Applied Data Science Capstone Project

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

Data description

The New York City population with borough and neighbourhood tabulation areas in CSV file format was downloaded from data.gov, which is the home of the U.S. Government's open data source. The data was originally collected from Census Bureaus' Decennial data dissemination. Neighbourhood Tabulation Areas (NTAs), are aggregations of census tracts that are subsets of New York City's 55 Public Use Microdata Areas. The data reveals change in population from 2000 to 2010 for each NTA. In the capstone project, population of year 2010 was used for exploration of the neighbourhoods. The link and the short print screen of the data are given below.

https://data.cityofnewyork.us/api/views/swpk-hqdp/rows.csv?accessType=DOWNLOAD

	Borough	Year	FIPS County Code	NTA Code	NTA Name	Population
0	Bronx	2000	5	BX01	Claremont-Bathgate	28149
1	Bronx	2000	5	BX03	Eastchester-Edenwald-Baychester	35422
2	Bronx	2000	5	BX05	Bedford Park-Fordham North	55329
3	Bronx	2000	5	BX06	Belmont	25967
4	Bronx	2000	5	BX07	Bronxdale	34309

Unlike New York City, the Greater London population with borough was not found in CSV file format. It was imported as a table from website address of: https://www.citypopulation.de/php/uk-greaterlondon.php. The table entails the population of each borough in Great London at different years. For the sake of fair comparative analysis with City of New York, in this project the population of the Great London which was estimated in June 2011 was adopted. Short print screen of the population table is illustrated below.

Name	Status	Population Estimate 1981-06-30	Population Estimate 1991-06-30	Population Estimate 2001-06-30	Population Estimate 2011-06-30	Population Estimate 2017-06-30
Barking and Dagenham	Borough	161,300	155,500	165,700	187,029	210,711
Barnet	Borough	295,200	297,700	319,500	357,538	387,803
Bexley	Borough	217,400	218,100	218,800	232,774	246,124
Brent	Borough	248,300	240,800	269,600	312,245	329,102
Bromley	Borough	299,200	293,500	296,200	310,554	329,391

In order to visualize the population of each borough of New York City and Greater London in choropleth map, boundaries of boroughs of both cities were imported. The polygon boundaries of boroughs of New York City in GEOJSON file format is downloaded from BetaNYC's Community Data Portal. The link location is: http://data.beta.nyc//dataset/68c0332f-c3bb-4a78-a0c1-32af515892d6/resource/7c164faa-4458-4ff2-9ef0-09db00b509ef/download/42c737fd496f4d6683bba25fb0e86e1dnycboroughboundaries.geojson. Whereas, the polygon boundaries of boroughs of Greater London in GEOJSON file format is accessed from: http://darribas.org/gds15/content/labs/data/london_boroughs.geojson.

Data implementation to solve the problem

After collecting the necessary population data of New York City and Greater London, the next step was data preparation. Data of both cities were arranged in the same manner. The population in each borough in New York City and Greater London was tabulated. All unnecessary data for this project are removed from the table. Fortunately, there were not null values in the source data. The data were visualized using choropleth map to get more insight about the population neighbourhood in the two big cities (New York City and Greater London).

As the main objective of this capstone project is to compare which city is more eco- and vegan-friendly for travellers, three search queries for each city centres using geocoder library and venues from Foursquare API were performed. The search queries were i) Vegan, ii) Metro Stations, and iii) Bike rental. Based on the results of the three quarries provided by the foursquare, the city which is more eco-and vegan-friendly was discovered.