Predicting the best neighborhood to open a vegan restaurant in NY

1 Introduction

1.1 Background

In recent years the interest in a vegan lifestyle has increased. There are a lot of reason as: people concerned with animal rights, animal mistreatment, climate change, health and others. Below are some statistics:

• If the world went vegan, it could save 8 million human lives by 2050, reduce greenhouse gas emissions by two thirds and lead to healthcare-related savings and avoided climate damages of $1.5 trillion.
(*https://www.oxfordmartin.ox.ac.uk/news/201603-plant-based-diets/*)

• Dairy cows have been bred to produce up to 10 times more milk than they naturally would. (*https://www.ciwf.org.uk/media/5235185/the-life-of-dairy-cows.pdf*)

• Male calves are of no use to the dairy industry and are less suitable for beef production. This means that every year around 95,000 male dairy calves are shot soon after birth and discarded as a by-product.
(*https://www.theguardian.com/environment/2018/mar/26/dairy-dirty-secret-its-still-cheaper-to-kill-male-calves-than-to-rear-them*)

• Animal agriculture contributes an estimated 18% to total greenhouse gas emissions from the five major sectors for greenhouse gas reporting. For the agriculture sector alone, farmed animals constitute nearly 80% of all emissions.
(*http://www.fao.org/3/a-a0701e.pdf*) [p. 112]

For more statistics about vegan, consult
https://www.vegansociety.com/news/media/statistics

1.2 Problem

As the number of vegan people is growing up, it is interest using data to decide where in the city of New York open an vegan restaurant. It is ideal to find some heighborhood with few restaurants of the kind and close to a lot of people.

2 Data acquisition and cleaning

2.1 Data sources and cleaning

Let's use the Foursquare API as a data source and the data neighborhood from *https://cocl.us/new_york_dataset*.

We analyse all the venues in Manhattan and drop all the is not a vegan or vegetarian restaurant. There are 24 vegan or vegetarian restaurants in the interest area.

3 Methodology

In this project we will direct our efforts on detecting areas of Manhattan that have low restaurant density, particularly those with low number of Vegan/Vegetarian restaurants. We will limit our analysis to area 5km around Manhattan.

In first step we have collected the required data: location and category of every restaurant within 5km from Manhattan. We have also identified Vegan/vegetarian restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Manhattan - we will use maps to identify a few promising areas close to center with low number of vegan/vegetarian restaurants and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Using KMeans Clustering we could grouping the vegan/vegetarian restaurants and so find the bests area to a new restaurant. These areas are Morningside Heights, Manhattan Valley, East Harlem, Lincoln Square and Lenox Hill.

4 Results

Our analysis shows that although there is a great number of restaurants in Manhattan, there are a low number of vegan/vegetarian restaurantes close to city center.

Highest concentration of restaurants was detected south from Central Park, so we focused our attention to areas north and east, corresponding to boroughs Morningside Heights, Manhattan Valley, East Harlem, Lincoln Square and Lenox Hill.

5 Discussion

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues. This, of course, does not imply that those zones are actually optimal locations for a new restaurant!

Purpose of this analysis was to only provide info on areas close to Manhattan but not crowded with existing vegan/vegetarian restaurants - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

6 Conclusion

Purpose of this project was to identify Manhattan areas close to center with low number of vegan/vegetarian restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new Vegan restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.