

## 1. Data Collection

A list of postal codes in Canada where the first letter is M is taken from wikipedia. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

Canada Post provides a free postal code look-up tool on its website, via its applications for such smartphones as the iPhone and Blackberry, and sells hard-copy directories and CD-ROMs. Many vendors also sell validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries. We shall gather basic information of borough , postcode and neighbourhood from this webpage

([https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)).

Besides we need the geospatial data of this postcode which we can get from Geocoder Python package. The problem with this Package is you have to be persistent sometimes in order to get the geographical coordinates of a given postal code. So, one can make a call to get the latitude and longitude coordinates of a given postal code and the result would be None, and then make the call again and you would get the coordinates. So, here I used a list of geographical coordinates of each postal code from this link (: [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)).

Once we collect all the data then we can go for further processing of these data sets. We have to analyse the data, process it, drop unnecessary columns, and use important column to extract information. However, we need to know in depth information of neighbourhoods of each postcode as well as borough in Toronto. For example, how many parks, gym, restaurants are there in each postcode. All of these, we would find at Foursquare API website (<https://developer.foursquare.com/>). Foursquare is a technology company that built a massive dataset of location data. This site will be used to find out the list of restaurants in different boroughs of Toronto. Then K-Means Clustering algorithm will be used to find out the best location. .