Final Capstone Project Report

Problem Statement

My family and I would like to go to Canada. We are justifying which city to choose from. Would it be Toronto, Canada or Montreal Canada. We need to see the similarities and dissimilarities of both cities. We will get to see the most common venues for both.

Project Background

Toronto and Montreal are two distinct flavor Canadian cities. Toronto is a financial capital of Canada and may be a modern historical cultural city as Montreal.

is seen as more historic, cultural and Europe-leaning, particularly regarding its French roots. As a data Scientist, we will see both city's venues.

Data & Resources used

List of postal codes of Canada
Wiki: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M for access to neighborhood data of Toronto region.

List of postal codes of Canada
Wiki https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_H for access to neighborhood data of Montreal region.

Geolocator or geographical coordinates of the neighborhoods: http://cocl.us/Geospatial_data for getting the longitude and latitude data for the neighborhoods.

Foursquare database: https://Foursquare.com to be used in order to explore the desired neighborhood data for various restaurant details and access the JSON files. This data shall be utilized to map the Indian restaurants in various locations. Problem solving strategy

Step-1: Using Beautiful Soup, web scraping of the neighborhood data from postal codes of Canada Wiki-link. Clean the Toronto data by removing the missing values and store the data in a python Dataframe consisting of three columns namely: PostalCode, Borough, and Neighborhood.

Step-2: Using Beautiful Soup, web scraping of the neighborhood data from postal codes of Canada Wiki-link. Clean the Montreal data by removing the missing values and store the data in a python Dataframe consisting of three columns namely: PostalCode, Borough, and Neighborhood.

Step-3: Either use geolocator or take the help of long-lat data from the geospatial data wikilink and append the geographical coordinates in the above dataframes to get new respective dataframes for further analysis.

Step-4: Getting location data using the Foursquare API. It will be used to retrieve information of the common venues in Toronto and Montreal neighborhoods. The API will return a JSON file which will be further converted into a Python Dataframe.

Step-5: Exploratory Data Analysis

Plot the prepared location data on the map for visualization purposes. Plot and find the relationship between Neighborhoods and Various Restaurants for both cities. List down the 10 most common

venues for both cities. Step-6: Lastly we will discuss the results based on the above findings and provide a snapshot of both cities which will help make a exploratory choice of our family vacation.

Results

The presentation and the jupyter notebook will show the results of my findings to determine my next vacation.

Conclusion

Pulling data from a Wikipedia website is challenging. Beautiful soup helped determined how to pull data and organize it the best way a data scientist can. Pandas helped tremendously to get the dataframes needed to complete the task.

Conecting to Foursquare API assists to get the venues for both cities. Folium maps assists in generating the visualization to see the hot spots.

General python coding is the great way to accomplish the goal as a data scientist needed to help determine the end goal. In my case the determination of the next big vacation.