**Task 6.1 Sourcing Open Data**

**Summary Data**

**First Data:** <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>

This is an open data, which is extracted from Johns Hopkins GitHub repository. The data is donated to Center for systems science and Engineering (CSSE). Also Supported by ESRI living Atlas Team and Johns Hopkins University Applied Physic Lab (JHU APL). At the JHU GitHub repository It was asked to cite Lancet Article for any use of this data in publication: [An interactive web-based dashboard to track COVID-19 in real time](https://doi.org/10.1016/S1473-3099(20)30120-1). As JHU data, you can verify this as a trustworthy data source.

Data**: Covid\_19 Data**, **Covid\_19 open line list**, Time\_Series\_Covid\_19 Confirmed\_US, Time\_Series\_Covid\_19 Confirmed, Time\_Series\_Deaths\_US, Time\_Series\_Deaths, Time\_Series\_Recovered. Some insights could be

* Changes in number of affected cases over time
* Change in cases over time **at country level**
* Latest number of affected cases

**Second Data**: <https://www.kaggle.com/sshikamaru/covid19-deaths-in-the-us>

This is an open data, which is provided by Centers for Disease Control and Prevention (CDC) through their National Center for Health Statistics. It was licensed by Public Domain U.S. Government. As government data, you can verify this as a trustworthy data source.

**Data: Covid\_19\_Death\_Counts\_by\_Sex\_Age\_State**

**Third Data:** <https://www.kaggle.com/tavishaggarwal/covid-19-dataset-on-country-level>

This data was extracted from the World Meter. Data: **Covid\_19 USA\_City**, **Covid\_19 Country**

This dataset opens the possibility of answering various questions like:

* **Does lockdown help in controlling the growth of COVID-19 cases**? If yes, which lockdown strategy works best?
* Future forecast of the cases for the country so that country can take precaution measures. And lot more

**Why chosen this data set?**

I was interest on the Topic of Covid\_19 and wanted to know more about it. I searched on Gaggle website and found that this data was updated and had all right requirements for my project.

**List of Questions as brainstorming**

1. **Covid\_19 Data**

* How is death rate of the countries change over time?
* How does recovery of patients change over time?

1. **Covid\_19 open line list**

* What are the top 10 countries affecting the most with the coronavirus?
* Which countries has the highest tests?
* Which countries has the highest death rate?

1. **Covid\_19 USA\_City**

* Which USA \_state has the highest death rate?
* Which USA state has the highest active case?
* Which USA\_state has the highest recover rate?
* Which USA\_state has the most cases?
* How does Death Rate changes over time in USA\_states?
* Why the death rate varies over time?

1. **Covid\_19 Country**

* Which country has the most case?
* Which country has the highest death rate and lowest death rate?
* Which country has the highest recovery rate?
* Which country has the highest active cases?
* Which country has the highest number of critical cases?
* Which continent has the highest death rate/ active case?
* Which country has the highest tests?
* How does active cases change over time?
* How does death rate changes over time?
* How does recovery rate changes over time?
* How does test rate changes over time?
* Why death rate varies by country?

1. **Covid\_19\_Death\_Counts\_by\_Sex\_Age\_State**

* Which age group has the death rate/lowest (vulnerable populations)?
* Which state has more vulnerable population?
* Which state has the highest death rate/lowest?
* Do males has the highest death or female?
* Which state has more female death?
* Which state has more male deaths?
* Which week has the highest death and lowest death rate?
* How death rate changes over the weeks?
* Does the gender impact the death rate?

1. Time\_Series\_Covid\_19 Confirmed\_US

* How does confirmed rate changed over the year in US?

1. Time\_Series\_Covid\_19 Confirmed

* How does confirmed rate changed over the year?

1. Time\_Series\_Deaths\_US

* How does death rate changes over year in US?

1. Time\_Series\_Deaths

* How does death rate changes over year?

1. Time\_Series\_Recovered

* How does recovering rate changes over year?

Key Analysis

* Explore the change in Confirmed and Deaths Cases of Covid-19 by States
* Explore the time series of Confirmed and Deaths Cases of Covid-19 by Months
* Does Gender play rule in Death Cases of Covid-19
* Which age groups are the vulnerable populations in Covid-19 Cases
* What are the top 10 states with high number of vulnerable populations
* Understand the relationship of Confirmed Cases and Deaths Cases by states

From [World Health Organization](https://www.who.int/emergencies/diseases/novel-coronavirus-2019) - On 31 December 2019, WHO was alerted to several cases of pneumonia in Wuhan City, Hubei Province of China. The virus did not match any other known virus. This raised concern because when a virus is new, we do not know how it affects people.

Coronavirus disease (COVID-19) is an infectious disease. These virus cases in the US is stacking up higher and higher. Understanding this virus is crucial to stopping it’s spread.

In this case study, we will be exploring ways to stop the spread of this virus.

**Hypothesis**:

If your vaccine vulnerable population the total deaths decrease by states

Where the photos were taken:

<https://womendeliver.org/covid-19-response/>

**Financial Aspect:**

* Cost of COVID lab tests (tests to determine if patient tests positive for the virus)
* Cost of COVID lab test - in the states where COVID death rate are the highest (different states might have different cost for the tests)
* Data: <https://www.beckershospitalreview.com/finance/the-average-cost-of-a-hospital-covid-19-test-in-each-state.html>
* Cost of vaccination: <https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html>