# Covid Data Analysis Project – Stage I

## Task – 2 (Individual)

**Description of Enrichment Dataset:**

**Presidential Election Results** – This dataset consists of the 2020 Presidential election results across different categories like Governors, House, President and Senate for each county. Each category has sub-datasets like the details of the total votes in the State, County, the number of votes secured by each candidate and the election result.

Below are the data dictionary tables:

**File Name: govenors\_county.csv**



**File Name: govenors\_county\_candidate.csv**



**File Name:** **governors\_state.csv**



**File Name:** **house\_candidate.csv**



**File Name:** **house\_state.csv**



**File Name:** **president\_county.csv**



**File Name:** **president\_county\_candidate.csv**



**File Name:** **president\_state.csv**



**File Name:** **senate\_county.csv**



**File Name:** **senate\_county\_candidate.csv**



**File Name:** **senate\_state.csv** 

**This enrichment data helps us identify how badly covid affected the 2020 election results.** However, the data does not show which candidate/party was previously elected for the respective position. Therefore, we will collect the previously elected candidate details of the particular state or find evidence of the political inclination for each county or state and combine it with the current dataset to draw proper conclusions.

**We can join this data to the covid 19 dataset based on either county or state** and see how the people replied to the politician/political party of the region. **Our initial hypothesis questions would be**

* **Did the regions that covid19 badly hit vote to the opposition?**
  + Generally, people impacted heavily by covid will blame the politician of the region and vote for the opposition.
* **Did the people look at covid19 impact on their county/state and vote, or did they look at the whole nation’s plight?**
  + In some situations, people might look at their county, and if the current state representatives worked hard to provide the best care, they would vote for the same person/party again. And in some cases, the entire blame goes to the President, and when they look at the dire situation of the entire nation, they might vote for the opposition

So to do the analysis, we will have to cluster the counties/states based on the covid19 intensity in those regions.We will need to create/identify the parameters that define the level of covid19 effect in a given region. In order to do this, if this data is insufficient, we will find relative evidence/dataset that can support our analysis.