# Coursera Applied Data Science Capstone Course
# Capstone Project - The Battle of Neighborhoods

**Phyo Shwe Yee Win**

**April 12, 2020**

## Introduction

New York City is the epicenter of the coronavirus pandemic in the United States. As the number of Covid-19 cases grew in NYC throughout March 2020, subway ridership plunged as businesses implemented remote working policies, and non-essential businesses closed. In this project, I will utilize the subway ridership data from the NYC Metropolitan Transportation Authority ("MTA") along with the FourSquare API to identify what kind of venues are most common near the subway stations that showed the greatest drop in ridership in March. Based on the results, we can infer what type of venues would have experienced the largest drop in the number of customers/visitors. We can also potentially infer what type of venues people would most likely visit if the New York State government were to relax the social distancing measures in the future. This analysis can be used by the New York State and New York City governments in planning support programs for businesses hardest hit by the pandemic, as well as provide them insights on how to start relaxing the New York State on PAUSE order when the time comes.

## Data

First, I will utilize the subway turnstile data from the MTA, available at http://web.mta.info/developers/turnstile.html. This dataset includes the cumulative number of entries and exits for each NYC subway station control area at various points in time, normally every four hours. From this dataset, I can calculate the daily number of entries and exits for each station, and identify the stations for which daily subway exits dropped the most from the beginning of March to the end of March.

Next, I will obtain the geographical coordinates for the identified stations using the NYC subway station geolocation data, available from the NYC Open Data website at https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49. I can obtain a GeoJSON file containing the names of the subway stations in New York City, as well as their latitude and longitudes.

Finally, I will use the FourSquare API to obtain the most popular venues around the identified stations and analyze whether the areas around the stations share particular characteristics, such as having mostly restaurants, fitness studios or coffee shops.