

DataEng: Project Assignment 3

Data Integration

Genevieve LaLonde
Data Engineering Winter 2021
Bruce Irvin

Assignment date: February 16

Due date: February 28, 2021 @10pm PT

Submit: [assignment submission form](#)

Submission

Make a copy of this document and update it to include the following visualizations. For each visualization extract from your database a list of {latitude, longitude, speed} tuples and then use the provided visualization code (see Section D above) to display bus speeds at all of the corresponding geographic coordinates. So, for example, if you are asked to visualize a “trip”, then you must query your database to find all of the {latitude, longitude, speed} tuples for that trip, and then display a map showing the recorded/calculated bus speed at each {latitude,longitude} location.

No need to produce software that neatly displays trips, routes, dates, times, etc. onto the visualization itself. Instead, just paste a screen capture of the map-based speed visualization into your submission document and then include a text description of the contents of the visualization. For example, text like this: “Bus Speeds for all outbound trips of route 65 between 9am and 11am on Sunday October 32, 2020.”

Visualization 1. A visualization of speeds for a single trip for any bus route that crosses the Glenn Jackson I-205 bridge. You choose the day, time and route for your selected trip. To find a trip that traverses this bridge, consider finding a trip that includes breadcrumb sensor points within this bounding box: [45.592404, -122.550711, 45.586158, -122.541270]. Any bus trip that includes breadcrumb points within that box either crosses the bridge or goes swimming in the Columbia river!

Visualization 2. All outbound trips that occurred on route 65 on any Friday (you choose which Friday) between the hours of 4pm and 6pm.

Visualization 3. All outbound trips for route 65 on any Sunday morning (you choose which Sunday) between 9am and 11am.

Visualization 4. The longest (as measured by time) trip in your entire data set. Indicate the date, route #, and trip ID of the trip along with a visualization showing the entire trip.

Visualization 5a, 5b, 5c, Three or more additional visualizations of your choice. Indicate why you chose each particular visualization.

I have not yet generated any visualizations. I will continue to work on this project and resubmit asap once I have generated them.

Your Code

Provide a reference to the repository where you store your code. If you are keeping it private then share it with Bruce (bruce.irvin@gmail.com), David and Aman (github references TBD).

So far I have modified my cron job to also collect the Stop Event Data.

https://github.com/coding-gen/dataeng/tree/main/project/3_integrate

During this time, I was finishing portions of project part 2 which were not completed by my group. I have just finished this work and it is available at:

https://github.com/coding-gen/dataeng/tree/main/project/2_validate_transform_enhance_store

I will be continuing to work on project part 3 alone, and will provide my code to my group as it becomes available.