**Presidential Election Results**

**Part 1:**

**Report describing the enrichment data and datatype – variable dictionary.**

**Table of contents**

* Introduction
* Overview
* Dataset Description
* Datatype – variable dictionary
* References

**Introduction**

For a more comprehensive analysis of the COVID-19 dataset from usafacts.org, we have enriched our primary dataset with additional information from the “Presidential Election Results (Political leaning) dataset”. This additional data provides insights into the political leanings of different regions in the United States during 2020 Presidential elections.

**Overview**

Source: Kaggle

URL: <https://www.kaggle.com/unanimad/us-election-2020>

Dataset overview:

Provides information like the state, county, candidate, party, total votes and the election result or candidate status (won or not).

**Dataset Description**

This dataset includes information related to the 2020 United States Presidential Election, specifically focusing on the vote counts and percentages for each major candidate at the county level. Additionally, it provides details about the political leaning of each county based on the candidate who received majority of votes in the county.

**Datatype – variable dictionary**

The data dictionary provides a clear understanding of the variables present in the Presidential Election Results – 2020 dataset, their definition, datatypes, possible values, and whether they include null data, or the field is required for analysis or not. It serves as a valuable reference for data analysis and interpretation.

**Geographic Information:** State, County

**Election Information:** Candidate, party, total votes, status (won or not)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Definition** | **Data type** | **Possible values** | **Any missing values?** | **Required?** |
| State | The name of the US state. | String object | Names of US states. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| County | The name of the county within US state. | String object | Names of US Sate counties. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| Candidate | A candidate running in the election for a county representing their specific party. | String | Person from list of people standing in elections. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| Party | A political organization that fields candidates. | String | Name of a party from list of political parties. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| total\_votes | Number of votes secured by won candidate. | Integer | Positive integers representing total votes count. | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
| won | It is a Boolean value representing the status of the participated candidate (won or not) | Boolean | Boolean values (True- the candidate won, False- if lost) | No, the data does not contain any null or missing values. | Yes, this is a mandatory field. |
|  |  |  |  |  |  |

**Part 2:**

**Identifying the individual variable which map between the datasets and merging the enrichment data with the primary COVID-19 dataset.**

To merge the enrichment data with the primary covid-19 dataset from usafacts.org, we can make use of the common variables in both the datasets ‘State’ and ‘County’ to create a unified dataset.

The query would be similar to:

Merge\_data= covid\_data.merge(election\_data, on=[’State’, ’County’], how=’left’)

**Part 3:**

**Describing how enrichment data can help in the analysis of COVID-19 spread.**

The enrichment data from the Presidential Election Result dataset can significantly contribute to the analysis of COVID-19 spread in several ways:

**Political Leaning vs. COVID-19 Response**: We can examine whether political leaning at the county level correlates with differences in COVID-19 case counts. This may lead to hypotheses about how political factors influenced public health responses.

**Voting Patterns and COVID-19 Impact:** By analyzing voting patterns alongside COVID-19 data, we can find whether areas that voted for a particular candidate or party experienced different levels of COVID-19 impact, potentially due to varying policies or behaviors.

**Initial Hypothesis Questions:**

Is there any correlation between the political leaning of a county and the number of covid cases reported?

Did counties that voted for a specific candidate in 2020 Presidential party experience a different trajectory of COVID-19 cases and deaths during the pandemic?

The enrichment data adds a political dimension to our analysis, allowing to explore potential relationships between political factors and the spread of COVID-19.