# COVID-19 Data Analysis Project Stage – I

# Presidential Election Results Report (Enrichment Dataset)

1.Describing the enrichment data and datatype - variable dictionary.

president\_county\_candidate.csv

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Definition | Data type | Possible values | Required? |
| State | It gives the name of the state | Text | Delaware, Georgia, Florida, Idaho,  Halwai. | Yes. |
| County | It defines the state county | Text | Ward 2, Kent County, Sussex County,  New Castle County | Yes. |
| Candidate | Name of the Candidate. | Text | Joe Biden, Donald Trump, Jorgensen, Howie Hawkins. | Yes. |
| Party | Name of the party | Text | DEM,REP,LIB,  GRN,WRI | Yes |
| Total vote | Number of votes the person received. | Integer | 44552, 41009,  195034, 56682 | Yes |
| Won | Did he won the election or not? | Boolean | True,False. | Yes |

Part 2: To integrate the enrichment data with the primary COVID-19 dataset from usafacts.org, we need to identify the common variables linking the datasets. In this case, we can use the 'State' and 'County' variables. However, there is a discrepancy in the format of state names between the enrichment dataset (full names) and the COVID-19 dataset (short form notations).

To address this disparity, we can add or replace a column in the enrichment dataset containing the short form notations of the state names. This additional column will align with the state column in the COVID-19 dataset. Subsequently, we can execute a merge operation on the datasets using the newly obtained short form notations of the state names and the 'County' column.

Part 3: Utilizing the enrichment data from the Presidential Election Result dataset can significantly enhance the analysis of COVID-19 spread in various ways:

1. **Political Leaning vs. COVID-19 Response:** We can investigate potential correlations between the political leaning at the county level and variations in COVID-19 case counts. This exploration may lead to hypotheses regarding how political factors influenced public health responses.
2. **Voting Patterns and COVID-19 Impact:** By examining voting patterns in conjunction with COVID-19 data, we can assess whether areas that supported a specific candidate or party experienced distinct levels of COVID-19 impact. This divergence might be attributed to differing policies or behaviors.

**Hypothesis Questions:**

* Is there a discernible correlation between a county's political leaning and the reported number of COVID-19 cases?
* Did counties that favored a particular candidate in the 2020 Presidential election exhibit a distinct trajectory of COVID-19 cases and deaths throughout the pandemic?

The inclusion of enrichment data introduces a political dimension to our analysis, enabling an exploration of potential relationships between political factors and the dynamics of COVID-19 spread.