Analysing Neighbourhoods of Toronto and Scarborough

Introduction:

My Friend is working in Canada and has her home in the city of Toronto. She loves her neighbourhood because of all the great amenities and venues that exist such as Schools, hospitals, pharmacies, parks and so on. She has recently received a job offer from another company which she considers her dream company which is in another city. It has great career prospects for her. However this requires her to move from the current city if she decides to accept the offer. She wants to explore the neighbourhoods that are closer to the new job and has similar amenities which she has now.

Business Problem:

- My friend stays in the city of Toronto and has amenities like Good Schools, hospitals, food joints, hospitals, parks, pharmacies and grocery stores.
- Her new job is in the city of Scarborough and wants the same amenities as she has now like
- Schools, grocery stores, hospitals, pharmacies, parks food joints nearby.
- She decides to contact one of his friend who works on data science projects and take help of him to explore the two cities and get recommendations to have her home in the new city.

Data Needs:

- Get Pincode and neighbourhood information for Toronto and Scarborough from the Wikipedia site "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M"
    - This will help us to get information on potal code and neighbourhood details which can then be enriched with the latitude and longitude information
- Enrich the neighbourhood information with Latitude and longitude information through geocoder data set. http://cocl.us/Geospatial_data
    - This will help us to enrich the neighbourhood details with latitude and longitude information and then can be passed to four square API to get the details.
- Use Foursquare data on venue categories, top tips, location data, ratings **Fousquare API**
    - This data set will be the heart of the entire analysis. By using this API we will get all the venues in each neighbourhood.
    - Once we get data we can cluster and analyse each neighbourhood and share the analysis to help her get the desired information

Methodology

- The basic requirement is to get the data about neighbourhood of both the cities and get the venue and rating information
- In order to get an accurate city segmentation, the whole city of Toronto and Scarborough are our population i.e. all the neighbourhoods of Toronto and Scarborough
- Scrap the Wikipedia page read the Post Code table for Canada and convert into data frame

    - Cleanse the data to ignore Not Assigned values and logically combine the rows to get the distinct information.
    - Filter the data for the Borough of Toronto and Scarborough
    - Read the Geocoder data to get the geographical coordinates information and join the two data sets
    - This exercise needs to be done for both Toronto and Scarborough to get borough and neighbourhood information for further analysis
    - Access Foursquare location data of each neighbourhoods, using one hot technique to get top 10 venues of each neighbourhood.
    - Apply K-means Cluster to segment all the neighbourhoods into clusters.
    - Filter the data frame and utilize Foursquare to find grocery stores, parks, pharmacies etc. Around each neighbourhood
    - Analyse the data to check if both the neighbourhoods are similar i.e. have same venues nearby

- Results
    - Neighbourhoods in both the cities are alike with little plus and minuses in terms of preferences based on data. hence there should not be issues for my friend to relocate and accept the job offer of her dream company
- Discussion :
    - Both the cities has similar neighbourhoods still recommendations and top ratings may differ. So the preference should be made in the order of priority and checked if the criteria is met. While there are parks and school stadium it can be assumed that the areas have good schools too but there are no ratings or recommendation obtained from the data. Further analysis is needed on that if school rating is the no 1 criteria.
- Conclusion:.
    - Both the cities has similar neighbourhoods still recommendations and top ratings may differ. So the preference should be made in the order of priority and checked if the criteria is met.