

Finding best neighbourhoods

Introduction

- ❖ The idea is to find neighborhood in Toronto city of Canada that has all the basic necessity shops within kilometers of the living place.
- ❖ The aim of the project is to divide the city neighborhoods in different categories according to shops and facilities available in the neighborhoods.
- ❖ The Foursquare API will be used to find all the nearby venues in neighborhoods and retrieve categories and count of shops in each category for each neighborhood.
- ❖ It would be very helpful for people new in the city as they will be able to identify neighborhood according to their needs.

Data acquisition and Cleaning

- ❖ Data of boroughs and neighborhoods of the Toronto City would be retrieved from Wikipedia (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto).
- ❖ The Geospatial data would be used to retrieve Longitude and Latitude of each neighborhood.
- ❖ Foursquare API would be used to retrieved nearby venues of each neighborhood.
- ❖ The Wikipedia data was retrieved by web scraping the page with BeautifulSoup library of python
- ❖ As a cleaning step neighborhood with less than 5 nearby venues were removed from the dataset.
- ❖ The rows with all zero values were removed.

Feature Selection

- ❖ Wikipedia Data: Columns Retrieved: Borough, Postal Code, Neighborhoods
- ❖ Foursquare Data: Latitude, Longitude, Venues, Category
- ❖ The categories selected for this project were basic necessity categories for ex., Gym, Grocery Store, Bank etc. These features can be selected according to user's need of facilities.
- ❖ The selected features would determine how clusters would be made of neighborhoods.

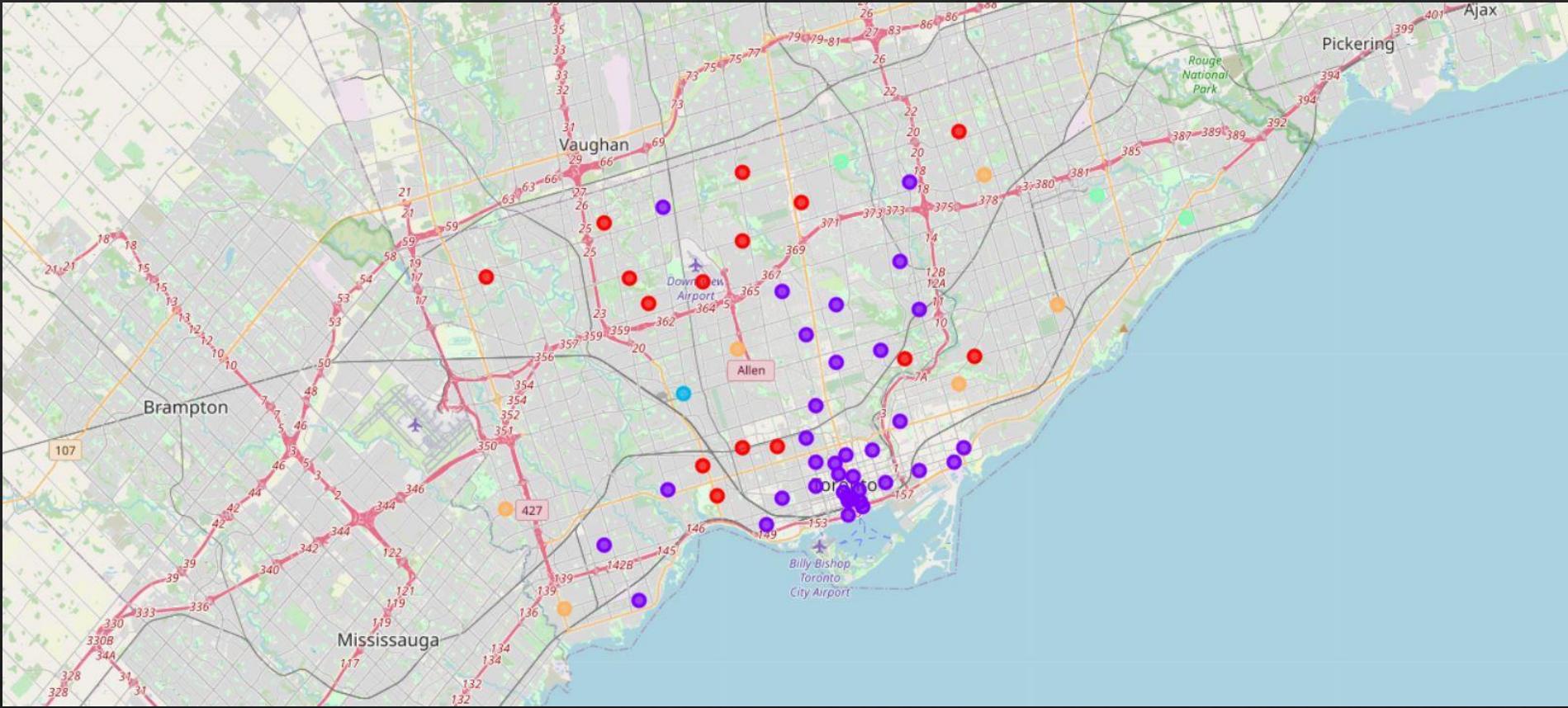
Methodology

- ❖ The explore request was used to get nearby venues. Limit of 100 was set for each neighborhood nearby venues
- ❖ The Foursquare API returned a JSON response of the explore query for all the neighborhoods.
- ❖ The new data frame containing Neighborhood name, longitude, latitude, Venue name, Venue Category, Venue latitude and Venue longitude.
- ❖ Categories data was one-hot encoded using pandas get_dummies function.
- ❖ K-means clustering with 5 clusters were used on the dataset.
- ❖ The frequency of occurrence of each category determined clusters of neighborhoods.

Map of Toronto



Clustered map of Toronto



Observations

- ❖ There are 5 different clusters of neighborhoods. Red and Purple clusters have more neighborhoods compared to other clusters.
- ❖ The red clusters are mostly on the airport side of the City which seems less populated.
- ❖ Purple neighborhoods are near University of Toronto and beach side. This side is more dense than other sides.
- ❖ The yellow cluster is of neighborhoods which are very far from main city area.
- ❖ The sea blue cluster has only one neighborhood in it which is inside city region but it is only one neighborhood in the area.
- ❖ The Cyan clusters are nearly on the border of the city.

Suggestions and Selection

- ❖ The sea blue cluster has only one neighborhood and it is very deserted area. This area does not have all the necessary facilities which makes it a very weak candidate for the selection of this neighborhood.
- ❖ The Cyan cluster is at very end of the city which makes it very obvious for having less amenities so it is also not good for selection.
- ❖ The yellow cluster has very similar properties as Cyan so it is also a very bad candidate.
- ❖ The red cluster has no ATMs. The purple has few ATMs but is scarce in terms of Gyms and Shopping Malls.
- ❖ The red cluster is very scattered and purple is very dense in the area.
- ❖ The decision of choosing neighborhood now depends on distance, area of choice and which facilities are more important than others.

Conclusion

- ❖ The project overall helps person select best neighborhood to live in.
- ❖ The one limitation I can identify of this approach is that some small shops in small cities may not be registered on Foursquare and it would become difficult to take them into consideration while finding best fit of neighborhood.
- ❖ The other aspect of the project may help shop owners and businessmen to determine what kind of shops would be required in the area.
- ❖ If a person could identify basic needs of people living in the neighborhood than one place with all those facilities can be built and would give guaranteed business.