MEHMET
KORKUT
mehmetkorkut77@gmail.com

Applied Data Science Capstone
Battle of Cities and Neighborhoods

2021-08-05

1 Introduction

This Project is based on a scenario. In this scenario, there is a man called Ismet. He is a fresh MSc Finance Graduate from one of the top universities of Europe. He finally found jobs after seeking about 6 months.Meanwhile, he was working on data analytics skills in order to be successful in this area. These jobs are in Manhattan,NY and Toronto,Ontario. Since these two job offers are almost the same in terms of salary and expectations, he decided to analyze and compare the cities. He has to decide in one week so he needs to be really quick on that task. The choice is too important for him because he will spend at least 4 years there. Nobody wants to live in a place that s(he) does not want
These cities are 2 big financial capitals of their countries and they are too big. He realized that he is capable of doing such an analysis based on data science skills on the area of office which they have to attend. Therefore, he can answer these type of questions;

1-Are these two cities similar?

2-Which city matches with my personal needs-hobbies?

3-Which city has more green area?

4-Which neighborhood should he choose to live after selecting the city?

In order to answer these questions, he needs location data. He uses the Foursquare API for gathering data to make graphical and statistical analysis of venues in the city.

The rest of the report is as follow;

After defining the problem in introduction, dataset and the source is presented. Then, in methodology section, there are explanatory data analysis inferential statistical testing. In addition, there is an explanation about which machine learning model is used and why. In results section, results of the model is discussed and come up with an answer to the questions. Then, I discuss the observations and drawbacks and make some recommendations for future work. Finally, I conclude the report with a conclusion part.

2 Data

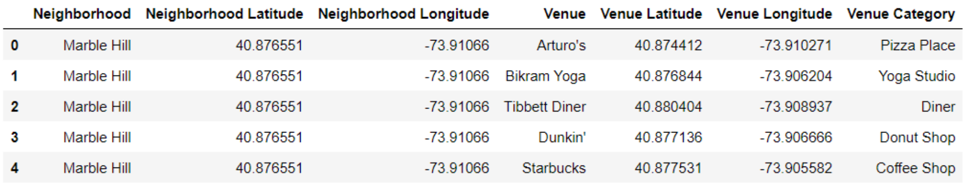
There are 3 sources of data in total. They are Foursquare API for venues data, Wikipedia for the Postal Codes,Borough and Neighborhood data of Toronto(gathered by BeautifulSoup) and IBM Cloud which has New York Postal Codes,Borough and Neighborhood data(csv format).
Sources :

- `https://developer.foursquare.com/docs/places-api/`
- `https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M`
- `https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS07`
  `labs/newyork_data.json`
- `https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS07`
  `labs_v1/Geospatial_Coordinates.csv`

The data formats are in csv and json format. Data are transformed into pandas dataframe and combined. The variables in datasets are borough,neighborhood name, postal code, latitude,longitude,venue and venue category.

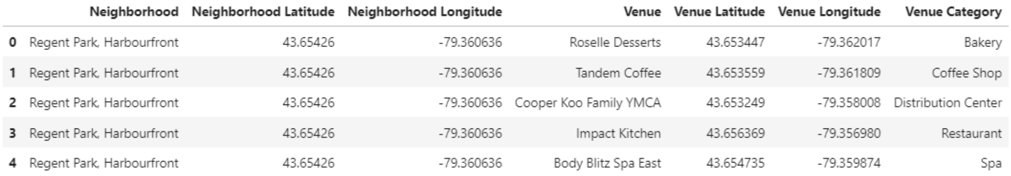
Note :The first 5 rows of datasets are presented in the next page

Manhattan Data



	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

Toronto Data



	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
4	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa