**Coursera Capstone Project**

**The Battle of Neighbourhoods**

**Angeletti Giovanni M.**

**26 May 2021**

1. **Introduction Description of the Problem and Background Scenario:**

I am a data scientist residing in Toronto. A business company in Manhattan, NY hire me for data analysis to open a restaurant. Both NY and Toronto cities are very different and are the financial capitals of their respective countries. One interesting idea would be to compare the neighbourhoods 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 is in the Manhattan area, New York.

• should be near to the metro stations.

• price of rent does not exceed $7,000 per month.

• desirable to have venues such as coffee shops, restaurants Asian Thai, wine stores, gym, food shops and so on.

Business Problem:

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

1. **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 neighbourhoods of Manhattan with their geodata (latitude and longitude)

• List of property for rent in Manhattan area with their addresses and price

• Venues for each Manhattan neighbourhood ( than can be clustered)

• List Subway metro stations

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 amenities 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 Geopydistance and Nominatim.

The processing of these DATA will allow to answer the key questions to decide:

• 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 restaurant? How the prices compare?

• How venues distribute among Manhattan neighbourhoods and around metro stations?

1. **Methodology section:**

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

The analysis and the strategy:

The strategy is based on mapping the above-described data, to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed: location near a subway, rental price, and similar venues to Toronto. This visual approach and maps with popups labels allow quick identification of location, price, and feature, thus making the selection very easy.

The processing of these DATA and its mapping will allow to answer the key questions to decide:

• What is the cost of available rental places that meet the demands?

• What is the cost of rent around a mile radius from each subway metro station?

• What is the area of Manhattan with best rental pricing that meets criteria established?

• How venues distribute among Manhattan neighbourhoods and around metro stations?

• Are there trade-offs between size and price and location?

• Any other interesting statistical data findings of the real estate and overall data.

**Methodology execution – Mapping Data**

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

Table

Description automatically generated

Map of Toronto with venues near CBD place – for reference

A picture containing text, bedclothes

Description automatically generated

**Manhattan Neighborhoods – Data and Mapping**

Build top 10 cluster dataframe for near by restaurants, cafè, gym, parks and so on, using Foursquare API.

I used the csv file with cluster neighborhood data that was produced with Foursquare during the course lab work and create a map.

A picture containing text, receipt

Description automatically generated

**Manhattan Rental Data**

Extract rental data from web scraping

Table

Description automatically generated

Price analysis and findings:

* Rental price mean value is $7000 per month as shown in the graph below

Chart, histogram

Description automatically generated

Manhattan rental data Map for price and geo location

Map

Description automatically generated

Merged venue cluster Map and rental data Map

A picture containing chart

Description automatically generated

Filter the cluster numbers to explore

Graphical user interface

Description automatically generated

Get Manhattan subway metro station List and mapped

Map

Description automatically generated

1. **Results:**

Now all maps are consolidated into one for final decision making

Map

Description automatically generated

Property Selection

Using the map above, I was able to explore all possibilities since the popups provide the information needed for a good decision.

Cluster 2 Sutton Place is the best choice for the rental for the requirements and budget

305 East 63rd Street rental cost is US7500 slightly above the US7000 budget but located 400 meters from "59 St-Lexington Av subway station" and very near to cluster 2, Sutton Place.

150 East 57th Street rent cost is USD5950, under the US7000 budget. The same distance from "59 St-Lexington Av Subway Station" but it is a bit more far from cluster 2, Sutton Place.

Based on all information 305 East 63rd Street is the best choice as proximity to Cluster 2 will make more traffic to the shop.

1. **Discussion**

In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course. I feel this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned. I have created a good project that I can present as an example to show my potential. I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases.

1. **Conclusion**

I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend for use in similar situations.