

CUSP-GX-6006.001: Data Visualization
SUMMER 2019**Final Challenge – Ride-pooling Simulation****Due Date:** 10:00 AM on July 31, 2019

Problem Statement: Ride-hailing services have become extremely popular in urban areas by providing a convenient and affordable alternative to owning a private car, with lower up-front costs, lower maintenance, and a smaller vehicle fleet, ride-hailing service provides an opportunity to reducing the cost of per-vehicle and sustainability of these services. More precisely, ride-hailing services have the potential to ensure that the vehicle supply meets the demand. This raises concerns about rebalancing has the same demand. This raises concerns about recently by multiple city planners in New York City. One strategy for addressing this is by encouraging passengers to share rides with ride-pooling services (e.g. Uberpool, Lyft) instead of providing more transit lines. This can further increase the efficiency of the system.

We would like to simulate the ride-pooling service in order to do this. The simulator will take as input the matching and output, which will be a visualization of the simulation. For this challenge, you will be asked to design a simulation that can be shared. In order to do this, you will need to design various service matching algorithms and input them into the simulator through visualization.

DATA SET

There will be two set of simulation results, where each produces all of the output files below.

vehicle_events.csv: a CSV file containing vehicle activities during the simulation, with the following fields:

- *Timestamp* – the time of the event in seconds. It is either since 1/1/1970, or since 00:00AM if it is a single day simulation.
- *Vehicle_ID* – this is an ID that uniquely identify a vehicle.
- *Stop_Intersection* – the intersection ID of the event. This ID should match with the intersection ID of the road network GeoJSON file.



- *Stop_Passengers* – the number of passengers involved at this stop. It is greater than 0 for picking up passengers; less than 0 for dropping off passengers; and equal 0 for waiting or rebalancing events.
- *Requested_Stop_Intersection* – this is the original requested intersection ID. For example, a user might request to pick up at intersection A, but the framework can only assigned a cab to pick the user at intersection B. In this case, A is the Requested_Stop_Intersection, and B is the Stop_Intersection. For waiting or rebalancing, Intersection_Stop and Requested_Stop_Intersection should be the same.

request.csv: contains a list of all trip requests, and matched-up locations, with the following fields:

- *Timestamp* – the time of the request, in YYYY-MM-DD HH:MM:SS format. If it is a single day, the date part can be omitted.
- *Simulated* – a boolean value indicating if the request is simulated or real.
- *Requested_Stop_Intersection* – the original requested intersection ID.
- *Requested_Stop_Latitude* – the original requested stop latitude.
- *Actual_Stop_Intersection* – the actual intersection where the cab picked up the user.
- *Actual_Stop_Latitude* – the actual stop latitude.

vehicle_paths.csv:

contains the path of each vehicle, it is guaranteed that the path is a continuous line (road segment) or a set of points. It only includes the following fields:

- *Timestamp* – the time of the request, in YYYY-MM-DD HH:MM:SS format. If it is a single day, the date part can be omitted.
- *Vehicle_ID* – the ID of the vehicle.
- *Latitude* – the latitude of the vehicle.
- *Longitude* – the longitude of the vehicle.
- *Num_Passengers* – the number of passengers in the vehicle.

In addition, we will provide a map of the New York City street network of Manhattan, where each vehicle's path is overlaid.

YOUR TASKS

You are asked to analyze the data and answer the following questions (and help us improve the system).

1. What is the serving rate for each scenario throughout the day, comparing to the overall serving rate? Serving rate is the number of successfully matched trip in a period of time. For example, is the serving rate higher in the rush hour or at night?
2. For those trips that could not be served, do they follow a spatial or temporal pattern? For example, are most of those trips originated in particular regions, and of certain times?
3. For the entire simulation, we limit vehicle speed to under 25mph. Could you see any vehicle travel exceeded those limits? If so, could you show us where (and potentially how) that happen?



4. We also limit vehicle capacity to at most 4 passengers. Were there vehicles violating this condition? If so, can you show any pattern about these vehicles? For example, how many of them were violating, and where were they distributed in both time and space?
5. What can we learn about the vehicle utilization? Are most of them empty, or with 1, 2, 3 or 4 passengers? Are there particular vehicles that tend to ride with more passengers than others? If so, how are they distributed in space and time.
6. Are most vehicle moving or idling? In which part of the city that we see vehicle idling/moving more often? For example, are there any “dead zone” where a vehicle just drops off passengers, and stay idling?
7. We want to know the average number of passengers per vehicle, the number of rebalance trips (stops) per vehicle, and the trends in these metrics over time.
8. If we analyze the data throughout the day?

YOUR SUBMISSION

You can pick your team size.
You are not required to work individually or in pairs (aka. roughly 10-15 minutes).

Your submission should include:

- Team name
- The task you are working on
- Link(s) to your work
- the time you spent on it
- A short summary of your findings



average number of passengers per vehicle, the number of rebalance trips (stops) per vehicle, and the trends in these metrics over time.

ies throughout the day?

tively tackle them. on this challenge to balance the workload

ossible throughout