**VA\_HUD\_USPS\_HISTORICAL\_ZIP\_CODE\_TRANSITIONS.R - ReadMe**

**Overview**

This script, VA\_HUD\_USPS\_HISTORICAL\_ZIP\_CODE\_TRANSITIONS.R, is designed to analyze the appearances and disappearances of ZIP codes across different quarters based on the HUD USPS data. The script processes the dataset to identify when ZIP codes first appear, disappear, or reappear, and summarizes these transitions.

**Datasets**

* **VA\_HUD\_USPS\_Zip\_Data\_Collapsed.xlsx:** This file contains the collapsed ZIP code data, including types and other relevant metrics, across multiple quarters.

**Script Functionality**

1. **Loading Necessary Libraries:** The script loads required R libraries such as readxl for reading Excel files, dplyr and tidyr for data manipulation, and writexl for writing Excel files.
2. **Loading and Preparing Data:**
   * The script loads the collapsed HUD USPS ZIP code data from the specified Excel file.
   * It extracts the year and quarter components from the YEAR\_QUARTER column and sorts the dataset accordingly.
3. **Tracking ZIP Code Presence:**
   * The script identifies all unique ZIP codes and all possible YEAR\_QUARTER values.
   * It creates a dataframe with each unique ZIP code and a column for each quarter, marking whether the ZIP code is "PRESENT" or "MISSING."
4. **Counting Appearances and Disappearances:**
   * The script counts the number of times each ZIP code appears (transitions from "MISSING" to "PRESENT") and disappears (transitions from "PRESENT" to "MISSING") across quarters.
   * It summarizes these counts into a dataframe that records the frequency of ZIP codes appearing or disappearing 0, 1, 2, 3, or 4 times.
5. **Saving the Results:**
   * The presence data, along with the appearance and disappearance counts for each ZIP code, are saved as an Excel file.
   * The summary of appearances and disappearances is also saved as a separate Excel file.
6. **Output Summary:**
   * The script prints a summary of the transition counts for quick reference.

**Dependencies**

* **readxl:** For reading Excel files.
* **dplyr:** For data manipulation.
* **tidyr:** For tidying data.
* **writexl:** For writing data to Excel files.

**File Output**

* **Data Outputs/VDSS\_Historical\_Zip\_Code\_Changes/VA\_HUD\_USPS\_Zip\_Appearance\_Disappearance\_Counts.xlsx:** This file contains the detailed counts of appearances and disappearances for each ZIP code.
* **Data Outputs/VDSS\_Historical\_Zip\_Code\_Changes/VA\_HUD\_USPS\_Zip\_Appearance\_Disappearance\_Summary.xlsx:** This file contains a summary of how many ZIP codes appeared or disappeared 0, 1, 2, 3, or 4 times.

**Usage**

1. Ensure all dependencies are installed and up to date.
2. Place the required dataset in the appropriate directory as specified in the script.
3. Run the script to analyze ZIP code appearances and disappearances across quarters.
4. The output files will be saved automatically in the Data Outputs/VDSS\_Historical\_Zip\_Code\_Changes/ directory.

**Notes**

* This script provides a detailed analysis of ZIP code transitions across quarters, which can be useful for understanding changes in ZIP code data over time.
* The accuracy of the results depends on the completeness of the input dataset.