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.