**VDSS\_HUD\_USPS\_Historical\_Zip\_Type\_ZipDataMaps\_Comparison.R - README**

**Author: Dylan Craig**

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

**File Name: VDSS\_HUD\_USPS\_Historical\_Zip\_Type\_ZipDataMaps\_Comparison.R**

**Purpose**

This script processes VA HUD USPS ZIP data to categorize ZIP codes based on their characteristics, compares these classifications with an external ZIP code dataset, and summarizes discrepancies in detailed tables.

**Overview**

The script, **VDSS\_HUD\_USPS\_Historical\_Zip\_Type\_ZipDataMaps\_Comparison.R**, consolidates and analyzes ZIP code data across multiple years and compares the results with an external ZIP code dataset to identify classification matches and mismatches.

**Datasets and Tools**

**Input Files:**

1. **VA\_HUD\_USPS\_Zip\_Data\_All\_Years.xlsx**: Contains ZIP code data from Q1 2010 to Q2 2024.
2. **VDSS\_Zip\_Type.xlsx**: External dataset providing ZIP code type classifications for comparison.

**Output Files:**

1. **VA\_HUD\_USPS\_Zip\_Data\_Collapsed.xlsx**: Collapsed ZIP data with type classifications.
2. **Comparison\_Table\_PO\_Box.xlsx**: Comparison table summarizing results for PO Box ZIP codes.
3. **VA\_HUD\_USPS\_ZipDataMaps\_Comparison.xlsx**: Comparison table summarizing results for Unique ZIP codes.

**Dependencies:**

* **R Libraries:**
  + tidyverse: Data manipulation and analysis.
  + pbapply: Function applications with progress bars.
  + writexl: Writing Excel files.

**Script Functionality**

**1. Data Loading:**

* Reads combined VA HUD USPS ZIP data and external ZIP code dataset.

**2. Data Collapsing:**

* Collapses ZIP data by ZIP and YEAR\_QUARTER, summing relevant ratio columns (e.g., RES\_RATIO, BUS\_RATIO).

**3. ZIP Code Classification:**

* Classifies ZIP codes based on:
  + **Standard:** RES\_RATIO > 0
  + **Unique:** RES\_RATIO == 0, OTH\_RATIO == 0, BUS\_RATIO > 0
  + **PO Box:** RES\_RATIO == 0, OTH\_RATIO > 0, BUS\_RATIO == 0
  + **Unique PO Box:** RES\_RATIO == 0, OTH\_RATIO > 0, BUS\_RATIO > 0
* Identifies whether each ZIP code has multiple types or a single type.

**4. Separate DataFrames:**

* Creates distinct datasets for ZIP codes with single or multiple types.
* Adjusts ZIP types for unique PO Box classifications.

**5. Comparisons with External Dataset:**

* Compares HUD USPS data with external ZIP data for:
  + **PO Box ZIP Codes:** Summarizes matches, mismatches, and missing data.
  + **Unique ZIP Codes:** Conducts similar comparisons.

**6. Summarizing and Exporting Results:**

* Saves results as:
  + **Collapsed ZIP Data**
  + **Comparison Tables** for PO Box and Unique ZIP codes.

**File Output**

1. **VA\_HUD\_USPS\_Zip\_Data\_Collapsed.xlsx:** Consolidated ZIP data with classifications.
2. **Comparison\_Table\_PO\_Box.xlsx:** Comparison results for PO Box ZIP codes.
3. **VA\_HUD\_USPS\_ZipDataMaps\_Comparison.xlsx:** Comparison results for Unique ZIP codes.

**Usage**

1. Place the input datasets in the specified directories.
2. Install required R libraries.
3. Run the script in the designated R project environment.
4. The output files will be saved in the specified directory.

**Notes**

* Ensure the input datasets are accurate and formatted correctly.
* This script is specific to the VA HUD USPS and external ZIP data criteria.
* Different datasets or criteria may require modifications to the script.