|  |  |
| --- | --- |
| **Order** | Generally contingent, i.e. results of earlier steps are necessary for subsequent steps. This is also the reason why trip linking and mode revision were added to Rulesy—they’re improved with fixes applied first, and subsequent steps also require them. |
| **Db Functions** | Spatial functions, regular expressions, and datetime functions are all available as part of any query clause—they can be combined however we want. This allows a lot of power for specifying conditions or manipulating the data. |
| **General Approach** | Apply solutions where supported by the bulk of evidence, but leave a revision code and keep the original tripid to trace what was done (i.e., manual revision may alter Rulesy edits). |
| **Continual improvement** | Glad to enhance or add to what’s there; again, the above tools create plenty of options.  I’ll spend more time investigating patterns to correct; whatever insight Toolsie can add, will also get incorporated. |
| **Intent: Feed Toolsie** | A view can provide all the fields in the format Toolsie prefers.  Rulesy not currently coded to publish the data—it is not the last stop. Some fields in question. |
| **Destiny: Elmer** | Easy import due to shared platform; haven’t yet coordinated with Chris on all aspects.  Additional geometric assignments can be easily added in Elmer; keeping mainly to reported fields and those necessary for cleaning. |

Rulesy overview:

STEP 1. **Load data from fixed format .csv files.**

1. Add additional fields, including geometry; create indices
2. Convert rMoves trip distances to miles (rSurvey records are already reported in miles)
3. Update tripnum to ensure it is sequential

STEP 2. **Parse/Fill missing address fields**

1. zipcode, city, county [anything else?]

STEP 3. **Corrections to purpose, etc fields**

1. Classify home & work destinations: a) with geographic proximity and purpose or destination name, b) proximity alone (no destinatation code)
2. Revise purpose field (to home or work, respectively) for the return portion of a single stop loop trip
3. Change code to pickup/dropoff when passenger number changes, and either duration is under 30 minutes or pickup/dropoff mentioned in dest\_name
4. Change code to 'family activity' when passenger number changes and duration is from 30mins to 4hrs
5. Update empty purpose code to 'school' when destination is school, person is < 17yrs or a student, and duration > 30 minutes.
6. Change 'Other' trip purpose when purpose is given in dest\_name field (see code)
7. For rMoves records that don't report mode or purpose:
   1. if traveling with another hhmember, take this from the most adult member with whom they traveled
   2. update mode if speed and distance make it obvious (far/fast = airplane, short/slow = walk)

STEP 4. **Trip linking**

1. Concatenate 4 modes into a single text field (also concatenate transit\_systems together, and transit\_lines together); used in STEP 5.
2. Remove trip components into another table:

* prior record does not end at home or work
* trip purpose is either the same, or 'change modes'
* duration between legs is either less than 15 minutes, or if purpose 'change modes', less than 30 minutes.
* either there is a mode change, or both modes are transit

1. denote trip component which have aggregate properties that disqualify them as a single trip:

* two instances of pool start, parking, park & ride, change vehicles
* over 5 components
* non-adjacent repeating transit lines (i.e. loop)

1. for legitimate linked trips, replace the initial component with the aggregate linked trip data

STEP 5. **Mode number standardization, including access and egress characterization**

1. Eliminate repeated values for modes, transit\_systems, and transit\_lines
2. Characterize access and egress trips, separately for 1) transit trips and 2) auto trips. (Bike/Ped trips have no access/egress)
3. Remove access/egress modes from 1) transit and 2) auto trip strings--not only at the ends, but also the middle.
4. Split the concatenated field into separate mode fields (same for transit systems & transit lines)

STEP 6. **Insert trips for those who were reported as a passenger** by another traveler but did not report the trip themselves

--Currently, using a tight constraint for overlap, this generates only 46 trips; more are expected, so we’ll scrutinize this.

STEP 7. **Flag inconsistencies** for further scrutiny

1. Underage driver: driver = 1, age < 4
2. Unlicensed driver: driver = 1, license = 3
3. Non-worker reporting work trip: worker = 0, dest\_purpose = 10, 11 or 14
4. Speed unreasonably high: walking > 20mph, biking > 40mph, driving >85mph, other > 600mph
5. No activity time prior to next departure: less than 1 minute between trips
6. Identical location as next trip: identical lat & lng
7. Time overlap with another trip: same person traveling two trips at same time
8. Same transit line listed multiple times in the same trip
9. Non-home trip purpose, destination home
10. Home or work trip purpose, destination elsewhere
11. Missing next trip link: destination of prior trip not the same as origin of subsequent trip
12. Starts from non-home location: first trip of day after 2am starts elsewhere
13. Unusually long duration at destination (see code)
14. Suspected drop-off or pick-up coded as school trip: purpose = 'school' but person not a student and >17yrs, or duration under 20 minutes