Applied Data Science Capstone Project Report

Introduction

According to the World Travel and Tourism Council, tourism industry contributes to around 10 percent of the world's economy. Though this industry contributes for significant share to the world's economy, it accounts for around eight percent of global greenhouse gas emissions. As the world gets richer and there is more demand for travelling, which in turn the tourism sector will grow faster than many other economic sectors. For instance, the global tourism industry as a whole has been growing at an annual rate of around five percent, outpacing the growth of international trade.

In present time, eco-friendly tourism is becoming a movement in order to mitigate global greenhouse gas emissions caused by the tourism sector. Some of the best way to discover a new place with minimal impact on the environment is by hiring a bicycle or using a public transport. Indeed, biking is not only better for the environment; it will also help to burn a few extra kilojoules after feasting on all that local grub.

In connection with tourism industry, there has been an increase in the demand for vegan travel. According to Intrepid Travel's Chief Purpose Officer, Leigh Barnes, "The vegan movement continues to grow, with millennials especially being more interested in vegan lifestyle and vegan tourism". Some popular tourist destinations have already seen a major explosion in the number of vegan restaurants in the past few years.

In this applied data science capstone project, two cities which are the top tourist destination in the world are explored. These cities are New York City, USA and Greater London, UK. The neighbourhoods of both cities are segmented and clustered. The ultimate goal of this project is to explore and get insight regarding which city is more eco- and vegan-friendly for health and environmental conscious travellers from all over the world.

Data

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. The data was 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 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.

	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.

Methodology

After collecting the necessary population data of New York City and Greater London, the next step was data preparation. Both data were arranged in the same manner. The population in each borough in New York City and Greater London was tabulated. Unnecessary data for the project are removed from the table. Unfortunately, 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 (Manhattan and City of London) using geocoder library and venues from Foursquare API were performed. The search queries are i) Vegan, ii) Metro Stations, and iii) Bike rental. The 2.5 km radius from the city centre was set criteria for search quarries. The limit for search result was set to 50. Based on the results of the three quarries provided by foursquare, the city which is more eco- and vegan-friendly was identified.

Results

New York City has five boroughs whereas Greater London has 33 boroughs ((including City of London). Indeed, City of London is not a borough but an independent county, but it this project it counts as borough). Among all borough of New York City, borough called 'Brooklyn' has the highest population size of 2,504,700 followed by 'Queens' with population size of 2,230,722. The borough named 'Staten Island' has the least population size which is 468,730. In case Greater London, borough called 'Croydon' has the highest population size of 364,815 whereas the borough named 'City of London' has the least population size which is 7,412. Though both cities have different number of borough and different population size in their respective borough, their total population size is quite comparable. The total population of New York City is 8,175,133 and Greater London has 8,204,407. Population of each borough in City of New York and Greater London is illustrated in Figure 1 and 2, respectively.
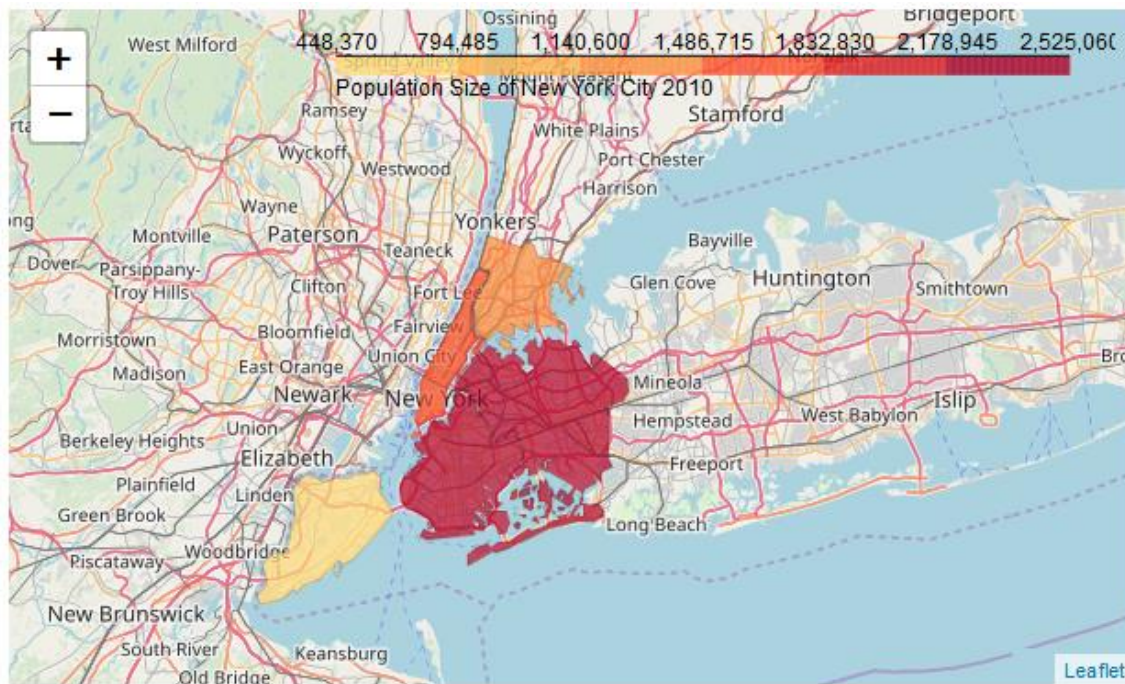
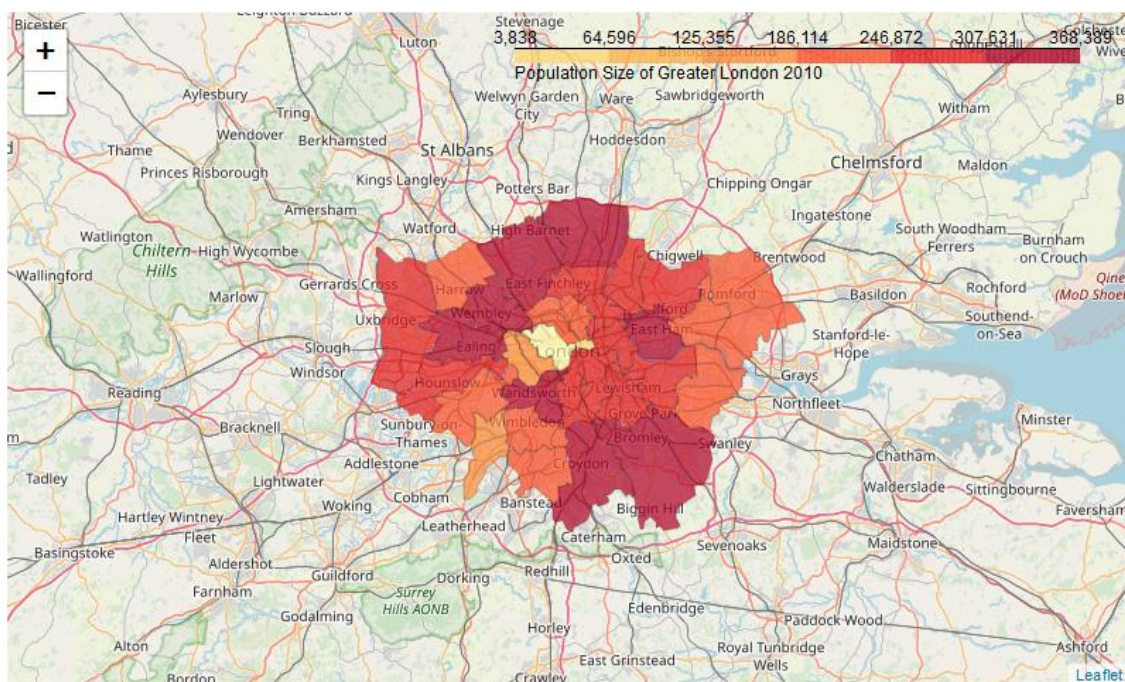Figure 1. Choropleth map of New York City's population.

Figure 2. Choropleth map of Greater London's population.

As discussed earlier, the ultimate goal of this project is to discover which city (New York City or Greater London) is eco- and vegan-friendly city for travellers. To examine this three search queries for each city centres (Manhattan and City of London) using geocoder library and venues from Foursquare API were

performed. The search queries are i) Vegan, ii) Metro Stations, and iii) Bike rental. Based on the results of the three quarries provided by foursquare, the city which is more eco-and vegan-friendly was identified. The search results of the three queries are given in Figure 3, 4 and 5 respectively. It can be clearly noticed from the figures that number of bike rental services within five kilometre diameter from Manhattan city centre is higher than City of London. The choice for vegan restaurants and assess to public transport in City of London are much better than Manhattan.

New York City
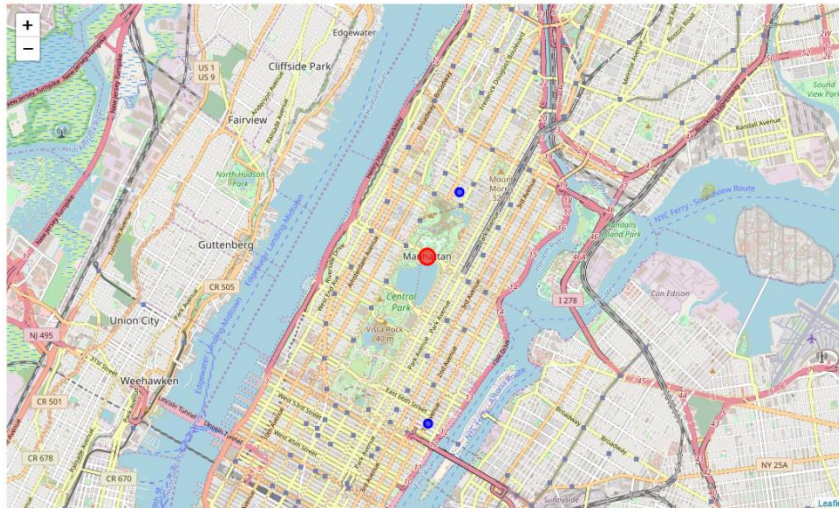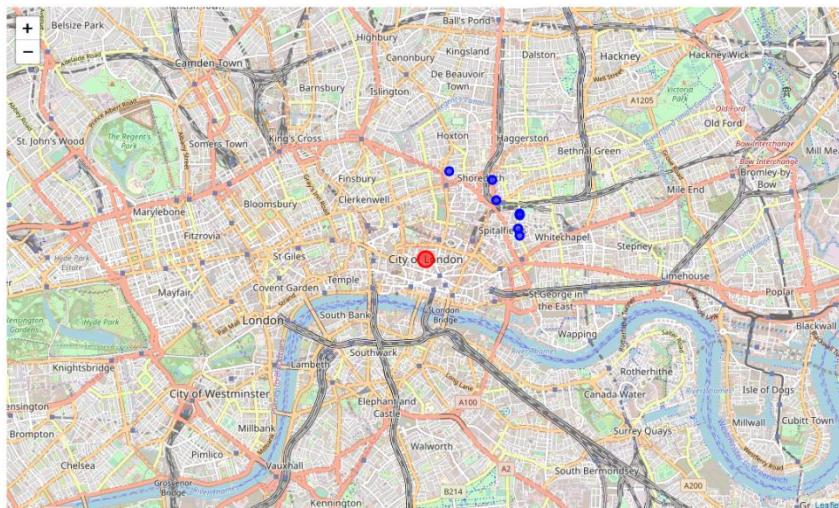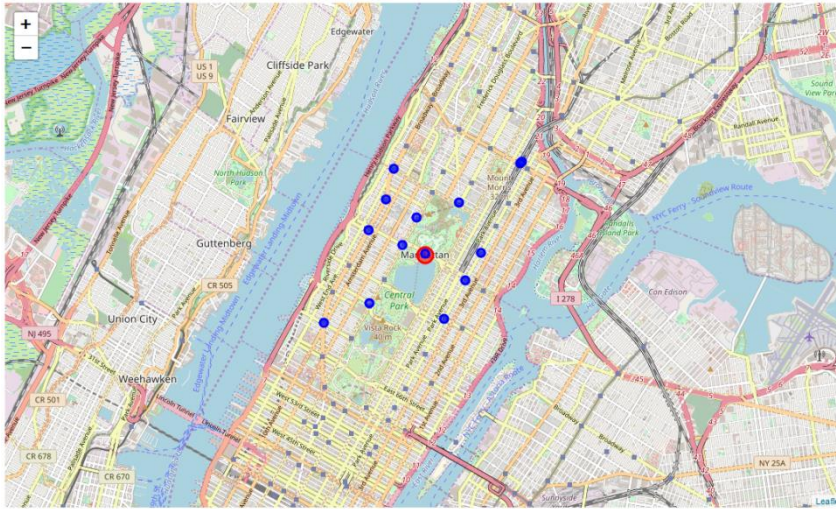

Greater London

Figure 3. Vegan restaurants stations in New York City and Greater London
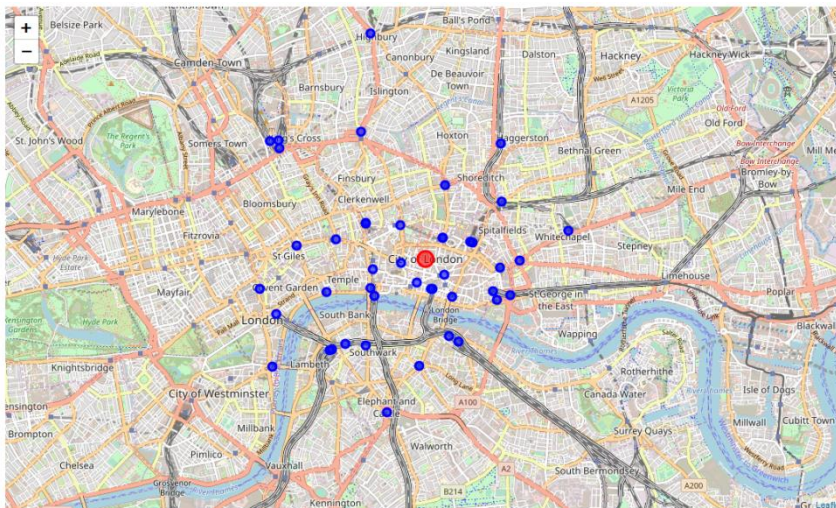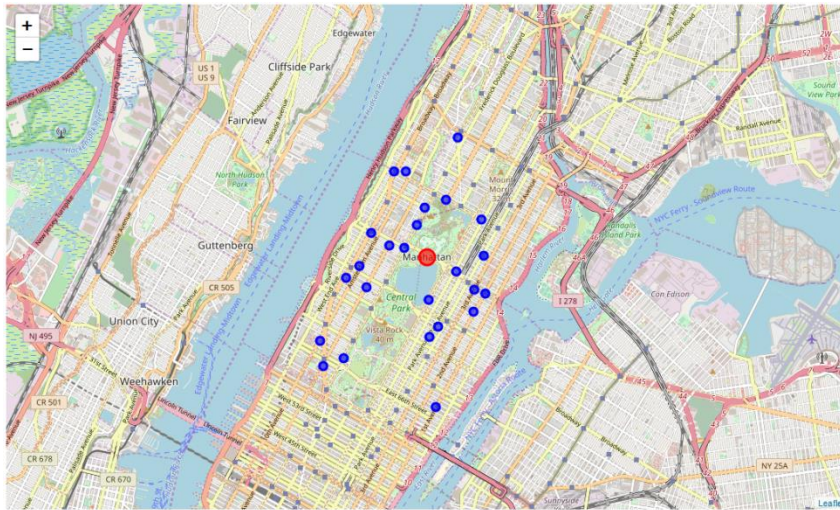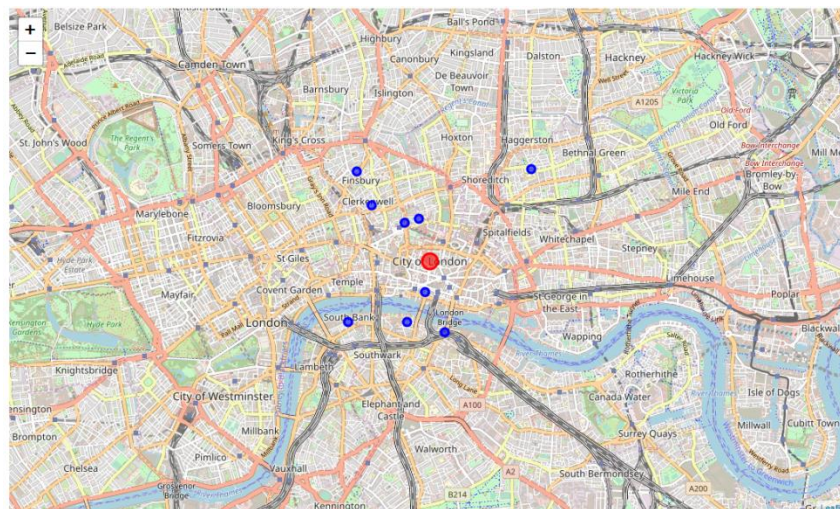
New York City

Greater London

Figure 4. Metro stations in New York City and Greater London

New York City

Greater London

'Figure 5. Bike rental services in New York City and Greater London

Discussion

The adopted Foursquare API tool is very helpful in exploring a geographical location and solves the described problem. Indeed, the Foursquare API yields not fully matched and even unrelated results for some search queries. For instance, for bike rental search, it provides such irrelevant results: Mobile Phone Shops, General Travel, Government Buildings, and Indian Restaurants. Hence, careful check of the results is vital in order to remove unrelated lists from the dataframe. It is also worth to mention that finding suitable data is somewhat challenging and time consuming.

Conclusions

From data segmentation and clustering, it was learned that New York City has five boroughs and Greater London has thirty three boroughs (including City of London). Though Greater London has many boroughs, both cities have almost equal population. From exploration of the geographical location, Greater London has a great number of accesses to public transport. The number of bike rental services within five kilometre diameter from Manhattan city centre is higher than City of London. The choice for vegan restaurants in City of London is much better than Manhattan. In general, both cities can be considered as eco-friendly cities but City of London is more vegan-friendly than New Your City.