Capstone Project - The Battle of Neighborhoods

INTRODUCTION: DISCUSS THE BUSINESS PROBLEM AND WHO WOULD BE INTERESTED IN THIS PROJECT

"Would you recommend a location in Toronto or New York City to open a new Peruvian restaurant?"

The increasing popularity of Peruvian food has been unstoppable since 2011 and the Culinary Institute of America named "the Year of Peruvian Cuisine" in 2014.

Therefore, my partner, who is a Peruvian entrepreneur and a passionate of food, wants to open a new Peruvian restaurant. He explains to me that the new restaurant should have shopping places nearby and transportation facilities within 5 minutes. He wants me concentrated on the selection of the restaurant's location according to its nearby environment. Restaurant facility and the rental price is not my concern.

This project will explore the similarities and dissimilarities between specific neighborhoods in the two cities, and determine which areas best fit the Peruvian restaurant customers

DATA SECTION: DESCRIBE THE DATA THAT WILL BE USED TO SOLVE THE PROBLEM AND THE SOURCE OF THE DATA

- Use data from the Toronto Neighborhood Wikipedia page to segment, cluster, and explore neighborhoods in Toronto. The dataset is found at https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
- Complete the information neighborhood table with latitude and longitude information
- Create a map of Toronto with neighborhoods. For this exercise, I will focus on the Downtown
- Define Foursquare credentials and version
- Analyze each Neighborhood
- Define top venues
- Cluster Neighborhood
- Use data from New York Neighborhood to segment, cluster and explore neighborhoods in NY
- The dataset is found at https://geo.nyu.edu/catalog/nyu_2451_34572
- Complete the information neighborhood table with latitude and longitude information
- Create a map of NY with neighborhoods - For this exercise, I will focus on Queens
- Define Foursquare credentials and version
- Analyze each Neighborhood
- Define top venues
- Cluster Neighborhood

METHODOLOGY: DISCUSS AND DESCRIBE ANY EXPLORATORY DATA ANALYSIS THAT YOU DID, ANY INFERENTIAL STATISTICAL TESTING THAT YOU PERFORMED, IF ANY, AND WHAT MACHINE LEARNINGS WERE USED AND WHY.

The Python packages that I used to explore and analyze the data were:

Nr.	Name of the package	Use of the package	Details of the package in the project
1	Requests	HTTP requests	HTTP requests would be made to this Foursquare API server using zip codes of Toronto and NY neighborhoods to find the latitude and longitude
2	Pandas	Data Analysis	Extensive analysis of two neighborhoods of Toronto and NY to get insights from the outcomes.
3	Numpy	Data in vectors	
4	Geopy	Retrieve location	Locate the coordinates of neighborhoods, addresses, cities, and landmarks across Toronto and NY.
5	Matplotlib	Plotting maps	Generate maps of data, vectors and lists of Toronto and NY
6	JSON	JSON files	Storage and exchange data
7	Folium	Map rendering	It would be used to visualize the neighborhoods cluster distribution of Toronto and NY city over an interactive leaflet map
	Sklearn	Machine Learning	Unsupervised ML algorithm K-mean clustering would be applied to form the clusters of different categories of places around the neighborhoods. These clusters would be analyzed individually, collectively and comparatively to derive the conclusions.

RESULTS SECTION: DISCUSS THE RESULTS

For Downtown Toronto Borough in Toronto, Canada, I used k-means to group the neighborhoods into 3 clusters:

- Cluster_0: It has 18 neighborhoods, and the most popular venues are coffee shops, international cuisine restaurants, and gyms.
- Cluster 1: It has one neighborhood, and the most popular venues are the coffee shops and gyms.
- Cluster 2: It has one neighborhood, and the most popular venues are the coffee shops and international cuisine restaurants.

For Queens Borough in New York, US, I used k-means to group the neighborhoods into 5 clusters:

- Cluster_0: It has two neighborhoods, and the most popular venues are mini markets and yoga studios.
- Cluster_1: It has four neighborhoods, and the most popular venue ar parks and yoga studios.
- Cluster_2: It has 161 neighborhoods, and the most popular venues are donut shops and international cuisine restaurants.
- Cluster_3: It has159 neighborhoods, and the most popular venues are the beach and a bakery.
- Cluster_4 has two neighborhoods, and the most popular venues are the beach and gyms

DISCUSSION SECTION: DISCUSS ANY OBSERVATIONS YOU NOTED AND ANY RECOMMENDATIONS YOU CAN MAKE BASED ON THE RESULTS.

Toronto has ten boroughs and 103 neighborhoods. Downtown Toronto borough has 1228 venues in 19 neighborhoods, and the most popular places are coffee shops and international cuisine restaurants that offer a menu for Toronto's famous financial district, which is similar to cities like New York or Chicago.

New York City has five boroughs and 612 neighborhoods. Queens borough has found 4160 venues in 162 neighborhoods, and the most popular sites determine the culinary offer of Queens, which is the largest and most diverse district in New York. Here the city's culture is being reinvented.

CONCLUSION SECTION: CONCLUDE THE REPORT.

In conclusion, based on the number of venues and a variety of places, I would suggest considering Queens over Downtown Toronto to locate a Peruvian restaurant because Queens, offers more options for extracurricular and outdoor activities for individuals and families. Besides, Queens is New York's largest borough, in which different cultures coexist, and it is the district that is evolving fast due to art, tourism, and cinema.