**KMeans clustering on Food places**

**Introduction**: People who lives away from their family and who have to do everything by their own faces many problems. One of them being where to eat every day and which places are popular in neighborhood. The main problem is to find places around any neighborhood based on how many people have rated particular place.

**Data**: For this problem to be solved, we will need data of different boroughs and their neighborhoods to find food places around it. The Latitude and Longitude of neighborhoods will be sent to Foursqaure API to find nearby places around them. We will also need number of likes to able to cluster similar places and also categories of places.

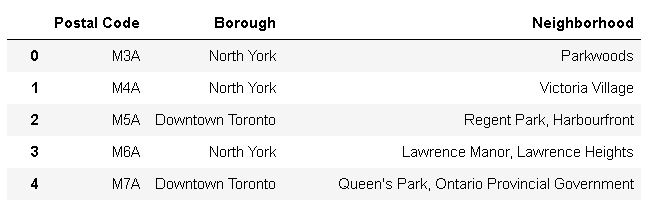
1) https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M – This page provides postal codes, name of boroughs and neighborhoods.

2) http://cocl.us/Geospatial\_data - It contains latitudes and longitudes of the boroughs.

Foursqaure API – This API provides information such as nearby venues with categories, location, likes etc.

**Methodology**: After collecting data from various sources, it was necessary to filter out the data that will be used to get solution.

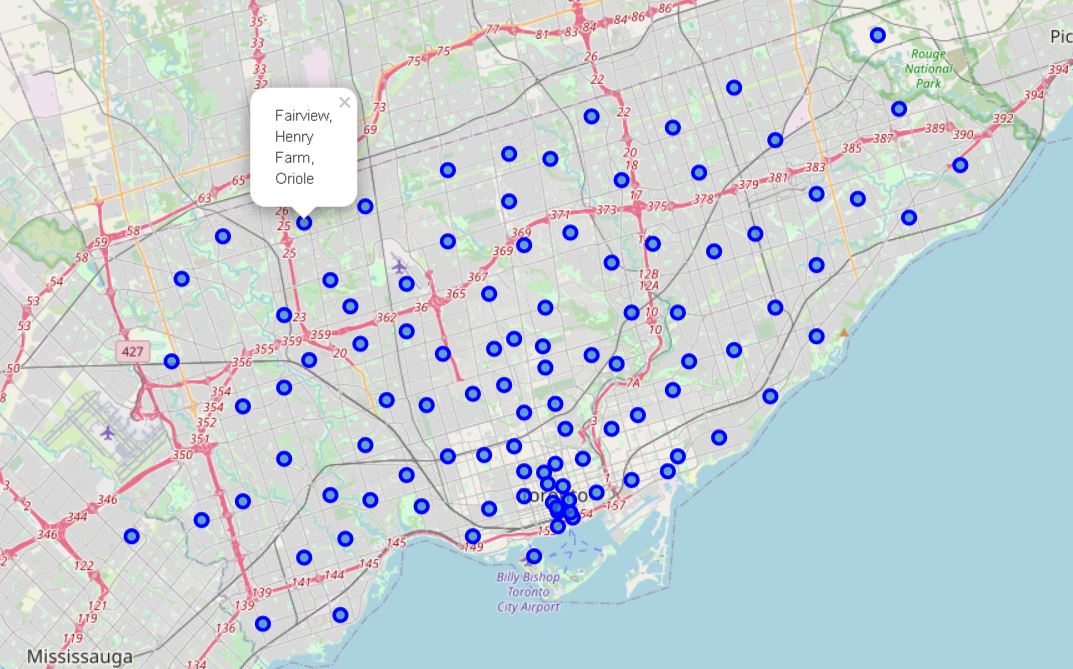
Data collected from Wikipedia page <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada> :



Data after adding location from <http://cocl.us/Geospatial_data> :

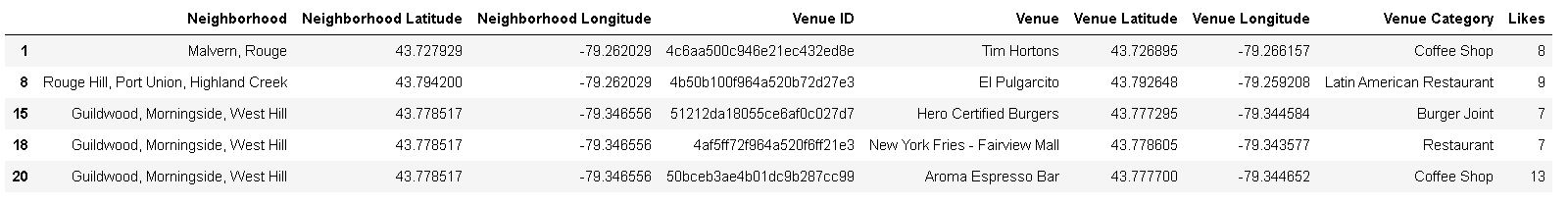


Data plotted on world map to visualize boroughs of Toronto, CA:

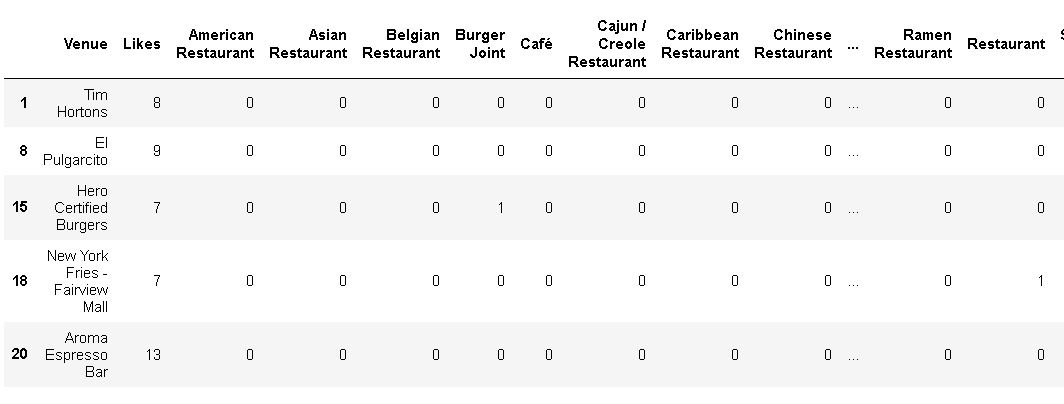


The labels on map shows neighborhoods for particular borough.

Using Foursquare API, we will now find out nearby venues with name, category, location and likes:



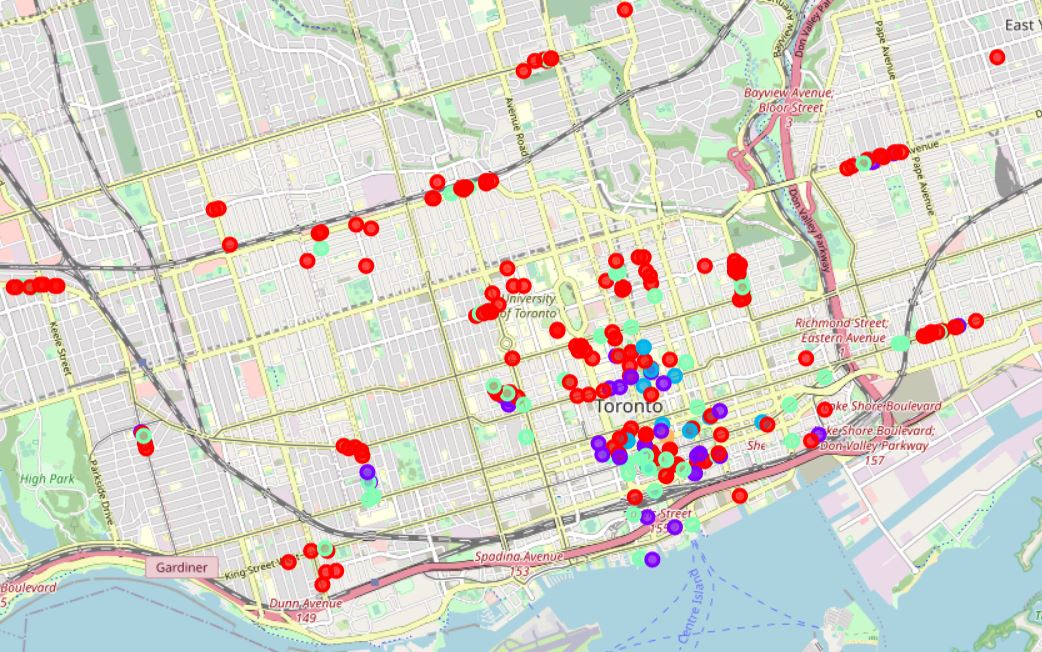
Applying one hot encoding on categories of venues:



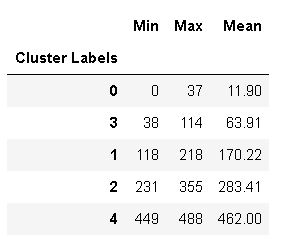
Data shown on above image is feed to KMeans algorithm in order to find places containing similar places in clusters. I have created 5 clusters which have likes in different ranges which categories them in separate clusters.

**Result:**

Visualizing clusters on a world map centered on Toronto, CA which shows cluster points with different colors.



Data of each clusters showing their minimum, maximum and mean values.



**Analysis:**

Here, we have categorized different food places into 7 different clusters based on their category and number of likes. The above dataframe shows different statistics for each cluster namely min, max, and mean. We can easily identify the different ranges od likes from clusters.  
  
1) Cluster\_0: In this cluster, we have range of 0-37. So, we can say that cluster contains places which are rated poorly based on the data we have used. Because there can be many factors such as how many people go there every day, what kind of people lives in neighborhood and so on. As there can be many factors to evaluate the same, but keeping in mind the data used in this notebook It contains places with very low likes compared to other places.  
  
2) Cluster\_3: We have places with number of likes 38-114. We can identify this cluster as slightly better rated places compared to Cluster\_0.  
  
3) Cluster\_1: This place has good rating as many users have liked so people may want to go here than the above two clusters.  
  
4) Cluster\_2: This contain places with high number of likes so they must be very popular around the neighborhood.  
  
5) Cluster\_4: This are very highly rated places and they should be at the top in the list of recommendation.

With above analysis we can easily find highly rated places in particular neighborhood in Toronto. It can also be used to analyze that in which area what kind of food places have low number of likes. So, if someone want to open a restaurant or food chain. They can do so and have good sells if they can provide good food as other places are not much popular in neighborhood.

**Conclusion:**

The problem mentioned in introduction is solved using data from various places, we found out similar places which can be recommended to anyone looking for a good place to eat. To conclude, system provides good places to eat as per the problem and can also be extended to add more places.

**Future extensions:**

A GUI can be developed which lets user entered any area and it can recommend places and also many other factors such as pricing of places, waiting time, open-close time can be added to have better recommendations.