**VDSS\_HUD\_USPS\_Historical\_Zip\_Type\_Missing\_Analysis.R - README**

**Author: Dylan Craig**

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

**File Name: VDSS\_HUD\_USPS\_Historical\_Zip\_Type\_Missing\_Analysis.R**

**Purpose**

This script analyzes ZIP code presence across various year-quarters and evaluates their missing statuses. It calculates percentages of missing versus present ZIP codes and provides further analysis based on NUMBER\_TYPES and TYPE distributions.

**Overview**

The script, **VDSS\_HUD\_USPS\_Historical\_Zip\_Type\_Missing\_Analysis.R**, examines the presence and absence of ZIP codes across the dataset and performs breakdowns by NUMBER\_TYPES (SINGLE or MULTIPLE) and TYPE. It outputs results in multiple Excel files for detailed reporting.

**Datasets and Tools**

**Input File:**

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

**Output Files:**

* **VA\_HUD\_USPS\_Zip\_Status.xlsx**: Labels ZIP codes as MISSING or PRESENT.
* **VA\_HUD\_USPS\_Zip\_Percent\_Missing\_Present.xlsx**: Provides the percentage breakdown of ZIP codes that are missing or present.
* **VA\_HUD\_USPS\_Missing\_Analysis\_By\_Number\_Types.xlsx**: Details missing ZIP codes categorized by NUMBER\_TYPES (SINGLE or MULTIPLE).
* **VA\_HUD\_USPS\_Missing\_Type\_Analysis\_Single.xlsx**: Focuses on SINGLE type ZIP codes, showing the percentage of each type that is missing or present.

**Dependencies:**

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

**Script Functionality**

**1. Loading Data:**

* Reads the historical ZIP code dataset from the input file.

**2. Identifying Missing ZIP Codes:**

* Labels each ZIP code as MISSING or PRESENT across all available YEAR\_QUARTER values.

**3. Performing Analyses:**

* **Overall Presence Analysis:**
  + Calculates the percentage of ZIP codes that are MISSING or PRESENT.
* **Analysis by NUMBER\_TYPES:**
  + Analyzes the distribution of missing and present ZIP codes by NUMBER\_TYPES (SINGLE vs MULTIPLE).
* **Type-Specific Analysis:**
  + Focuses on SINGLE type ZIP codes to analyze the percentage of each TYPE that is missing or present.

**4. Exporting Results:**

* Saves the analysis outputs to Excel files in the specified directory.

**File Output**

1. **VA\_HUD\_USPS\_Zip\_Status.xlsx:** Labels ZIP codes as MISSING or PRESENT.
2. **VA\_HUD\_USPS\_Zip\_Percent\_Missing\_Present.xlsx:** Shows the percentage breakdown of ZIP code presence.
3. **VA\_HUD\_USPS\_Missing\_Analysis\_By\_Number\_Types.xlsx:** Breaks down missing ZIP codes by SINGLE or MULTIPLE types.
4. **VA\_HUD\_USPS\_Missing\_Type\_Analysis\_Single.xlsx:** Analyzes SINGLE type ZIP codes by TYPE and presence.

**Usage**

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

**Notes**

* Ensure the input dataset is complete and formatted correctly.
* This script is tailored to analyze ZIP code presence across quarters and may require adjustments for different datasets or criteria.
* Missing data in the input file may affect the accuracy of the results.