A description of the data and how it will be used to solve the problem

Gaby Bou Tayeh

March 2020

1 Description of the used data

In order to take a final decision on which apartment to rent in Manhattan NY, the following data is required:

- Information on Manhattan's neighborhoods along side the coordinates of each neighborhood(latitud,longitud).
- Information on the subway metro stations in Manhattan (location, coordinates).
- a List of apartments that are published for rent in Manhattan area with descriptions (price, location, address, etc..)
- List of venues in the Manhattan neighborhoods (e.g. top 10)

mh_neigh_data.tail():

	Borough	Neighborhood	Latitude	Longitude
35	Manhattan	Turtle Bay	40.752042	-73.967708
36	Manhattan	Tudor City	40.746917	-73.971219
37	Manhattan	Stuyvesant Town	40.731000	-73.974052
38	Manhattan	Flatiron	40.739673	-73.990947
39	Manhattan	Hudson Yards	40.756658	-74.000111

Figure 1: mh_neigh_data.csv

1.1 Data sources

The list of Manhattan neighborhoods is worked out during Lab exercise of the applied data science course (A csv file containing the final results was saved locally which will be retrieved later and read into a Pandas Dataframe). a part of the csv file 'mh_neigh_data.csv' content is illustrated in Figure 1.

A list of Manhattan subway metro station was prepared and it was completed by scrapping data from this wikipedia page(https://en.wikipedia.org/wiki/List_-of_New_York_City_Subway_stations_in_Manhattan). Additional information where also scrapped from NY Transit authority and Google maps (https://www.google.co-m/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,-11z/data=!3m1!4b1). The geolocation of the station were obtained using Nominatim. Details will be shown later in the Notebook. The subway csv file is "MH_subway.csv"' and below is an illustration of a part of it:

mhsub.tail(): sub_station sub_address lat long

17	190 Street Subway Station	Bennett Ave, New York, NY 10040, USA	40.858113	-73.932983
18	59 St-Lexington Av Station	E 60th St, New York, NY 10065, USA	40.762259	-73.966271
19	57 Street Station	New York, NY 10019, United States	40.764250	-73.954525
20	14 Street / 8 Av	New York, NY 10014, United States	40.730862	-73.987156
21	MTA New York City	525 11th Ave, New York, NY 10018, USA	40.759809	-73.999282

Figure 2: MH_subway.csv.csv

A list of appartements for rent was collected by web-browsing real estate companies in Manhattan : http://www.rentmanhattan.com/index.cfm?page= -search&state=results https://www.nestpick.com/search?city=new-york&page=1&ord-er=relevance&district=manhattan&gcid=CjwKCAiAjNjgBRAgEiwAGLlf2hkP3A-cPxjZYkURqQEswQK2jKQEpv_MvKcrIhRWRzNkc_r-fGi0lxoCA7cQAvD_BwE&t-ype=apartment&display=list https://www.realtor.com/apartments/Manhattan_NY A csv file ("MH_flats_price.csv") was compiled indicating: location of the apartment (Area), address, number of beds, and monthly rental price. Nominatim is used again to retrieve the coordinates (lat, long) of each apartment.

1.2 How the Data can serve our objective?

The data will be used as follows:

The Nominatim (from geopy) and the Foursquare API will be used to retrieve the coordinates of the clustered neighborhoods, apartments, metro station and display them on an interactive map using Folium. Information about the name of each neighborhood and borough, the price of each apartment, and the distance between apartments and their nearest metro station will be easily accessible using the interactive map. All this information will help us decide which apartment to rent.