**VA\_HUD\_USPS\_ZIP\_DATA\_ALL\_YEARS.R - ReadMe**

**Overview**

This script, VA\_HUD\_USPS\_ZIP\_DATA\_ALL\_YEARS.R, is located in the R project directory structure at Scripts/VDSS\_Historical\_Zip\_Code\_Changes/. The script processes HUD\_USPS ZIP code data across multiple years and quarters, specifically from Q1 2010 to Q2 2024. It filters the data to include only Virginia ZIP codes as identified by the USPS ZIP code data, and creates a combined dataset for further analysis of HUD\_USPS ZIP code data.

**Datasets**

* **HUD\_USPS\_ZIP\_CROSSWALK Files:** These files contain ZIP code and tract information for each quarter from 2010 to 2024. Each file is named in the format ZIP\_TRACT\_<MMYYYY>.xlsx, where <MMYYYY> represents the quarter and year.
* **USPS\_Zip\_County.xlsx:** This file contains ZIP code data filtered specifically for Virginia (VA), including the corresponding county information.

**Script Functionality**

1. **Loading Necessary Libraries:** The script loads required R libraries such as readxl for reading Excel files, dplyr for data manipulation, writexl for writing Excel files, pbapply for progress bars, and ggplot2 for potential plotting.
2. **Generating Year-Quarter Strings:** The function get\_year\_quarter() creates a string in the format "Q<Quarter> <Year>" for each combination of year and quarter.
3. **Loading and Merging Data:** The script iterates over each combination of year and quarter, checks if the corresponding file exists, and reads the data into a list. The data for each quarter is tagged with the corresponding year-quarter label.
4. **Converting Column Names:** All column names are converted to uppercase and any duplicate names are made unique to ensure consistency.
5. **Filtering for VA ZIP Codes:** The combined dataset is filtered to include only ZIP codes present in the USPS ZIP County file for Virginia.
6. **Converting Ratios to Numeric:** Columns related to residential, business, other, and total ratios are converted from text to numeric format to facilitate further analysis.
7. **Saving the Output:** The processed data is saved as an Excel file in two specified directories within the R project format:
   * Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_ZIP\_DATA\_ALL\_YEARS.xlsx
   * Data Outputs/VDSS\_Historical\_Zip\_Code\_Changes/VA\_HUD\_USPS\_ZIP\_DATA\_ALL\_YEARS.xlsx

**Dependencies**

* **readxl:** For reading Excel files.
* **dplyr:** For data manipulation.
* **writexl:** For writing data to Excel files.
* **pbapply:** For displaying progress bars.
* **ggplot2:** (Optional) Included for potential plotting needs, although not directly used in this script.

**File Output**

* **VA\_HUD\_USPS\_ZIP\_DATA\_ALL\_YEARS.xlsx:** This file contains the combined and processed HUD\_USPS ZIP code data for all years from 2010 to 2024. It is saved in both VDSS\_Historical\_Zip\_Type\_Changes and VDSS\_Historical\_Zip\_Code\_Changes directories.

**Usage**

1. Ensure all dependencies are installed and up to date.
2. Place all required data files in the appropriate directories as specified in the script.
3. Run the script to process the HUD\_USPS ZIP code data.
4. The output files will be saved automatically in the Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/ and Data Outputs/VDSS\_Historical\_Zip\_Code\_Changes/ directories.

**Notes**

* The script relies on the presence of specific files in the Raw Data/VDSS\_Historical\_Zip\_Type\_Changes/HUD\_USPS\_ZIP\_CROSSWALK/ directory. Missing files will result in incomplete data for the affected quarters.
* The accuracy of the output depends on the correctness and completeness of the input datasets.