BUSINESS ANALYSIS FOR OPENING A RESTAURANT IN MANHATTAN OF NEW YORK CITY

By – Pratyush Sinha

Introduction/Business Problem

Manhattan often referred to locally as the City, is the most densely populated borough of New York City, its economic and administrative center, cultural identifier and historical birthplace. The borough is coextensive with New York County, one of the original counties of the U.S. state of New York. The borough consists mostly of Manhattan Island, bounded by the Hudson, East, and Harlem rivers; several small adjacent islands; and Marble Hill, a small neighborhood now on the U.S. mainland, physically connected to the Bronx and separated from the rest of Manhattan by the Harlem River.

A UK restaurants chain plans to open a new restaurant in Manhattan in New York City, since it is the most densely populated borough of New York City. Due to its dense population there is a huge increase in demand for restaurant business and is Ideal place to open a restaurant. The vendor does not have an idea of location that would be optimal for opening a restaurant hence they have reached out to me.

Can I suggest an optimal location for opening a new restaurant to the vendor by doing data analysis on current Restaurants and Geographical factors?

Data Description

The dataset used for this project is the json data of New York City. This data is made available by IBM and it can be found here and is ready to use. The data contains info about all the Borough, its neighborhood and their Latitude, Longitude info. Below are the two rows of above dataset

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279

Also, we will use FourSquare developer API to extract data for all the restaurants in Manhattan. We will combine the Manhattan data with data extracted from FourSquare API and create a new dataframe. This dataframe would be available for Data Science modeling.

Methodology

The location data of Manhattan, New York City is downloaded in the format of JSON files, which is then converted to a panda's data frame. By importing Folium library, a map of Manhattan city is drawn and all the neighborhoods were marked.
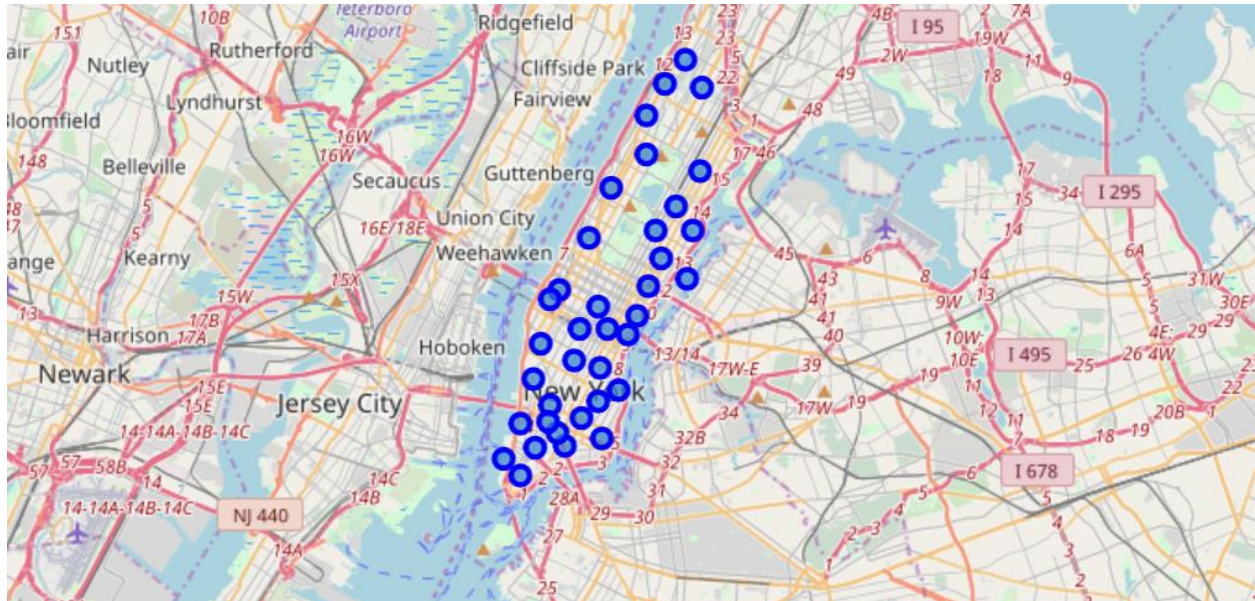
Next, I gathered data of all the existing Restaurants of Manhattan with the help of FourSquare API and this data is stored in a panda's data frame.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs
0	3 Guys Restaurant	Diner	49 E 96th St	US	New York	United States	Madison Ave	570	[49 E 96th St (Madison Ave), New York, NY 1012...	[{'label': 'display', 'lat': 40.78744262250426... 4
1	Junior's Restaurant & Bakery	American Restaurant	1515 Broadway	US	New York	United States	at W 45th St	4168	[1515 Broadway (at W 45th St), New York, NY 10...	[{'label': 'display', 'lat': 40.758539, 'lng':... 4
2	Carmine's Italian Restaurant	Italian Restaurant	2450 Broadway	US	New York	United States	btwn W 90th & W 91st	1198	[2450 Broadway (btwn W 90th & W 91st), New Yor...	[{'label': 'display', 'lat': 40.7910963, 'lng'... 4
3	Malecon Restaurant II	Latin American Restaurant	764 Amsterdam Ave	US	New York	United States	btw 97th St & 98th St	987	[764 Amsterdam Ave (btw 97th St & 98th St), Ne...	[{'label': 'display', 'lat': 40.79493159833159... 4
4	Gabriela's Restaurant & Tequila Bar	Mexican Restaurant	688 Columbus Ave	US	New York	United States	at 93rd St.	761	[688 Columbus Ave (at 93rd St.), New York, NY ...	[{'label': 'display', 'lat': 40.79123991711048... 4

Then based on postal codes, restaurant were groupd to determine the density. Below is the current restaurants of Mnahattan city.
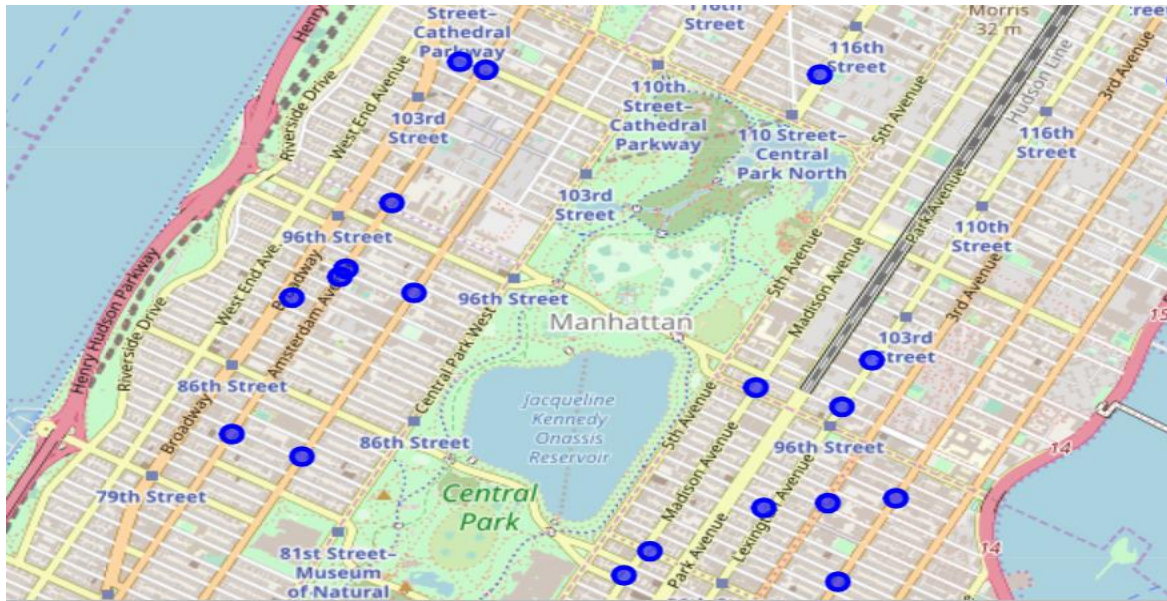
Then one hot encoding is used to convert the categorical variable(for our scenario, restaurant name variable is used).

	postalCode	3 Guys Restaurant	525LEX Restaurant & Lounge	88 Chinese Restaurant	Accra Restaurant	Acosta Deli - Restaurant	Agora Turkish Restaurant	Astro Restaurant	Italian Restaurant & Brick Oven Pizza	Carmine's Italian Restaurant	Carmine Italia Restaurar
0	10128	1	0	0	0	0	0	0	0	0	
1	10036	0	0	0	0	0	0	0	0	0	
2	10024	0	0	0	0	0	0	0	0	1	
3	10025	0	0	0	0	0	0	0	0	0	
4	10025	0	0	0	0	0	0	0	0	0	

The mean of frequency of occurrence of restaurant are grouped by postal codes.

```
manhattan_new_grouped = manhattan_new_onehot.groupby('postalCode').mean().reset_index()
manhattan_new_grouped
```

	postalCode	3 Guys Restaurant	525LEX Restaurant & Lounge	88 Chinese Restaurant	Accra Restaurant	Acosta Deli - Restaurant	Agora Turkish Restaurant	Astro Restaurant	Bellini Italian Restaurant & Brick Oven Pizza	Carmine's Italian Restaurant	R
0	10001	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	
1	10010	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	
2	10017	0.0	0.5	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	
3	10019	0.0	0.0	0.0	0.0	0.000000	0.000000	0.166667	0.000000	0.000000	
4	10021	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	
5	10022	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	
6	10023	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	
7	10024	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.333333	0.333333	

Our goal is identify the current restaurant distribution of Manhattan. I used for K-mean clustering on the preprocessed data. Since K-mean clustering is an efficient model for pattern recognition, data segregation and comparison.

Results

k-mean clustering is performed using 5 clusters. The results showed that some of the locations have a higher density of cafes while some other has lower density of restaurants. Hence, places with medium density will be ideal for opening restaurants.

Observations and recommendations

Based on the business analysis of the restaurant density. The neighborhood regions under postcode 10024, 10029, 10027 are optimal for opening a restaurants.

Conclusion

In this project we aimed to do a business analysis to find a suitable location opening a new restaurant in Manhattan of New York City using the Manhattan data and FourSquare data. I choose k-mean which clustered the restaurant according to postal code. After analyzing the cluster, we found the best place to open a restaurant.