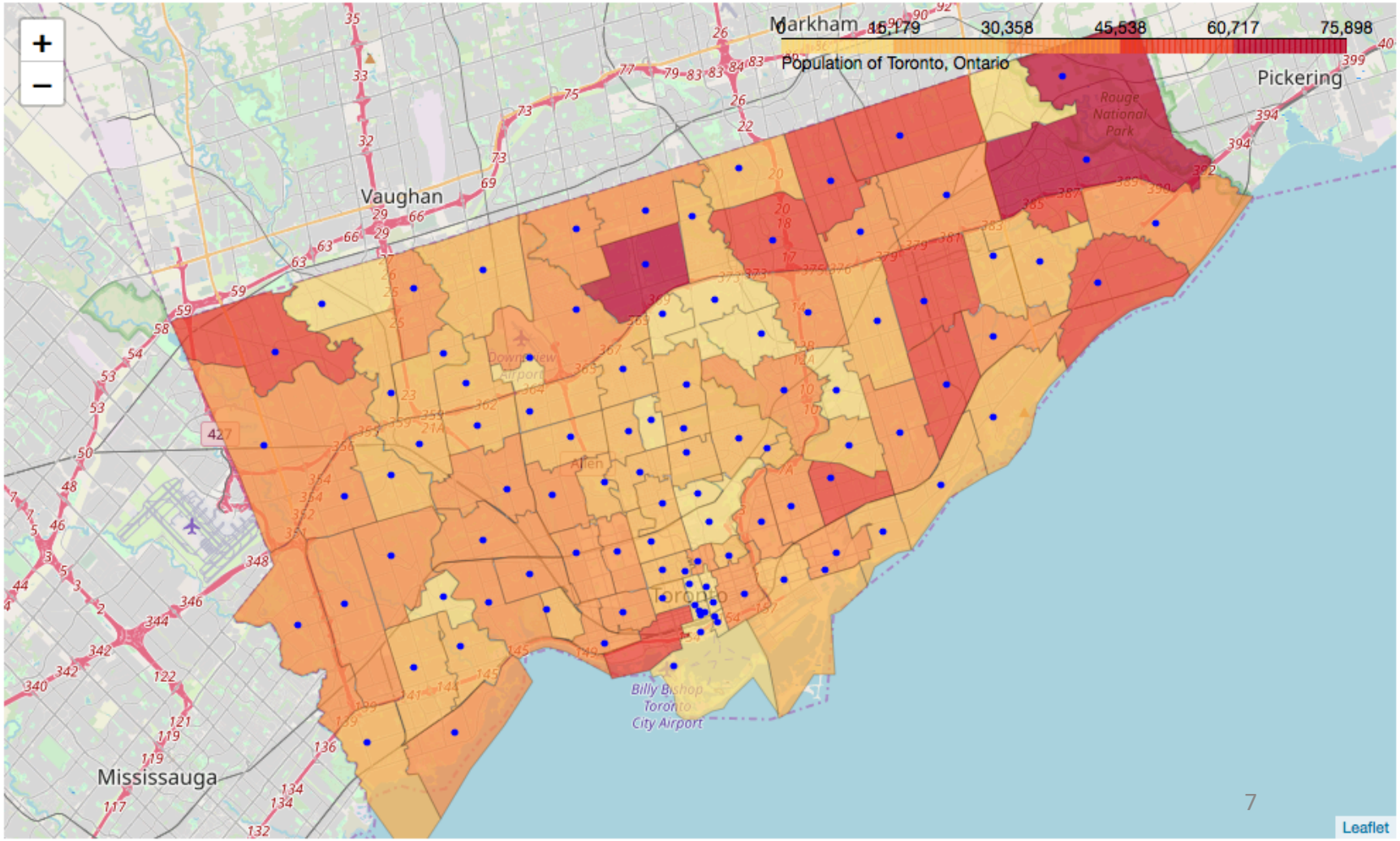
- Large density of restaurants in the Toronto Downtown compared to the borders of Toronto.



Exploratory Data Analysis

Population in each neighbourhood

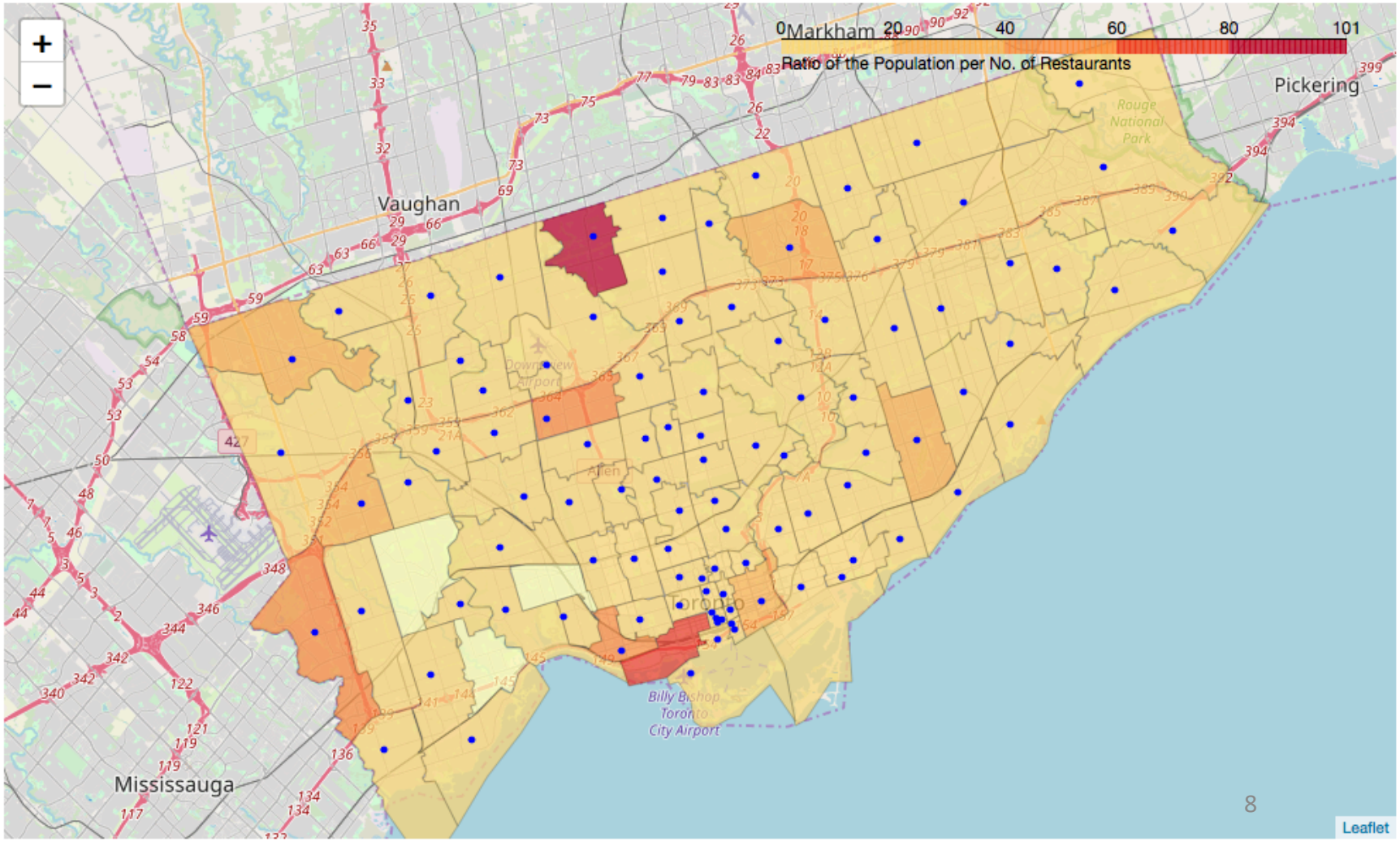
- Large population centers near the borders of Toronto.
- Many small neighbourhoods near Toronto Downtown with relatively smaller populations.



Results and Discussion

Lack of restaurants in some neighbourhoods

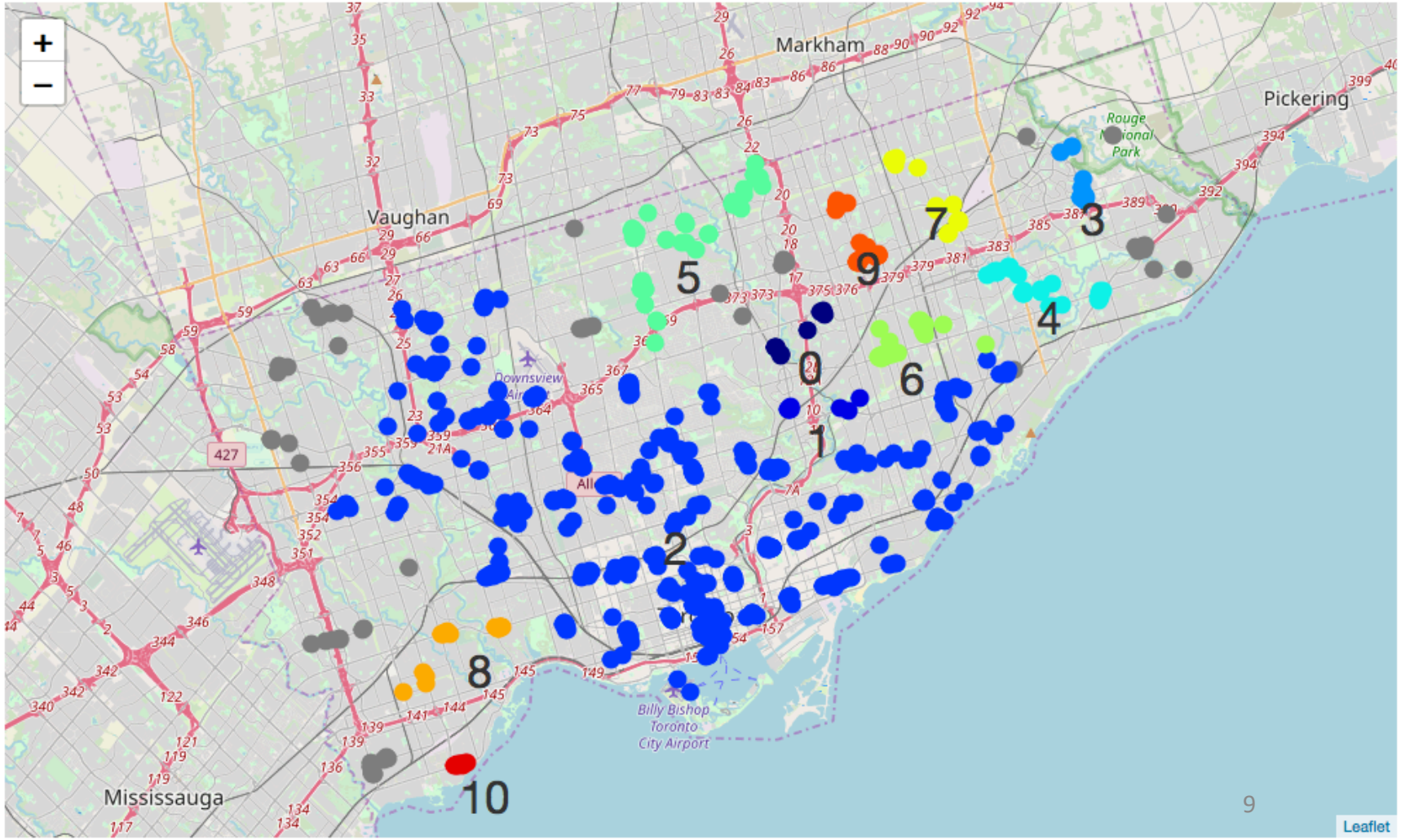
- The ratio between the population and number of restaurants in each neighbourhood.
- Willowdale West in North York has only 1 restaurant and a population greater than 40,000.



Results and Discussion

Clustering the restaurants based on their location

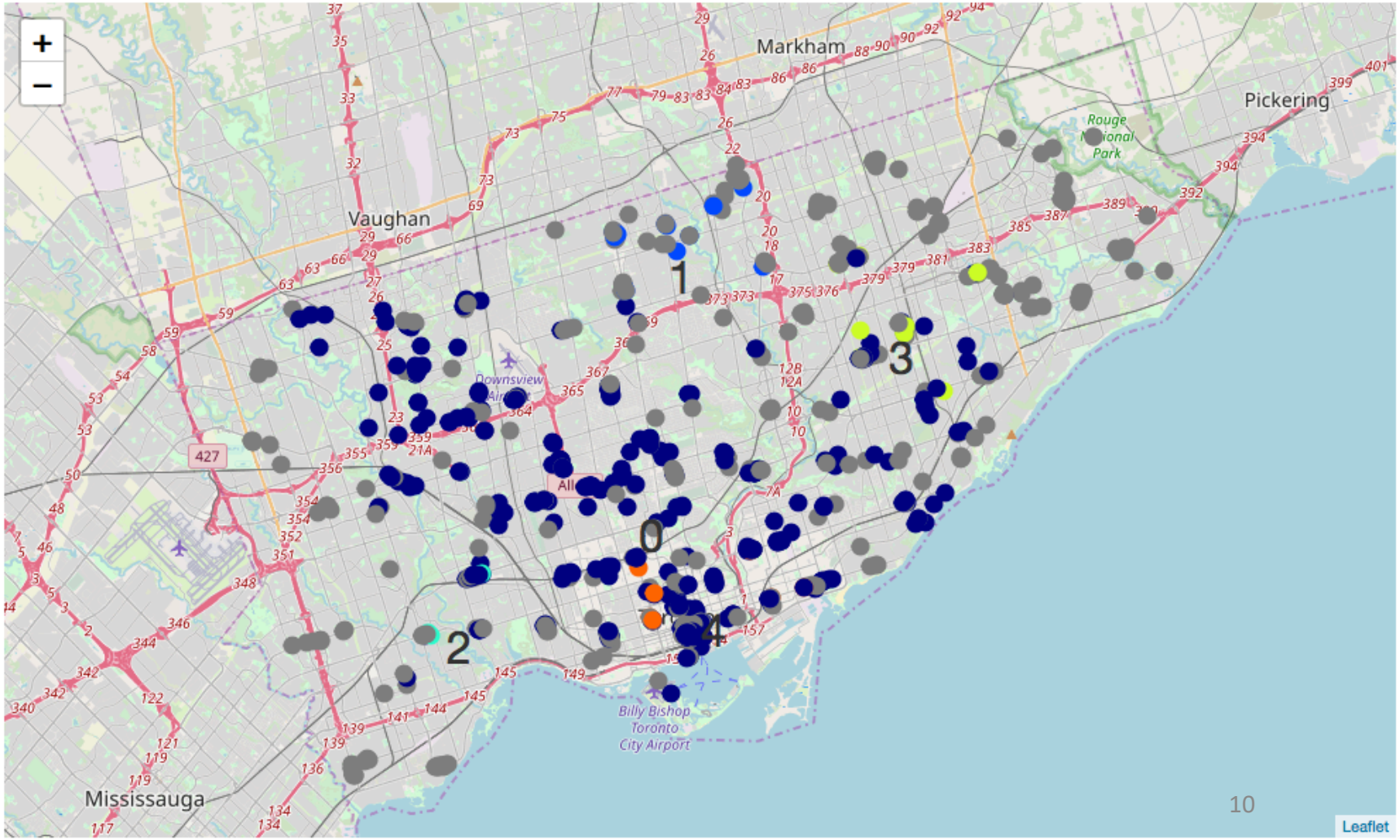
- A large cluster (#2) in the middle of Toronto but smaller ones near the borders and the north east area.



Results and Discussion

Clustering the restaurants based on their location and their rating, like count, tip count and price tier

- A large cluster (#0) in the middle of Toronto sharing the same features.
- Table in the next slide showing the features of each cluster.



Results and Discussion

- Cluster #4 is located near the Toronto Downtown area and has very competitive features such as high rating, like count, tip count and low price.
- Cluster #3 is located near the Scarborough Centre area and has weak competitive features such as low rating, like count and tip count. This could potentially be a good place to have a new restaurant.

Cluster #	Avg Rating	Avg Like Count	Avg Tip Count	Avg Price Tier
0	6.83	1.51	1.03	2.0
1	6.49	1.4	0.53	2.0
2	7.42	0.0	0.0	2.0
3	5.12	0.6	0.3	2.0
4	6.92	1.9	1.2	1.0

Conclusions

- Different datasets from Foursquare and Canada Statistics were explored to identify optimum locations to start a new restaurant.
- The optimum locations were defined as either those that have very low number of restaurants compared to their population, or those that have some restaurants, but they are of low quality/unpopular.
- The results suggested that Willowdale West in North York has relatively a very few restaurants compared to its population. Also, the restaurants located near Scarborough Centre area have relatively low ratings and are less liked by their visitors.