

Coursera Capstone project

Coursera IBM Data Science Certification

BOUTEMINE
MOUNA

March 2020

Data Section

Data Requirements

- Geodata for current residence in Singapore with venues established using Foursquare.
- List of Manhattan (MH) neighborhoods with clustered venues established via Foursquare (as in Course Lab).
https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods
- List of subway metro stations in Manhattan with addresses and geo data (lat,long):
https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan , (<https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/data=!3m1!4b1>)
- List of apartments for rent in Manhattan area with information on neighborhood location, address, number of beds, area size, monthly rent price and complemented with geo data via Nominatim (<https://www.nestpick.com/new-york/>)
- Place to work in Manhattan (Park Avenue and 53rd St) for reference

Data Sources, Data Processing and Tools used

- Singapore data and map is to be created with use of Nominatim , Foursquare and Folium mapping
- Manhattan neighborhoods were obtained from Wikipedia and organized by Neighborhoods with geodata via Nominatim for mapping with Folium.
- List of Subway stations was obtained via Wikipedia, NY Transit web site and Google map,
- List of apartments for rent was consolidated from web-scraping real estate sites for MH. The geolocation (lat,long) data was found with algorithm coding and using Nominatim.
- Folium map was the basis of mapping with various features to consolidate all data in ONE map where one can visualize all details needed to make a selection of apartment