**Analyzing COVID-19 Data For New York City**

**Nilesh Koratpallikar**

**April 2020**

1. **Introduction**

**1.1 Background**

COVID-19 has been the most devastating pandemic that has affected the entire world, since 1918. Several countries are struggling to effectively fight against the pandemic, which has already led to a tremendous loss of lives, across the globe. In US, several states have been already affected. In particular, New York City has been the most severely impacted. It is important to understand how this pandemic has affected people within various boroughs of New York City across a spectrum of gender, age, demographics and derive meaning insights from the data.

* 1. **Problem**

The goal of this project is to analyze the COVID-19 data for New York City and identify potential correlations between the loss of lives and other variables including gender, age, demographics, availability of hospital/medical cities, across various Boroughs of New York City. Also, given the availability of data such as number lives lost for each calendar day since the beginning of this pandemic, would like to predict the curve for modeling future loss of lives.

* 1. **Interest**

Analyzing the COVID-19 data for New York City, would be very beneficial and the key insights could help several groups including the medical community to come up with remediation steps in order to minimize the impacts of this pandemic, potentially in other places.

1. **Data Acquisition**
   1. **Data needs & Data sources**

* Most of the data sets are available at [COVID-19 NY website](https://www1.nyc.gov/site/doh/covid/covid-19-data.page) which is updated on a daily basis.
* For Choropleth Data Visualization, the latitude and longitudes of various New York City Boroughs would be needed. This geojson dataset is available at [New York City Boroughs](https://geo.nyu.edu/catalog/nyu_2451_34572).
* Foursquare location API would be used to search/explore and to get the points of interest such as medical facilities, within each New York City Borough.
* Additional data sets such as New York City demographic data would also be sourced.

The data sets would be cleansed, as required for exploratory data analysis and described in detail, within subsequent sections of this report.