Coursera Data Science Capstone Project

May 01, 2020

1.2 Data

For analyzing potential Hispanic and Asian restaurants and the place to open one, different data sources will be used. For a profound decision multiple factors influence the decision:

- Distribution of different restaurants in all boroughs of NYC
- Demographic distribution of Hispanics and Asian in the boroughs of NYC

The following data sources will be used:

- A Geojson file to get the location data of NYC. This file is important for visualization and the coordinates of the boroughs in NYC.
- The number, type, and location of all restaurants will be obtained using the **Foursquare API.** The Foursquare API makes it possible to receive information and coordinates of a venue. [7] Only the latitude and longitude are needed to get the data from Foursquare. The complete list is after that filtered for restaurants, to get only relevant venues for this problem.
- Census data of NYC to get the demographic distribution of Hispanics and Asians. The data will be obtained from **Kaggle.** The Census data is built up of two CSV files. One contains the percentage of ethnic demographic information (for example Hispanics) for a specific latitude/longitude. The other CSV file contains the coordinates with Census numbers. These two files will be cleaned and merged, to get one dataframe. [6]

Bibliography

[1]
http://demographia.com/db-worldua.pdf

[2]
http://www.ameredia.com/resources/demographics/asian_american.html

[3]
https://www.statista.com/statistics/259776/number-of-people-who-went-to-restaurants-in-new-york-by-type/

[4]
https://www.asanet.org/news-events/footnotes/jun-jul-aug-2019/features/latino-population-new-york-city

[5]
https://python-visualization.github.io/folium/

[6]
https://www.kaggle.com/muonneutrino/new-york-city-census-data

[7]
https://developer.foursquare.com/