Problem

Where is the proper location in New York to open a restaurant?

Using Data science analytics, location data and map of the area we help business to solve their problem to find a better place to open a new restaurant which turns out to be profitable to the clients.

Target

- The clients who reached us for a solution
- Startups who want to open a food place in the city of New York
- Enthusiasts who want to acquire knowledge on a small scale data science project

Although the main target and purpose of this project is to considering the business problem of the client who came up with "What is the proper place to open a restaurant in the city of New York?"

To answer this question we need data and information about the food chains and restaurants in the city of New York. For this data we collect it from the freely available datasets from the internet.

The details of the data collected is given below

1. <u>Collection</u>

   <u>Data</u>

   We collected the data from

   - NYU spatial data repository.
   - Foursquare location and venue data.

   We collected the names of the neighborhoods and communities present in the city with the help of the NYU spatial data repository.

   We collected the names of restaurants and venues all around the city of New York with the help of the Foursquare location data.

   With the help of these two combines dataset and the visualization and analysis techniques we can find a best place in the city that meets the needs of the problem.

   With all the needed data collected we can proceed to the analysis part of the problem where find the best possible solution to the problem.

   Samples of the data are given below

	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

Sample of the neighborhoods

```
Co-op City
Eastchester
Fieldston
Riverdale
Kingsbridge
Marble Hill
Woodlawn
Norwood
Williamsbridge
Baychester
Pelham Parkway
City Island
Bedford Park
University Heights
Morris Heights
Fordham
East Tremont
West Farms
High   Bridge
Melrose
Mott Haven
Port Morris
Longwood
Hunts Point
Morrisania
Soundview
Clason Point
Throgs Neck
Country Club
Parkchester
Westchester Square
Van Nest
Morris Park
Belmont
Spuyten Duyvil
North Riverdale
Pelham Bay
Schuylerville
```

Sample of the venues

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

Combined dataset of neighborhoods and venues

2. Analysis

- Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

Now after retrieving the data we perform the statistical analysis on the data.

➢ Firstly we obtain the data from the above mentioned websites and import them to our working environment.

➢ We import the libraries we need to work, as they are very useful in the course of the work.

➢ After the import, we refine the data we obtained. This includes removing the content we don't need. Correcting the data types of the columns. Identifying the Not Assigned values. Removing the NaN values as they produce wrong results.

- After that we append the dataset into a pandas data frame
- Now after appending the neighborhood details, we obtain the foursquare location data
- Using foursquare data we create a new data frame with the help of pandas.
- We merge the neighborhood data and the foursquare location data to get a single data frame containing the *Name of neighborhood, venue, venue category, latitude and longitude of the neighborhood and the venue.*
- Now we find the most common venues visited by the people.
- Then we use *K-clustering means* method to create 5 clusters.
- These clusters are based on factors like distance and similarity of venues.
- After dividing the area into clusters we can now see and analyze which cluster is suitable for the business problem stated.
- The five clusters are obtained to further perform the decision analysis and find a better solution to the problem.

- <u>Results</u>

  In the process of finding the best profitable place suitable to open a restaurant in New York, we considered these factors to give our solution

  - ❖ Competition
  - ❖ Reach
  - ❖ Publicity

  Observing the 5 clusters we can say that the clusters 1, 3 and 4 are not the clusters we are searching for. There is no apparent benefit if any restaurant business is started in the areas which fall under the clusters 1, 3 and 4.

  Examining the clusters 2 and 5 we can easily say that this is the best place to open a restaurant.

- **<u>Discussion</u>**

When we discuss about the problem given through highlighting the areas found best suitable for the business startup, we see that cluster 4 is a good option as the place is centre for a lot of restaurants with a lot of cuisine options. This is a good choice if we want to open a new restaurant. But opening a restaurant here has its own advantages and disadvantages.

<u>Cluster-5</u>

An advantage being a new restaurant is a good option for the people to break the routine and try a new flavor and ambience. This attracts the attention and good word of mouth publicity from the customers if proven good.

Disadvantage is the fierce competition. Due to the high number of restaurants and its loyal customer base, generally people don't often move to new restaurant. This is not generally the case but need to be aware of.

To conclude about cluster 5 we can say that it's a good area to open with its own pros and cons. It's the client needs which can finally take a decision on this area.

Cluster-2

Coming to cluster -2, this is not as populated as cluster -5 in restaurants implying a good chance of development of a new startup. Cluster-2 is in abundance in facilities of public transport adding extra benefit. Frequent movement of people in the area attracts a significant amount of people to the restaurant. With the word of mouth publicity it's easy on the restaurant to reduce on the advertising bills. Lack of the heavy competition business can easily break the ice with the customers and try something new than to follow a strict pattern.

To conclude about cluster 2, we can say it is the best of the available options followed by cluster 5.

3. Conclusion

On concluding we can say that the cluster 2 is the best option to open a restaurant business. Cluster 5 also satisfies the needs but the heavy competition, less public transport facilities make it less chosen option.

The final decision is taken by the client based on their resources and choices.
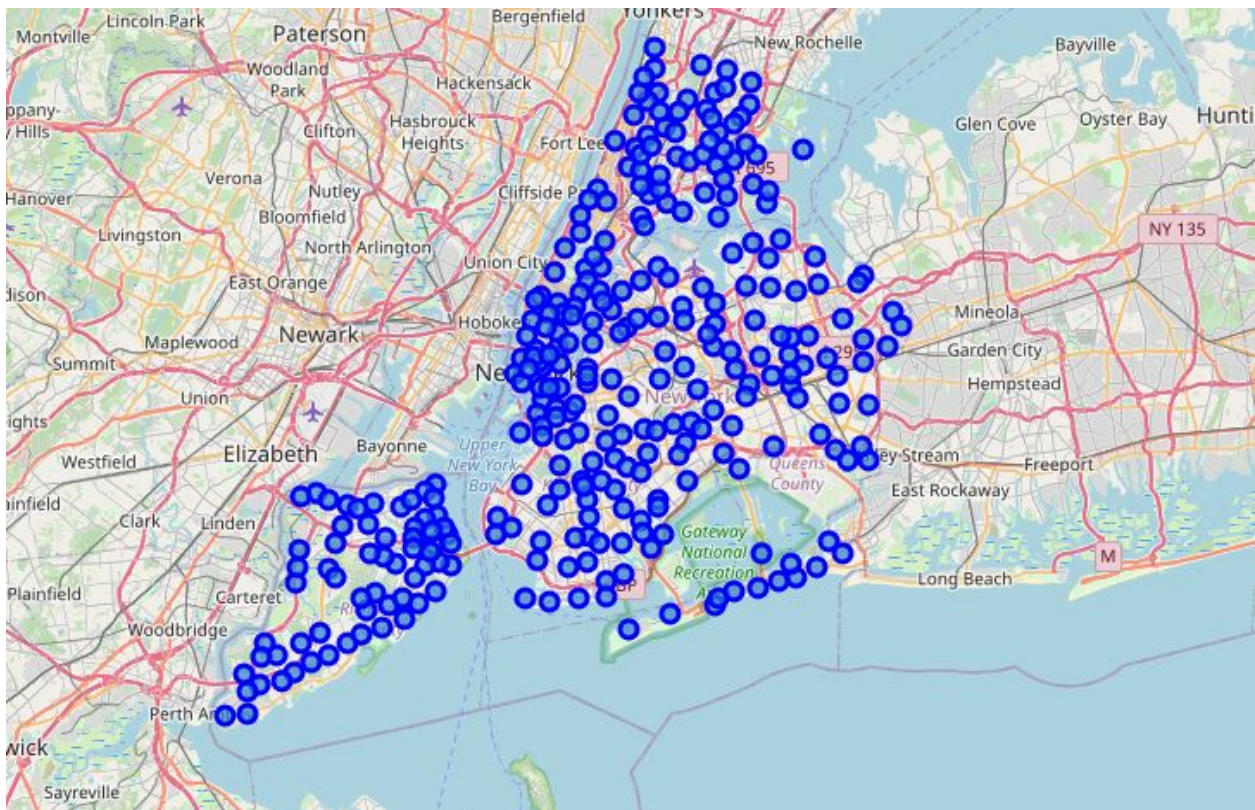
Images from the analysis

Co-op City
Eastchester
Fieldston
Riverdale
Kingsbridge
Marble Hill
Woodlawn
Norwood
Williamsbridge
Baychester
Pelham Parkway
City Island
Bedford Park
University Heights
Morris Heights
Fordham
East Tremont
West Farms
High Bridge
Melrose
Mott Haven
Port Morris
Longwood
Hunts Point
Morrisania
Soundview
Clason Point
Throgs Neck
Country Club
Parkchester
Westchester Square
Van Nest
Morris Park
Belmont
Spuyten Duyvil
North Riverdale
Pelham Bay
Schuylerville
Edgewater Park

Sample of list of venues

```
print(ny_venues.shape)
ny_venues.head()
```

(9826, 7)

number of venues considered

Clusters map

****