Site Selection of Frontline Warehouse for Inventory Preposition

Business Problem

LSC is an innovative distributor in NYC. Their main business is arranging coffee bean distributions to local coffee shops. However, since their distribution center is too far away from the NYC, their distributions are suffered from a high order to delivery time. LSC's new director, Jeff, who has a strong background in logistics and supply chain management, wants to try out a practice in military called inventory preposition in which inventory are loaded to a predefined area which are closed to the point of use so that the delivery can be made timely and thus improve the customer experience.

After a thorough research about the cost, the company decide to locate a frontline warehouse in Manhattan and use it as a proof of concepts. And now the company needs to choose a site for the warehouse.

This project can be taken as a pilot study on using machine learning to help the process of site selection, a typical task in supply chain management. In this project, both K-Means Clustering and DBSCAN are explored to find a ideal location grouping to do the site selection. Traditional center of gravity method is also applied based on the site grouping selected

Data

In this case, a location data consisting of the quantity of coffee shops in Manhattan area is required. Foursqure will be a good data source for this project as it has both neighborhood data in Manhattan and how many coffee shops in each neighborhood. Besides that, all coffee shops' location is also fetched in order to perform a better analysis

After collecting the data, a clustering approach will be applied to find the best coffee shop grouping, i.e. the ideal location for a frontline warehouse.

Data Descriptions

The data acquired from Foursquare consists of 7 types of data: Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude and Venue Category. The column venue category provides a perfect reference for us on identifying café in Manhattan area. And five categories Neighborhood, Café, Coffee Shop, Cantonese Restaurant, Pet Café, Convenience Store, are selected. And please be aware that all Cantonese restaurants are also considered as a café as they do provide perfect coffee experience. Other location data is used to plot the map and perform the clustering.

Methodology

There are two ways of doing site selection in traditional industry practice, one is by existing grouping in which a comprehensive ecosystem has already been built, and it would be easy to build a new supporting function with less administrative efforts. The other one is by exploring new groupings for selected or focused existing functions.

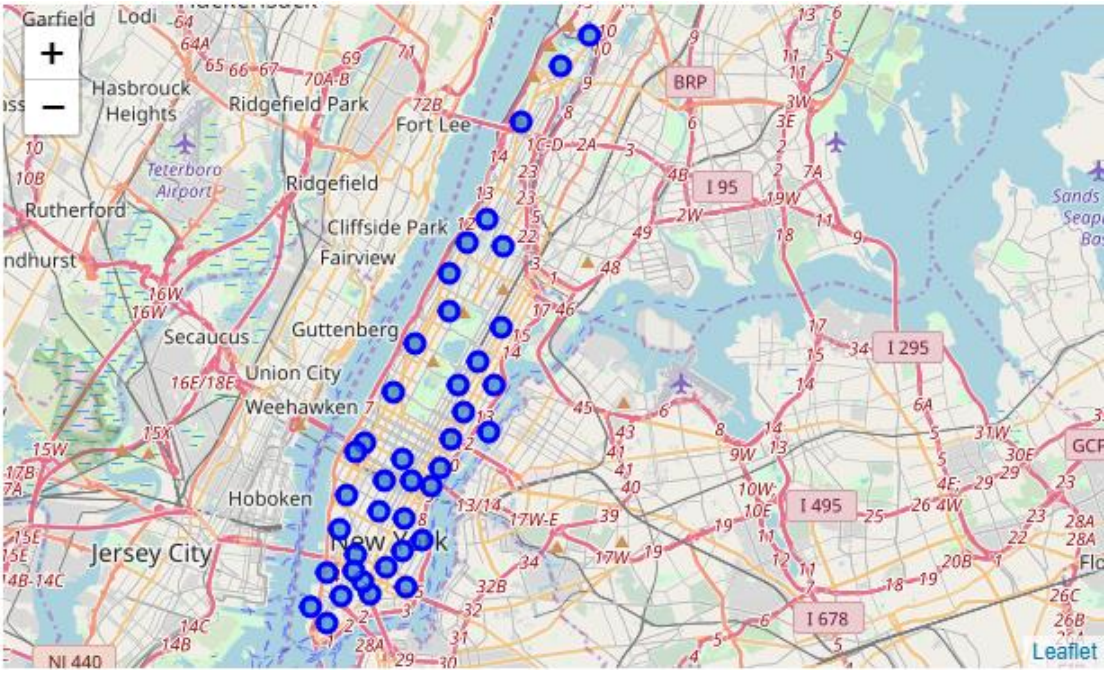
In the first approach, the first method can be described as explore the existing neighborhoods and their corresponding current venue. In this case, descriptive analysis is performed to form a basic understanding of the current neighborhoods' status, such as how many cafés on average exists in each neighborhood.

After the descriptive analysis, a K-means clustering will be taken and groups the existing neighborhoods into 3 classes showing different level of café shop quantities. If a certain pattern is discovered, a center of gravity method can be calculated and thus identified an ideal location for the new warehouse.

In the second approach, a new grouping based on café shops should be identified. To do so, DBSCAN can be applied to find the optimal grouping. Then a center of gravity method can be calculated and thus identified an ideal location for the new warehouse.

Results

We first acquired the data required, and plot them against the map of Manhattan:



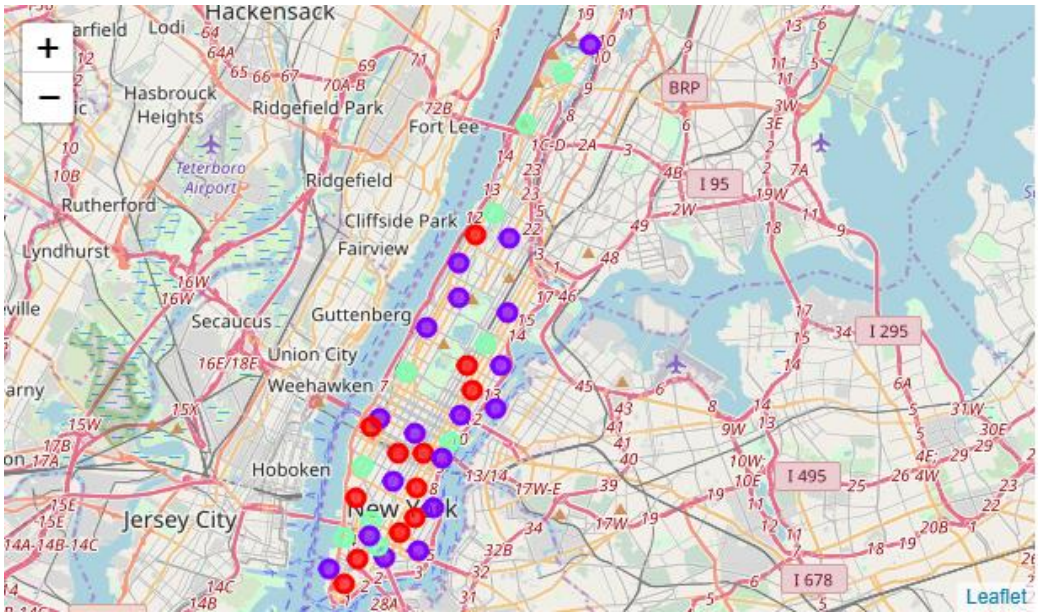
As showed from the map, the majority of the communities locate in the downtown area of Manhattan. So, it might be a hint for us to focus on the downtown area.

Then a statistical analysis is conducted to calculate the number of cafés in the area.

	Café	Coffee Shop	Cantonese Restaurant	Pet Café	Convenience Store
count	40.000000	40.000000	40.000000	40.000000	40.000000
mean	0.900000	1.275000	0.025000	0.025000	0.025000
std	1.057331	1.280775	0.158114	0.158114	0.158114
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000
50%	1.000000	1.000000	0.000000	0.000000	0.000000
75%	1.250000	2.000000	0.000000	0.000000	0.000000
max	3.000000	6.000000	1.000000	1.000000	1.000000

In the 40 neighborhoods of Manhattan, there is around 1 café shop type in each area, with a relatively large variance. It suggests that coffee is very popular in Manhattan area, so it is generally a great place for the company to locate a frontline warehouse.

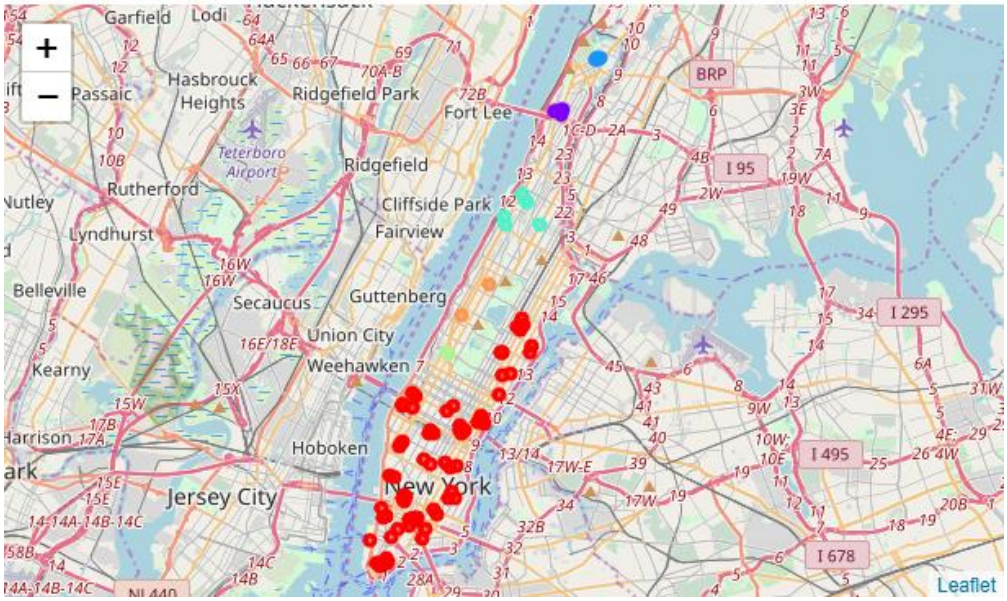
Then the neighborhoods are clustered based on their number of cafés there, and the results are shown as follows:



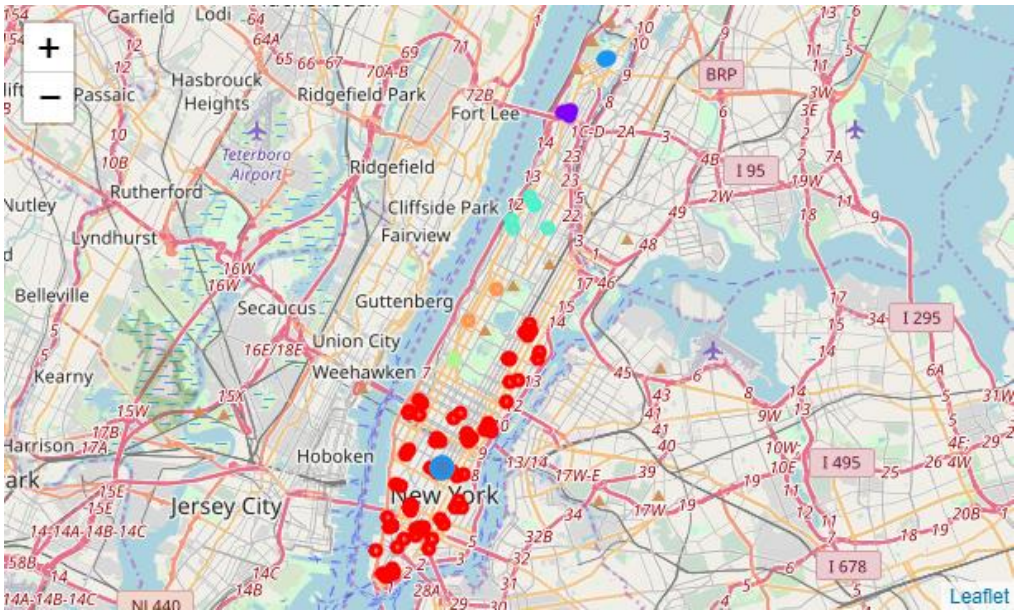
The map suggests that, there is no pattern on neighborhood level data, in which case a method with higher administrative effort has to be selected. In this case, we need

to plot each café in the map and tries to identity a patter.

In the second approach, DBSCAN are used to perform the clustering. And the result shows as follows:



The red dots are the recommended groupings for locating the new warehouse. The center of gravity method is used to find the center of the grouping. And it is calculated by the average of location data in the selected grouping



Discussion

The selection based on existing neighborhoods is not applicable in this case, as the

clustering shows roughly no notable pattern rather scatters among the map. So a more expensive approach should be applied, and based on the output of DBSCAN algorithms, the new experimental warehouse should be located in the downtown area of Manhattan in which many café shops located. And the optimal location should be in (40.74, -73.99)

Conclusion

To conclude, a location in or near (40.74, -73.99) should be considered for the new warehouse, as it is in the center of highlighted groupings, and machine learning algorithms can enhance the performance of traditional operations research algorithms to better perform the assigned tasks. However, further cost analysis and effort of implementation should be involved in the further analysis of the selected location. Further adjustments may need to be applied as new information may come out.