**New York City neighbourhoods and their susceptibility to COVID-19**

**Introduction**

The coronavirus disease 2019 (COVID-19) is an infectious disease that has resulted in a global pandemic that has affected the lives of millions of people around the world. While the impact of this disease has been felt everywhere, some neighbourhoods have been hit harder than others. In this project, I would like to investigate how the number and types of venues in neighbourhoods in New York City (NYC) affects the number of cases in that neighbourhood. This investigation will be important to many stakeholders like the government, hospitals and ordinary people. With the results of this study, the government will be able to identify the types of neighbourhoods which require more attention with regards to the pandemic, while hospitals in such neighbourhoods can prepare themselves for a large influx of patients and ordinary people living in such neighbourhoods can take the necessary precautions to avoid exposure.

**Data**

I will be using three main sources of data for this project. Firstly, I will use the data from the New York City government, <https://www1.nyc.gov/site/doh/covid/covid-19-data.page>, to obtain the number of COVID cases in each zip code. I will also use publicly available data to obtain the latitude and longitude for each zip code which is available in [https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/.](https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/)

Finally, I will use the Foursquare API to obtain the data for the venues in each zip code neighbourhood using the latitude and longitude. In particular, I will collect the number of venues in each category like restaurants, parks etc. This will be used to identify what types of venues increase the risk of a neighbourhood having a large number of cases.