Nick Smart

IBM Data Science Professional Certificate: Capstone Project

Introduction

The City of New York, New York is one of the most cultural, compact, and diverse regions of the entire United States. New York is a "hot-spot" for finance, entrepreneurship, along with visual and performing arts. This major city has created a highly competitive culinary market. New players looking to open bars, restaurants, etc. should conduct sufficient geospatial research prior to making any legal commitments.

Farm Fresh Foods, a Farm-to-Table restaurant concept is seeking to build their first location somewhere within New York. It is understood that Farm Fresh Foods is looking to find an ideal location for their flagship restaurant. In order to investigate the Farm Fresh Food locations, the following information has been gathered and presented. This document will provide an introductory-level understanding of New York demographics and population, along with basic market research regarding cuisine preferences.

Business Problem

Farm Fresh Foods must identify a suitable location to place their new restaurant concept. To adequately research the market, several data sources must be extracted, cleansed, transformed, and then analyzed to gain geospatial insight. Recommendations can then be provided to management regarding placement of the restaurant.

The following criteria must be explored:

- "Cuisine-Deserts": What cuisines are oversaturated / undersaturated in New York City
- Low Competition Areas / Areas where existing, similar restaurants are underperforming
- What menu options are New Yorkers looking for?

Required Data

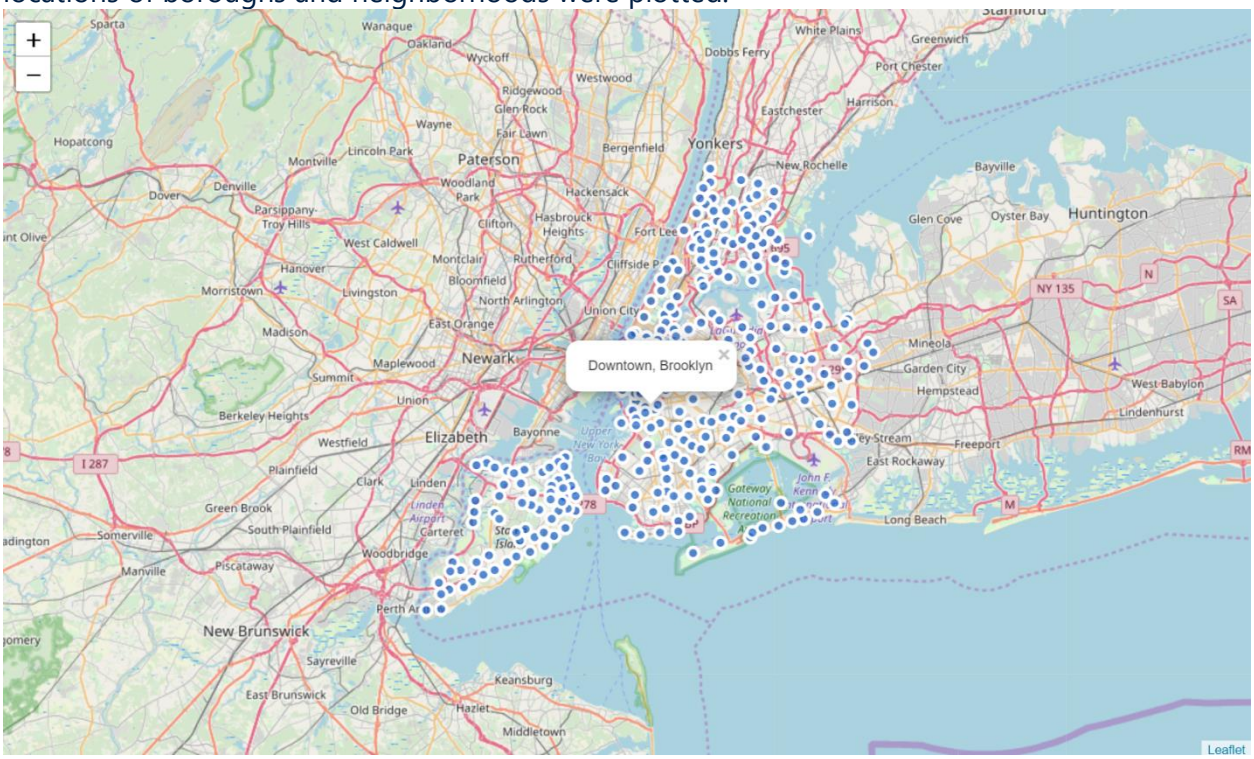
For this assignment, the following data sources are required:

- New York Neighborhoods
    - List of five boroughs and 306 neighborhoods.
- "Scraped" Wikipedia Websites
    - New York Population
- ~~Foursquare API Data~~
    - ~~Pre-existing venue information for each neighborhood~~
    - ~~Exploratory purposes of New York neighborhoods~~
- DOHMH
    - Data set related to Local Farmers Market

Nick Smart

Methodology

To gain a sense of bearings to borough and neighborhood locations, a public json containing borough and neighborhood locations in a Latitude and Longitude format. Data exploration was conducted and revealed that the data frame contained a total of five boroughs and 306 neighborhoods. Using the provided Latitude and Longitude and the folium package, the locations of boroughs and neighborhoods were plotted.



The next stage of exploration involved utilizing the BeautifulSoup web-scraping tool. This tool was used to extract information from a Wikipedia page about borough and county population totals and population density. This is what the original data set looked like after being populated into a pandas df.

	New York City's five boroughsvte\n	Jurisdiction\n	Population\n	Land area\n	Density\n	Borough	County
0	The Bronx\n	\n Bronx\n	1,471,160\n	19,570\n	42.10\n	109.04\n	34,653\n
1	Brooklyn\n	\n Kings\n	2,648,771\n	23,900\n	70.82\n	183.42\n	37,137\n
2	Manhattan\n	\n New York\n	1,664,727\n	378,250\n	22.83\n	59.13\n	72,033\n
3	Queens\n	\n Queens\n	2,358,582\n	31,310\n	108.53\n	281.09\n	21,460\n
4	Staten Island\n	\n Richmond\n	479,458\n	23,460\n	58.37\n	151.18\n	8,112\n
5	City of New York	8,622,698	806.863	302.64	783.83	28,188	10,947\n
6	State of New York	19,849,399	1,547.116	47,214	122,284	416.4	159\n
7	Sources:[14] and see individual borough articl...	NaN	NaN	NaN	NaN	NaN	NaN

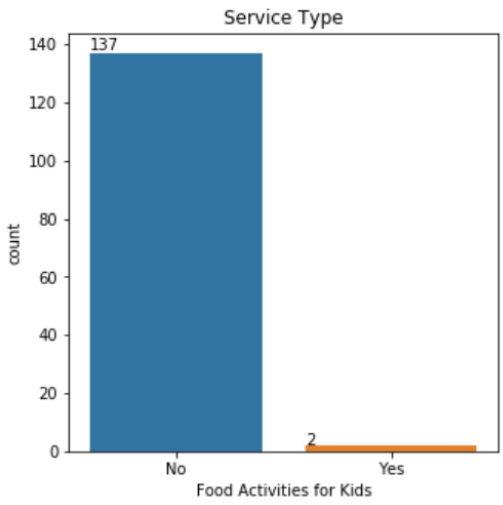
Clearly, data cleansing, sorting, and filtering needs to take place to gain actual insight into this raw data set. Renaming items and deleting empty fields yielded the following result:

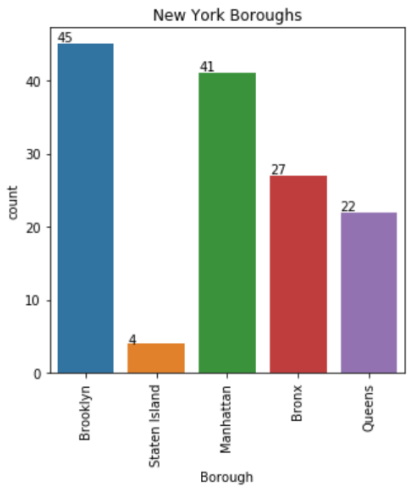
	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi	persons_sq_km
0	The Bronx	Bronx	1,471,160	19,570	42.10	109.04	34,653
1	Brooklyn	Kings	2,648,771	23,900	70.82	183.42	37,137
2	Manhattan	New York	1,664,727	378,250	22.83	59.13	72,033
3	Queens	Queens	2,358,582	31,310	108.53	281.09	21,460
4	Staten Island	Richmond	479,458	23,460	58.37	151.18	8,112
5		City of New York	8,622,698	806.863	302.64	783.83	28,188
6		State of New York	19,849,399	1,547.116	47,214	122,284	416.4
7		Sources:[14] and see individual borough articles					

The next necessary step is to identify Farmers Markets within the area by importing public data, cleansing it, and analyzing it to extract insights.

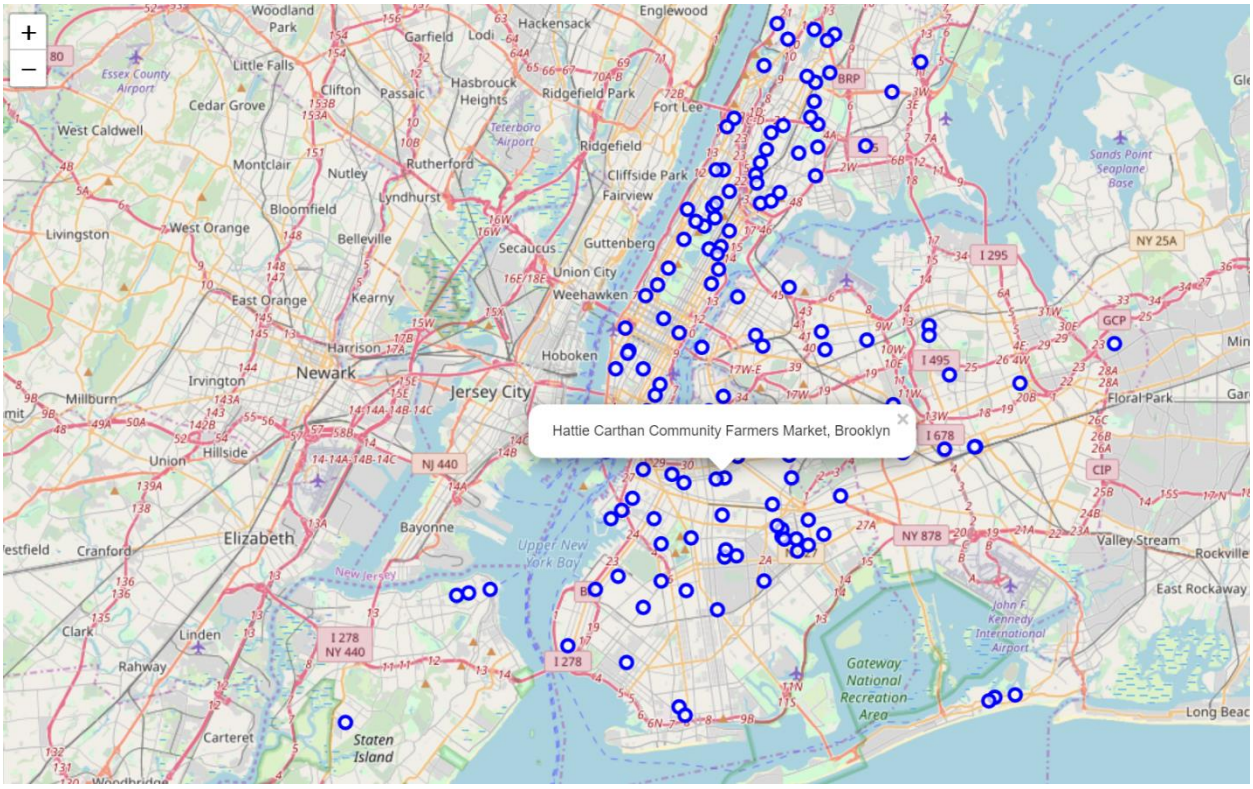
	Borough	Market Name	Street Address	Latitude	Longitude	Days of Operation	Hours of Operations	Season Dates	Accepts EBT	Open Year-Round	Stellar Cooking Demonstrations	Food Activities for Kids	Location Point
0	Brooklyn	Urban Oasis Farmers Market	681 Clarkson Ave	40.656255	-73.936608	Wednesday	2 - 5:30 p.m.	06/26/2019-11/06/2019	No	No	No	No	(40.656255, -73.936608)
1	Staten Island	Staten Island Mall Greenmarket	Marsh Ave & Ring Rd	40.583804	-74.161245	Saturday	8 a.m. - 3 p.m.	Year-Round	Yes	Yes	No	No	(40.583804, -74.161245)
2	Manhattan	Mount Sinai Hospital Greenmarket	E 99th St bet Madison & Park Aves	40.789169	-73.952743	Wednesday	8 a.m. - 5 p.m.	06/12/19-11/27/19	Yes	No	No	No	(40.789169, -73.952743)
3	Bronx	170 Farm Stand	E 170th St & Townsend Ave	40.839882	-73.916783	Wednesday	2:30 - 6:30 p.m.	07/10/2019-11/27/2019	Yes	No	No	Yes	(40.839882, -73.916783)
4	Manhattan	Grass Roots Farmers Market	W 145th St bet Edgecombe & Bradhurst Aves, at ...	40.823647	-73.943844	Tuesday &\nSaturday	9 a.m. - 4 p.m.	07/11/2019-11/21/2019	Yes	No	No	No	(40.823647, -73.943844)

With this raw data, several charts and a map were produced to provide a visual aid to better understand the data being presented in the Farmers Market Data set from DOHMH.





Nick Smart



Findings

Our first data set was able to provide spatial context to where boroughs and counties were located in the greater New York area. However, no real insights were able to be made with this data aside from getting acquainted with familiar region names.

The second data set used, extracted from Wikipedia, was able to bring real insight to this geospatial question. Farm Fresh Foods is looking to find a market space that is densely populated to increase customer interaction and "presence" in the greater New York area. From these findings, Queens County seems to be a prime area to investigate with an average of 281.09 people per square mile, with Kings County coming in second at an average of 183.42.

The Foursquare API was originally planned to be a useful tool for data exploration, but after testing with the API, no applicable insights were able to be made to resolve this problem. Time and resources were better spent on the Farmer's Market data-set.

The Farmer's Market data set provided several useful bits of information. The most important to Farm Fresh Foods was the realization that there is a serious lack of Food Activities for Kids in the New York area. This single piece of information has been incredibly helpful to Farm Fresh Foods: we have now made the decision to partner with Farmer's Markets in Queens to interact with families and understand what kind of cuisine they are looking for in their area. Other notable finds include that although Queens County is the densest area, they have the second least amount of Farmer's Markets in their area.

Nick Smart

Conclusion

Farm Fresh Foods set out with the idea of identifying a suitable location to place their new restaurant concept. To adequately research the market, several data sources were extracted, cleansed, transformed, and then analyzed to gain geospatial insight.

Recommendations are as follows:

- Engage with Queens County Farmer's Markets by implementing family-friendly events to immediately fill a clear need in the market.
- While interacting with those in the Queens area, understand what types of cuisine locals are looking for.
- Partner with Farmer's Markets in Queens to supply consistent and sustainable produce to the restaurant.