

Table of Contents

The following is a directory of the files contained within the CIC Student Challenge project repository.

1. **Data** – datasets (raw, cleaned, final), methods (cleaning and report downloading), and variable dictionaries.
 - a. cleaned_compiled_data_and_keys (Datasets used for analysis), data downloads that needed to be cleaned, variable dictionary, and cleaning methods.
 - i. CDF_KFF_altered_raw_download – contains 4 excel datasets that were cleaned before being added to the datasets used for analysis.
 - ii. Prepared_Data_for_Python_Analysis – contains the prepared excel datasets used for OLS/LASSO regression.
 - iii. Methods_Data_Prep_Cic_Challenge.pdf – Breakdown of steps taken to compile data used in the cic_challenge_data sets used for OLS/LASSO regression.
 - iv. Variable_Dictionary_Cic_Challenge_Data_Sets.pdf – An explanatory list of all the variables included in the final data set. Including disclaimers and notes from the download sources.
 - b. raw_downloads
 - i. CDC – Contains raw data sets and the original variable dictionary downloaded from the CDC Socrata data tool.
 1. CDC_Socrata_Variable_Dictionary.xlsx – used with the CDC Socrata SQL no code reporting tool.
 2. Jan_Sept_2021_Covid-19_Vaccinations_in_the_United_States, County_20250831.csv – First half of a large download of Covid vaccination data.
 3. Sept_2021_2023_COVID-19_Vaccinations_in_the_United_States,Count (1).csv – Second half of a large download of Covid vaccination data.
 - ii. KFF – Contains raw data sets downloaded from the KFF Covid19 Monitor Custom State Reporting tool.
 1. Adult_poverty_rate_by_family_structure.xlsx – poverty rate measured by adults with/without children and overall totals per state and year (2021, 2022, 2023).
 2. Disparities.xlsx – Poverty Rate by Race/Ethnicity per state and year. (2021, 2022, 2023)
 3. Economic_Measures.xlsx – Total Gross State Product (GSP), Seasonally adjusted unemployment rate by month, state, SNAP (Supplementary Nutritional Assistance Program) Total Benefits distributed, and SNAP monthly average participants, by state and year.

Table of Contents

4. Mental_health_status_by_days.xlsx – number of poor mental health days per month, by state and year.
 5. Population_Distribution.xlsx – Total US Population, Total US population by Race/Ethnicity, state, and year (2021, 2022, 2023).
 - iii. Centroids.xlsx – contains longitude and latitude columns downloaded from Google Dataset Publishing Language
 - iv. Download_Steps_Heirarchical_Files.pdf – this is a step-by-step guide on how I utilized the reporting tools to obtain the raw_downloads data sets including links to reporting tools.
2. **Outputs** - contains .png used for visual, tables, figures, the project white paper.
 - a. .png – contains images of statistical tables and visuals used in the CIC_Tables_and_Figures PDF.
 - b. CIC_Tables_and_Figures.pdf – Correlation heatmap, regression outputs, coefficient tables and charts, plot and geographic visuals.
 - c. CIC_White_Paper.pdf – This is the paper that was submitted for the CIC_Student_Research Challenge.
 - d. VIF_OLS_LASSO_tables.xlsx – contains the table output from the analysis scripts.
3. **Scripts** – contains analysis code and html links to the JSON geographic bubble maps.
 - a. cic_challenge_regression.ipynb
 - b. bubble_map_with_legends.html (2021, 2022, 2023)