The Battle of the Neighborhoods - Report

1. Introduction and Business problem

Problem Background:

New York is a big city, is the financial capital of USA and the most populous city. It is diverse and multicultural with a lot of kind of business for any segment.

For all those things, is highly competitive and the cost of doing business is also one of the highest and risky. Thus, any new business venture needs to be analyzed carefully. The results from analysis will give us good understanding of the business environment for taking the best decisions in the market.

Problem Description:

A vet is a business which who protects the health and well-being of both animals. They diagnose and control animal diseases and treat sick and injured animals. They also advise owners on proper care of their pets. The City of New York have many people who lives with pets and need to have close to his house a vet in any case.

It is necessary that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on the location such as :

New York Population
New York City Demographics
Who are the competitors in that location?
Segmentation of the Borough
Untapped markets

Success Criteria:

The success criteria of the project will be a good recommendation of borough/Neighborhood choice to the vet based on lack of such places in that location.

2. Data

Data 1 : Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood.
This dataset exists for free on the web. Link to the dataset is :

https://geo.nyu.edu/catalog/nyu_2451_34572

Data 2 : For the below analysis we will get data from wikipedia as given below :
New York Population
New York City Demographics

https://en.wikipedia.org/wiki/New_York_City
https://en.wikipedia.org/wiki/Economy_of_New_York_City
https://en.wikipedia.org/wiki/Portal:New_York_City

Data 3 : New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896521	-73.844680	Pharmacy
2	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898283	-73.850478	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
4	Wakefield	40.894705	-73.847201	Dunkin Donuts	40.890631	-73.849027	Donut Shop

3. Methodology:

Business understanding

The main goal is to get the optimum location for the vet.

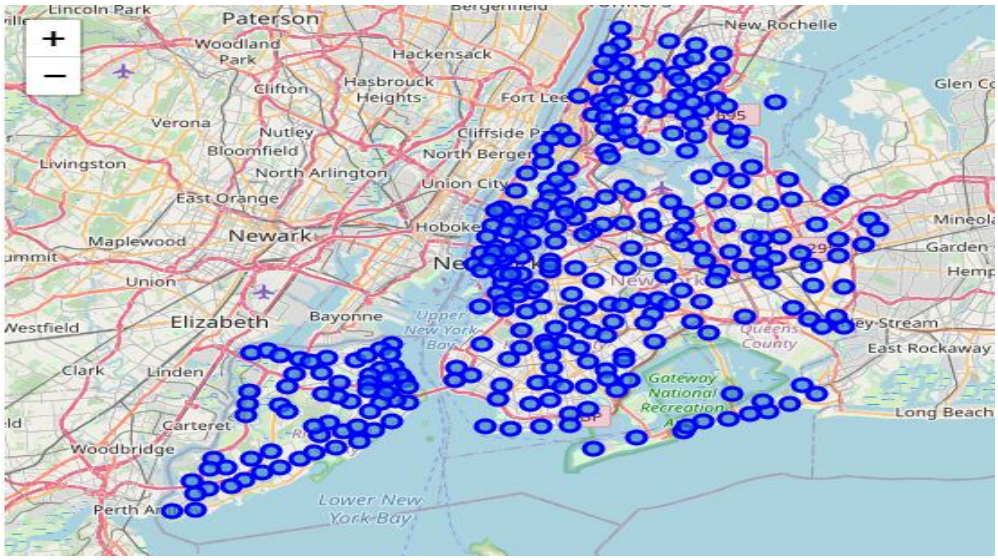
Analytic approach

New York city has 5 boroughs and 306 neighborhoods, we are going to cluster all the boroughs in this project.

Exploratory data analysis

Data 1 : New York city geographical coordinate data

We load the data and explore it from newyork_data.json file and transform the data with pandas dataframe. This data will be use to get venues data from Foursquare using their API. Finally we used geopy and folium libraries for creating a map of the city with neighborhoods on top.



Data 2 : New York city venues

To analyze population, we scrapped the data from Wikipedia pages, we used BeautifulSoup Python library, this library help us to parser html to xml

	Borough	County	Estimate_2019	square_miles	square_km	persons_sq_mi	persons_sq_km
0	The Bronx	Bronx	1,418,207	42.10	109.04	33,867	13,006
1	Brooklyn	Kings	2,559,903	70.82	183.42	36,147	13,957
2	Manhattan	New York	1,628,706	22.83	59.13	71,341	27,544
3	Queens	Queens	2,253,858	108.53	281.09	20,767	8,018
4	Staten Island	Richmond	476,143	58.37	151.18	8,157	3,150
5		City of New York	8,336,817	842.343			783.83
6		State of New York	19,453,561	1,731.910			122,284

Data 3: New York city geographical coordinates

New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision veterinarians in the city. We will use the Foursquare API to explore neighborhoods in New York City.

Veterinarians Visualization:

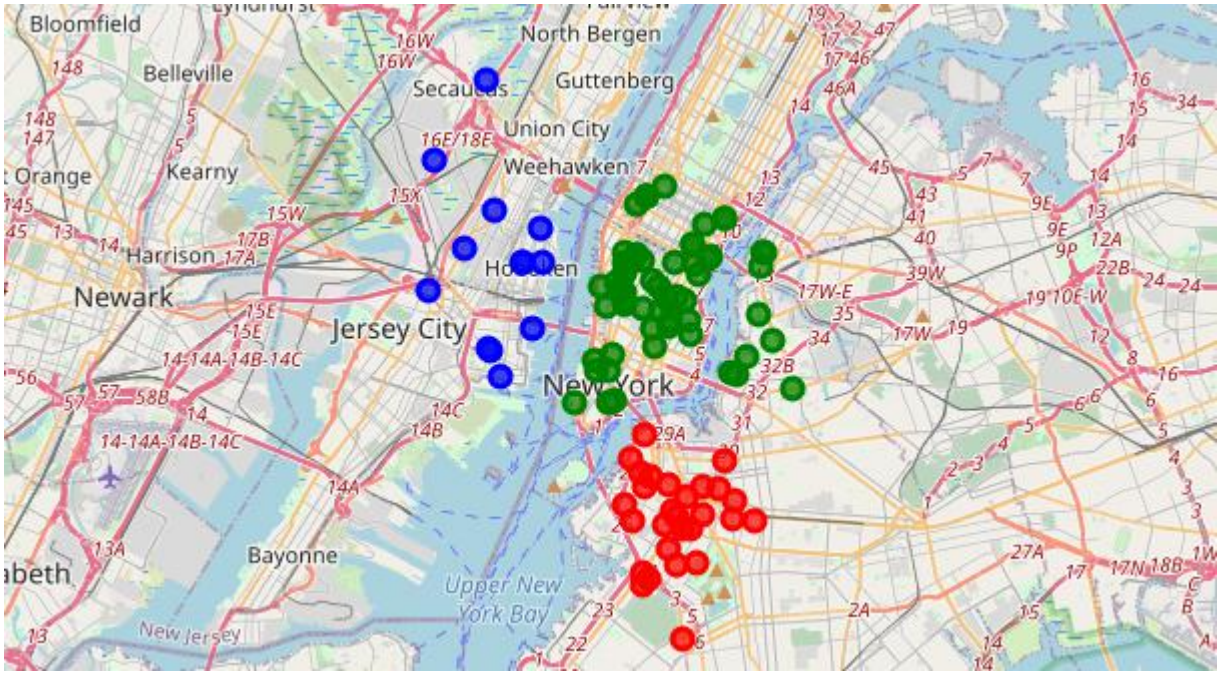
Using the geographical coordinates of New York in Foursquare API calls are made to get top 200 venues in a radius of 10000 meters. The venues data is:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	New York	40.7127281	-74.0060152	Seaport Animal Hospital	40.708821	-74.003925	Veterinarian
1	New York	40.7127281	-74.0060152	West Village Veterinary Hospital	40.739605	-74.002964	Veterinarian
2	New York	40.7127281	-74.0060152	Hope Veterinary Clinic	40.686878	-73.985526	Veterinarian
3	New York	40.7127281	-74.0060152	Pure Paws Veterinary Care of Clinton Hill	40.682521	-73.963018	Veterinarian
4	New York	40.7127281	-74.0060152	Eastside Animal Hospital	40.755886	-73.966446	Veterinarian

4. Results:

From this venues data we filtered and used only veterinarians for New York clustering.

Neighborhoods K-Means clustering based on mean occurrence of veterinarian category.
In the below Map Visualization, we can see the cluster created by using K-Means for New York.



Cluster0: The total of the sum of cluster0 has smallest value, it shows that the market is not saturated.

5. Discussion:

- We could increase the veterinarians increasing the radio
- Is more frequently that people is living with pets in big cities

6. Conclusion:

There are a few veterinarians in some places of New York then could be a low risk to start a new business like that.