**Project - Covid 19 Data Analysis and Comparison**

**Context**

The novel coronavirus, also known as SARS-CoV-2, is a contagious respiratory virus that first reported in Wuhan, China. On 2/11/2020, the World Health Organization designated the name COVID-19 for the disease caused by the novel coronavirus. This new strain of virus has strike fear in many countries as cities are quarantined and hospitals are overcrowded.

This project aims at exploring COVID-19 through data analysis and visualization.

.

**Milestone 1** – Data Gathering

For this project , I have selected data from the following sources:

1. **CSV** – The Covid 19 data is scrapped from John Hopkins University github repo : <https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series> , this has
   1. Daily time series summary tables, including confirmed, deaths and recovered. All data is read in from the daily case report. The time series tables are subject to be updated if inaccuracies are identified in our historical data.
   2. Two time series tables are for the US confirmed cases and deaths, reported at the county level.
   3. Three time series tables are for the global confirmed cases, recovered cases and deaths. Australia, Canada and China are reported at the province/state level.
   4. Data is updated at a daily basis.
2. **API** – I have used the api : [https://services.arcgis.com/lQySeXwbBg53XWDi/arcgis/rest/services/Map/FeatureServer/0/query?where=1%3D1&outFields=\*&outSR=4326&f=json](https://services.arcgis.com/lQySeXwbBg53XWDi/arcgis/rest/services/Map/FeatureServer/0/query?where=1%3D1&outFields=*&outSR=4326&f=json)

which will provide the demographic info (Age, employment, sex, ethnicity etc ) for US at a County level for all the corona virus cases, this information is vital to understand the rate of spread across communities in United States. I will use this data to deep dive into corona cases in US and generate some interesting facts.

1. **Web** – I am scrapping data from https://www.worldometers.info / website for more insights at a world level like population, density, area and other factors.

The datasets 1 and 2 are related by county and country code, whereas the dataset 1 and 3 are related by Country code, this will help me join the 3 datasets into one and perform any slicing dicing for visualization.