

| Date | Day of the week | Sensor readings | DB insertions |
|------------|-----------------|-----------------|---------------|
| 2021-02-03 | | 381854 | 282020 |
| 2021-02-04 | | 380519 | |
| 2021-02-05 | | 384540 | |
| 2021-02-06 | | 175810 | |
| 2021-02-07 | | 134608 | |
| 2021-02-09 | | 371613 | |
| 2021-02-10 | | 370452 | |

Data Validation Assertions:

- The ACT_TIME field will be between 0-86400
- The GPS_LATITUDE field will never be an empty string
- The GPS_LONGITUDE field will never be an empty string
- The GPS_SATELLIETS field will never be an empty string
- The DIRECTION field will never be an empty string
- The OPD_DATE field will be a valid datetime
- The VELOCITY field will never be an empty string
- The METERS field will never be an empty string
- The VEHICLE_ID field will never be an empty string
- The GPS_LONGITUDE field will always be between -180 and 180
- The GPS_LATITUDE field will always be between -90 and 90
- The SCHEDULE_DEVIATION field will never be an empty string
- The DIRECTION field will be between 0 - 359
- The ACT_TIME field will never be an empty string
- The OPD_DATE field will never be an empty string
- If the VELOCITY field is empty, then the DIRECTION field must also be empty
- The GPS_HDOP field will never be an empty string
- The SCHEDULE_DEVIATION field will be higher for the shortest routes
- The VELOCITY field will be higher on average for longer routes
- Buses spend a majority of their time travelling north and south

Data Transformations:

- Empty values get converted to "0" for certain fields (gps_satelliets, radio_quality, meters, velocity, schedule_deviation, gps_hdop). This choice was largely driven by wanting to ensure there were default values for fields, even if they may not be used.
- Lat/Long get converted to zero if they fall outside of expected values (ex, -90/90/-180/180). Ideally, they should be removed from the dataset but I chose to keep them for the sake of having the complete dataset.

Example Queries

1. How many vehicles are there in the C-Tran system?
- query: `select count(distinct(vehicle_id)) from trip;`
2. How many bread crumb reading events occurred on October 2, 2020?
- query: `select count(*) from breadcrumb where DATE(tstamp) = '2020-10-02';`
3. How many bread crumb reading events occurred on October 3, 2020?
- query: `select count(*) from breadcrumb where DATE(tstamp) = '2020-10-03';`