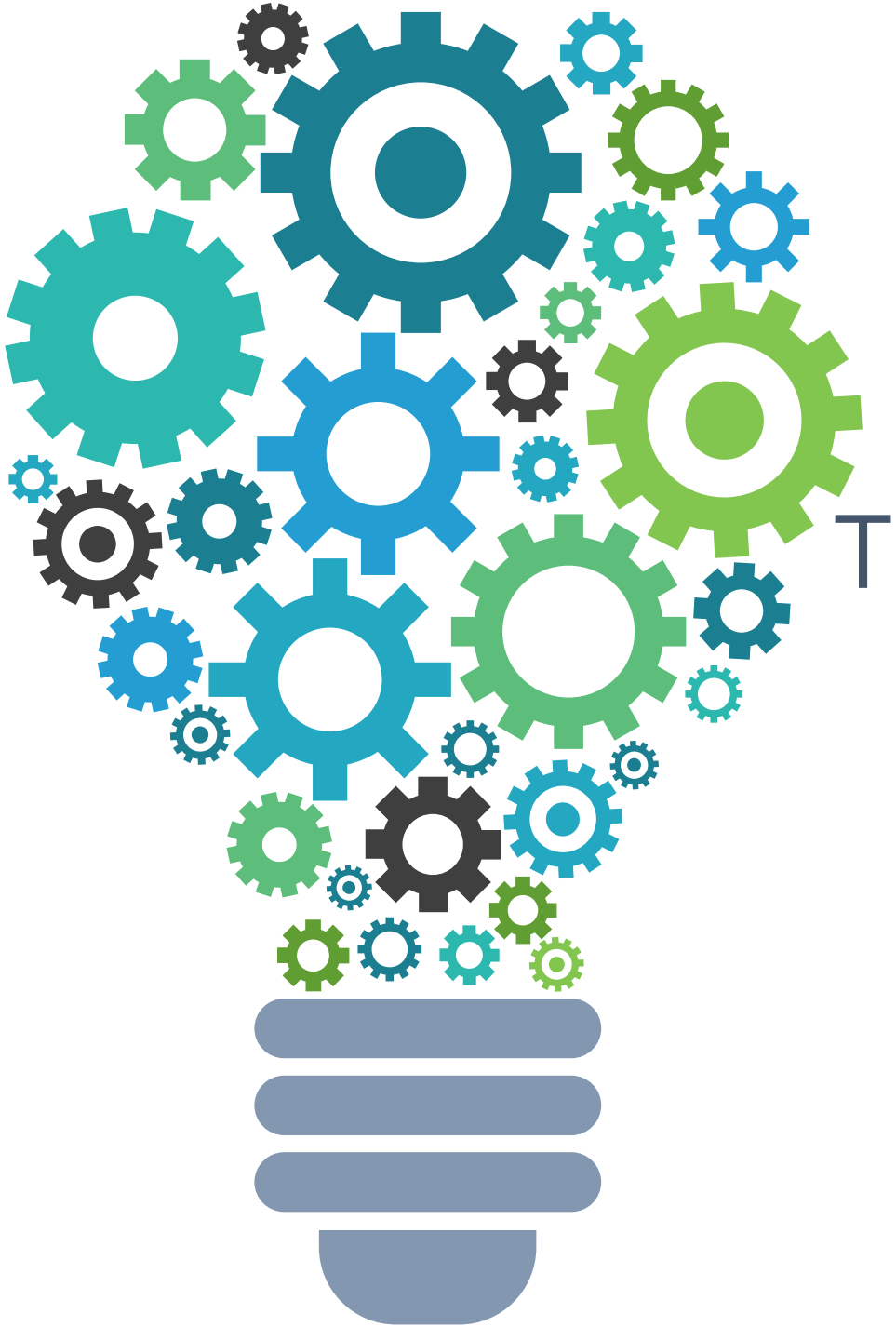


Coursera Capstone Project

The Battle of Neighborhoods

Toronto

Potential areas to open a New Indian Restaurant

Ravi Kant

Business Problem

Toronto is a highly multicultural city and has residents from various nationalities. One of the most important time out for the people is eating out with friends and family.

People from different nationalities like to visit the restaurant of their ethnicity as it makes them feel closer to home and their culture, apart from satisfying their taste buds.

Restaurant owners and want-to-be owners are always on constant lookout for places where they can set up the restaurants.

Toronto has a high Indian diaspora and the people are scattered through out the city. Although, some areas cater well to the diaspora in terms of restaurant offerings, others lack them.

The purpose of the analysis is to find potential areas where an Indian restaurant can be set up.

Data for the analysis

Following data is required:

- List of neighbourhoods in Toronto. The scope of this report is throughout the Greater Toronto Area, which combines neighbour cities and have great inter-connectivity.
- Latitude and longitudes of the neighborhoods.
- Venue data, which provides different available amenities.

Sources of data:

- Postal Codes, Borough and Neighborhood Data is available on Wikipedia.
- Latitude and Longitude information is downloaded.
- Venue information can be gathered using Foursquare API.

Links for the sources of data:

- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- https://cocl.us/Geospatial_data
- https://api.foursquare.com/

Methodology

List of neighborhoods in Toronto is gathered from Wikipedia page. The data is cleaned and sorted.

All the unassigned and null values are removed.

The data is then grouped as per different postal codes.

Latitude and longitude data is downloaded.

The dataframe with neighborhoods is combined with coordinates dataframe to create a new dataframe with all the information.

The neighborhoods are superimposed over a map of Toronto for visualization.

Foursquare API is used to retrieve information of venues in the different neighborhoods in a 3 km radius.

It is converted to dataframe with neighborhood details and venue details.

Each neighborhood is grouped according to mean of frequency of different neighborhoods and data for only 'Indian restaurant' is selected.

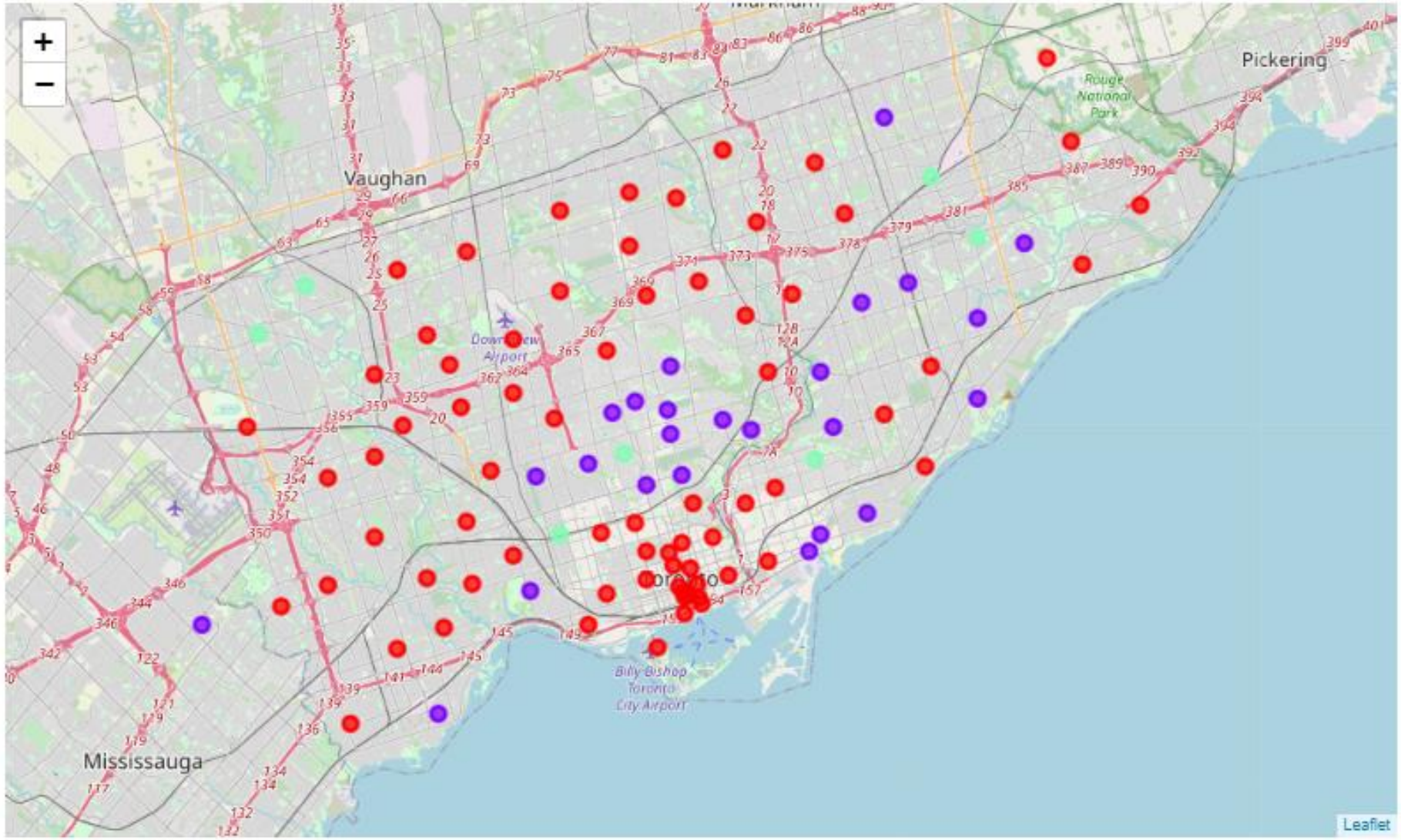
Clustering is done using K-means algorithm.

Initially 5 clusters were selected. Optimum number of clusters found to be 3.

Map with different clusters superimposed was populated.

All the clusters were analysed.

Results



The 3 clusters are:

- Red – Low density of Indian restaurants
- Blue – Medium density of Indian restaurants
- Aqua - High density of Indian restaurants

Only a few areas have high density of Indian restaurants.

Medium density areas are much higher in number.

Low density areas are pretty spread out around Toronto.

Discussion

1st Clsuter

Lowest density of Indian restaurants

Some areas with high Indian diaspora has low density of Indian restaurants.

Presents good opportunity to further look into those areas.

3rd Cluster

Highest density of Indian restaurants

These areas are probably saturated.

New restaurants in these areas will face stiff competition

Limitations

This report also highlights the potential areas and further evaluation requires careful evaluation with other variables.

Further analysis of the low density areas with regards to costs, business volume and other aspects need to evaluated before a final decision is made.

The results such provided only aid in making business decisions. The map provides a good idea which localities to look for a start.

Conclusion

This report highlights the potential areas where a new Indian restaurant can be set up.

Through the analysis, pandas and numpy provide great tools to extract and manipulate data.

Folium is a great way to visualise information on a map.

K-means clustering is a fast and efficient method to cluster the neighborhoods. However, iteration is required to find out the optimum number of clusters. In this case 3 clusters seem to be optimum.

Overall, the analysis presents good insights into distribution of Indian restaurants throughout Toronto. It is not only beneficial for setting up new restaurants, but also for people who want to explore different restaurants in an area and hence may look for high density areas.


Thank You