Restaurant Location – Recommender System

_____

Submitted by

Ruthvic Punyamurtula

IBM Data Science Professional Coursera Capstone

Introduction/Business Problem

Restaurant Location – Recommender System

Manhattan is the most densely populated of New York City's 5 boroughs. It's mostly made up of Manhattan Island, bounded by the Hudson, East and Harlem rivers. Among the world's major commercial, financial and cultural centers, it's the heart of "the Big Apple". New York City has been called both the most economically powerful city and the leading financial center of the world, Manhattan is home to the world's two largest stock exchanges, the New York Stock Exchange and NASDAQ. Manhattan has high population with a population density of 72,918.4/sq mi as of 2017.

So a city with high population and fast developing resources requires more food outlets. Due to its dense population there is a huge increase in demand for restaurant business and Manhattan is an ideal place to open a restaurant. A newbie in restaurants chain plans to open their branch in Manhattan, New York City and the vendor does not have any idea in picking an ideal location which is economic as well as profitable to them. So I being a data science practitioner would like to analyze the data and help them pick a right place.

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 at https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json and is ready to use. The data contains info about all the Borough, its neighborhood and their Latitude, Longitude info. Below is the head of the dataframe displaying 5 rows

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

      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 used for clustering and further analysis.

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.
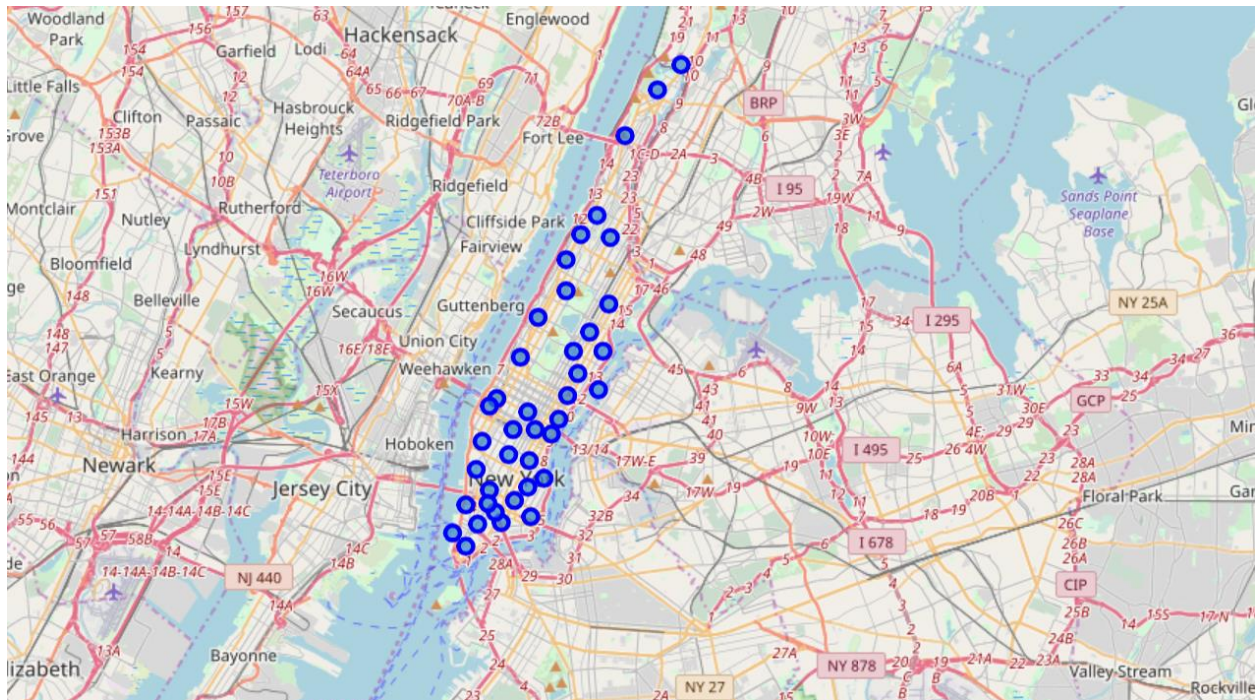
Next, I have gathered data of all the existing Restaurants of Manhattan with the help of FourSquare API and this data is stored in a panda's data frame.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs
0	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...
1	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...
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'...
3	Junior's Restaurant & Bakery	American Restaurant	1515 Broadway	US	New York	United States	at W 45th St	4188	[1515 Broadway (at W 45th St), New York, NY 10...	[{'label': 'display', 'lat': 40.75834934354621...
4	Fred's Restaurant	American Restaurant	476 Amsterdam Ave.	US	New York	United States	at W 83rd St	1492	[476 Amsterdam Ave. (at W 83rd St), New York, ...	[{'label': 'display', 'lat': 40.78565778922316...

Then based on postal codes, restaurant were grouped to determine the density. Below is the current restaurants of Manhattan 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	Bellini Italian Restaurant & Brick Oven Pizza	Carmine's Italian Restaurant	...	Sylvia's Restaurant	The Carlyle Restaurant	The F Restau
0	10025	0	0	0	0	0	0	0	0	0	...	0	0	
1	10128	1	0	0	0	0	0	0	0	0	...	0	0	
2	10024	0	0	0	0	0	0	0	0	1	...	0	0	
3	10036	0	0	0	0	0	0	0	0	0	...	0	0	
4	10024	0	0	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	...	Sylvia's Restaurant	The Carlyle Restaurant
0	10001	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	...	0.0	0.0
1	10017	0.0	0.333333	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	...	0.0	0.0
2	10019	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.2	0.000000	0.000000	...	0.0	0.0
3	10021	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	...	0.0	0.0
4	10022	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	...	0.0	0.0
5	10023	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	...	0.0	0.0

Our goal is to 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 restaurant.

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.