Memorandum

To: Derek Wu, Neil Cholli  
From: Dylan Craig  
Date Created: 12/20/2024  
Subject: Data Management Memo on Office Distances (CommonHelp VA) (ZCTA)

**Purpose**

To document the data cleaning, management, and geospatial analysis process for calculating distances and travel times between ZCTA codes where residents live and VDSS local offices they may need to visit based on their county of residence. The analysis also includes an all-years dataset for broader insights into trends over time.

**Refer to READ\_ME files and script directory for technical details on file names and logic.**

**Step 1: Load and Prepare Data**

1. **ZCTA and ZIP Code Data:**
   * ZCTA crosswalk files (HRSA ZCTA dataset) were used to link ZIP codes to ZCTAs.
   * USPS ZIP-FIPS matching data provided mapping between ZIP codes and their corresponding counties.
2. **VDSS Office Data:**
   * Loaded VDSS office location data for specific years (e.g., 2012) and for all years.
   * Ensured FIPS codes matched across datasets for consistency.
3. **Geospatial Data:**
   * TIGRIS shapefiles provided boundary data for ZCTAs in Virginia.
   * Centroid coordinates (latitude and longitude) were extracted for each ZCTA.

**Step 2: Merge Datasets**

1. **ZCTA Centroid Integration:**
   * Coordinates for ZCTA centroids were merged with the crosswalk and USPS ZIP-FIPS data to create a comprehensive dataset.
2. **Complete Merged Dataset:**
   * Each row in the dataset includes:
     + VDSS local office coordinates.
     + ZCTA midpoint coordinates.
     + Associated ZIP codes, FIPS codes, and counties.

**Step 3: Calculate Geospatial Distances and Travel Times**

1. **Haversine Distances:**
   * "As-the-crow-flies" distances were calculated for each ZCTA-office pair using geospatial formulas.
2. **Google Maps API Integration:**
   * Driving and public transit distances and travel times were computed using the Google Maps API.
   * Scripts include error handling and progress tracking to ensure robust API calls.
3. **Unit Conversion:**
   * All distances were converted from meters to miles, and times from seconds to minutes.

**Step 4: Save and Analyze Results**

1. **Year-Specific Results:**
   * Separate Excel outputs for the 2012 analysis, including all calculated metrics (e.g., Haversine, driving, and transit distances and times).
2. **All-Years Analysis:**
   * Consolidated dataset containing distance and time calculations for all available years.
3. **Error Checking:**
   * Rows with missing or unusual values were flagged for review.

**Key Outputs**

1. **2012-Specific Analysis:**
   * **File Name:** VDSS\_Office\_ZCTA\_GeoCode\_2012\_Distances.xlsx
   * Includes metrics for Haversine, driving, and transit distances/times.
2. **All-Years Analysis:**
   * **File Name:** VDSS\_Office\_ZCTA\_GeoCode\_All\_Distances.xlsx
   * Consolidated geospatial data for broader temporal analysis.

**Key Enhancements from Updated Scripts**

1. **Expanded Distance Metrics:**
   * Integration of driving and transit distances/times in addition to Haversine calculations.
2. **Robust Data Processing:**
   * Enhanced merging and data quality checks to handle multi-year datasets.
3. **Interactive Mapping Potential:**
   * The datasets enable downstream mapping and visualization efforts.

**Recommendations**

* Use the all-years dataset for longitudinal studies to assess changes in accessibility over time.
* Leverage the Google Maps API-enhanced metrics for more precise policy evaluations.
* Continue refining centroid-based approaches for improved spatial accuracy.