Restaurant Supply Co. expansion in NYC

Fabio Nozoy

April 28, 2021

1. Introduction

1.1 Background

"Restaurant Supply Co" is an enterprise dedicated to supply a numerous items of material for restaurants. The company wants to ingress in the New York City to dispute with various competitors.

1.2 Problem

Location and salary in New York City are very expensive. Those 2 resources have to be optimal.

1.3 Interest

First locations have to be in neighborhoods with most restaurants.

And the full implementation of the Sales Teams in NYC must balance the distribution of the teams in the city. Prerequisites are 40 restaurants per salesperson and 5 salesperson per Sales Team location.

2.Data acquisition and cleaning

2.1 Data Source

For this case, it makes sense to use the newyork_data.json data available on Coursera. It has NYC information's like 'Borough', 'Neighborhood', 'Latitude' and 'Longitude'.

foursquare.com has the list of points of interest that is assessable using explore API. For this case it is required the Asset Category, Latitude and Longitude.

2.2 Data Cleaning

Data collected from multiple sources were combined into one DataFrame, that were populated with a lot of undesired NYC points of interest. As there were not a single category such as restaurant or food, the categories were filtered using the following words: restaurant, bakery, coffee, café, bar, breakfast, burger, creperie, food, gastro, hot dog, pub and steakhouse.

3. Exploratory data analysis

3.1 Neighborhood with most restaurants

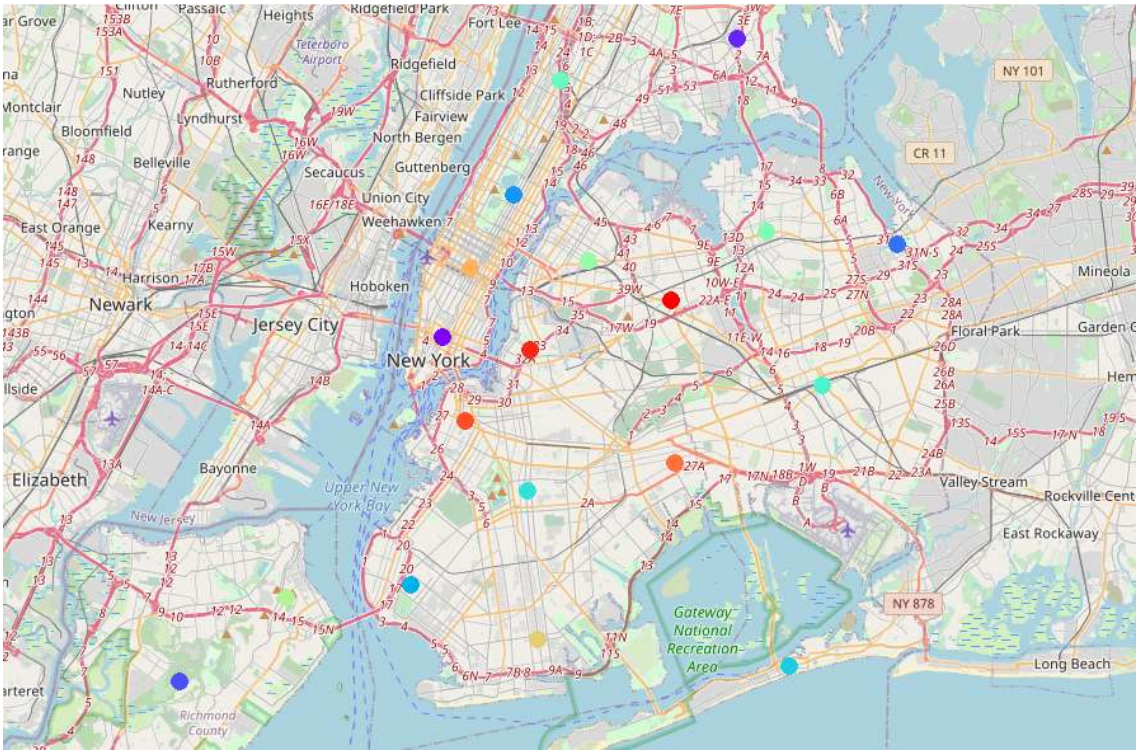
From the DataFrame it was obtained a sub product with the neighborhood name and the count() of restaurants, descending order by the quantity.

Neighborhood	counts
Murray Hill	88
Astoria	65
Upper West Side	63
South Side	61
East Village	60
Turtle Bay	57
Greenwich Village	55
Chinatown	55
Noho	55
North Side	54
Prospect Heights	53

Murray Hill neighborhood has plenty of restaurants with total of 88. From the 40 neighborhood of NYC the TOP 10 represents 15% of the restaurants of the city.

3.2 Best Sales team locations in NYC

Based on the latitude and longitude of each of the 4111 restaurants and the requirements of Sales team size and coverage, it is necessary 21 locations very well distributed in the city. It was used the method K-mean to determine 21 areas of responsibility with a list of restaurants that exclusively belongs to only one area, this list also determine center location of the area by a calculated latitude and longitude.



4. Conclusion

Purpose of this project was to identify NYC areas high number of restaurants in order to aid stakeholders in narrowing down the search for optimal location for restaurant supplies sales. By calculating restaurant density distribution from Foursquare data we have first identified to 10 Neighborhoods with Murray Hill at the top with 88 restaurants, 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 balance Sales Team centers, and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.