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 reliable alternative to personal vehicles. They provide a service that is comparable to owning a private car, while eliminating some of the more burdensome aspects of ownership, e.g., high up-front costs, maintenance, searching for parking etc. Moreover, by collectively time-sharing the use of a vehicle fleet, ride-hailing services are able to distribute fixed costs over a large user base, providing an opportunity to reducing the cost of personal mobility. While ride-sharing offers a number of benefits, the scalability and sustainability of these services is limited by the endemic inefficiency caused by the asymmetry of user demand. More precisely, ride-hail vehicles must be constantly rebalanced or moved (without passengers) to ensure that the vehicle supply is aligned with user demand. In contrast to a private car ownership model, rebalancing has the unfortunate side-effect of increasing the total vehicle mileage driven in the system to satisfy the same demand. This raises concerns about increasing congestion on city streets, as has been expressed by recently by multiple city planners including NYC. Given that ride-hailing vehicles can carry multiple passengers simultaneously, one strategy for mitigating this effect is to promote ride-pooling into the service, encouraging passengers to share rides with other commuters. Most major ride-hailing companies already provide a pooling service (e.g. Uberpool, Lyftline, and Via), where passengers pay a reduced fee to compensate for the inconveniences (additional travel time, sharing a vehicle etc.) associated with pooling. More recently, there have also been more transit like ride-pooling services that require users to walk to pickup points (and from drop off points) instead of providing a door-to-door service (e.g. Uber Express Pool, Lyft Shared Saver). This can further increase the efficiency of the ride-hailing system and reduce its negative externalities.

We would like to understand the impact of ride-pooling, particularly if all rides are meant to be shared. In order to do this, we have developed a framework to simulate ride-pooling scenarios including various service matching and rebalancing strategy. Then, to verify and study each scenario, we need to look into the simulator output, which logs all vehicle positions and activities. Your task is to analyze the logs through visualization. For this challenge, the simulation is limited to only Manhattan, and using 500 vehicles.

DATA SETS

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_Intersetion* – 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, Interesevtion_Stop and Requested_Stop_Interesection 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. It is either since 1/1/1970, or since 00:00AM if it is a single day simulation.
- *Requested_Pickup* – the intersection ID of the requested pickup
- *Requested_Dropoff* – the intersection ID of the requested dropoff
- *Actual_Pickup* – the actual pickup intersection ID (or -1 if the trip cannot be served)
- *Actual_Dropoff* – the actual dropoff intersection ID (or -1 if the trip cannot be served)

vehicle_paths.csv: a CSV file containing fine-grain trajectories of all vehicles. For each vehicle, it is guaranteed that consecutive pings listed in this file is a direct travel path (aka. an actual road segment) or idling position. Its timestamps should match with those in **vehicle_events.csv**. However, it only includes the following fields due to space:
- *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.
- *Latitude* – the latitude of the vehicle
- *Longitude* – the longitude of the vehicle
- *Num_Passengers* – the number of passengers in the vehicle at that moment

In addition, we also have a GeoJSON file (**manhattan.geojson**) consisting the road network of Manhattan, where each vertex ID is also listed along with their locations.

YOUR TASKS

You are asked to use visualization (e.g. constructing a plot, or an interactive visualization) and help us answering the following questions.

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 forever?

7. We want to understand the trends among vehicles in term distance travel, average number of passengers, the number of revenue trips (a pair of pickup/dropoff), the number of rebalance trips (stops with 0 passengers but positive duration), and idling time. How can you help us explore this trends?

8. If we are to pick a particular vehicle ID, how can we effectively illustrate its activities throughout the day?

YOUR SUBMISSIONS

You can pick one or several of the tasks above, and build a set of visualizations to effectively tackle them. You are not expected to address all tasks, but the more the better. You can work on this challenge individually or of team at most 2 persons. Of course, a 2-person team is expected to double the workload (aka. roughly twice the number of tasks).

You submissions should include the following:
- Team member names (both should submit if you work in a 2-person team)
- The tasks that you're choosing
- Link(s) to your visualization(s) that tackle the tasks. Your visualization must be accessible throughout the time of grading, which is at least until August 10th.
- A short justification of how your visualization can effectively solve the tasks.