Transit Count User Guide

Step 1: Downloading GTFS Files

Scripts were developed to automate the process of downloading GTFS data from the TransitFeeds website. These scripts are customized for each province, using URLs specific to transit agencies.

To select a representative day for transit usage, the following considerations were applied:

1. **Exclusion of Statutory Holidays and Weekends**: These days were excluded to avoid atypical transit usage patterns.
2. **Exclusion of Certain Non-Statutory Holidays**: Some non-statutory holidays were also excluded, as transit service and usage may vary due to different policies across different organizations or companies.
3. **Priority for High-Usage Days**: The scripts aim to select days with the highest potential for transit usage, representing typical weekday commuting patterns.

Process Overview:

- The code iterates through a predefined date range, checking each date for eligibility based on the criteria above.
- If a valid date is identified, the corresponding GTFS data is downloaded into a date-specific folder.
- All GTFS outputs are later consolidated into a central folder for further analysis.

Outcome: The extracted GTFS files include stop files, which provide the locations of transit stops across different transit agencies in the city. These stop locations form the basis for subsequent analyses, such as calculating transit accessibility and stop density.

Step 2: Organizing and Categorizing GTFS Data

Once the GTFS files and stop files are downloaded, a script was developed to streamline the process of organizing and unzipping these files. This script performs the following tasks:

1. **Unzipping Files and Organizing by Province**
   The downloaded GTFS files are unzipped and stored in province-specific folders. Each folder contains subdirectories named after the respective transit agencies, ensuring clear organization and ease of access for further analysis.
2. **Categorizing Transit Agencies by CMA and Non-CMA Areas**
   The script identifies transit agencies operating within Census Metropolitan Areas (CMAs). If a transit agency serves any part of a CMA, it is categorized as a CMA agency. This categorization is essential for potential future analyses, allowing researchers to separate CMA and non-CMA transit data depending on the research objectives.

3. **Outputs**
    - A folder containing all transit stops across all transit agencies for each province.
    - A separate folder for each province, containing stops exclusively from CMA transit agencies.

Step 3: Calculating Transit Stop Counts within Dissemination Area Buffers

In this step, the script calculates the number of transit stops within a 1-kilometer buffer around the centroid of each Dissemination Area (DA). Below is a detailed explanation of the process:

1. **Preparation of Dissemination Area Data**
    - Dissemination Area shapefiles are loaded, and the dataset is filtered for each province separately based on their PRUID.
    - Centroids of the filtered DAs are calculated, and 1-kilometer buffers are created around each centroid to define the area of analysis for transit stop counts.
2. **Processing GTFS Stop Files**
    - The script identifies and processes all stops.txt files from the GTFS dataset for different provinces. These files contain geographic coordinates of transit stops (latitude and longitude).
    - Transit stop coordinates are converted into spatial features and reprojected to the same coordinate system as the DA buffers (NAD83 / Statistics Canada Lambert, EPSG:3347).
3. **Counting Transit Stops within Buffers**
    - Using spatial intersection, the script identifies and counts the number of transit stops that fall within the 1-kilometer buffer of each DA centroid.
4. **Output**
    - The final results are compiled into a table where each row represents a DA, along with the corresponding count of transit stops within its buffer (transit_stop_count).
    - The results are saved as Excel files for different provinces (transit_stop_counts_alberta_all.xlsx) for further analysis.