

# The battle of neighborhoods in the city of Toronto

by Tomas Stankevičius

2021



## Introduction / Business Problem

The company wants to start a restaurant business and is trying to decide which area in Toronto could be a right spot to do that. An area should have the biggest potentiality in developing restaurant business. Location should help company to reach as wide customers auditory as possible to make business profitable. As company does not possess a knowledge of Toronto districts and neighborhoods, it decided to hire a business consultant company which provides various consultant services for companies which want to establish new business in Toronto area. A consultant company utilizing Data Science methods is going to cluster Toronto city area and provide recommendations where is the best place to start a restaurant business.

## Explanation about Data

Consultant company is going to use Foursquare location data services to build a data base of various venues in every neighborhood of Toronto city. Then this data base will enriched with additional data provided by Statistics of Canada. After that, company is going to build a clustering model based on unsupervised Machine learning clustering model – K-Means.

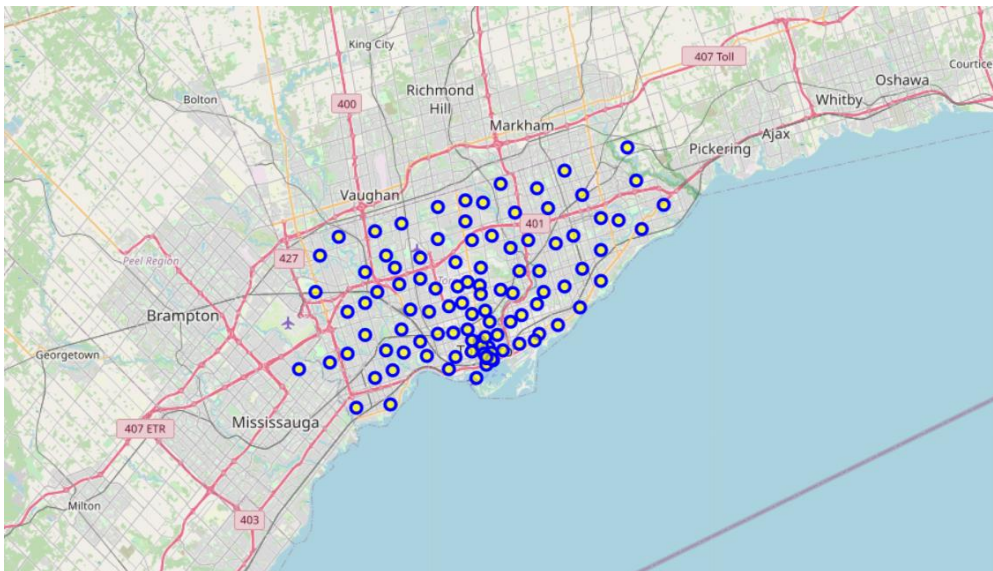
At first we are going to gather information from wiki page [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) containing Toronto neighborhoods and postal codes. After parsing and building a data frame in pandas the fraction of that data looks like this:

	PostalCode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...

Next, we will add coordinates for every postal code, in order, later to build a data base of venues in every neighborhood of Toronto using Foursquare API. This the example of data frame in pandas after merging mentioned data:

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724

Using Folium package we can map those neighborhoods on Toronto map to get better understanding of their locations:



A Foursquare API explore endpoint provides a list of recommended venues near the current location. We are going to use this endpoint to build our venues data base. A fraction example of response code looks like this:

```

{
  "meta": {
    "code": 200,
    "requestId": "5ac51ef86a607143de8eg5cb"
  },
  "response": {
    "warning": {
      "text": "There aren't a lot of results near you. Try something more general, reset your filters, or"
    },
    "suggestedRadius": 600,
    "headerLocation": "Lower East Side",
    "headerFullLocation": "Lower East Side, New York",
    "headerLocationGranularity": "neighborhood",
    "totalResults": 230,
    "suggestedBounds": {
      "ne": {
        "lat": 40.724216906965616,
        "lng": -73.9896507407283
      },
      "sw": {
        "lat": 40.72151724718017,
        "lng": -73.98693222860872
      }
    },
    "groups": [
      {
        "type": "Recommended Places",
        "name": "recommended",

```

Lastly, a data from Canada Statistics is going to be used too. Particularly a “Population and Dwelling Count Highlight Tables, 2016 Census” data set taken from <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Tables/CompFile.cfm?Lang=Eng&T=1201&OFT=FULLCSV>

After all data pre-processing, we are going to use K-means clustering algorithm from Python’s scikit-learn package to cluster neighborhoods. After that we will analyze clusters by their features and come up with conclusion – which neighborhoods are the best place to start a new restaurant business.