**Research Proposal**

# Background Context:

Coronavirus is a family of viruses that can cause illness, which can vary from common cold and cough to sometimes more severe disease. Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV) were such severe cases with the world already has faced.

SARS-CoV-2 (n-coronavirus) is the new virus of the coronavirus family, which first discovered in 2019, which has not been identified in humans before. It is a contiguous virus which started from Wuhan in December 2019. Which later declared as Pandemic by WHO due to high-rate spreads throughout the world.

Currently (on the date 13 Jan 2021), this leads to a total of over 1.9M Deaths across the globe. Pandemic is spreading all over the world; it becomes more important to understand about this spread.

# Data

The dataset used in research is published worldometers website, which can be accessed using the link below:

<https://www.worldometers.info/coronavirus/>

This dataset is an effort to summarise relevant data associated with confirmed, death, recovered cases, and number of tests performed across different geographic regions to date, to provide some insight to the pandemic situation over the world.

# Research Design

## Research Question:

There are significant number of cases reported across the globe since the start of the Pandemic, with some countries reported significantly higher number of cases than others.

But how accurate are those data presented? For countries with lower number of confirmed cases, can we rely on those data to represent the true situation? Or are those low numbers simply due to insufficient number of testing performed? How serious is the true situation if the number of testing is increased?

## Hypothesis Testing:

The number of testing performed per million population by different countries are divided into three groups: High Level Testing, Medium Level Testing, and Low Level Testing.

Null Hypothesis: The level of testing is NOT associated with the number of confirmed cases per million population. Therefore, the mean difference in the total number of confirmed cases per million population amongst those three groups is zero.

ANOVA analysis will be performed amongst those three different level testing groups against the total number of confirmed cases per million population, to determine if the hypothesis is true or not.

# Audience

This research will help Disease Control and Prevention Agencies such as WHO, CDC, as well as Health Professionals/Advisors to gain better understanding of the true situation of the pandemic and identify any improvement can be made towards the testing strategy across the globe.