**Applied Data Science Capstone Report**

**The Battle of Neighborhoods in Nashville: Opening a New Pet Store**

Karly Rezac

**Introduction/Business Problem**

The American Pet Products Association (APPA) estimates that within the year, $75.38 billion will be spent on pets in the United States alone. The pet services market, which includes grooming, boarding, training, and care has been the fastest-growing product segment under the industry umbrella. As such, an eCommerce-based Pet retailer has decided to open their first brick and mortar store in the Nashville metro area as part of their new business strategy.

The retailer has invested funding for a deep dive into location data analysis to determine which neighborhood in Nashville would be the most suitable for the new store. As the retailer is leaning into the pet services market, they feel that their best strategy is to seek a location in higher density areas where consumers typically frequent for shopping and entertainment. This analysis will include Foursquare location data, where we will dig deeper into locations with a high frequency of restaurants and grocery stores, as well as look into locations of other pet stores to avoid opening too close to a competitor. We will utilize these tools to make data-driven decisions about these neighborhoods.

**Data**

As stated in the introduction, listed below are the characteristics of a neighborhood that will influence the location chosen:

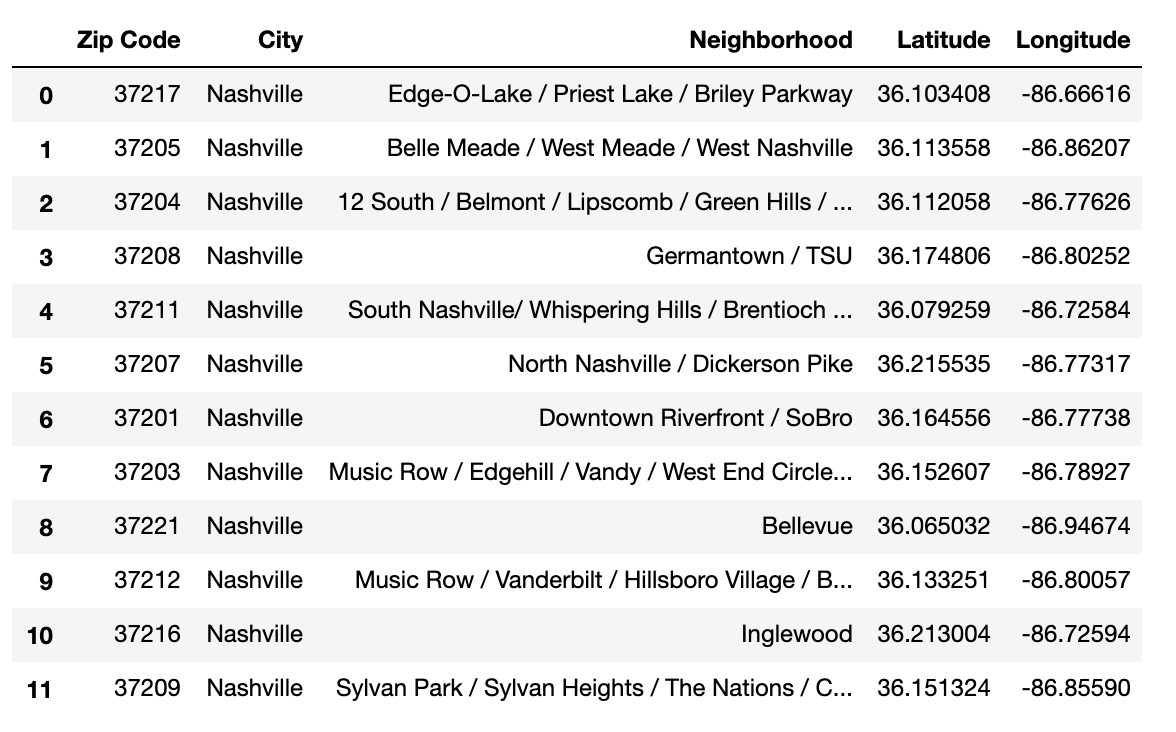
* The frequency of existing restaurants in the neighborhood
* The frequency of existing grocery stores in the neighborhood
* Any nearby pet stores, which would be competition

We will use the below data sources to extract, generate, and analyze the required information:

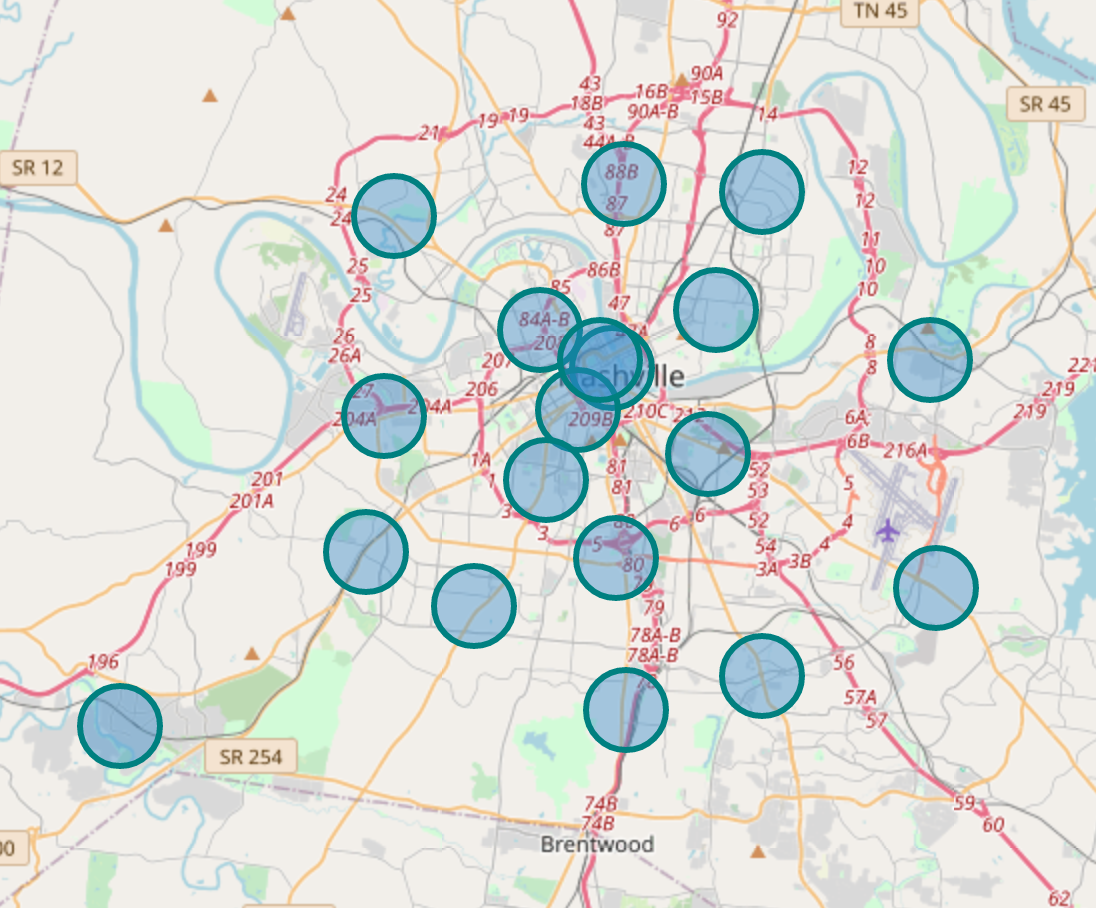
* List of all neighborhoods in the Nashville metro area, and their corresponding ZIP Codes and coordinates
* Source: <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=TN> – Includes ZIP codes, city, state, and coordinate data
* Source: <https://nestinginnashville.com/buying-a-home-in-nashville/zip-code-map/> - Includes Nashville neighborhood names and ZIP codes
* Frequency of restaurants, grocery stores, and pet stores in each neighborhood (Foursquare API)
* Coordinates of all venues (GeoPy Nominatim)

**Methodology and Analysis**

For this project, the first step was to collect location data on the neighborhoods of Nashville. I found sources for this data on the internet. For the ZIP Code and coordinates data, I was able to find a downloadable dataset. For the neighborhood names, I had to scrape the data from a webpage.



We used GeoPy Nominatim for coordinates of Nashville and then created a Folium map of Nashville with an overlay of the Nashville neighborhoods, using the coordinates from the dataset.



Next, we explored each neighborhood and their corresponding venues using Foursquare API location data. Below is a screenshot of a dataframe where we took the mean frequency of occurrence for each venue category.

A screenshot of a cell phone

Description automatically generated

From there, we pulled the most common venues in each neighborhood and put them into a dataframe, screenshot below.

A screenshot of a cell phone

Description automatically generated

Then, we visualized the frequency of occurrence of venues for the neighborhoods in Nashville and isolated the categorical venues. The ideal pet store locations will have a high density of grocery stores and restaurants, and a low density of other pet stores that may be competitors.

The below data shows a violin plot from the seaborn library, which visualizes a density estimation of the underlying distribution.

A screenshot of a cell phone

Description automatically generated

As shown in the above visualization, we narrowed down the neighborhoods based on the frequency of nearby venues such as grocery stores and restaurants, which the retailer specified as higher density areas where consumers typically frequent for shopping and entertainment. The visualization also shows neighborhoods with a greater frequency of pet stores, which would likely want to be avoided due to the competition.

Fortunately, none of the neighborhoods had a high frequency of pet stores that also had a high frequency of restaurants and grocery stores. This allows us to narrow down our neighborhoods to the three with high frequencies of both restaurants and grocery stores.

As such, we narrowed down the neighborhoods to:

* Inglewood
* Sylvan Park/Sylvan Heights/The Nations/Charlotte Park
* Music Row/Edgehill/Vandy/West End Circle/Downtown/The Gulch/Wedgewood Houston

The final neighborhoods visualized on a Folium map:

A picture containing text, map

Description automatically generated

**Results and Discussion**

The result of this analysis is three neighborhoods in Nashville that contain a low frequency of pet stores, but high frequency of restaurants and grocery stores: Inglewood, Sylvan Park/Sylvan Heights/The Nations/Charlotte Park, and Music Row/Edgehill/Vandy/West End Circle/Downtown/The Gulch/Wedgewood Houston.

These three neighborhoods meet the criteria specified by the retailer. However, this does not necessarily mean that these neighborhoods are ideal locations for a new pet store. The purpose of this analysis is to narrow down potential neighborhoods based on the criteria provided. Realistically, it's possible that there are reasons for the low frequency of existing pet stores in those neighborhoods, which would rule them out regardless of the lack of competition in that area. This analysis should be a good starting point to guide the next part of the process for the retailer to finalize a store location.

**Conclusion**

By calculating grocery store, restaurant, and pet store density distribution from Foursquare data, we were able to identify three neighborhoods that the retailer can investigate further. Realistically, this analysis could have been performed with many different methodologies and based on many different data sources. For this project, I wanted to stick to a straight-forward approach to narrowing down options for the retailer.

This analysis and recommendations have achieved the objective of narrowing down the best neighborhood options based on the available data. The final determination for the store location will be made by the retailer, using their own expertise in the pet industry, and taking into consideration additional factors such as characteristics of each neighborhood, pricing, economic variables, and others.