



The Best-fit Positions for Food Delivery Sites Analysis

IBM APPLIED DATA SCIENCE CAPSTONE

March 2019

Authored by: Jie Wu



Table of Contents

- Introduction
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

“Solutions for a small planet”

Introduction

This is a data-driven site selection analysis for food delivery start-ups in Manhattan to select the best-fit positions where they could provide the fastest service and take less operation risk.

This analysis is based on some assumptions:

- 1) Customers prefer the high rating merchandises
- 2) Food delivery order cannot be canceled after the food is picked up from the Merchandise
- 3) The distance between site location and high rating merchandises is the dominated factor of business operation risk.

According to the assumptions we should group the all high rating merchandises and find every point that has the shortest distance to all of the merchandises in the group. Those points will be the best locations for food delivery sites.

Data

Now we could decide what kind of data we will need:

- 1) New York city neighborhoods data
- 2) The position data of all food, drink and grocery stores in Manhattan area

So we are going to use the Foursquare location data API to get the data we need and will use the centroid model of clustering analysis to find the centroid of each group in the data. The K-means clustering algorithm of this model works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. The results of this algorithm are the centroids of the K clusters, which can be used as our best-fit positions of delivery sites.

1. Neighborhood Candidates

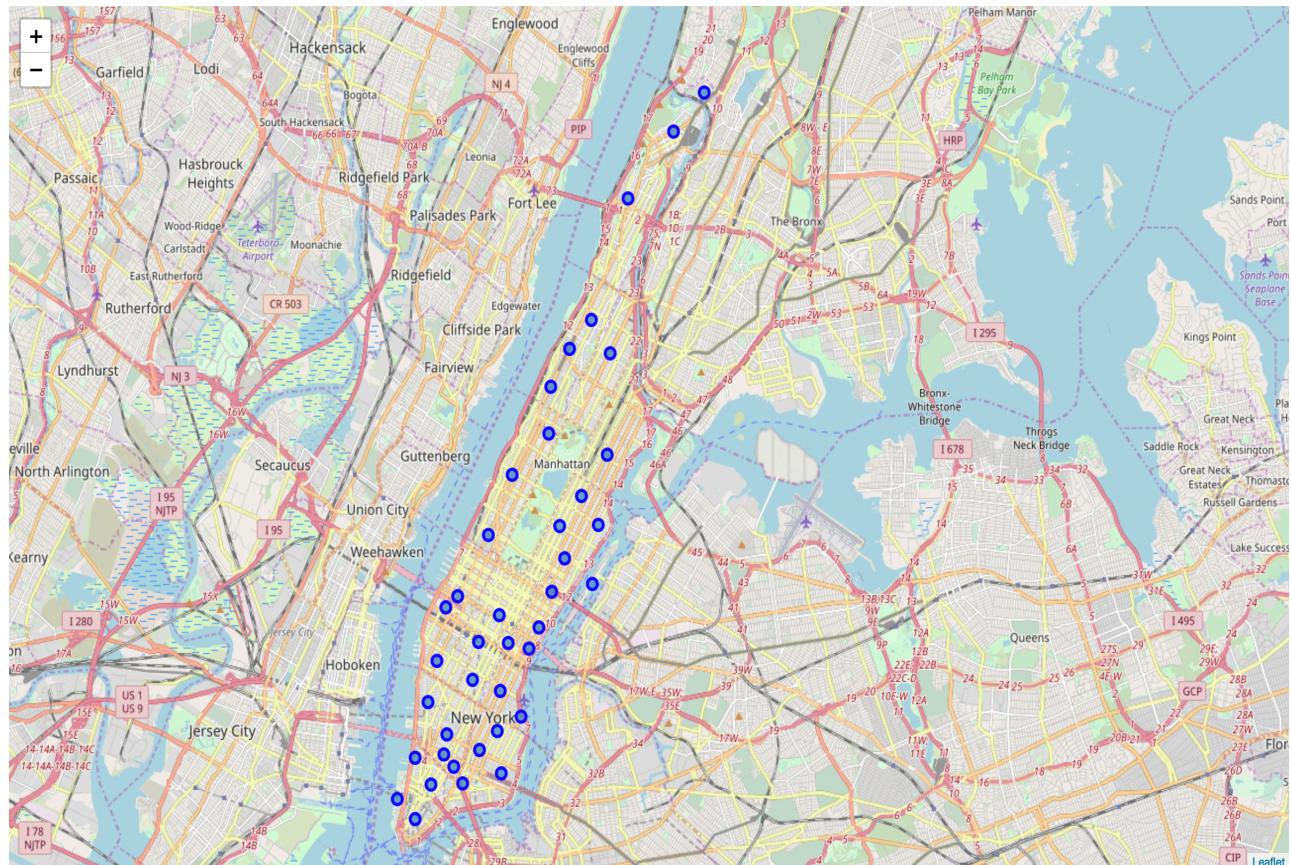
First of all, we need to find the New York City neighborhoods data. We download the data from https://geo.nyu.edu/catalog/nyu_2451_34572. After some cleaning, we get a dataset like this:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Next, we slice the New York City data and create a new dataset of Manhattan data.

Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	-73.910660
1	Manhattan	Chinatown	-73.994279
2	Manhattan	Washington Heights	-73.936900
3	Manhattan	Inwood	-73.921210
4	Manhattan	Hamilton Heights	-73.949688

Then we visualize the Manhattan with the 40 neighborhoods in it.



2. Foursquare

After have the neighborhoods candidates, we use Foursquare API to get information on food stores in Manhattan area in order to explore the stores and segment them.

	Venue	Longitude	Latitude
0	Arturo's	-73.910271	40.874412
1	Tibbett Diner	-73.908937	40.880404
2	Land & Sea Restaurant	-73.905873	40.877885
3	Dunkin' Donuts	-73.906507	40.876993
4	Parrilla Latina	-73.906073	40.877473
5	Boston Market	-73.905412	40.877430
6	Subway Sandwiches	-73.909586	40.874667
7	SUBWAY	-73.905385	40.878493
8	Subway	-73.905380	40.877720
9	Hernandez Grocery	-73.912591	40.875897
10	Terrace View Delicatessen	-73.912746	40.876476
11	Subway Sandwiches	-73.905308	40.878270
12	Spicy Village	-73.993530	40.717010
13	The Fat Radish	-73.991950	40.715323
14	Kiki's	-73.992036	40.714476
15	Super Taste (百味蘭州拉面) (Super Taste)	-73.993511	40.715198
16	Cheeky Sandwiches	-73.991529	40.715699
17	Scarr's Pizza	-73.991649	40.715335
18	Wah Fung Number 1 Fast Food 華豐快飯店	-73.994177	40.717278
19	Lan Zhou Handmade Noodle & Dumpling	-73.996742	40.715543

There are more than 2800 top-rated food stores in this dataset. We plot them on the Manhattan map.



Now we have all the top-rated food stores in Manhattan area. This concludes the data gathering phase. Now we're ready to use this data for analysis to produce the report on optimal locations for food delivery site!

Methodology

In this project we will put our efforts into finding the locations in Manhattan that have the nearest distance between them and the top-rated food stores.

At the first we have collected the required data: location and category of every top-rated food store in Manhattan.

Then We will 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.

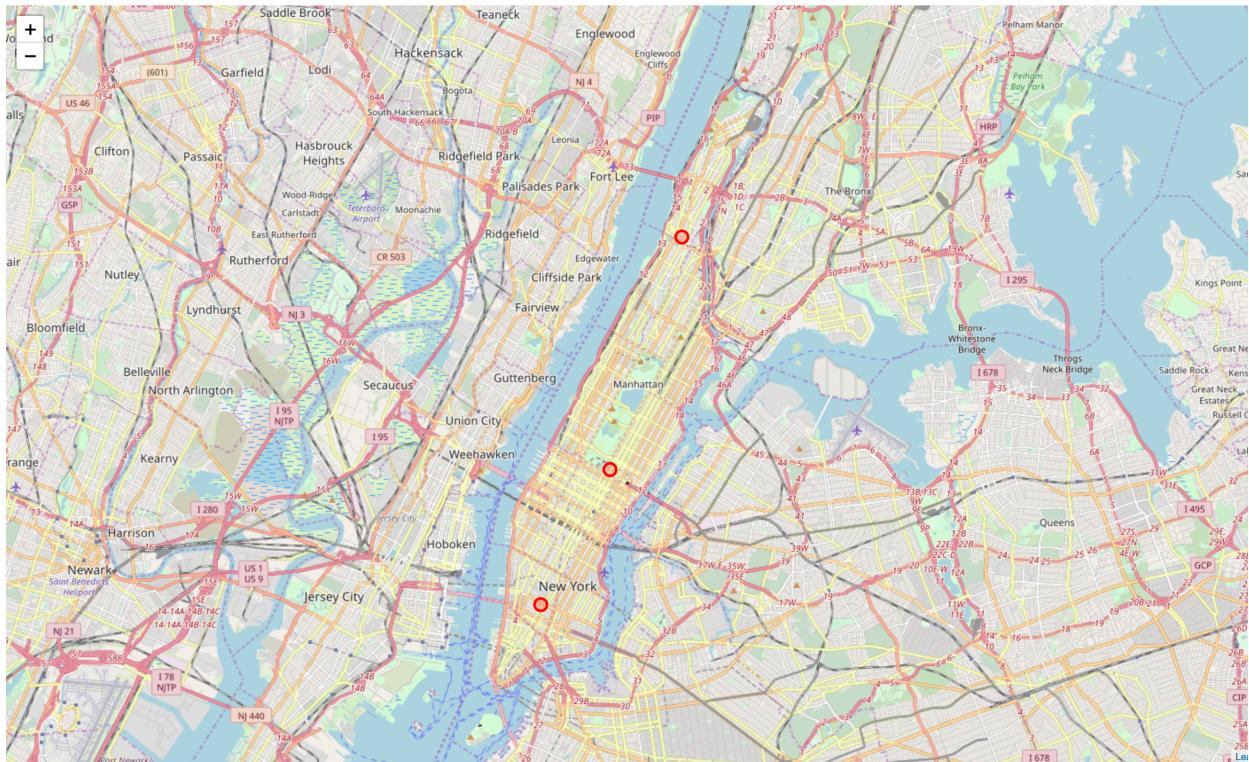
Analysis

Next, we can use K-means clustering analysis to segment the food stores. As we know Manhattan area has uptown, midtown and downtown three parts, so we are going to set our k as 3.

Finally, we find the positions we want and list as below:

	Latitude	Longitude
0	40.765089	-73.970822
1	40.833244	-73.943007
2	40.725485	-73.997637

We can find these locations on the map of Manhattan:



Results and Discussion

Our analysis tries to cluster more than 2800 top-rated food stores in Manhattan and find the centers of these clusters. First of all, we get all address data and handle them into onehot dataframe. Then we use K-means clustering algorithm to cluster the stores. According to the three areas of Manhattan: Uptown, midtown and downtown, we set the K as 3 and get the three centers of clusters. These centers will be the recommended best-fit positions for food delivery sites.

Conclusion

This project is based on some business assumptions and focus on finding best-fit positions of delivery sites for start-ups. Final decision on optimal locations will be made by our start-ups customers based on specific characteristics of the locations in every recommended zone, taking into consideration additional factors like cost of each location, proximity to major roads, real estate availability etc.

“THINK”