Nationwide: Telematics Assessment Exercises

Jason Barkeloo

July 18, 2020

Table of Contents

Code location for further fleshed out examples

- All code for these exercises can be found via these links as ipython/jupyter notebooks located on my github in addition to attachments sent with with the presentation
  - Part 1: github: BarkelooNationwideAssessmentPart1.ipynp
  - Part 2: github: BarkelooNationwideAssessmentPart2.ipynp

Tasks to be Completed

Analysis Task:

- 1: Data Cleaning
- 2: Setting of hard braking and acceleration tresholds based on the data
- 3: Trip-by-trip Analysis and Summary

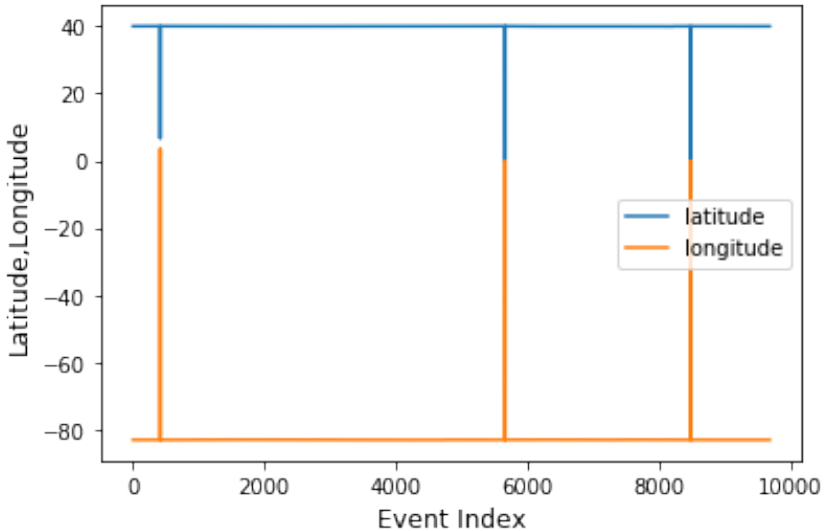
Data Set Overview:

- 9687 rows of 4 variables including:
  - trip_id: a trip number identifier
  - local_dtm: a datetime timestamp of the event entry
  - latitude: latitudinal coordinate
  - longitude: longitudinal coordinate

Datasets are loaded into pandas dataframes for further analysis

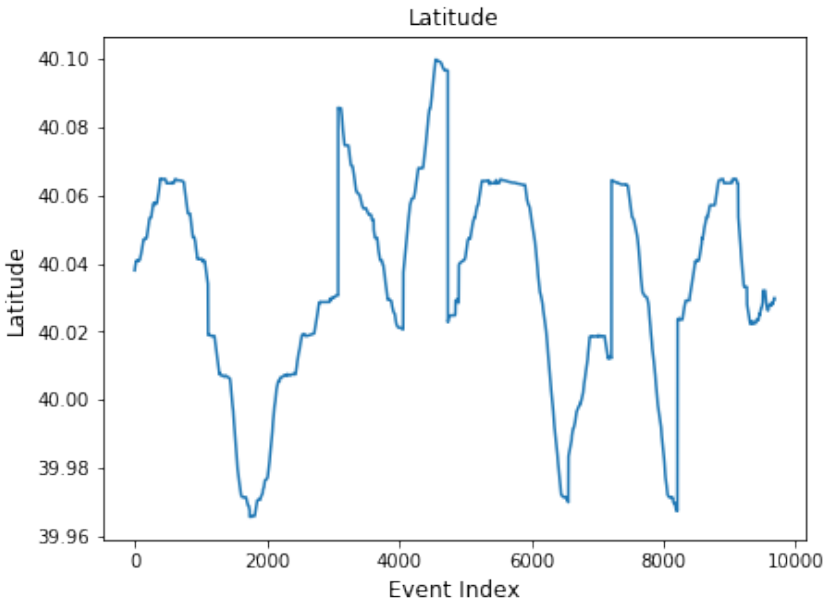
Data Cleaning, Gross Features

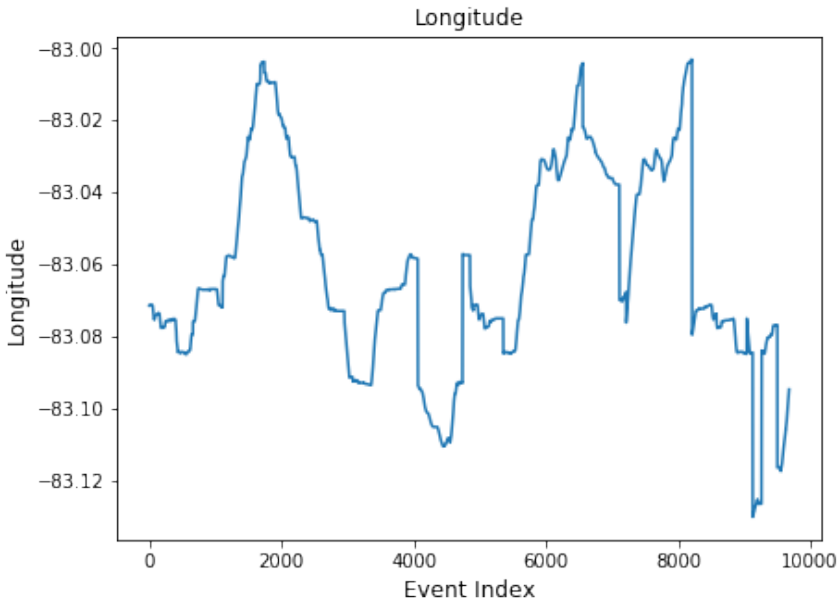
- 3 Large unphysical features occur in the dataset (teleportation across the globe for 2-4 seconds)
- These events are pruned by requiring the latitude and longitude are within 2° of the median for the data set.
- This includes an area on the order of the state of Ohio
  - Assumption: The sensors are used for checking daily driving habits and not long, rare, road trips.
  - No other points are removed under this cut just these large outliers but if this assumption is false (i.e. long-haul truck drivers use these) this would need to be adapted



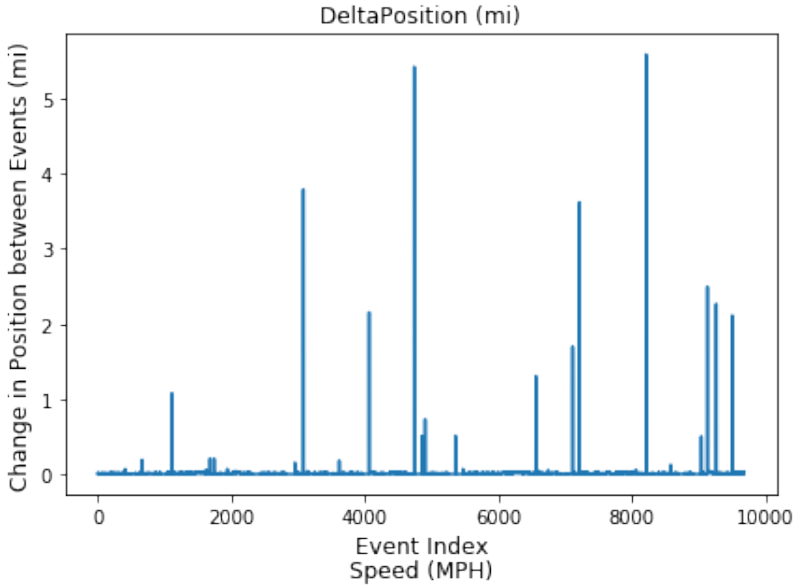
Result of Gross Cleaning

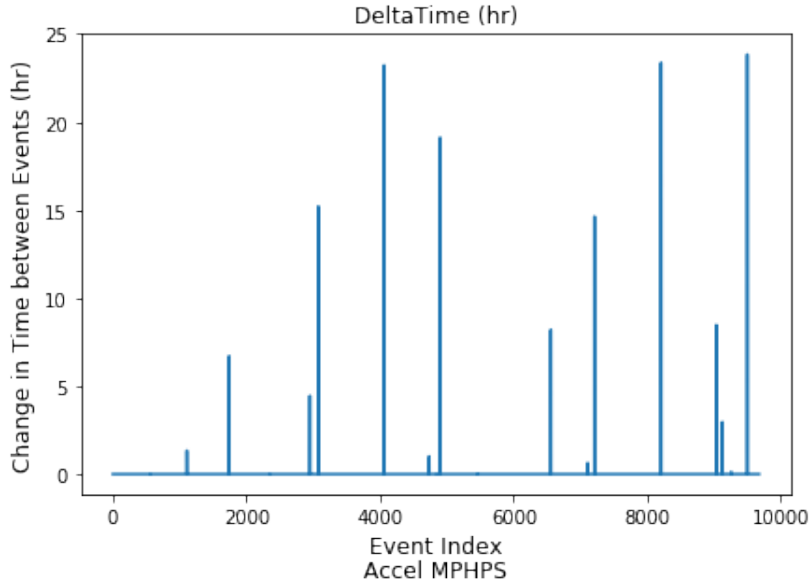
- The median cut before leaves the longitude and latitude plots in a reasonable state.
- Still some very fast jumps which are coincident, typically, with a change in trip_id (GPS drift while off)
- From this data and corresponding timestamps in local_dtm plots of the speed $s = \frac{\Delta \text{Positions}}{\Delta \text{Time}}$ and acceleration $a = \frac{\Delta \text{Speed}}{\Delta \text{Time}}$ can be made

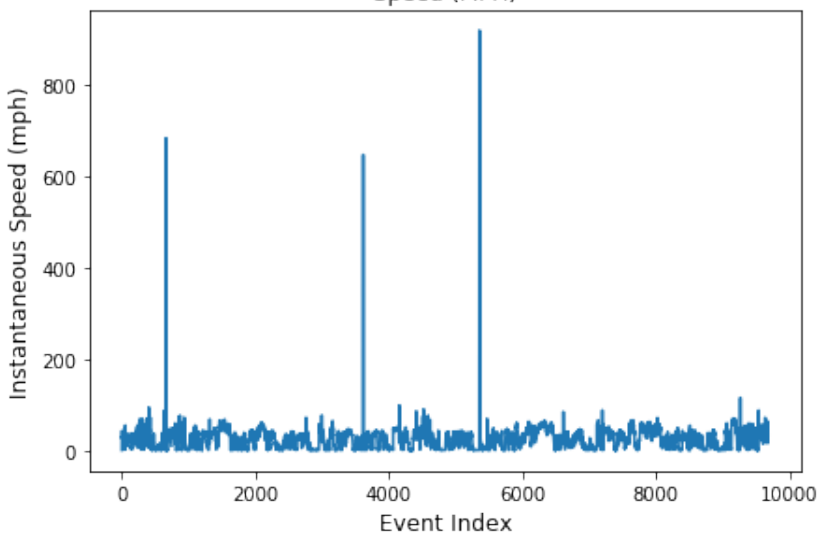


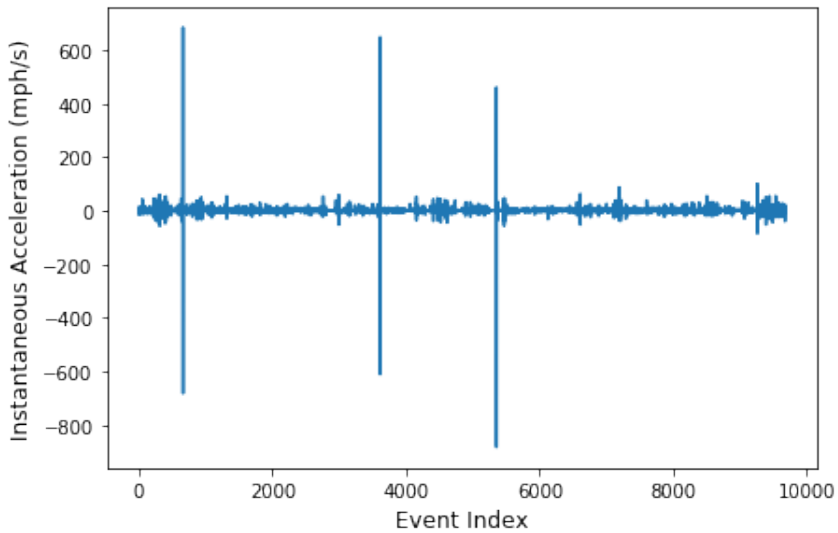


Further Cleaning - ΔPosition, ΔTime, Speed, Acceleration



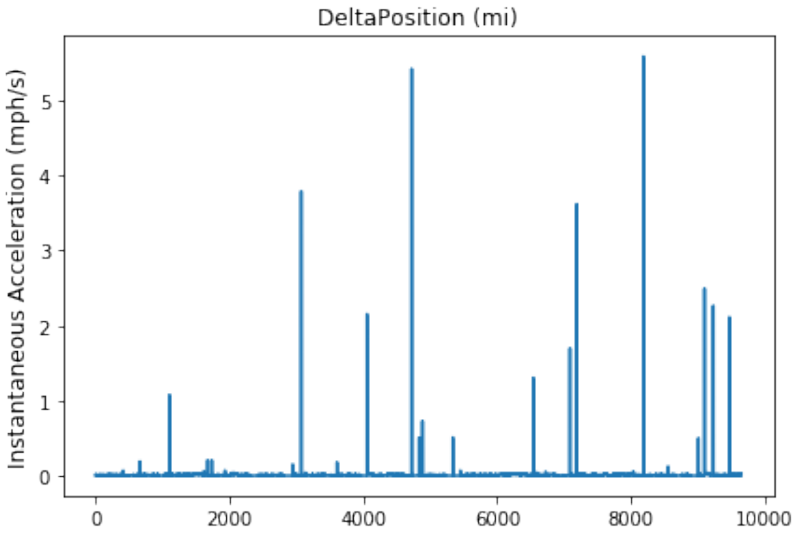


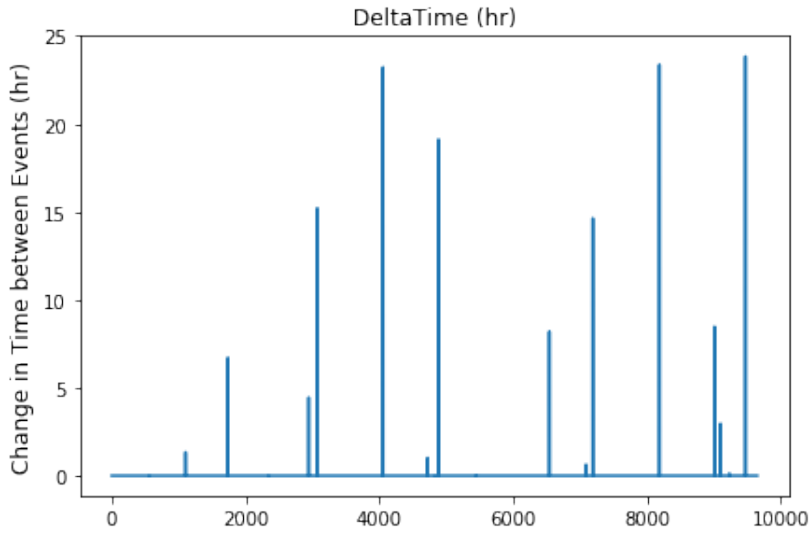


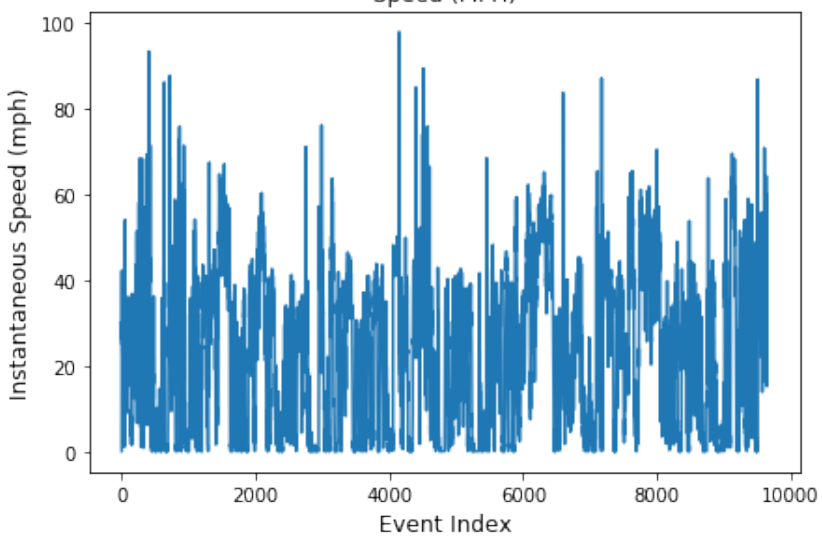


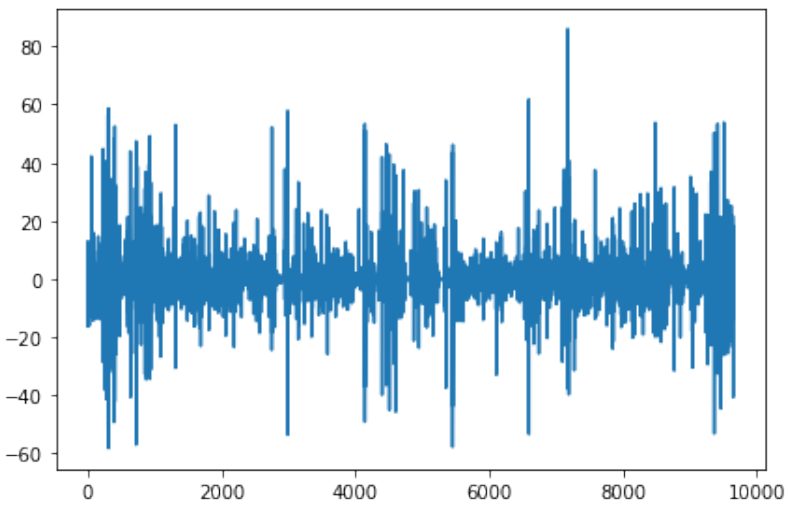
More Features to be Cleaned

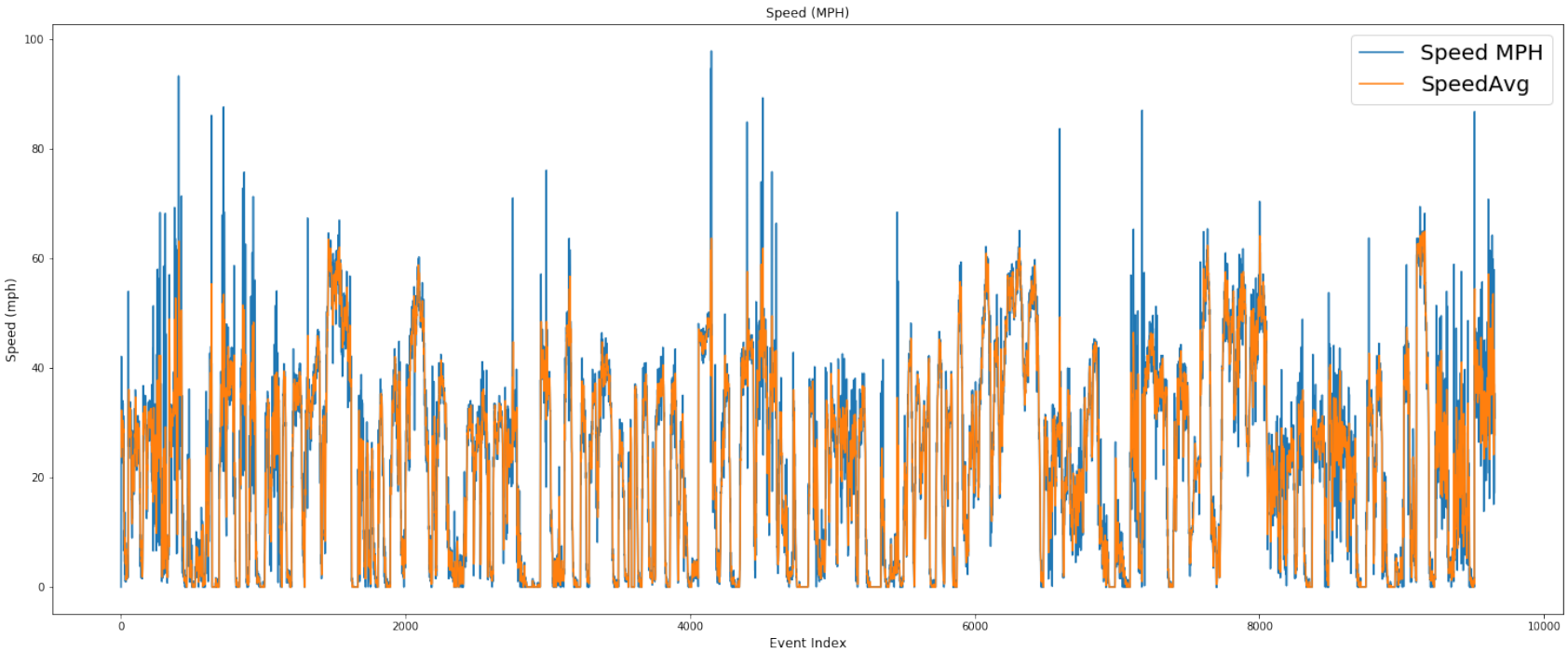
- From ΔPosition, ΔTime we see the large number of drifts which account for the gps drift from trip differences
- These jumps will not be an issue when analyzing trip by trip as the change in position starts from the first point of the trip
- Speed and Acceleration plots show an additional 3 further unphysical events. These are resultant from small gps errors for a few seconds and need to be dealt with
- Another issue comes when ΔTime between two events is 0 i.e., if the frequency drops below 1Hz and two readings are taken within a second.
  - A 0th order approach is taken to these points and only the first is kept. An alternative would be averaging the latitude/longitude for those points. This would be a change within the same second and as such will not have much of an effect that isnt then averaged out in the acceleration

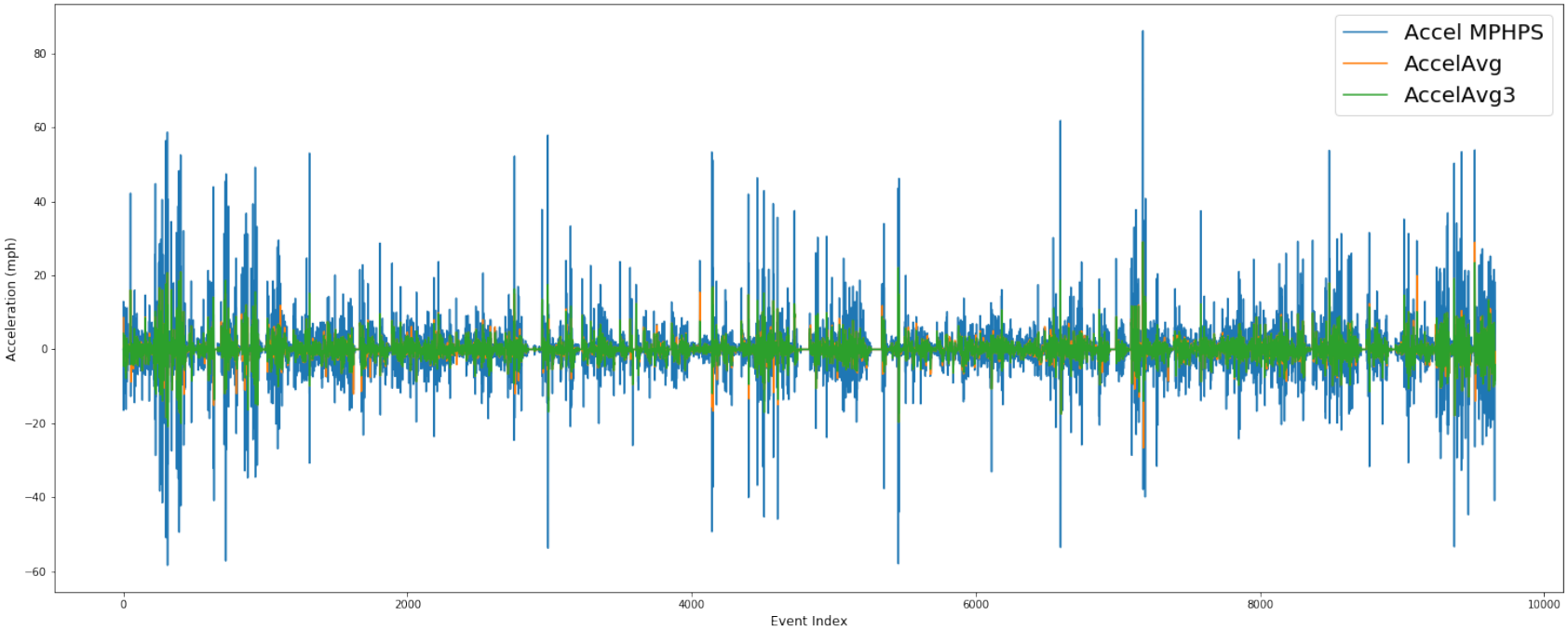

DeltaPosition (mi)
Instantaneous Acceleration (mph/s)
Event Index

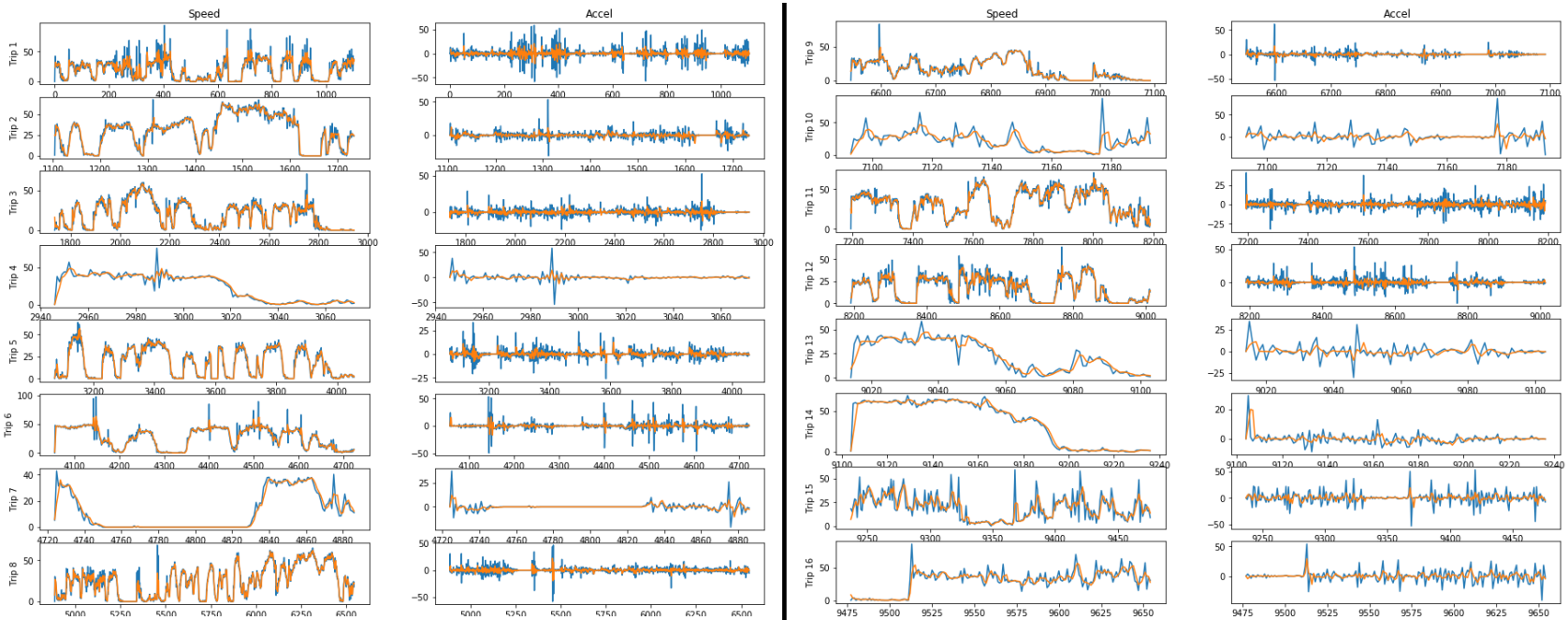

DeltaTime (hr)
Change in Time between Events (hr)
Event Index

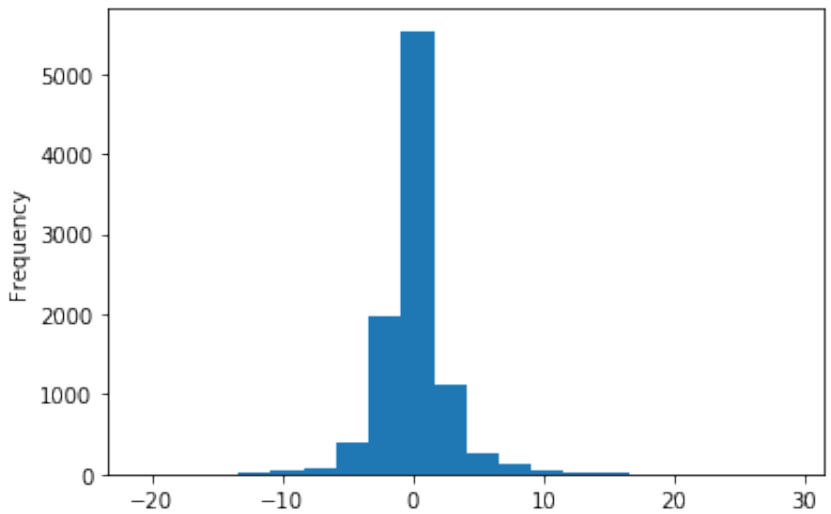

Speed (MPH)
Instantaneous Speed (mph)
Event Index


Accel MPHPS


Speed (MPH)
Speed MPH
SpeedAvg
Speed (mph)
Event Index
0
20
40
60
80
100
0
2000
4000
6000
8000
10000


Accel MPHPS
AccelAvg
AccelAvg3
Acceleration (mph)
Event Index
80
60
40
20
0
−20
−40
−60
0
2000
4000
6000
8000
10000


Speed
Accel
Speed
Accel
Trip 1
Trip 2
Trip 3
Trip 4
Trip 5
Trip 6
Trip 7
Trip 8
Trip 9
Trip 10
Trip 11
Trip 12
Trip 13
Trip 14
Trip 15
Trip 16


Frequency
5000
4000
3000
2000
1000
0
−20
−10
0
10
20
30

```
Trip:  1
        Hard Accel Events: 47
        Hard Brake Events: 40
        Idle Time: 3.05 min,    Total Time: 22.25 min
        Distance Traveled: 6.83 mi

Trip:  2
        Hard Accel Events: 7
        Hard Brake Events: 5
        Idle Time: 1.15 min,    Total Time: 13.03 min
        Distance Traveled: 7.29 mi

Trip:  3
        Hard Accel Events: 10
        Hard Brake Events: 11
        Idle Time: 3.45 min,    Total Time: 24.45 min
        Distance Traveled: 7.77 mi

Trip:  4
        Hard Accel Events: 3
        Hard Brake Events: 1
        Idle Time: 0.23 min,    Total Time: 2.33 min
        Distance Traveled: 1.06 mi

Trip:  5
        Hard Accel Events: 10
        Hard Brake Events: 5
        Idle Time: 3.07 min,    Total Time: 18.70 min
        Distance Traveled: 9.75 mi

Trip:  6
        Hard Accel Events: 18
        Hard Brake Events: 10
        Idle Time: 0.90 min,    Total Time: 12.60 min
        Distance Traveled: 7.83 mi

Trip:  7
        Hard Accel Events: 3
        Hard Brake Events: 1
        Idle Time: 1.48 min,    Total Time: 4.03 min
        Distance Traveled: 6.55 mi

Trip:  8
        Hard Accel Events: 11
        Hard Brake Events: 13
        Idle Time: 4.22 min,    Total Time: 33.73 min
        Distance Traveled: 14.31 mi

Trip:  9
        Hard Accel Events: 4
        Hard Brake Events: 7
        Idle Time: 1.60 min,    Total Time: 10.97 min
        Distance Traveled: 4.28 mi

Trip:  10
        Hard Accel Events: 8
        Hard Brake Events: 9
        Idle Time: 0.02 min,    Total Time: 1.93 min
        Distance Traveled: 2.34 mi

Trip:  11
        Hard Accel Events: 11
        Hard Brake Events: 11
        Idle Time: 0.62 min,    Total Time: 19.20 min
        Distance Traveled: 13.82 mi

Trip:  12
        Hard Accel Events: 12
        Hard Brake Events: 14
        Idle Time: 2.90 min,    Total Time: 16.15 min
        Distance Traveled: 10.06 mi

Trip:  13
        Hard Accel Events: 3
        Hard Brake Events: 3
No Idle Time for this Trip
        Idle Time: 0.00 min,    Total Time: 1.65 min
        Distance Traveled: 1.13 mi

Trip:  14
        Hard Accel Events: 1
        Hard Brake Events: 0
        Idle Time: 0.15 min,    Total Time: 2.48 min
        Distance Traveled: 4.04 mi

Trip:  15
        Hard Accel Events: 17
        Hard Brake Events: 13
No Idle Time for this Trip
        Idle Time: 0.00 min,    Total Time: 4.47 min
        Distance Traveled: 3.69 mi

Trip:  16
        Hard Accel Events: 7
        Hard Brake Events: 10
        Idle Time: 0.32 min,    Total Time: 3.37 min
        Distance Traveled: 3.80 mi
```

FCNC: What are we looking for? $t\bar{t} \rightarrow W(\rightarrow l\nu)b + q\gamma$

- Final state topology
  - One Neutrino, from W
  - One Lepton, from W
  - One B-jet, SM top
  - **One Photon, FCNC Top**
  - One Jet, FCNC Top

Background Processes

- Due to all of the processes at hadron colliders it is important to model similar event topologies well.
- Major backgrounds include $t\bar{t}$, W+Jets, Z+Jets, + processes with an associated photon

Monte Carlo Generation

- All of our MC data is put through a showering algorithm for propagation from final decay states
- Various showering algorithms are used at ATLAS - Pythia, Herwigg, etc.
- All of these will add radiative photons
- These events can be contained in other samples with explicit photons originating from the hard interaction
- Need to remove these events or risk double counting events

Object Preselection

- We preselect events with objects that look like our expected topology
- Reminder that I require:
  - Exactly one lepton (e or μ) $\geq$ 28 GeV
  - Exactly one Good photon $\geq$ 25GeV
  - Missing Transverse Energy $\geq$ 30GeV
  - $\geq$ 2 Jets (at least one being b-tagged)
- All following plots will have signal scaled to 0.2% of nonallhadronic $\sigma_{t\bar{t}}$, MC scaled to 36.07fb^{-1}
- Only electron channel shown. Similar results for the muon channel are seen.

Background Processes

- Due to all of the processes at hadron colliders it is important to model similar event topologies well.
- Major backgrounds include $t\bar{t}$, W+Jets, Z+Jets, + processes with an associated photon

Background Processes

- Due to all of the processes at hadron colliders it is important to model similar event topologies well.
- Major backgrounds include $t\bar{t}$, W+Jets, Z+Jets, + processes with an associated photon

Background Processes

- Due to all of the processes at hadron colliders it is important to model similar event topologies well.
- Major backgrounds include $t\bar{t}$, W+Jets, Z+Jets, + processes with an associated photon

Conclusion, Outlook

- Orthogonal validation/control regions are in development
- Data grid run complete, need to incorporate into CR/VR plots
- Next grid run will include a couple of looser regions for CR/VRs
  - 0 Photon Samples for Backgrounds with no Real Photons
  - 0 BJet Samples - possibly for WJets region
- Top Group - Pushing for MVA, want to start investigations using these techniques

Backup