

Name: Deepika Dasaroju
Student ID: 888078129
Subject: CSC-605 - Data Science

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`

| Column Name | Definition | Data type | Possible values | Required? |
|---------------|---|-----------|-------------------------------------|-----------|
| state | Name of the state | String | Delaware,Indiana etc... | Yes |
| county | Name of the county | String | Kent County,New Castle County etc.. | Yes |
| current_votes | Number of votes polled for governor's election the county | Integer | 1000,2000 etc.... | Yes |
| total_votes | Total number of votes in the county | Integer | 1000,2000 etc.... | Yes |
| percent | Percentage of votes polled | Float | 80.0,90.0,95.0 etc.... | Yes |

File Name: `govenors_county_candidate.csv`

| Column Name | Definition | Data type | Possible values | Required? |
|-------------|--|-----------|-------------------------------------|-----------|
| state | Name of the state | String | Delaware,Indiana etc... | Yes |
| county | Name of the county | String | Kent County,New Castle County etc.. | Yes |
| candidate | Name of the candidate contesting for governor's role | String | John Carney,Kathy DeMatteis,etc.. | Yes |
| party | Name of the Candidate's Party | String | DEM,REP,IPD etc... | Yes |
| votes | Votes secured by the candidate | Integer | 500,600,632 etc.. | Yes |
| won | Did the candidate win in the county? | Boolean | TRUE,FALSE | Yes |

File Name: `governors_state.csv`

| Column Name | Definition | Data type | Possible values | Required? |
|-------------|---|-----------|-------------------------|-----------|
| state | Name of the state | String | Delaware,Indiana etc... | Yes |
| votes | Number of votes polled for governor election in the state | Integer | 10000,20000 etc.... | Yes |

File Name: `house_candidate.csv`

| Column Name | Definition | Data type | Possible values | Required? |
|-------------|--|-----------|-----------------------------------|-----------|
| district | Name of the district | String | Delaware at large,Florida's 18th | Yes |
| candidate | Name of the candidate contesting for house | String | John Carney,Kathy DeMatteis,etc.. | Yes |
| party | Name of the Candidate's Party | String | DEM,REP,IPD etc... | Yes |
| total_votes | Number of votes secured by the candidate | Integer | 10000,20000 etc.... | Yes |
| won | Did the candidate win? | Boolean | TRUE,FALSE | Yes |

Name: Deepika Dasaroju
Student ID: 888078129
Subject: CSC-605 - Data Science

File Name: house_state.csv

| Column Name | Definition | Data type | Possible values | Required? |
|---------------|--|-----------|--|-----------|
| district | Name of the district | String | Delaware at large,Florida's 18th district etc... | Yes |
| current_votes | Number of votes polled for house election in the state | Integer | 10000,20000 etc.... | Yes |
| total_votes | Total number of votes in the state | Integer | 10000,20000 etc.... | Yes |
| percent | percentage of votes polled | Float | 80.0,90.0,95.0 etc.... | Yes |

File Name: president_county.csv

| Column Name | Definition | Data type | Possible values | Required? |
|---------------|---|-----------|-------------------------------------|-----------|
| state | Name of the state | String | Delaware,Indiana etc... | Yes |
| county | Name of the county | String | Kent County,New Castle County etc.. | Yes |
| current_votes | Number of votes polled for president election in the county | Integer | 10000,20000 etc.... | Yes |
| total_votes | total number of votes in the county | Integer | 10000,20000 etc.... | Yes |
| percent | percentage of votes polled | Float | 80.0,90.0,95.0 etc.... | Yes |

File Name: president_county_candidate.csv

| Column Name | Definition | Data type | Possible values | Required? |
|-------------|---|-----------|-------------------------------------|-----------|
| state | Name of the state | String | Delaware,Indiana etc... | Yes |
| county | Name of the county | String | Kent County,New Castle County etc.. | Yes |
| candidate | Name of the candidate contesting for the president's position | String | John Carney,Kathy DeMatteis,etc.. | Yes |
| party | Name of the party | String | DEM,REP,IPD etc... | Yes |
| total_votes | total number of votes secured by the candidate in the county | Integer | 10000,20000 etc.... | Yes |
| won | Did the candidate win in the county? | Boolean | TRUE,FALSE | Yes |

File Name: president_state.csv

| Column Name | Definition | Data type | Possible values | Required? |
|-------------|--|-----------|-------------------------|-----------|
| state | Name of the state | String | Delaware,Indiana etc... | Yes |
| total_votes | total number of votes polled for the president election from the state | Integer | 10000,20000 etc.... | Yes |

File Name: senate_county.csv

| Column Name | Definition | Data type | Possible values | Required? |
|---------------|--|-----------|-------------------------------------|-----------|
| state | Name of the state | String | Delaware,Indiana etc... | Yes |
| county | Name of the county | String | Kent County,New Castle County etc.. | Yes |
| current_votes | Number of votes polled for the senate election in the county | Integer | 10000,20000 etc.... | Yes |
| total_votes | total number of votes in the county | Integer | 10000,20000 etc.... | Yes |
| percent | percentage of votes polled | Float | 80.0,90.0,95.0 etc.... | Yes |

Name: Deepika Dasaroju
Student ID: 888078129
Subject: CSC-605 - Data Science

File Name: senate_county_candidate.csv

| Column Name | Definition | Data type | Possible values | Required? |
|-------------|---|-----------|-------------------------------------|-----------|
| state | Name of the state | String | Delaware,Indiana etc... | Yes |
| county | Name of the county | String | Kent County,New Castle County etc.. | Yes |
| candidate | Name of the candidate contesting for the senate | String | John Carney,Kathy DeMatteis,etc.. | Yes |
| party | Name of the Candidate's Party | String | DEM,REP,IPD etc... | Yes |
| total_votes | total number of votes secured by the candidate | Integer | 10000,20000 etc.... | Yes |

File Name: senate_state.csv

| Column Name | Definition | Data type | Possible values | Required? |
|-------------|---|-----------|-------------------------|-----------|
| state | Name of the state | String | Delaware,Indiana etc... | Yes |
| total_votes | total number of votes polled for the senate election in the state | Integer | 10000,20000 etc.... | Yes |

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?**
 - o 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?**
 - o 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.

Name: Deepika Dasaroju
Student ID: 888078129
Subject: CSC-605 - Data Science