**Introduction: Business Problem**

The year 2020 has brought along with it many challenges related to the current pandemic. Individuals living in cities such as New York have limited living space and often rely on gyms as a place to stay in shape, unplug from work and family life, and socialize with like-minded individuals. This demand has been exacerbated by temporary gym closures due to government mandated lockdowns in an attempt to decelerate the spread of Covid-19. People generally would rather pay monthly fees to get access to a wide variety of equipment and classes, instead of paying for at-home equipment that is expensive, takes up a lot of space, and is now often out-of-stock due to demand created by the pandemic.

NYC storefront vacancies and decreased foot traffic have led to lower rents. This, coupled with the pent-up demand related to people wanting to get out of their apartments and experience pre-pandemic life again, make 2021 a great year to launch a fitness business. Manhattan is the most densely populated of New York City's 5 boroughs with about 1.6m people calling it home and, pre-pandemic, another 1.6m people commuting into the borough daily. It is mostly made up of Manhattan Island, bounded by the Hudson, East and Harlem rivers. The island is among the world's major commercial, financial, and cultural centers.

This report will focus on finding recommended locations that can be used to open a gym in the borough of Manhattan. First, we will identify already established gyms. Next, we will use a clustering model to find similar areas in Manhattan using demographic data from each borough and region. Ideally, we will find a suitable location that is not near any existing gyms. To accomplish our goal, we will use a variety of data science tools to fetch the raw data, visualize it, and identify a few of the most promising areas based on the above criteria. We will also explain why specific locations were identified, along with describing their advantages and traits so the different options can be debated amongst the business’ stakeholders.

**Data**

In this project we’ll first fetch and extract a json dataset comprised of each neighborhood in the five boroughs of New York City. We will then take a quick look at the dataset in our notebook to ensure the output is as planned. After reviewing the dataset’s structure, we will set up an empty dataframe to store only Manhattan’s neighborhoods, define the dataframe fields we need to pull from our original json dataset, then use a for loop to loop through the original dataset and populate our new dataframe row by row. Finally, we’ll preview our new dataframe by using the Pandas dataframe.head() method to ensure our loop code worked correctly. This dataframe will then be used in conjunction with data procured from a Foursquare API. This API connection will enable us to identify neighborhood business clusters. Neighborhoods without a high concentration of gyms, but that have characteristics of other neighborhoods with concentrations of gyms, will be recommended locations for opening one of our own.