Coursera Capstone Project
The Battle of Neighborhoods (Week 2)
By: Sabanah Salvatierra
Introduction and Data Sections

Introduction Section :

Discussion of the business problem and the interested audience in this project.

Data Section: - Description of the data that will be used to solve the problem and the sources.

Methodology section - Discussion and description of exploratory data analysis carried out, any inferential statistical testing performed, and if any machine learnings were used establishing the strategy and purposes.

Results section - Discussion of the results.

Discussion section - Elaboration and discussion on any observations noted and any recommendations suggested based on the results.

Conclusion section - Report Conclusion.

1. Introduction Section : Discussion of the business problem and the audience who would be interested in this project. Description of the Problem and Background Scenario: Let's say I am a data scientist residing in Toronto. A business firm in Manhattan, NY hire me for data analysis to open a restaurent. Both NY and Toronto cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighborhoods of the two cities and determine how similar or dissimilar they are. In order to make a comparison and evaluation of the rental options in Manhattan NY, I must set some basis, therefore the property in Manhattan must meet the following demands:

desired location must be business property desired location is in the Manhattan area, New York should be near to the metro stations price of rent not exceed $7,000 per month desirable to have venues such as coffee shops, restaurants Asian Thai, wine stores, gym and food shops

Business Problem: The challenge is if someone is looking to open a restaurant in Manhattan area, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office?

2. Data Section: Description of the data and its sources that will be used to solve the problem Description of the Data:¶ The following data is required to answer the issues of the problem:

List of Boroughs and neighborhoods of Manhattan with their geodata (latitud and longitud) List of apartments for rent in Manhattan area with their addresses and price Venues for each Manhattan neighborhood (than can be clustered) How the data will be used to solve the problem The data will be used as follows:

Use Foursquare and geopy data to map top 10 venues for all Manhattan neighborhoods and clustered in groups (as per Course LAB) Use foursquare and geopy data to map and identify the venues and ammenities near each metro station Use Foursquare and geopy data to map the location of rental places create a map, the average rental price per square ft Addresses from rental locations will be converted to geodata(lat, long) using Geopy-distance and Nominatim. The procesing of these DATA will allow to answer the key questions to make a decision: what is the cost of rent (per square ft)? what is the area of Manhattan with best rental pricing that meets criteria established? What are the venues of the two best places to open a restaurent? How the prices compare? How venues distribute among Manhattan neighborhoods and around metro stations?

Reference of venues around current residence in Toronto for comparison to Manhattan place

neighborhood_latitude = 43.6532

neighborhood_longitude= -79.3832

Reference: 'https://api.foursquare.com/v2/venues/explore?&client_id=MDCU1FDAMJDPK1 WSLUJE32I0GVXHSKQX23DXCYY0K0BLVJQY&client_secret=LGELKAGOO5ULDQ W4WDXKRUL13GIQCT3EZVDB12NNU5QVBT2Q&v=20180604&ll=43.6532,-79.3832 &radius=500&limit=100'

	name	categories	lat	lng
0	Downtown Toronto	Neighborhood	43.653232	-79.385296
1	Nathan Phillips Square	Plaza	43.652270	-79.383516
2	LUSH	Cosmetics Shop	43.653557	-79.380400
3	Indigo	Bookstore	43.653515	-79.380696
4	CF Toronto Eaton Centre	Shopping Mall	43.654447	-79.380952
5	M Square Coffee Co	Coffee Shop	43.651218	-79.383555
6	Crepe Delicious	Fast Food Restaurant	43.654536	-79.380889
7	Poke Guys	Poke Place	43.654895	-79.385052
8	UNIQLO ユニクロ	Clothing Store	43.655910	-79.380641
9	John & Sons Oyster House	Seafood Restaurant	43.650656	-79.381613

Data Description We require geographical location data for Toronto. Postal codes in each city serve as a starting point. Using Postal codes we use can find out the neighborhoods, boroughs, venues and their most popular venue categories.

3. Methodology

We will be creating our model with the help of Python so we start off by importing all the required packages.

Package breakdown: Pandas : To collect and manipulate data in JSON and HTMl and then data analysis requests : Handle http requests matplotlib : Detailing the generated maps folium : Generating maps of Toronto sklearn : To import Kmeans which is the machine learning model that we are using. The approach taken here is to explore each of the cities individually, plot the map to show the neighbourhoods being considered and then build our model by clustering all of the similar neighbourhoods together and finally plot the new map with the clustered neighbourhoods. We draw insights and then compare and discuss our findings.

5. Results and Discussion The neighbourhoods of Toronto are very mulitcultural. There are a lot of different cusines including Indian, Italian, Turkish and Chinese. Toronto seems to take a step further in this direction by having a lot of Restaurants, bars, juice bars, coffee shops, Fish and Chips shop and Breakfast spots. It has a lot of shopping options too with that of the Flea markets, flower shops, fish markets, Fishing stores, clothing stores. The main modes of transport seem to be Buses and trains. For leisure, the neighbourhoods are set up to have lots of parks, golf courses, zoo, gyms and Historic sites. Overall, the city of London offers a multicultural, diverse and certainly an entertaining experience. Paris is relatively small in size geographically. It has a wide variety of cusines and eateries including French, Thai, Cambodian, Asian, Chinese etc. There are a lot of hangout spots including many Restaurants and Bars. Paris has a lot of Bistro's. Different means of public transport in Paris which includes buses, bikes, boats or ferries. For leisure and sight seeing, there are a lot of Plazas, Trails, Parks, Historic sites, clothing shops, Art galleries and Museums. Overall, Paris seems like the relaxing vacation spot with a mix of lakes, historic spots and a wide variety of cusines to try out.

6. Conclusion The purpose of this project was to explore the cities of Toronto and see how attractive it is to potential tourists and migrants. We explored both the cities based on their postal codes and then extrapolated the common venues present in each of the neighbourhoods finally concluding with clustering similar neighbourhoods together. We could see that each of the neighbourhoods in both the cities have a wide variety of experiences to offer which is unique in it's own way. The cultural diversity is quite evident which also gives the feeling of a sense of inclusion. Toronto seem to offer a vacation stay or a romantic getaway with a lot of places to explore, beautiful landscapes, amazing food and a wide variety of culture. Overall, it's upto the stakeholders to decide which experience they would prefer more and which would more to their liking.