DATA SCIENCE CAPSTONE: BATTLE OF NEIGHBORHOODS IN TORONTO

BUSINESS PROBLEM

Toronto is the heart of Ontario, Canada with a population of 2.93 Million people. It is the most densely populated and urban place of Canada, and known best for the CN tower, Scotiabank Arena, Toronto Islands and much more. Due to the popularity of this city, not only is there a massive tourist attraction, but a strong hold for respected businesses, universities, and residential areas as well. This makes it very challenging for coming entrepreneurs to choose locations in a large, developed city.

To this day, the city continuously expands with new venues creating a very competitive atmosphere for entrepreneurs looking to open location-based business whether it belongs to cuisine, retail, industrial or residential. The location of the venue tremendously impacts the strength of the business.

This project aims to aid entrepreneurs looking to open such businesses by exploring patterns into multiple neighborhoods situated in Toronto, by categorizing them into clusters in order to explore the strength and weakness of similar types of locations. Doing so can benefit entrepreneurs looking to open businesses by guiding them towards choosing the cluster that best suits their business.

DATA DESCRIPTION

The data used to analyze in this project has been collecting using the scraping method from the wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M along with using Four Square API's to get information of **venues**, and geocoder to obtain **coordinates**. A file is generated beforehand containing the coordinates of each **postal code**, named 'Geospatial Coordinates.csv'.

The data will be used to generate **clusters** to group similar neighborhoods together based on their characteristic, namely **venue types** and their concentration. This will help deduce the pattern between **neighborhoods**.

Here is an example of what the data looks like:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
5	Victoria Village	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place
6	Victoria Village	43.725882	-79.315572	Cash Money	43.725486	-79.312665	Financial or Legal Service
7	Regent Park, Harbourfront	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
8	Regent Park, Harbourfront	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
9	Regent Park, Harbourfront	43.654260	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot