Course: Applied Data Science Capstone (IMB/Coursera)

Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods (Week 1)

Author: Amedeo Amiti

Date: 07 February 2020

Introduction/Business Problem

For this final assignment I imagined to be a data analyst hired by a fictional company named "CIBITALIA".
CIBITALIA is an Italy-based business operating in the field of hospitality/food & beverage. They specialize in providing consultancy services to Italian restaurants. Their services encompass all the aspects of business and restaurant management including (but not limited to):
- marketing and advertisement
- styling/restyling
- rebranding
- web & social media management
- food supply chain
- sourcing/procurement
- recruitment
- accountancy

From their HQ in Italy, the company has successfully expanded abroad and is currently active in some of the main cities of the world: Paris, London, Tokyo, Dubai, Sydney, Rio de Janeiro and Berlin, just to name a few.
CIBITALIA intends now to enter the USA market and have targeted the city of New York as primary area of interest.
As a first step, they would like to capture some basic information before proceeding with their project.
They asked me to run an exploratory analysis of New York City and its Italian restaurant. The expected result would be a snapshot of the current situation in New York as regards the number, the location and the distribution of all Italian restaurants. In particular, the company is interested in some specific data, upon which they will make decision and structure their preliminary business plan. Therefore, they asked me to investigate and generate insights about the following points:
- How many Italian restaurants are there in New York City?
- How they distribute across the 5 boroughs of New York City?
- Considering the demography of New York City, how Italian restaurants distribute in relation to the local population?
- Which are the local communities (i.e.: neighbourhoods) with the highest number of Italian restaurant?
My task is to retrieve, leverage, manipulate and analyse various datasets in order to come up with answers to the above queries.

The company also asked me to interpret the results and, based on my findings, to make recommendations on how to best approach their entrance to this new market.

Data

I will be using several datasets containing different types and formats of data. Here is the list of all datasets I will be drawing insights from.

Dataset name:	NYC_neighbourhood_names
Source:	https://cocl.us/new_york_dataset
Format:	.json
Description:	A dataset made available by Cognitive Class. This dataset contains the names of all the neighbourhoods of New York City, together with the name of the boroughs each neighbourhood belongs to and the geographic coordinates (latitude, longitude) of each neighbourhood. These data will be sorted and used as the primary base for further processing as they relate each neighbourhood with the relevant borough. They will also be used for map visualization as they provide coordinates for each neighbourhood.

Dataset name:	Italian restaurant of New York City
Source:	.json
Format:	Foursquare
Description:	This dataset is the result of a call to Foursquare API which returns the names of all the venues within a certain range. Together with the name of the venues, the call will also return an unique venue ID and the venue category (e.g.: bar, coffee shop, fast food, Italian restaurant, etc.). These results will be duly filtered, consolidated and saved (for practical reasons) as a .csv file named "ita_rest_nyc", which will serve as a base for further processing.

Dataset name:	NYC_borough_population
Source:	https://data.cityofnewyork.us/City-Government/NYC-Population-by-Borough/h2bk-zmw6
Format:	.csv
Description:	This dataset is provided by NYC OpenData and contains the population for each borough in New York City. Data updated to August 2016. These data will be combined with previous results in order to obtain useful insights on the ratio between restaurants and population.

Dataset name:	NYC_borough_boundaries
Source:	https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

Format:	.geojson
Description:	This dataset is provided by NYC OpenData and contains GIS data to spatially reference the boundaries of New York City boroughs. In other terms, the data from this dataset are polygons whose vertices define (on a map) the borders of each NYC borough. This dataset is essential for the production of choropleth maps.