



Opening a food and supply store for dogs in Manhattan

IBM Applied Data Science Capstone Project

Carlotta Brunetti - 6 December 2020

Introduction - Fact and figures on the pet industry market

- The pet industry expenditures (\$99 billions by the end of 2020) in US have **steadily increased over the past 30 years**
- **High-quality food and treats for pets make up the highest selling portion of the pet market**
- The pet market industry is **recession-resistant**
- In US, there are more households with dogs than with children. Around **63.4 million US households own a dog** compared to 42.7 million who own a cat
- **New York City is among the 25 most pet friendly cities in America** with 8.5 million residents and more than 600,000 dogs

Introduction - Aim of the project

Business problem

Opening a food and supply store for dogs in Manhattan would be a profitable business as well as a service very appreciated by the residents who own a dog. The location of the store is the key factor in the success of the business. The aim of the project is to answer the following business questions:

1. Which are the best neighbourhoods to live with a pet dog in Manhattan?
2. Where should an entrepreneur open a food and supply store for dogs in Manhattan?

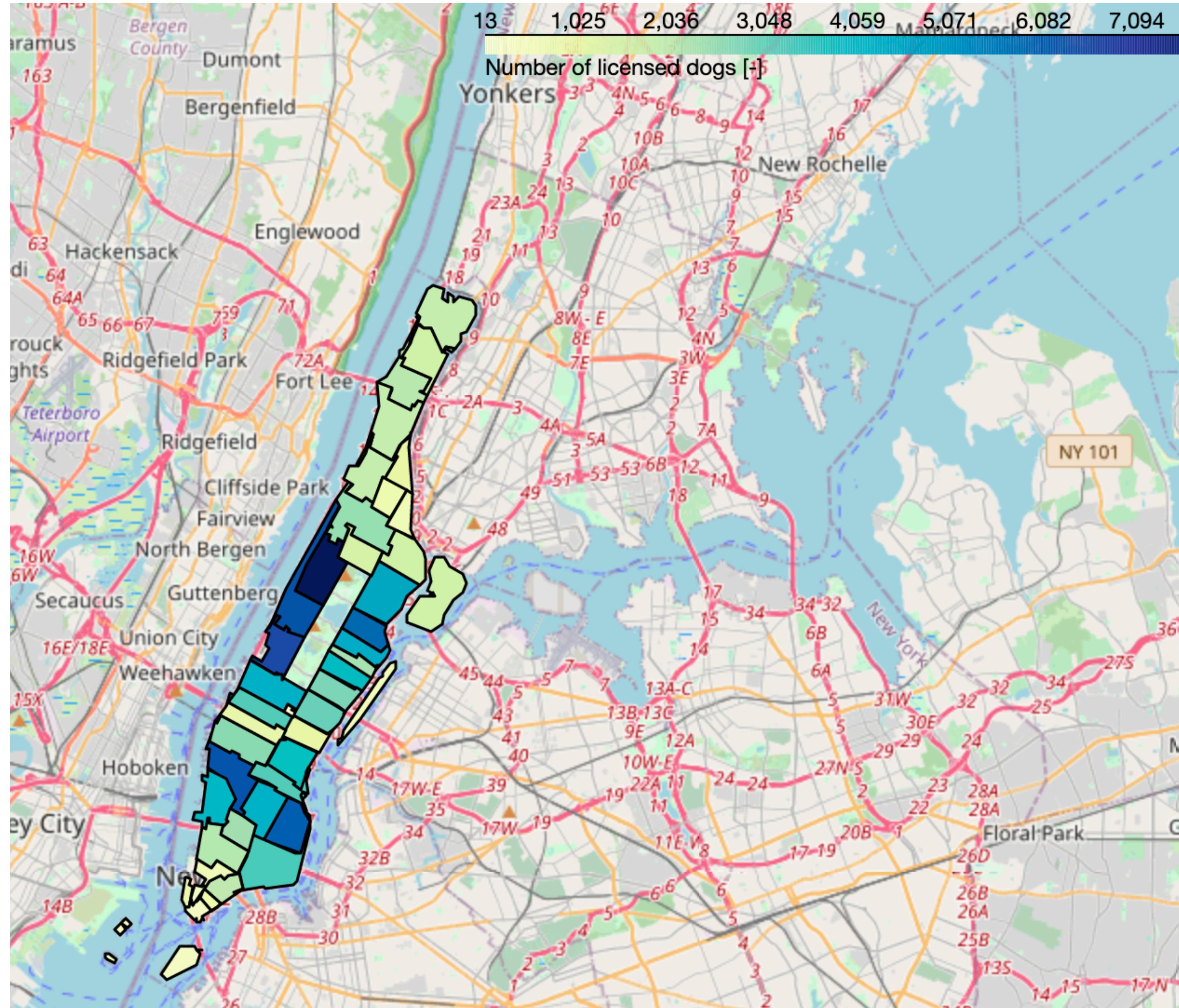
Target stakeholders

- Dog owners moving to Manhattan who are interested in finding the best neighbourhood to live with their furry-friends.
- Dog-lover entrepreneur seeking to open a successful business during the COVID-19 pandemic

Data description

- Zip codes and corresponding neighbourhood names in Manhattan (data source: <https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm>)
- Latitude and longitude of the zip code areas in Manhattan (obtained using the pgeocode library)
- NYC Dog Licensing Dataset (data source: <https://data.cityofnewyork.us/Health/NYC-Dog-Licensing-Dataset/nu7n-tubp>)
- Venue data and venue category ID (data source: Foursquare APIs, <https://developer.foursquare.com/docs/build-with-foursquare/categories/>)
- Polygons of the zip code areas in New York City (data source: https://jsspina.carto.com/tables/nyc_zip_code_tabulation_areas_polygons/public/map)

Methodology - Zip code areas and dog count in Manhattan

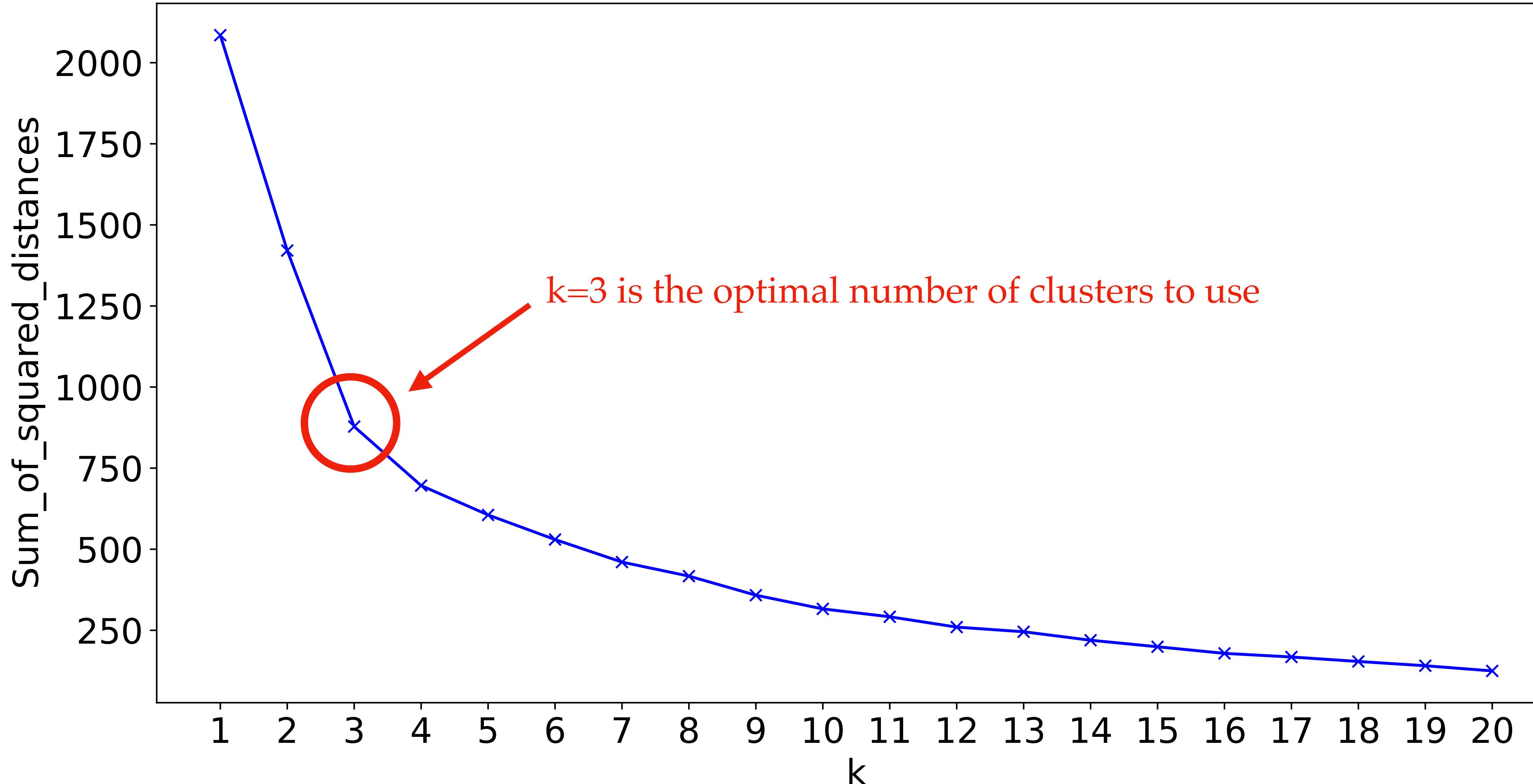


Methodology - Venue data from Foursquare APIs

- Venues within an area of 500 m around the latitude and longitude coordinates of each zip code area are explored. The search is limited to the following 9 venue categories:
 1. Supermarket
 2. Pet service (e.g., pet wash, spa, hotel, daycare)
 3. Pet store
 4. Veterinarian
 5. Dog run
 6. Park
 7. Trail
 8. Pet café (i.e., café completely dog friendly)
 9. Animal shelter

Methodology - Elbow method for optimal k-means clustering

Elbow Method For Optimal k



Methodology - Data analysis to answer the first business question

- Analyse the clustering results and **select the clustered area(s) that is(are) more relevant** to answer the business question.
- **Compute the total number of venues** (excluding the category "Animal Shelter" that is not relevant to answer the first business question) for each zip code area in the cluster.
- **Analyse the distribution of the venues** (stacked bar chart) for the first 5 zip code areas with the highest total number of venues.

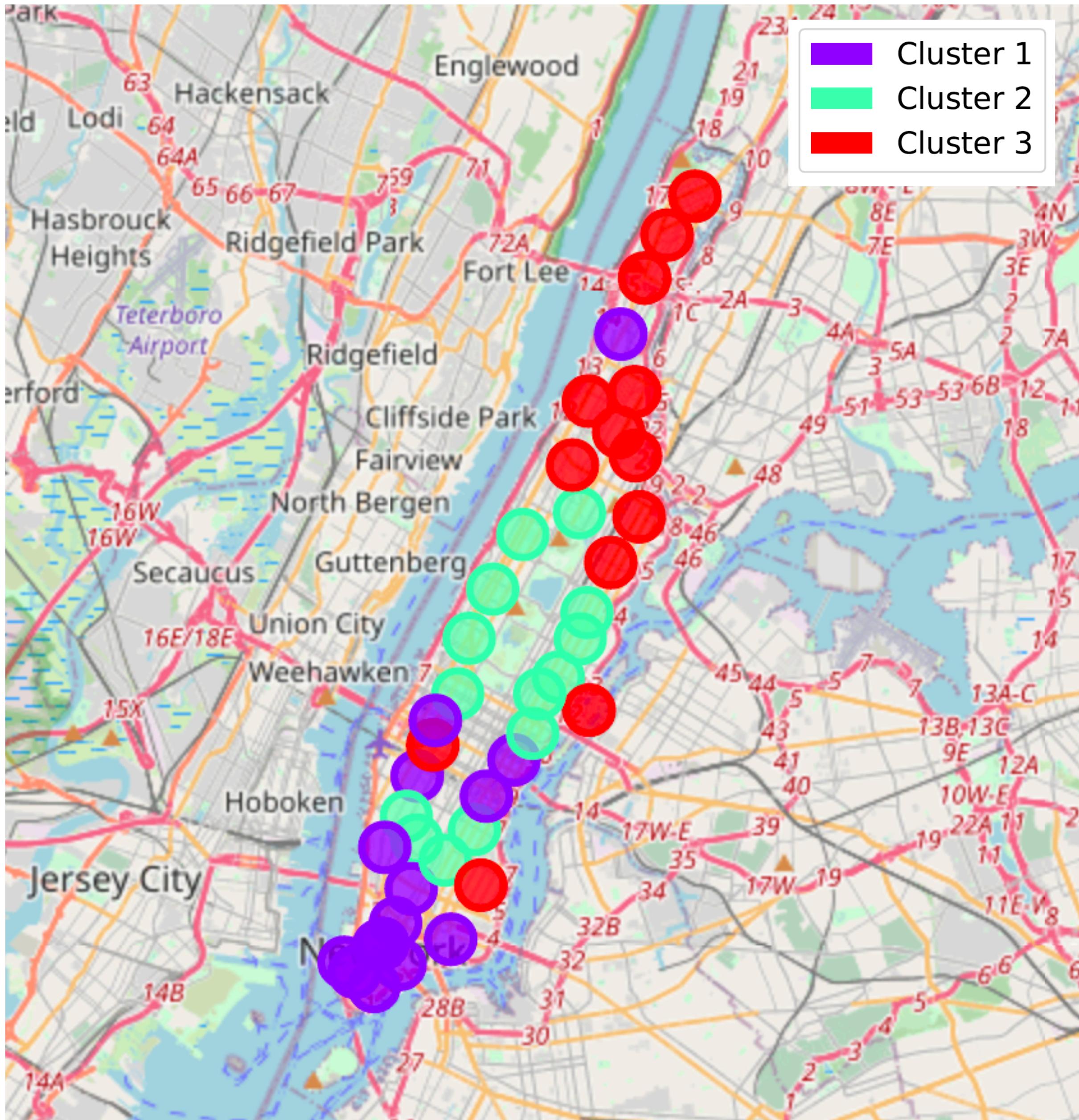
Methodology - Data analysis to answer the second business question

- Compute the total number of venues where food and supplies for dogs are sold that is the total number of pet stores and supermarkets (N_{fs}) in each zip code area.
- Compute the number of pet stores and supermarkets relatively to the number of licensed dogs (N_d) that live in each zip code area:

$$\alpha = \frac{N_{fs}}{N_d}$$

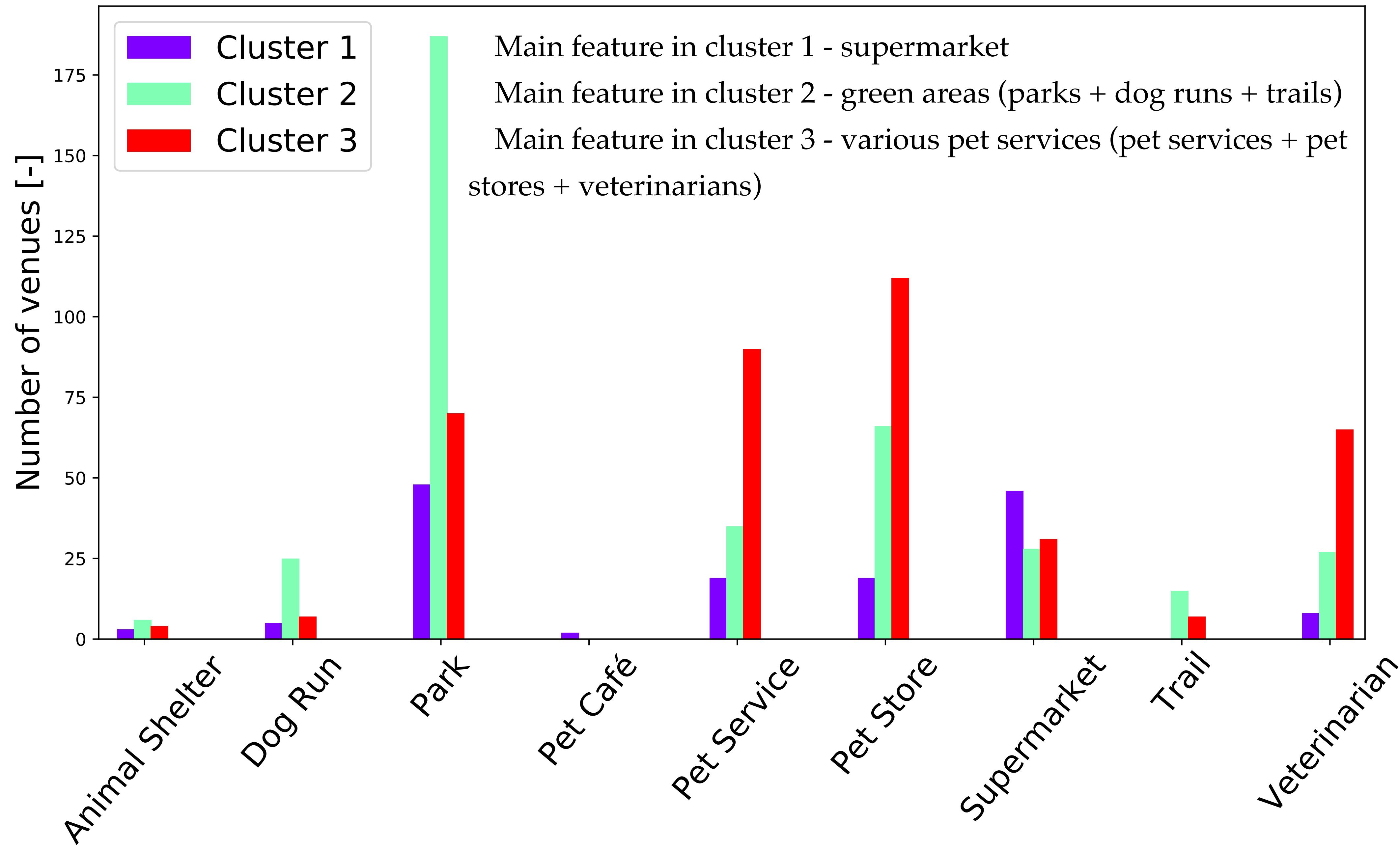
- The zip code area with the lowest ratio (few food and supply stores compared to the number of dogs) = good spot in Manhattan where to open a food and supply store for dogs

Results - Clustering the zip code areas in Manhattan



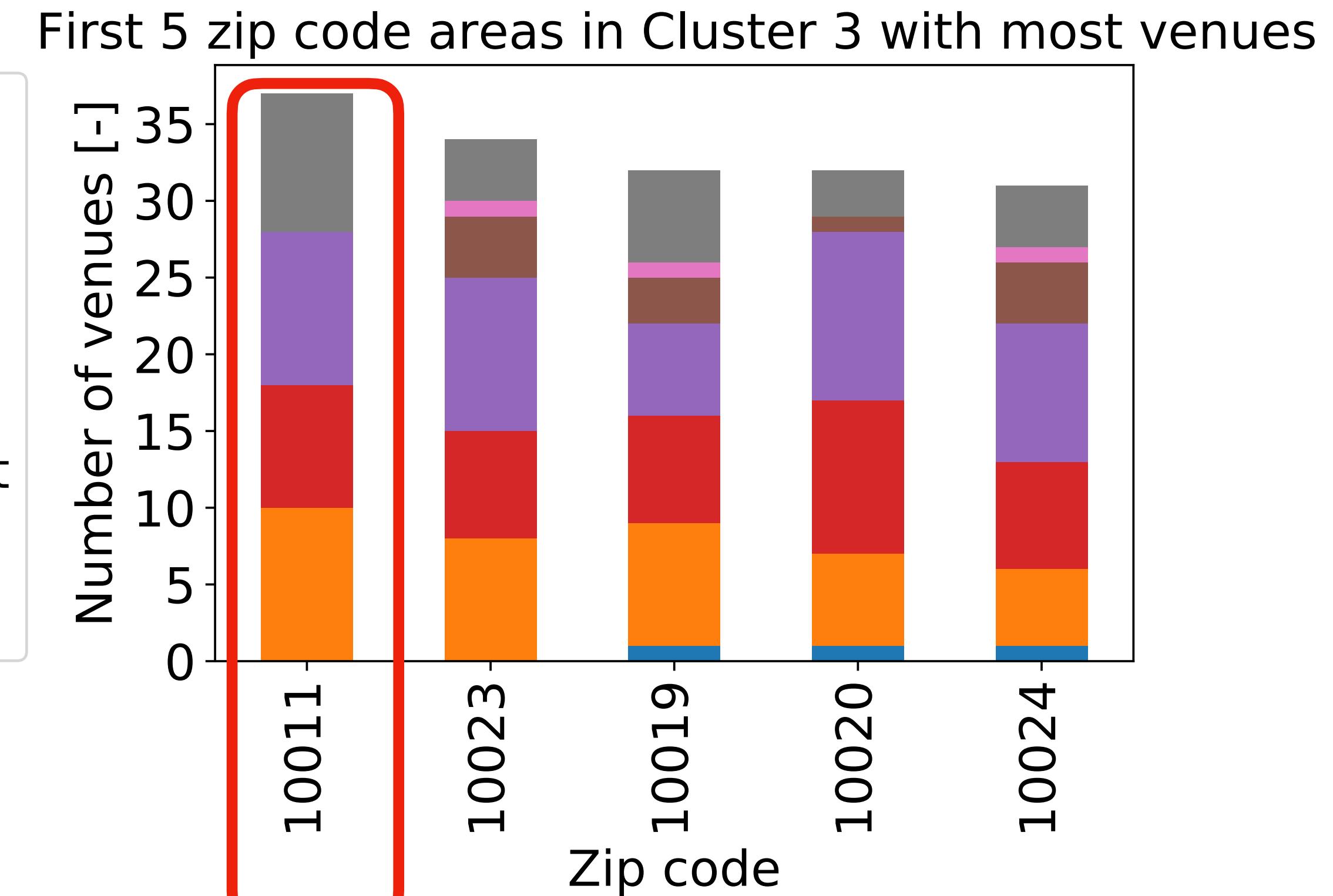
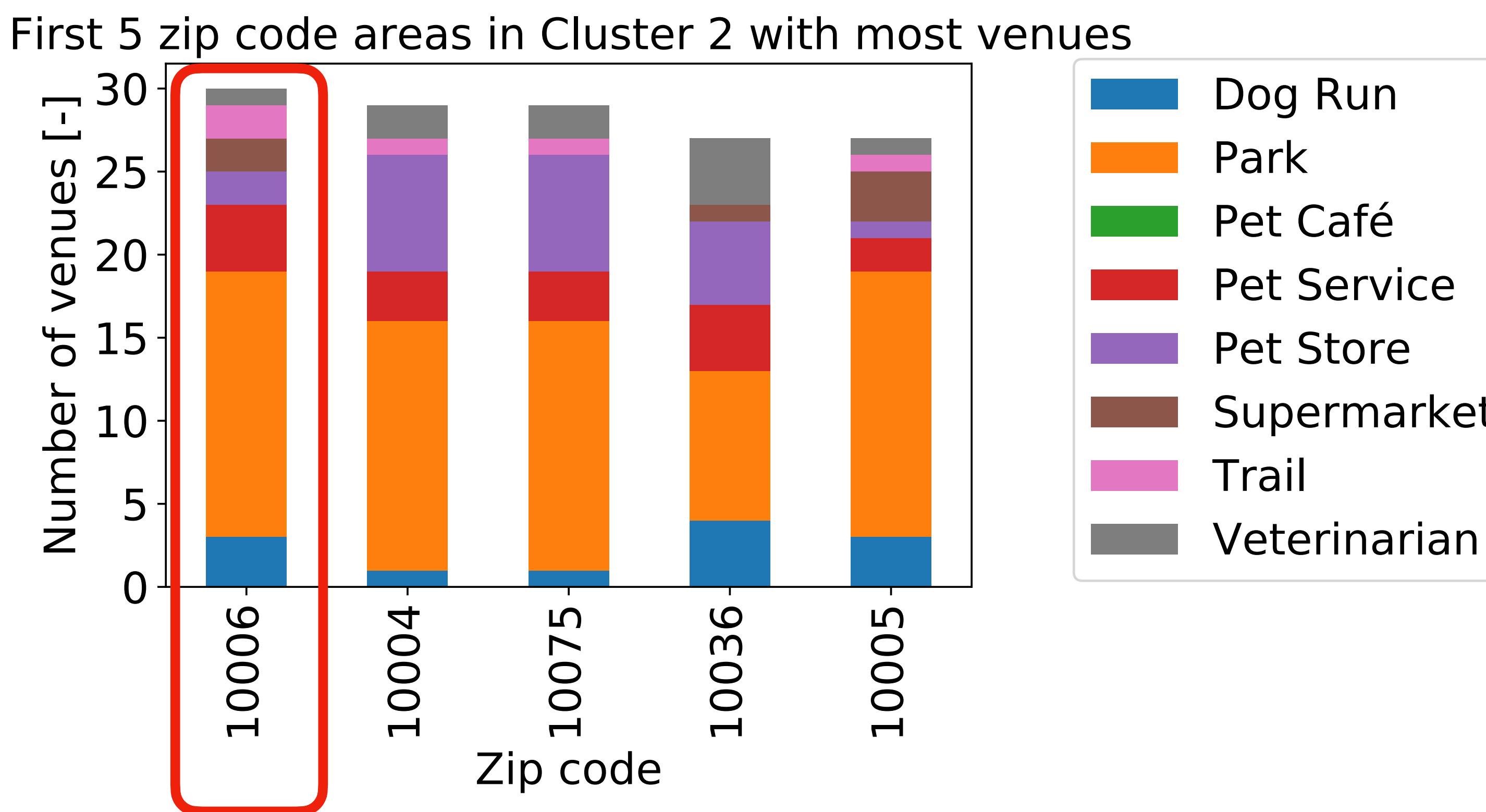
K-means clustering method applied to cluster the zip code areas in Manhattan based on the venue data from Foursquare APIs

Results - Clustering the zip code areas in Manhattan

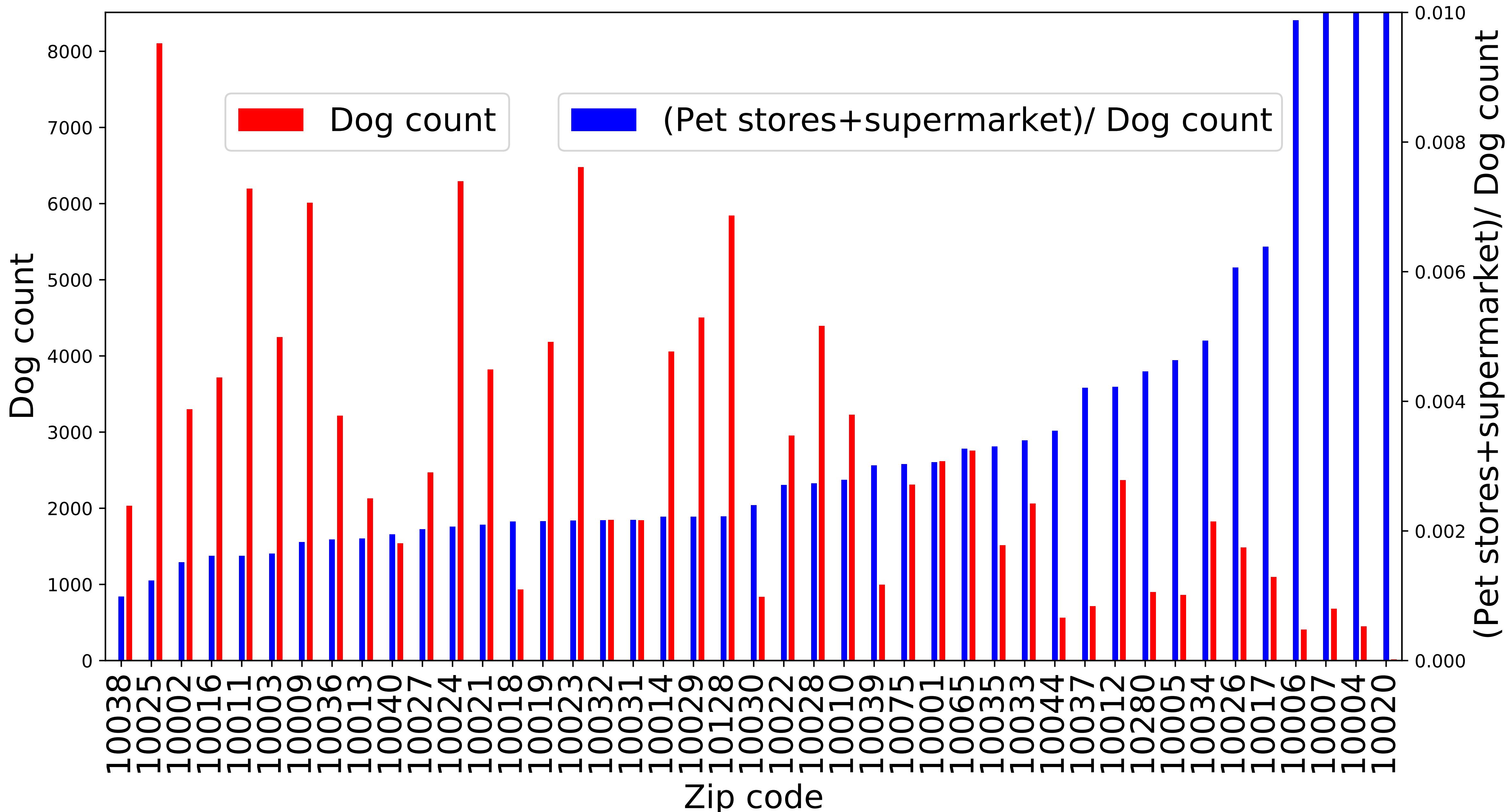


Results - Venue distribution in the zip code areas in Manhattan

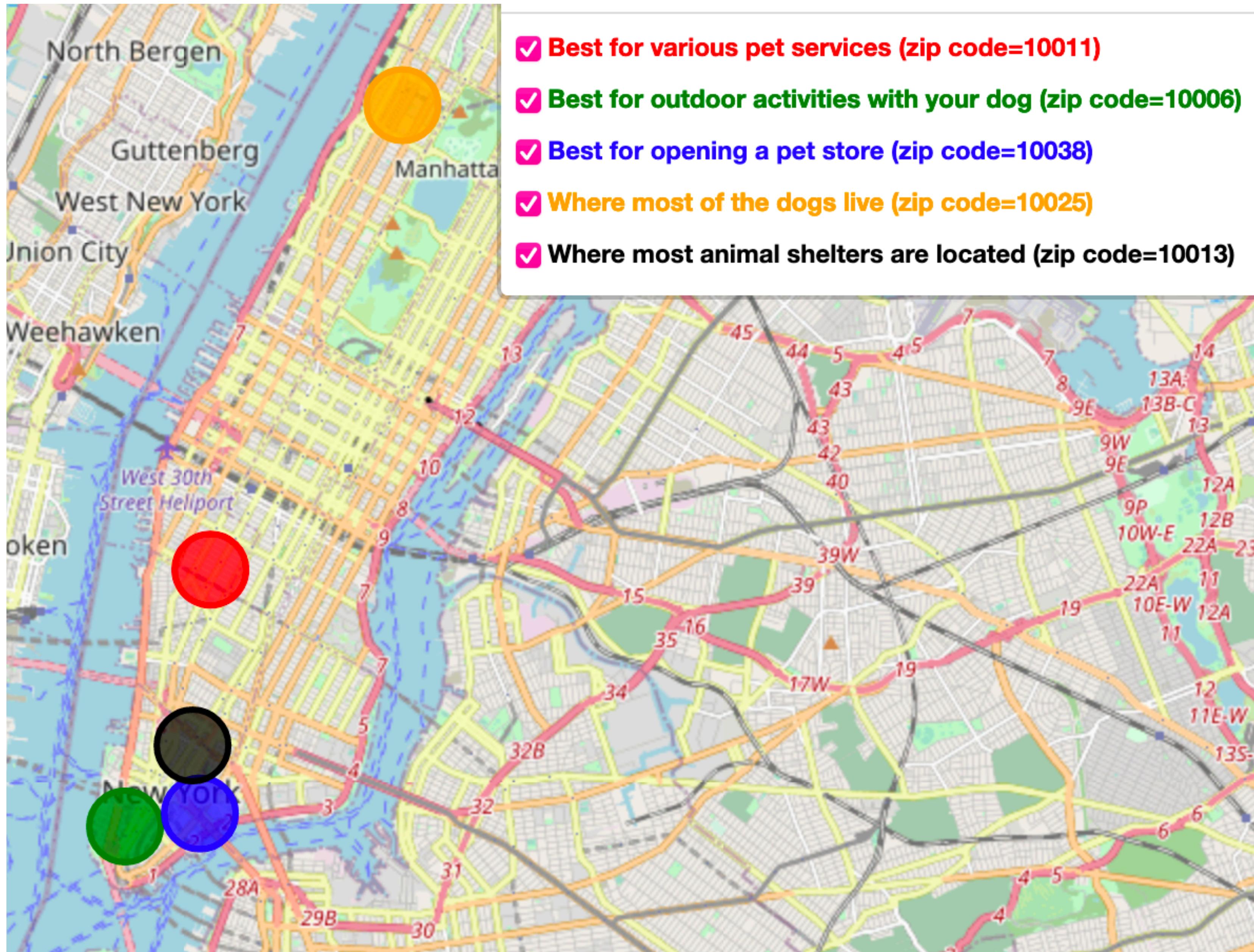
- Cluster 1 is not so relevant to answer the first business questions.
- Distribution of the venues for the first 5 zip code areas with the highest total number of venues in cluster 2 and cluster 3 are investigated



Results - Number of food and supply stores and dog count



Results - Summary plot



Conclusion

Q1: Which are the best neighbourhoods to live with a pet dog in Manhattan?

A1: Lower Manhattan (zip code 1006) and Chelsea (zip code 10011)

Zip code area 10006 = best for active dogs and dog owners who love to spend most of their time outdoor.

Zip code area 10011 = best for dog owners who love to have a wide choice of pet services and vet care facilities.

Q2: Where should an entrepreneur open a food and supply store for dogs in Manhattan?

A2: In Lower Manhattan (zip code 10038)

Zip code area 10038 = the one with the smallest number of pet stores and supermarket per number of licensed dogs; next to the zip code area 10013 where there are the highest number of animal shelters that might need a lot of food and supplies; next to the zip code area 10006 where most of the green areas are located and likely very frequented by dog owners.

Discussion and outlook

Observations to properly evaluate the results in the correct prospective:

- The **size of the areas** used to explore the Foursquare venues (circle with a 500 m radius) **does not exactly match the polygons** representing the zip code areas
- The **Foursquare database** is prone to **incorrect classification of the venues** since it is based on the input from the users
- **Not all the dogs** living in Manhattan are **licensed**
- **Include other datasets** such as the average rent prices, average gross incomes and crime statistics in each zip code area