Rktv’s Final Project Capstone

**Introduction/Business Problem:**

Yoga has become a part of the fitness culture of the New York City and widely accepted across the multi-cultural communities of the vibrant city. Yoga is what everyone needs in a city that never sleeps, like New York. Even with its high sprung culture, you will be sure to find a class you love and a yoga community that welcomes you. One can find many different lineages of **yoga classes in New York**. Whether in Midtown, downtown, or suburbs. One can find plenty of workshops and certified teacher training courses spread across Manhattan and Brooklyn compared to other boroughs.

Now, if someone is looking to open a new yoga studio in NYC, can we use data science to help to find out whether a specific area is suitable for business or not?

This is the **business problem** I would like to address through this project. The **target audience / stakeholders** are those who plan to open a new yoga studio in NYC, whether as a start-up or a new franchisee of an existing brand, and their investors.

The goal of this analysis is to explore the NYC to identify the number of Yoga studios available in each area to decide on the best possible location to start a new yoga studio.

Is it advisable to open a new studio in an untapped locality or does it make sense to open in a clustered area where the density of the studios is more. What other indicators we can derive from the data that makes it a justifiable recommendation.

**Data:**

The freely available NYC dataset will be used for this study.

We will convert addresses into their equivalent latitude and longitude values. Also, we will use the Foursquare API to explore neighborhoods in New York City. We will use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. K-means clustering algorithm will be used for this task. Finally, the Folium library will be used to visualize the neighborhoods in New York City and their emerging clusters to identify the availability of number of Yoga studios in each area.

Data Preparation and cleaning:

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

The data set is transformed to Pandas dataframe and populated one row at a time.



The success of the Yoga studio will largely depend on one’s ability to understand and relate to the neighborhood of choice than on the amount of foot traffic passing the studio.

It is important to analyze the neighborhoods. Some neighborhoods, like those with fitness clubs and health food stores, work much better than others.

This project doesn’t look in to the nature of services to be offered but only the location.

Neighborhood names, their latitudes and longitudes and top venues are obtained for exploration.

The venues details are filtered for the key word Yoga to identify the number of Yoga studios in each neighborhoods.

The clusters are ccreated using K-means .



Conclusion and future work:

This project provides initial insights about the existing Yoga studios in each neighbourhood and acts as in put to the prospective Business team that is considering opening a new Yoga studio business in Manhattan.

The location is one of the important things ,however this project doesn’t investigate or recommend the nature of the services to be offered as a competitive strategy.That can be addressed in future work.