How Machine Learning Can Help a Food Truck Manager to Find the Best Neighbourhood to Sell in Manhattan

Israel Pinheiro de Siqueira

israelps@gmail.com

1 Introduction

Find a profitable neighborhood for a food truck is not an easy task. Today, finding a profitable area is done by experimentation and observation and requires some kind of expertise, and obviously, luck.

With the advance of computational resources, machine learning algorithms became more popular and can be used to solve a vast class of problems that can't be solved a few years ago.

Using machine learning algorithms to find and group neighborhoods based on the residents preferred restaurants and based on the idea that we, as human beings, tend to group with others that are similar in some sense (tastes, lifestyle, thoughts, ideas, etc) [1],[2] we will show in this work how machine learning can be used to make good indications of a profitable neighborhood based on the type of food (or foods) he offer.

2 Data

The principal idea is to cluster and segment neighborhoods based on locals' favorite restaurants.

To do this we need:

- Names and geographic coordinates(latitude, longitude) of all neighborhoods of Manhattan;
- Trending restaurants around each neighborhood;

To get the names and geographic coordinates of all neighborhoods of Manhattan we resorted to the dataset: 2014 New York City Neighborhood Names available for free in the New York University Spatial Data Repository (https://geo.nyu.edu/) and filter the Borough to match Manhattan. In figure [1] we have all neighborhoods in New York and in red we have only the ones that belong to Manhattan and will be used in this work.

To get the Trending restaurants we resort to the Foursquare API available in (https://developer.foursquare.com/). First, we need to create a free developer account. After that we create a python script to get the top 40 venues for each neighborhood. An example of the dataset is shown in table [01].

Given that we are now able to procced with the metodology.

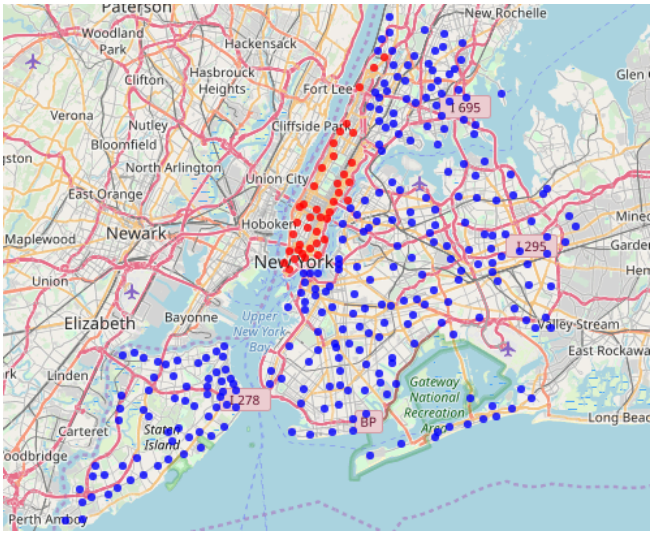

Fig. 1: New York neighborhoods in red. In blue we have the neighborhoods of Manhattan that will be used in this work.

Name	Category	Lat.	Lon.
Cold Cut City	Sandwiches	40.88	-73.91
Kam Wah Kitchen	Asian	40.88	-73.91
Gold Mine Cafe	Café	40.88	-73.90
Q'Kachapa	Spanish	40.88	-73.90
Sam's Pizza	Pizza	40.88	-73.91

Table 1: Example of Marble Hill trending venues

3 Methodology

In this work we will help a food truck of Mexican food find the best neighborhood in Manhattan to sell their tacos. Using K-Means to cluster the data based on the 5 (five) trending restaurants in each neighborhood category we will be able to find the most profitable neighborhood for the food truck.

4 Results

As we can see in table [02], we have an example of the dataset after group the 5 most trending restaurant categories in each neighborhood. With this dataset, we make a K-Means clustering using the best k-parameter of 4 (four). This k-parameter was chosen by experimentation and give the best results to clustering. In figure [02] we can see in the map the clusters segmented by color.

Neighborhood	1st	2nd	3rd	4th	5th
Battery Park City	Pizza	Italian	Chinese	Sandwiches	BBQ
Carnegie Hill	Bakery	Pizza	Coffee	French	Sushi
Central Harlem	Deli / Bodega	Fried Chicken	Chinese	African	Southern / Soul
Chelsea	Italian	Bakery	Coffee	Pizza	Sushi
Chinatown	Chinese	Dumplings	Malay	Vegetarian / Vegan	Italian

Table 2: The 5 most trending restaurants category in each neighborhood

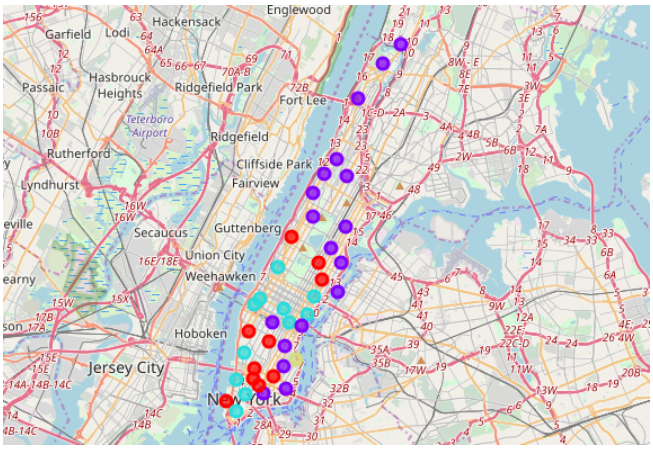

Fig. 2: Manhattan neighborhoods clustered by the restaurant category using the k-means clustering algorithm with k-parameter = 4

5 Discussion

In table [03] we highlight all neighborhoods that are suitable to our candidate Mexican food truck sell their tacos. We can see that in this cluster we have 37 neighborhoods that can be profitable to our business manager candidate. East Harlem and Manhattan Valley, for example, are the neighborhoods that have Mexican food as the most favorite food and can be a good place to sell. On the other hand, we have almost all other neighborhoods with Mexican food with all levels of preference and that can be used to determine the most profitable neighborhood to sell, maybe a neighborhood with less Mexican restaurants as favorite food (but no zero) like Lower East Side or Washington Heights can be a most profitable place to start.

6 Conclusion

We can see that the use of Foursquare API together with the spatial data from the New York University Spatial Data Repository in addition to good data preparation and machine learning clustering algorithms, like K-Means, can be used to determine the best place to start a new food business. In fact, this is a powerful methodology can be used to determine not only to start a food business but for any other type of business.

Neighborhood	1st	2nd	3rd	4th	5th
Marble Hill	Sandwiches	Deli / Bodega	Donuts	Seafood	Diner
Chinatown	Chinese	Dumplings	Malay	Vegetarian / Vegan	Italian
Washington Heights	Deli / Bodega	Pizza	Mexican	Chinese	Restaurant
Inwood	Pizza	Coffee	Restaurant	Mexican	Bakery
Hamilton Heights	Pizza	Deli / Bodega	Coffee	Mexican	Chinese
Manhattanville	Chinese	Deli / Bodega	Seafood	Mexican	Sandwiches
Central Harlem	Deli / Bodega	Fried Chicken	Chinese	African	Southern / Soul
East Harlem	Mexican	Pizza	Deli / Bodega	Bakery	Latin American
Yorkville	Italian	Pizza	Deli / Bodega	Sandwiches	Chinese
Roosevelt Island	Deli / Bodega	Sandwiches	Coffee	Pizza	Greek
East Village	Pizza	Japanese	Italian	Vegetarian / Vegan	American
Lower East Side	Pizza	Deli / Bodega	Chinese	Mexican	Ramen
Manhattan Valley	Mexican	Pizza	Indian	Thai	Coffee
Morningside Heights	Food Truck	Deli / Bodega	Pizza	American	Coffee
Gramercy	Deli / Bodega	Thai	Italian	Pizza	Mexican
Carnegie Hill	Bakery	Pizza	Coffee	French	Sushi
Midtown South	Korean	Japanese	Bakery	Deli / Bodega	Restaurant
Tudor City	Coffee	Deli / Bodega	Mexican	American	Food Truck

Table 3: The cluster (purple on map) with the most profitable neighborhoods to sell tacos and the favorite restaurant categories of their residents.

7 Acknowledgment

A special thank's to all instructors from IBM Data Science Specialization for the great work providing us with invaluable content through all specialization courses making this project (and all other) possible.

References

1. Castano, E., Yzerbyt, V., Paladino, M.P., Sacchi, S.: I belong, therefore, i exist: Ingroup identification, ingroup entitativity, and ingroup bias. Personality and Social Psychology Bulletin **28**(2), 135–143 (2002)
2. Ellemers, N., Spears, R., Doosje, B.: Sticking together or falling apart: In-group identification as a psychological determinant of group commitment versus individual mobility. Journal of personality and social psychology **72**(3), 617 (1997)