

Find a place for the restaurant

By Anna Budilova

Introduction/Business Problem

My choice of problem is to find out best places to open a restaurant in Toronto.

Good locations for restaurant is a key to success. What is the good location for restaurant? Some place with no restaurant or less restaurants then in other places but people go by very often.

Data

Using data from Foursquare and data about postal codes of Toronto, i going to find out list of best places to open restaurant. To find locations I'm going to use set of postal codes of Toronto to split city for regions.

Then using Foursquare data, which contains information about different venues in Toronto, I get data about restaurant's locations . Then analyze venues categories assign every venue category to venue groups as «restaurant», «food», "entertainment» and so on.

List of Toronto's postal codes is taken from wikipedia, converted into table view (there were different types of table view, and they changed it every day so I downloaded it)

Sample of Postal code's data

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront

Sample of Toronto's venue data

Venue	Venue Latitude	Venue Longitude	Venue Category
Brookbanks Park	43.751976	-79.332140	Park

KFC	43.754387	-79.333021	Fast Food Restaurant
Variety Store	43.751974	-79.333114	Food & Drink Shop
Victoria Village Arena	43.723481	-79.315635	Hockey Arena
Tim Hortons	43.725517	-79.313103	Coffee Shop

To find regions with less restaurants we are going to Intersect foursquare data with postal codes of Toronto. This operation will give us information about how many venues of every group are located in every postal code region. With this information i'm going to find regions where less restaurants among other regions.

Methodology

Whole dataset consists of list of venues with venue group and regions or postal code where this venue is located. To analyze regions base on this dataset I need to convert this data in dataset of postal codes with some characteristics. characteristics of regions for this projects are: how many venues of every group contains region:

1. how many restaurants
2. how many other food venues like cofeeshops or cafe
3. how many entertainment venues like theatre or concert halls
4. how many outdoor venues like parks
5. how many business venues like coworking places
6. how many shops or stores
7. how many hotels

Those are factors to find place for restaurant. List of venues group was much longer, but they have very small amount of venues, so venue group with more impact are left.

Using these factors dataset of venues is transformed to dataset of regions. As I don't need prediction, only clustering, so k-means algorithm is good for my purpose. Apparently regions share some similarities in their characteristics, based on which, regions can be grouped into clusters.

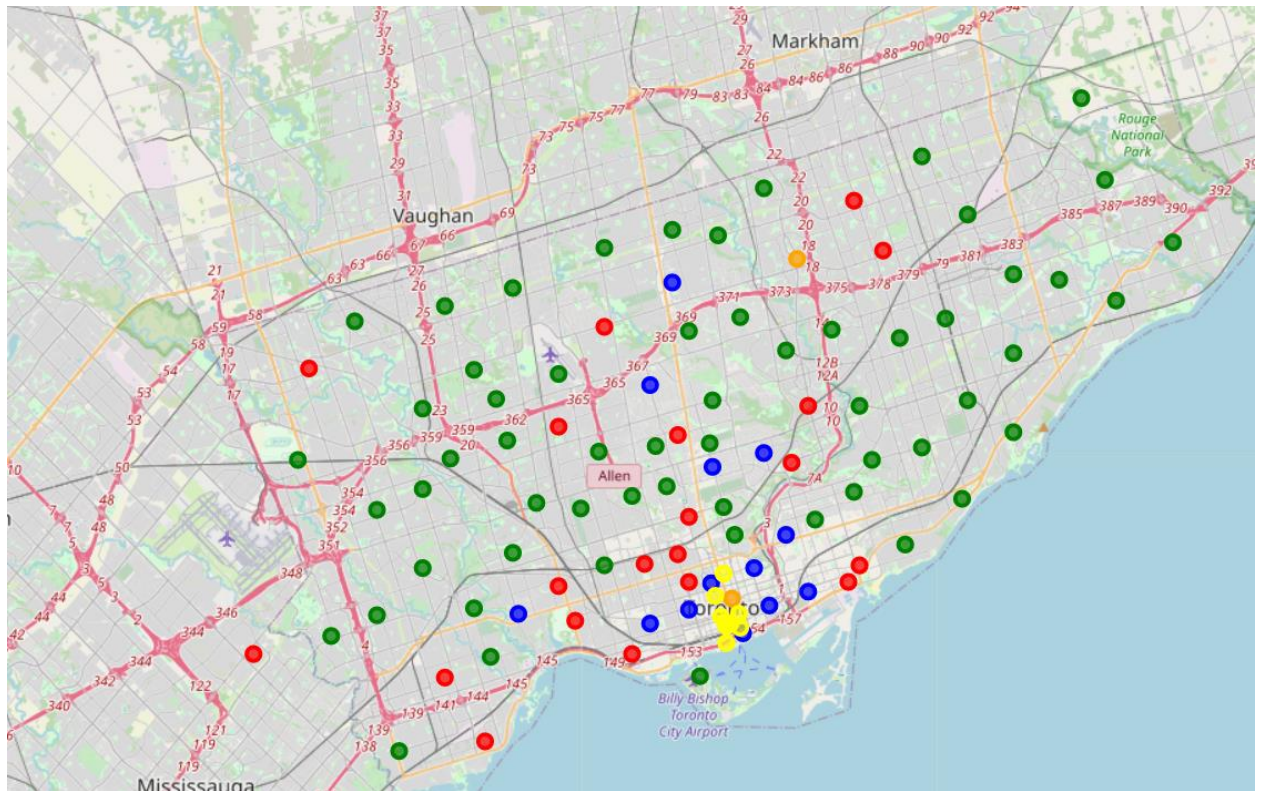
Results

Splitting dataset gave 5 clusters of regions.

- cluster 0 (blue dots) consists of central regions and has average value of 10 restaurants per region
- cluster 1 (red dots) consists of different regions with average value of 3 restaurant per region.

- clusters 2, 4 (yellow, orange dots) consist of central region with average value of 18-23 restaurant per region
- cluster 3 (green dots) most popular cluster where no more then 0.5 restaurant per region in average, and 75% of regions in this cluster has no restaurants at all.

My choice is cluster 3 (green dots on the map)



Discussion

Next step, perhaps, is to analyze data inside cluster with additional information about neighborhood, buildings, check prices, population. May be with this information and information about future restaurant's idea cut list of places much more.

Conclusion

Recommendation for choose place I would start with regions with no restaurants at all, but of course there are a lot of questions except amount of other restaurants, for example rent price, nice street view and so long. However with list of regions we have a good start to find places.