In these two weeks I mainly focus on data cleaning & formatting. The original data table has 25 columns and 11.5M rows. I deleted some unnecessary columns and only kept 9 columns.

v_Class and location are categorical variables. So I filtered these two variables by selecting v_Class = 2 (2 refers to “auto”) and location = “peachtree” (peachtree refers to Peachtree Street in Atlanta, GA). In this way we can narrow down the scope of this huge data and make the study more focused on a certain area, which makes the calculations in Python faster.

The variable “Global_time” refers to the elapsed time in milliseconds since Jan 1, 1970. I defined a function to transform and split this time into day_num, month_num, year_num, week_num, weekday_num, hours, minutes, seconds.

The variable “Global_X” and “Global_Y” is based on Georgia State Plane Coordinate System in feet (EPSG: 2240). I transformed it into EPSG:4326 to get human-readable decimal degrees.

Then I will use Folium to visualize the attributes like velocity and acceleration. I am still working on this part.

Challenges:
1. The time is not consecutive in seconds. I need to find a way to make the trip visualization look more make sense on the basemap. (maybe use Google “Snap to Roads” API)
2. The dataset does not have “heading angle” so I am not able to plot the direction of the acceleration in each second. I am still finding a method to figure this out.

Next Step:
1. Try to make the trip look consecutive and smooth on the road network
2. Try to calculate the heading(direction of acceleration) in each second of time and visualize these directions as a field of arrows
3. Make color-coded lines for velocity and acceleration.