ENRICHMENT DATASET -PRESIDENTIAL ELECTION

A. The Presidential Election Results Dataset

The presidential election results dataframe provides valuable information that can be used in the analysis of COVID-19 data. Presidential dataset is the table containing the outcomes of the 2020 presidential poll in the USA. This dataset provides detailed information on the 2020 USA presidential election results. The data has six fields namely State, County, Candidate, Political Party, Total votes per candidate, and the winner per county.

The data dictionary outlines the key columns in the enrichment data and their definitions. Here's how each column can contribute to the analysis:

1. **State:**
    - **Definition:** Name of the state.
    - **Use in COVID-19 Analysis:** Allows for the grouping and analysis of COVID-19 data at the state level. The state-level analysis can provide insights into regional patterns, variations in the spread of the virus, and the effectiveness of state-level policies.
2. **County:**
    - **Definition:** Name of the county.
    - **Use in COVID-19 Analysis:** Enables the analysis of COVID-19 data at the county level. This can be valuable for understanding localized trends, assessing the impact of the pandemic on specific communities, and identifying areas that may require targeted interventions.
3. **Candidate:**
    - **Definition:** Name of the contestant (presidential candidate).
    - **Use in COVID-19 Analysis:** While the candidate names themselves may not have a direct impact on COVID-19, the data can be leveraged to explore potential correlations or differences in COVID-19 outcomes based on political affiliations or leadership.
4. **Party:**
    - **Definition:** Name of the political party associated with the candidate.

- **Use in COVID-19 Analysis:** Allows for the categorization of candidates based on their political party. This information can be used to study whether COVID-19 outcomes differ based on the political leadership in each area.

5. **Total-Vote:**
    - **Definition:** Vote count for the candidate.
    - **Use in COVID-19 Analysis:** While the total vote count is directly related to the election, it may not have a direct impact on COVID-19. However, it can be used in combination with other variables to explore relationships or patterns.
6. **Won:**
    - **Definition:** Winner in the polls (Boolean - True or False).
    - **Use in COVID-19 Analysis:** Understanding which candidate won in a particular area can be valuable when assessing the political landscape and potential correlations with COVID-19 outcomes. It provides context for the political environment in each region.

The data dictionary is presented in Table 1. The data dictionary outlines the key columns in the enrichment data and their definitions.

Table I : Description of Presidential Dataset (Data Dictionary)

Name	Definition	Data Type	Possible Values	Required?
State	Name of State	Text	Florida, Geaogia Hawaii, Idaho, Illinois	Yes
County	Name of county	Text	Kent County, Ward 2, Duval County, Hardee County, Highlands County	Yes
Candidate	Name of Contestant	Text	Joe Biden, Donald Trump, Jo Jorgensen, Howie Hawkins	Yes
Party	Name of Political Party	Text	DEM, REP, LIB, GRN	Yes
Total-Vote	Vote Count	Integer	420, 1282, 1003, 259, 56682	Yes
Won	Winner in the Polls	Boolean	True, False	Yes

B. Merging with the Full Covid Dataset and the Presidential Election Dataframe

To merge the two dataframes, common variables which exists in both dataframes are essential. In this case, the common variables are "State" and "County" since they are present in both the presidential results table and the COVID-19 cases table. So, the variable that can be used to join the two tables would be a combination of "State" and "county". When performing the merging operation, the rows in the presidential results dataframe will be matched with the corresponding rows in the COVID-19 cases dataframe based on the values in these two columns. In other words, ['State', 'County'] is used for merging by:

i. Understand the Columns:
ii. Ensure Consistency:
iii. Clean Data (if needed)

iv. Merge Datasets:

v. Check the Result:

C. Hypothesis

Several hypothesis statements could be formulated. A few of them are:

- Is there a positive correlation between the **political affiliation of U.S. states** during the 2020 presidential election and the incidence of **COVID-19 cases and deaths reported** in the states from January 2020 to December 2021?
- Is there a positive correlation between the **political affiliation of U.S. states** during the 2020 presidential election and the incidence of **COVID-19 deaths reported** in the states from January 2020 to December 2021?
- Is there a correlation between the victory of a specific presidential candidate in the 2020 election and the spread of the covid-19 cases/deaths in the corresponding states during the period from January 2020 to December 2021??
- Is there a correlation between the total votes cast in U.S. states during the 2020 presidential election and the occurrence of COVID-19 cases and deaths reported in those states from January 2020 to December 2021?
- Does the number of voters in the 2020 election lead to an increase in the covid cases and deaths?