**VDSS\_HUD\_USPS\_Helmuth\_Historical\_Analysis.R - README**

**Author: Dylan Craig**

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

**File Name: VDSS\_HUD\_USPS\_Helmuth\_Historical\_Analysis.R**

**Purpose**

This script analyzes historical ZIP type changes for Bailey-Helmuth ZIP codes based on HUD USPS data. It evaluates characteristics such as the type, presence, and timeline of ZIP codes in the dataset and classifies them as old or new for further analysis.

**Overview**

The script, **VDSS\_HUD\_USPS\_Helmuth\_Historical\_Analysis.R**, processes ZIP codes associated with the Bailey-Helmuth dataset to determine their classification using HUD USPS ZIP code data. It identifies key attributes such as the presence of data across quarters, type consistency, and changes over time.

**Datasets and Tools**

**Input Files:**

1. **VA\_HUD\_USPS\_Zip\_Data\_Collapsed.xlsx**: Contains ZIP code data across multiple quarters, including type and timeline.

**Output File:**

* **Bailey\_Helmuth\_Zips\_Analysis.xlsx**: Provides detailed analysis of old and new Bailey-Helmuth ZIP codes.

**Dependencies:**

* **R Libraries:**
  + dplyr: Data manipulation
  + readxl: Reading Excel files
  + writexl: Writing Excel files

**Script Functionality**

**1. Loading and Preprocessing Data:**

* Reads the collapsed HUD USPS ZIP code dataset.
* Extracts all unique YEAR\_QUARTER values for timeline analysis.

**2. Defining Bailey-Helmuth ZIP Codes:**

* Creates a dataframe listing old and new ZIP codes for analysis.

**3. Analyzing ZIP Codes:**

* Defines a function to evaluate ZIP codes for:
  + **NUMBER\_TYPES:** Whether the ZIP has multiple or single types in the dataset.
  + **TYPE:** The type(s) associated with the ZIP code.
  + **MISSING:** If data is consistently present or missing across all quarters.
  + **POST\_2010:** Whether data exists beyond Q1 2010.
* Applies the function to both old and new ZIP codes.

**4. Combining Results:**

* Integrates the analysis results for old and new ZIP codes into a consolidated dataframe.

**5. Saving Results:**

* Outputs the combined analysis as **Bailey\_Helmuth\_Zips\_Analysis.xlsx**.

**File Output**

**Bailey\_Helmuth\_Zips\_Analysis.xlsx**:

* Provides analysis of:
  + Type classification of ZIP codes.
  + Presence across all quarters.
  + Timeline of data availability relative to Q1 2010.

**Usage**

1. Place the input dataset in the specified directory.
2. Install required R libraries.
3. Run the script within the designated R project environment.
4. The output will be saved automatically to the specified folder.

**Notes**

* Ensure input files are correctly formatted and located in the appropriate directory.
* The analysis focuses specifically on Bailey-Helmuth ZIP codes.
* Changes to the list of analyzed ZIP codes require updates to the script.
* The accuracy of results depends on the completeness of the input dataset.