**VDSS\_Final\_Data.R - ReadMe**

**Overview**

This script, **VDSS\_Final\_Data.R**, is designed to merge various datasets to create a comprehensive Excel sheet that assigns multiple attributes to each ZIP code. These attributes include Border Status, Type (PO Box, Unique, Standard), ZCTA match, Treatment Status, and Bad Zip County Rate. The final output provides a unified dataset for further analysis.

**Datasets and Tools**

* **VDSS\_Zip\_Border\_Status\_Final\_Designation.xlsx**: Contains the border status classification for each ZIP code.
* **Zip\_Residential\_Status.xlsx**: Provides the residential or non-residential status for each ZIP code.
* **VDSS\_Zip\_Type.xlsx**: Contains the type classification (PO Box, Unique, Standard) for each ZIP code.
* **Zip\_ZCTA\_Crosswalk.xlsx**: Matches each ZIP code to its corresponding ZCTA using the Germuska ZIP to ZCTA crosswalk.
* **zip\_treat.dta**: Contains treatment status data for each ZIP code.
* **bad\_zipcounty\_rate.dta**: Provides data on ZIP codes, including the rate of incorrect services by the wrong VDSS office.

**Script Functionality**

1. **Loading Necessary Libraries**: The script loads essential R libraries, including readxl for reading Excel files, dplyr for data manipulation, haven for reading .dta files, and writexl for writing the final output to an Excel file.
2. **Loading and Preprocessing Data**:
   * The script reads in the datasets listed above and ensures consistency by converting all ZIP columns to character type.
3. **Merging Datasets**:
   * The datasets are merged sequentially by ZIP code to create a comprehensive dataset that includes all relevant attributes.
4. **Deduplication**:
   * The merged data is deduplicated by ZIP and FIPS to ensure that each ZIP code has a unique entry.
5. **Handling Missing Data**:
   * The script identifies any missing data for each attribute by comparing the merged data against the original datasets.
6. **Saving the Output**: The final deduplicated and merged dataset is saved as an Excel file for further use.

**Dependencies**

* dplyr: For data manipulation.
* readxl: For reading Excel files.
* haven: For reading .dta files.
* writexl: For writing data to Excel files.

**File Output**

* **VDSS\_Final\_Data.xlsx**: This Excel file contains the final merged and deduplicated dataset, with all relevant attributes assigned to each ZIP code.

**Usage**

1. Ensure all dependencies are installed and up to date.
2. Run the script to merge the datasets and create the final comprehensive dataset for ZIP codes.
3. The output file will be saved automatically in the **Data Outputs/VDSS\_Final\_Data/** directory.

**Notes**

* The accuracy of the final dataset depends on the correctness and completeness of the individual datasets used in the merging process.
* The script prints the number of missing rows for each dataset to help identify any potential issues with data completeness.