**VDSS\_ZIP\_TYPE.R - README**

**Author: Dylan Craig**

**Date Last Updated: 12/29/2024**

**File Name: VDSS\_ZIP\_TYPE.R**

**Purpose**

This script assigns a type (Standard, Unique, or PO Box) to each ZIP code in Virginia by merging USPS ZIP-County data with ZIPDATAMAPS data. It also identifies and saves any ZIP codes that failed to merge, allowing for troubleshooting and further analysis.

**Overview**

The script processes ZIP code data by:

1. Loading USPS ZIP-County and ZIPDATAMAPS datasets.
2. Merging the datasets to assign types to ZIP codes.
3. Identifying and saving ZIP codes that did not successfully merge.
4. Saving the merged dataset and unmatched ZIP codes as separate Excel files.

**ZIP Code Types Explained**

* **Standard:** Common ZIP codes for residential or business areas with routine mail delivery.
* **Unique:** ZIP codes assigned to specific organizations, buildings, or entities such as universities or hospitals.
* **PO Box:** ZIP codes designated for areas with mail delivery to post office boxes only.

**Datasets and Tools**

**Input Files:**

1. **USPS\_Zip\_County.xlsx:** Contains ZIP code data and corresponding county information.
2. **VDSS\_ZIPDATAMAPS\_TYPE.xlsx:** Provides ZIP code type classifications (Standard, Unique, PO Box).

**Output Files:**

1. **VDSS\_Zip\_Type.xlsx:** Merged dataset with assigned ZIP code types.
2. **VDSS\_Zip\_Type\_No\_Type.xlsx:** ZIP codes that did not match any type.

**Dependencies:**

* **R Libraries:**
  + readxl: For reading Excel files.
  + dplyr: For data manipulation.
  + openxlsx: For writing Excel files.

**Script Functionality**

**1. Loading Data:**

* Reads ZIP-County data and ZIPDATAMAPS data using readxl.

**2. Merging Datasets:**

* Performs a full join on the ZIP column to merge the datasets.
* Identifies ZIP codes that failed to merge due to missing matches in either dataset.

**3. Saving the Outputs:**

* Saves the merged dataset to VDSS\_Zip\_Type.xlsx.
* Saves the unmatched ZIP codes to VDSS\_Zip\_Type\_No\_Type.xlsx.

**File Output**

1. **VDSS\_Zip\_Type.xlsx:**
   * Contains all ZIP codes and their assigned types.
   * Includes data from both input datasets.
2. **VDSS\_Zip\_Type\_No\_Type.xlsx:**
   * Contains ZIP codes with no assigned type.
   * Useful for identifying issues in the input datasets.

**Usage**

1. Ensure the input files are placed in the Raw Data/VDSS\_Zip\_Type/ directory.
2. Install the required R libraries.
3. Execute the script in an R project environment.
4. Access the output files in the Data Outputs/VDSS\_Zip\_Type/ directory.

**Notes**

* The accuracy of the results depends on the completeness and correctness of the input datasets.
* Unmatched ZIP codes should be reviewed to address potential data gaps or discrepancies.