**VDSS\_ZCTA\_ACS\_Characteristics\_Aggregated.R - README**

**Author: Dylan Craig**

**Date Last Updated: 12/28/2024**

**File Name: VDSS\_ZCTA\_ACS\_Characteristics\_Aggregated.R**

**Purpose**

This script processes and aggregates ZIP Code Tabulation Areas (ZCTA) data from multiple demographic and socioeconomic variables, transforming raw data into usable metrics and generating calculated rates and weighted averages for analysis.

**Overview**

The script, **VDSS\_ZCTA\_ACS\_Characteristics\_Aggregated.R**, performs advanced data processing on ZCTA-level data, extracting key indicators such as population, poverty rates, employment status, race/ethnicity, and education levels. It calculates rates for each category and aggregates weighted averages for deeper insights.

**Datasets and Tools**

**Input Files:**

* **VDSS\_ZCTA\_Characteristics\_Aggregated.csv:** Contains raw ZCTA-level data with demographic and socioeconomic variables for processing.

**Output Files:**

1. **VDSS\_ZCTA\_Characteristics\_Cleaned.csv:** Processed dataset including calculated rates and validation checks.
2. **VDSS\_ZCTA\_Characteristics\_Cleaned.xlsx:** Excel version of the cleaned dataset.

**Dependencies:**

* **R Libraries:**
  + dplyr: Data manipulation.
  + readr: Reading CSV files.
  + stringr: String operations.
  + tidyr: Data tidying and reshaping.
  + writexl: Writing Excel files.
  + tidyverse: Comprehensive data manipulation and analysis.

**Script Functionality**

**1. Data Loading and Preparation:**

* Loads the input dataset from CSV format.
* Configures options to disable scientific notation for better readability.

**2. Variable-Specific Processing:**

* **Population Data:** Extracts and renames total population columns.
* **Employment Data:** Processes employment status variables and calculates derived metrics for employed, unemployed, and non-labor force populations.
* **Industry Data:** Cleans and combines industry-related variables, calculating the proportion of the workforce in each industry.
* **Poverty Data:**
  + Calculates child poverty counts and rates.
  + Computes overall poverty rates based on population estimates.
* **Median Income:** Renames and processes median income data.
* **Race and Ethnicity:** Extracts Hispanic and non-Hispanic racial categories, aggregating smaller groups into a single column for clarity.
* **Sex and Age Groups:**
  + Processes male and female populations.
  + Combines age ranges into consistent categories (e.g., 0-4 years, 5-9 years).
* **Education Levels:** Aggregates education attainment variables into four levels (No High School, High School Graduate, Some College, College Graduate).
* **Marital Status:** Processes marital categories (Never Married, Married Spouse Present, Widowed, etc.).

**3. Rate Calculations:**

* Calculates rates for each demographic and socioeconomic category, such as:
  + Employment and unemployment rates.
  + Proportions of populations by industry.
  + Poverty and child poverty rates.
  + Proportions for race/ethnicity, age, and education levels.

**4. Validation Checks:**

* Ensures population-based categories sum correctly to total population estimates.
* Flags discrepancies for further review.

**5. Weighted Averages:**

* Computes weighted averages for key variables across ZCTAs, including poverty rates, median income, race/ethnicity proportions, and more.

**6. Data Export:**

* Saves the processed dataset in CSV and Excel formats.

**File Output**

1. **VDSS\_ZCTA\_Characteristics\_Cleaned.csv:** Includes calculated rates, validation checks, and weighted averages.
2. **VDSS\_ZCTA\_Characteristics\_Cleaned.xlsx:** Excel file with the same content as the CSV output for easier data review.

**Usage**

1. Place the input CSV file in the specified directory.
2. Install the required R libraries.
3. Execute the script in an R project environment.
4. Retrieve processed outputs from the output folder.

**Notes**

* Ensure the input dataset is complete and correctly formatted.
* Review validation flags to address any potential data issues.
* Adjust column names in the script if the input file structure changes.