Capstone Project - The Battle of Neighborhoods

❖ **Introduction:**

The City of New York, is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States. This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

❖ **Business Problem:**

The City of New York is famous for its entertaining culture. The dance culture comprises of various international clubs inclined by the city's immigrant history. Dance studios became very popular in the United States and it spread across various location across the geographical area. The rich culture with lively people makes thing more attractive for international travels.
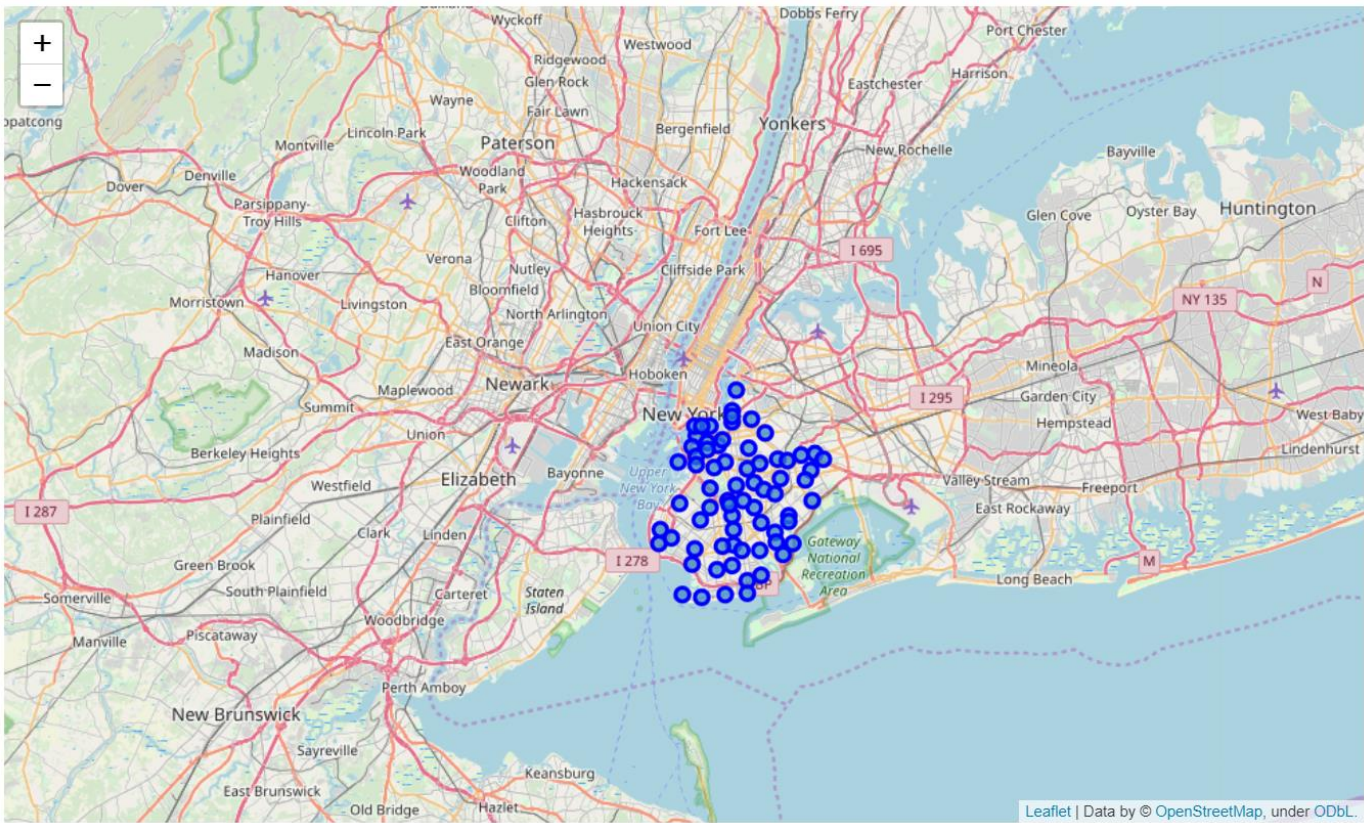
The need of the hour is to get a dance studio nearby accompanied with security, low cost and hygienic environment. The Business problem depicted in this project is to address a user-friendly searching app which encapsulated with easy searching techniques and highly trained machine learning models.

❖ **Data Selection:**

	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

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

❖ New York Map:

❖ Methodology:

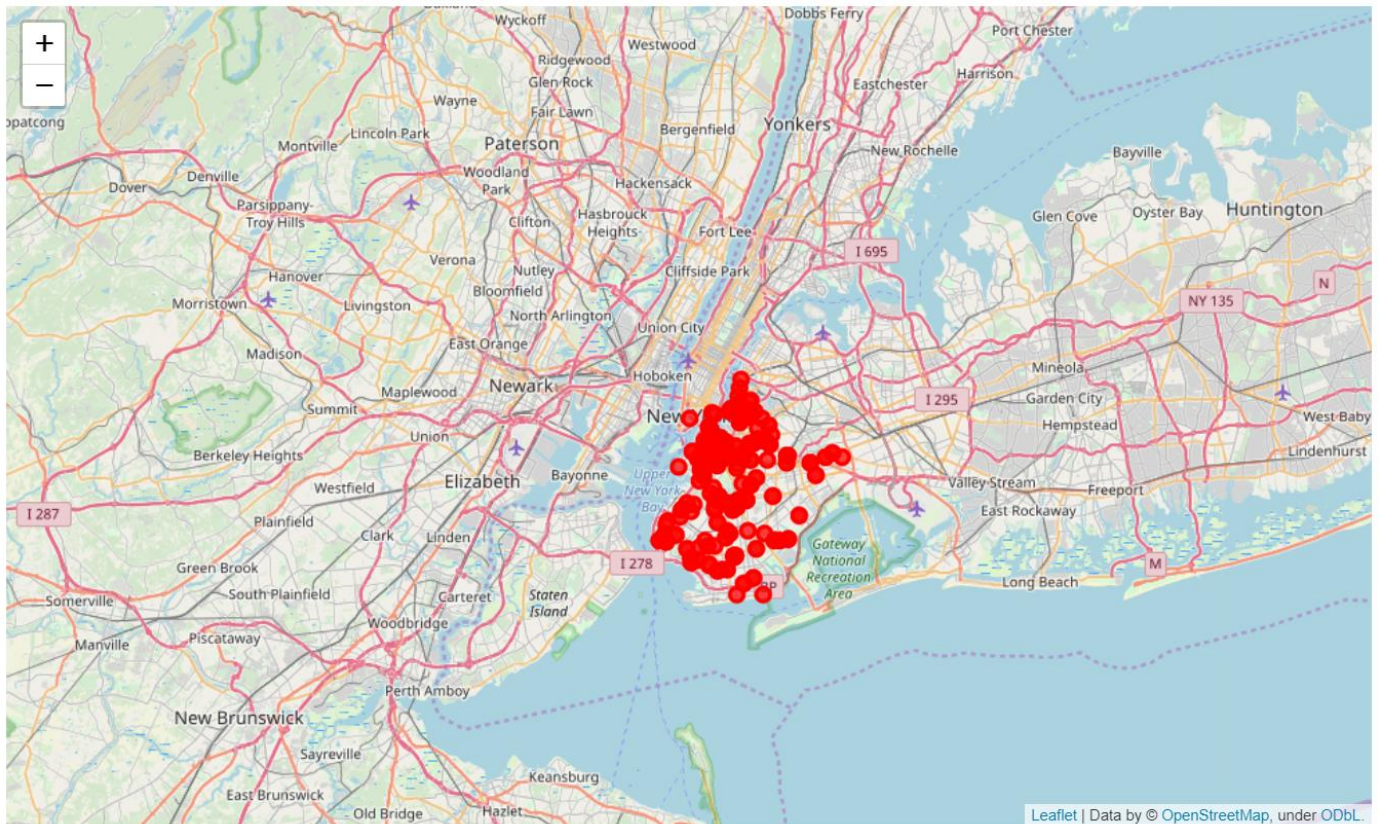|	Neighborhood	Art Gallery	Country Dance Club	Dance Studio	Event Space	General Entertainment	Gymnastics Gym	Office	Performing Arts Venue	School	Yoga Studio
0	Bay Ridge	0	0	1	0	0	0	0	0	0	0
1	Bay Ridge	0	0	1	0	0	0	0	0	0	0
2	Bay Ridge	0	0	1	0	0	0	0	0	0	0
3	Bay Ridge	0	0	1	0	0	0	0	0	0	0
4	Bay Ridge	0	0	1	0	0	0	0	0	0	0

	Neighborhood	Art Gallery	Country Dance Club	Dance Studio	Event Space	General Entertainment	Gymnastics Gym	Office	Performing Arts Venue	School	Yoga Studio
0	Bath Beach	0.000000	0.000	0.666667	0.000	0.000000	0.166667	0.000000	0.000000	0.000000	0.166667
1	Bay Ridge	0.000000	0.000	1.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Bedford Stuyvesant	0.000000	0.125	0.875000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Bensonhurst	0.000000	0.000	0.714286	0.000	0.000000	0.142857	0.000000	0.000000	0.000000	0.142857
4	Bergen Beach	0.000000	0.000	1.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	Boerum Hill	0.040000	0.000	0.960000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Borough Park	0.000000	0.000	1.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Brighton Beach	0.000000	0.000	1.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	Broadway Junction	0.000000	0.000	1.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	Brooklyn Heights	0.076923	0.000	0.846154	0.000	0.000000	0.000000	0.000000	0.076923	0.000000	0.000000
10	Bushwick	0.000000	0.000	1.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bath Beach	Dance Studio	Yoga Studio	Gymnastics Gym	School	Performing Arts Venue	Office	General Entertainment	Event Space	Country Dance Club	Art Gallery
1	Bay Ridge	Dance Studio	Yoga Studio	School	Performing Arts Venue	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club	Art Gallery
2	Bedford Stuyvesant	Dance Studio	Country Dance Club	Yoga Studio	School	Performing Arts Venue	Office	Gymnastics Gym	General Entertainment	Event Space	Art Gallery
3	Bensonhurst	Dance Studio	Yoga Studio	Gymnastics Gym	School	Performing Arts Venue	Office	General Entertainment	Event Space	Country Dance Club	Art Gallery
4	Bergen Beach	Dance Studio	Yoga Studio	School	Performing Arts Venue	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club	Art Gallery

Top 'N' Neighborhood

Dance Studio in Brooklyn

❖ **Results:**

Cluster - 1

```
In [31]: brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 1, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape
         [1]))]]
```

Out[31]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
19	Cobble Hill	Dance Studio	School	Art Gallery	Yoga Studio	Performing Arts Venue	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club
20	Carroll Gardens	Dance Studio	School	Art Gallery	Yoga Studio	Performing Arts Venue	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club

Cluster - 2

```
In [32]: brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 2, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]]
```

Out[32]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Bensonhurst	Dance Studio	Yoga Studio	Gymnastics Gym	School	Performing Arts Venue	Office	General Entertainment	Event Space	Country Dance Club	Art Gallery
33	Bath Beach	Dance Studio	Yoga Studio	Gymnastics Gym	School	Performing Arts Venue	Office	General Entertainment	Event Space	Country Dance Club	Art Gallery

Cluster - 3

```
In [33]: brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 3, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]]
```

Out[33]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
15	Williamsburg	Dance Studio	General Entertainment	Yoga Studio	School	Performing Arts Venue	Office	Gymnastics Gym	Event Space	Country Dance Club	Art Gallery
17	Bedford Stuyvesant	Dance Studio	Country Dance Club	Yoga Studio	School	Performing Arts Venue	Office	Gymnastics Gym	General Entertainment	Event Space	Art Gallery
18	Brooklyn Heights	Dance Studio	Performing Arts Venue	Art Gallery	Yoga Studio	School	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club
22	Gowanus	Dance Studio	School	Yoga Studio	Performing Arts Venue	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club	Art Gallery
40	Downtown	Dance Studio	Performing Arts Venue	Art Gallery	Yoga Studio	School	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club
50	North Side	Dance Studio	Office	General Entertainment	Yoga Studio	School	Performing Arts Venue	Gymnastics Gym	Event Space	Country Dance Club	Art Gallery
51	South Side	Dance Studio	Office	General Entertainment	Yoga Studio	School	Performing Arts Venue	Gymnastics Gym	Event Space	Country Dance Club	Art Gallery
61	Fulton Ferry	Dance Studio	Performing Arts Venue	Yoga Studio	School	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club	Art Gallery
62	Vinegar Hill	Dance Studio	Performing Arts Venue	Yoga Studio	School	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club	Art Gallery
65	Dumbo	Dance Studio	Performing Arts Venue	Yoga Studio	School	Office	Gymnastics Gym	General Entertainment	Event Space	Country Dance Club	Art Gallery

Cluster - 4

```
In [34]: brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 4, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]]
```

Out[34]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Greenpoint	Dance Studio	Office	Event Space	Yoga Studio	School	Performing Arts Venue	Gymnastics Gym	General Entertainment	Country Dance Club	Art Gallery

❖ **Discussion:**

It is clear now that k-means is a simplistic, yet powerful algorithm and it can be useful for many different types of problems that may arise in analytics. With that said, it may not always be the best choice for our problem and there are some assumptions that the algorithm makes which we need to be aware of if we are going to use it. Probably the biggest assumption and limitation of k-means is that it assumes that the clusters are spherical.

❖ **Conclusion:**

This presentation is developed for IBM data science capstone project assignment to keep in view the optimal usage of machine learning model for better outcome. Although dataset used is very small and the accuracy of the outcome is compromised but is it a good opportunity to showcase the learning of the course by adhering the use case.