# ENGI E4800 - Data Science Capstone & Ethics

# Progress Report II – DiDi Reinforcement Learning

Danyang Han (dh2990), Anita Pinto (ap3650), Elizabeth Yum (eyy2106)

Mentors: Zhiwei (Tony) Qin and Vineet Goyal

April 3, 2021

## I. Progress Overview

In the first Progress Report (https://github.com/eyy4/Spr-2021-DiDi-RL/blob/master/reports/DiDi-ReinforcementLearning-FirstReport.pdf), we outlined the overall project goals and methodology and presented a proposed method to tackle the problem of driver repositioning through an environment simulator and generalized policy iteration. Figure 1 below summarizes the overall project methodology.

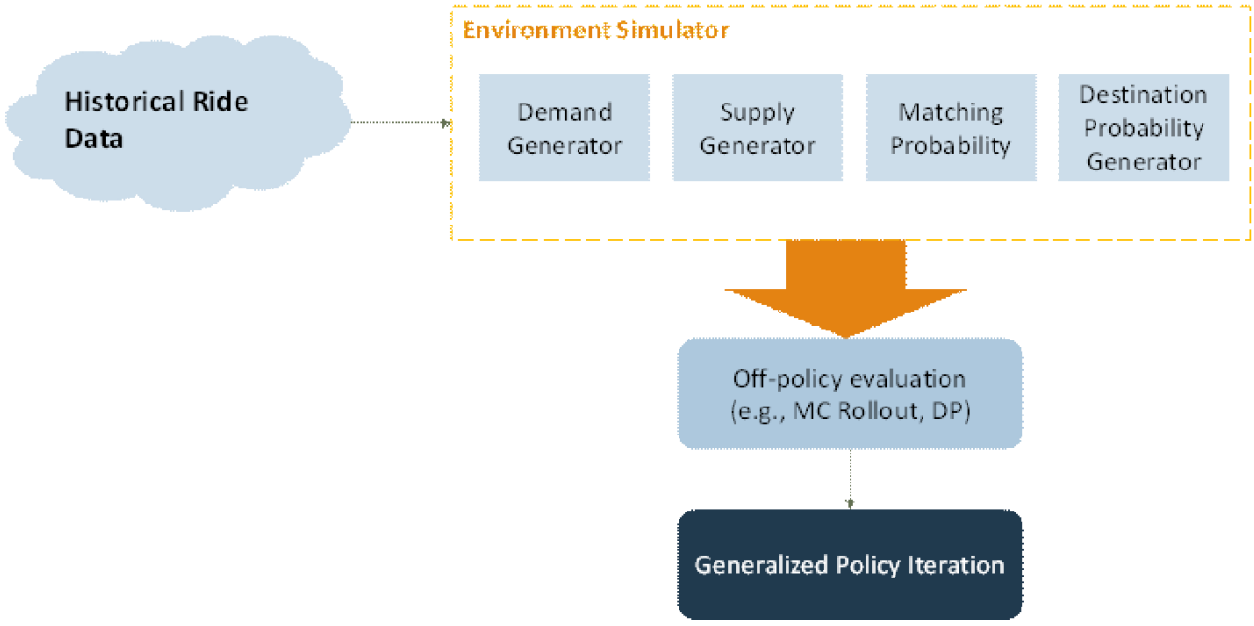



**Figure 1: Project Methodology**

In this second report, we will detail the progress made on key aspects of the environment simulator.

## II. Demand Generator

As discussed in Progress Report I, the demand generator was the crux of the environment simulator but subject to many challenges. The three problem statements

outlined in Progress Report I are below with a brief description of the solution we arrived at. A longer description of the solution follows below.

**Challenge #1**: Demand highly sensitive to zone, time, and day

- **Solution**: Demand was estimated separately for each zone, time interval, and based on weekday/weekend indicator

**Challenge #2**: Historic data only contains observed demand

- **Solution**: We followed an approach to estimate the demand lost via censoring (see below for more details).

**Challenge #3**: In attempting to adapt the approach for censored demand in the newsvendor problem, we must be able to identify a proxy to a true "stock-out" indicator

- **Solution**: We experimented with and settled on a stock-out indicator, and based on the presence of the stock-out indicator applied an optimization function to arrive at an estimate.

## Detailed Description of our Chosen Approach

### General Approach

For a given taxi zone, and time interval -> 30 min,
We have 30 days of arrival data

Assuming → Poisson Distribution with parameter $\lambda$

Under normal circumstances, the $\lambda$ would be the mean of these 30 days observations ( $x_1$, $x_2$ -- $x_{30}$) in this time interval for a given zone.

To tackle the problem of estimating lost demand, we need to first calculate the stockout indicator similar to the newsvendor problem in which we do not have access to the underlying demand distribution. Considering no censored demand, the probability of passengers arrival rate for a given taxi zone per delta time interval would just be (1 - the probability of no passengers arriving at the zone in that time interval).

Pr(at least one passenger) = 1 - Prob(no passengers)

Pr(arrival of K passengers ) = $\frac{\lambda^k e^{-\lambda}}{k!}$

Pr(at least one passenger) = 1 - exp(-λ ) [k =0]

We assume that distribution of K passengers arriving at a taxi zone in a given time interval is a Poisson Distribution of parameter $\lambda$ which is also the mean of arrivals. We set the demand

estimate equal to the maximum likelihood estimate given the stockout indicator which we would call as censored parameter $\delta$

For uncensored arrival rate, the $\lambda$ would be the mean of the 30 days arrival observations ( $x_1$, $x_2$ -- $x_{30}$) in this time interval for a given taxi zone.

For estimating censored demand, we can calculate the censored parameter or stockout ( $\delta_1$, $\delta_2$, .. $\delta_{30}$) Where $\delta$ = 0 means no censoring, $\delta$=1 means censoring present

Our goal is to find the parameter $\lambda$ given pairs of data ( $x_1$ $\delta_1$) … ($x_{30}$,$\delta_{30}$)

$$\lambda_{ML} = argmax(\ \sum_{i=1}^{n}\ (1-\delta_i)\ log(f(x_i,\lambda) + \delta_i log(1-F(x_i,\lambda)\ )$$

Where,

$$f(x_i,\lambda) = \frac{e^{-\lambda} * \lambda^{-x_i}}{x_i!}$$

$$F(x_i,\lambda) = \sum_{i=1}^{n} \frac{e^{-\lambda} * \lambda^{-x_i}}{x_i!}$$

## Choice of Stock-out Indicator

To find Censor Indicator for a taxi zone in a given time interval, we analyzed the pickup-time intervals. A pick time interval is the max time difference between 2 pick-ups or rides in a taxi zone on a given time delta for a given day. We assign censor indicator=1 if the pick-up interval observed is much larger than the average pickup intervals. Also, this censoring detection was performed separately for weekday and weekend data. By not setting a strict pick-up interval cut-off, we have provided flexibility and independence for different taxi zones for different time intervals, as pickup interval for a busy taxi zone on a weekend morning hour may vary when compared against pick up interval for a quieter taxi zone on a weekday at night hours.

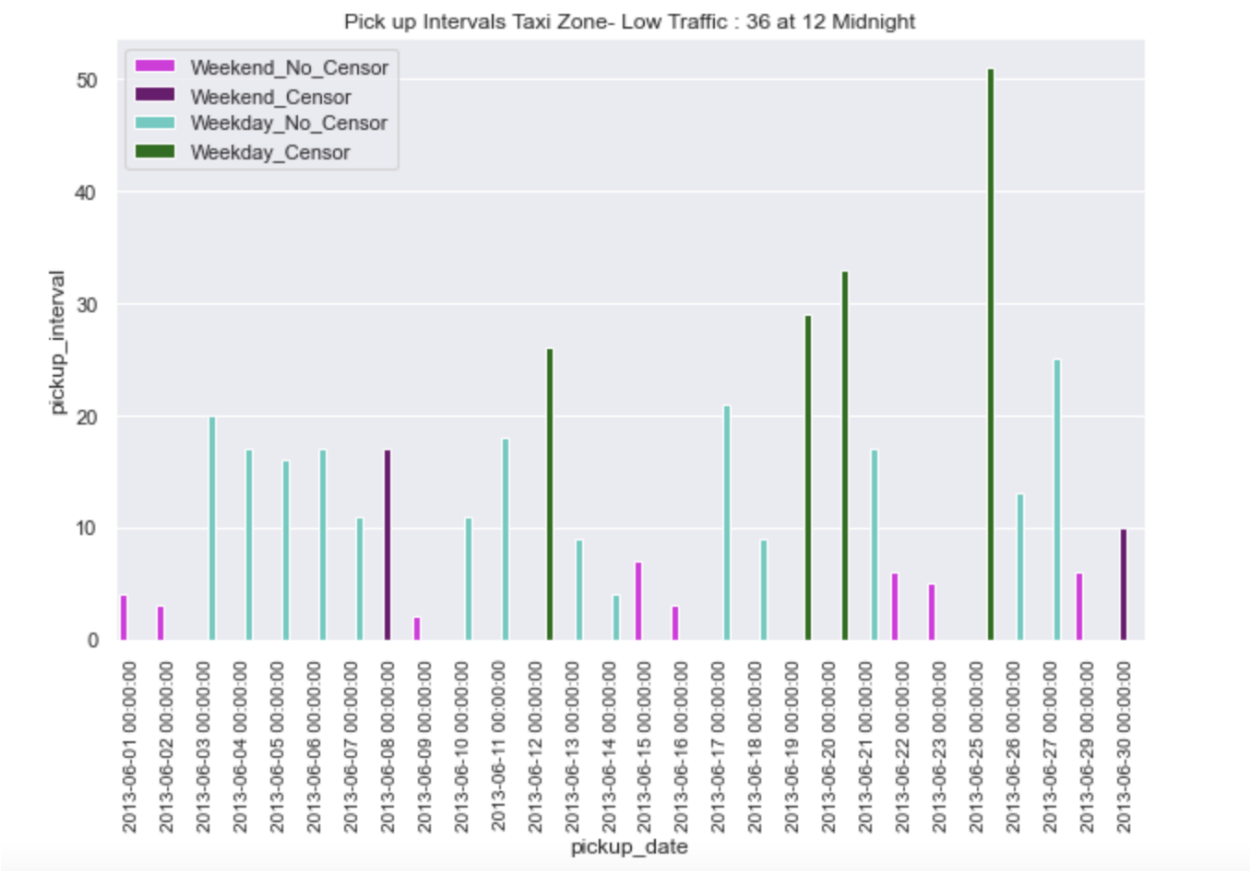



**Fig shows pickup intervals for taxi zone with low level traffic, notice how the censoring threshold for weekend and weekdays differ based on how the pickup interval ranges for those groups.**


**Figure shows pickup intervals for taxi zone with mid level traffic, notice that most dates the pick-up intervals are zero.**

One challenge we came across when estimating censor indicators based on pickup intervals is that for high traffic zones, the pickup intervals tend to be almost zero. In such cases, we used the difference in pickup and drop-off counts to estimate the censoring indicator. We used the same methodology that if the difference is pickup and drop off counts lie in extreme values, we indicate these are censored time intervals.

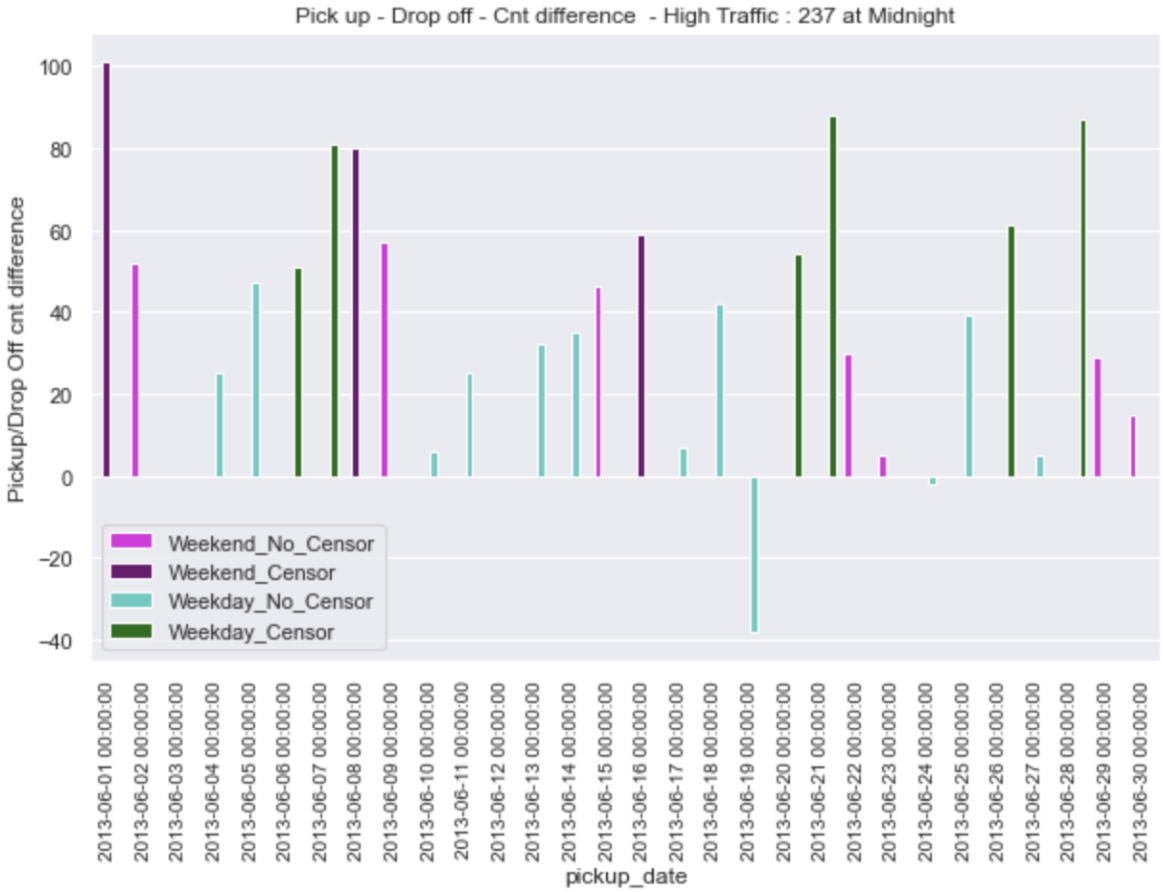



**Figure shows pickup intervals for taxi zone with high traffic levels, censoring is based on the difference between Pick and Drop off counts.**

## Optimization

We used scipy's optimize functions to calculate the maximum likelihood parameter for the censored poisson distribution.

| | taxi_zone | time_stamp | time_interval | arr_weekday | arr_weekend | avg_weekday | avg_weekend |
|---|---|---|---|---|---|---|---|
| **0** | 36 | 1 | 00:00:00 | 4.660944 | 6.817303 | 4.388889 | 6.500000 |
| **1** | 36 | 2 | 00:30:00 | 4.955894 | 14.080490 | 4.473684 | 13.500000 |
| **2** | 36 | 3 | 01:00:00 | 4.994004 | 18.597219 | 4.350000 | 17.800000 |
| **3** | 36 | 4 | 01:30:00 | 4.502053 | 16.696456 | 3.894737 | 15.900000 |
| **4** | 36 | 5 | 02:00:00 | 4.489099 | 21.521196 | 3.833333 | 19.300000 |
| **5** | 36 | 6 | 02:30:00 | 3.267948 | 19.500000 | 2.833333 | 19.500000 |
| **6** | 36 | 7 | 03:00:00 | 2.105988 | 16.900000 | 1.875000 | 16.900000 |
| **7** | 36 | 8 | 03:30:00 | 2.017914 | 15.000000 | 1.687500 | 15.000000 |
| **8** | 36 | 9 | 04:00:00 | 2.665017 | 14.200000 | 2.153846 | 14.200000 |
| **9** | 36 | 10 | 04:30:00 | 2.244193 | 11.122935 | 1.769231 | 10.400000 |
| **10** | 36 | 11 | 05:00:00 | 1.322275 | 5.000000 | 1.181818 | 5.000000 |
| **11** | 36 | 12 | 05:30:00 | 1.433553 | 3.660361 | 1.166667 | 3.100000 |
| **12** | 36 | 13 | 06:00:00 | 1.937500 | 2.000000 | 1.000000 | 2.000000 |
| **13** | 36 | 14 | 06:30:00 | 1.937500 | 1.583724 | 1.000000 | 1.333333 |
| **14** | 36 | 15 | 07:00:00 | 1.000000 | 1.554688 | 1.000000 | 1.000000 |
| **15** | 36 | 16 | 07:30:00 | 2.663965 | 1.554688 | 1.500000 | 1.000000 |

**Figure shows how the estimated demand considering censoring compares with that of average demand for a taxi zone ( 36 - low traffic region )**

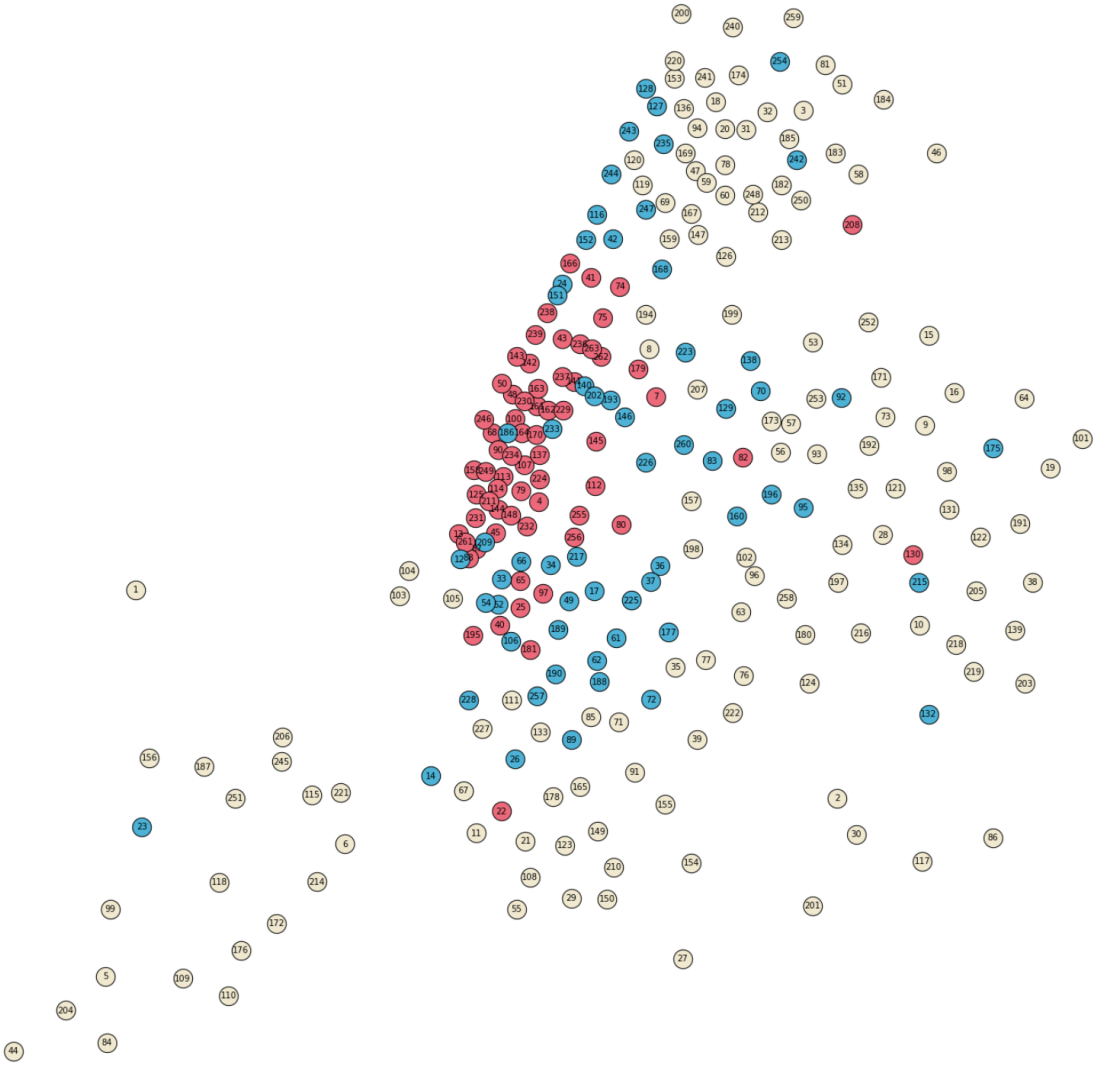


**Figure illustrates, taxi zones across around midnight on the first day of the month that displayed signs of censored demand**

Potential Additions in the Future

The team discussed and experimented with multiple potential stock-out indicators, before arriving at the chosen indicator which tested well and made broad intuitive sense. However, as the choice of the correct indicator was not clear-cut, there is the potential after the productionizing of the overall environment simulator, for this or future teams to review other iterations of a stock-out indicator. In addition, for this iteration, the priority was largely in creating a working demand estimation model. For corner cases (e.g., low traffic area zones during low traffic area times), data may be sparse and there is the potential to apply manual solutions for zones (e.g., grouping, etc.) with data sparsity problems.

## III. Supply Generator

For the environment simulator and policy evaluation, there were two approaches that were considered. One was that of having a single agent learning within a broader environment that is simulated by historical data. The second approach was that of having multiple agents learning and following a specified policy with a dynamically changing environment as a part of that. The implications behind these two approaches are also discussed in further detail in following section on Matching Probability. However, for the initial productionizing environment, it was decided to go with a single learning agent.

Within the single agent framework, the supply would be determined as follows:

- The single agent will begin either a morning or evening shift on the weekday or weekend. As shown in the figure below, nearly all taxi drivers start their shift at around the same time, at around 7 AM for weekday and around 9 AM morning shift drivers and approximately 5 PM for weekday and weekend evening shift drivers. For consistency, our single driver will begin every shift consistently at the same time -- the average starting time for drivers of that shift for the month of June and end every shift in the same manner. Given the June dataset, the average start times and end times of the drivers on the morning and evening shift can be found in Appendix C.

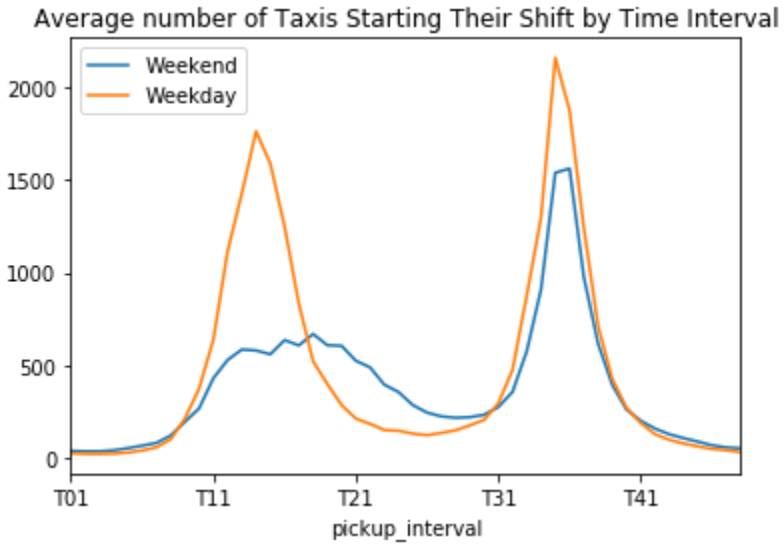



**Figure 1: Average Number of Taxis Picking Up Their First Passenger by Time Interval**

- the single agent would be assigned a starting location randomly based on the historical starting points of taxis in that assigned shift during the month of June

- the single agent would be assigned a matching probability and destination probability and move correspondingly; after drop-off the agent would take action based on the policy they were assigned

For learning, the total supply of taxis will be considered only inasmuch as they impact our single driver's matching probability. Due to computational complexity, rather than track all drivers, the average number of available taxis by zone will be taken from historical data to reflect the available supply of taxis in the relevant and neighboring zones.

## IV. Matching Probability

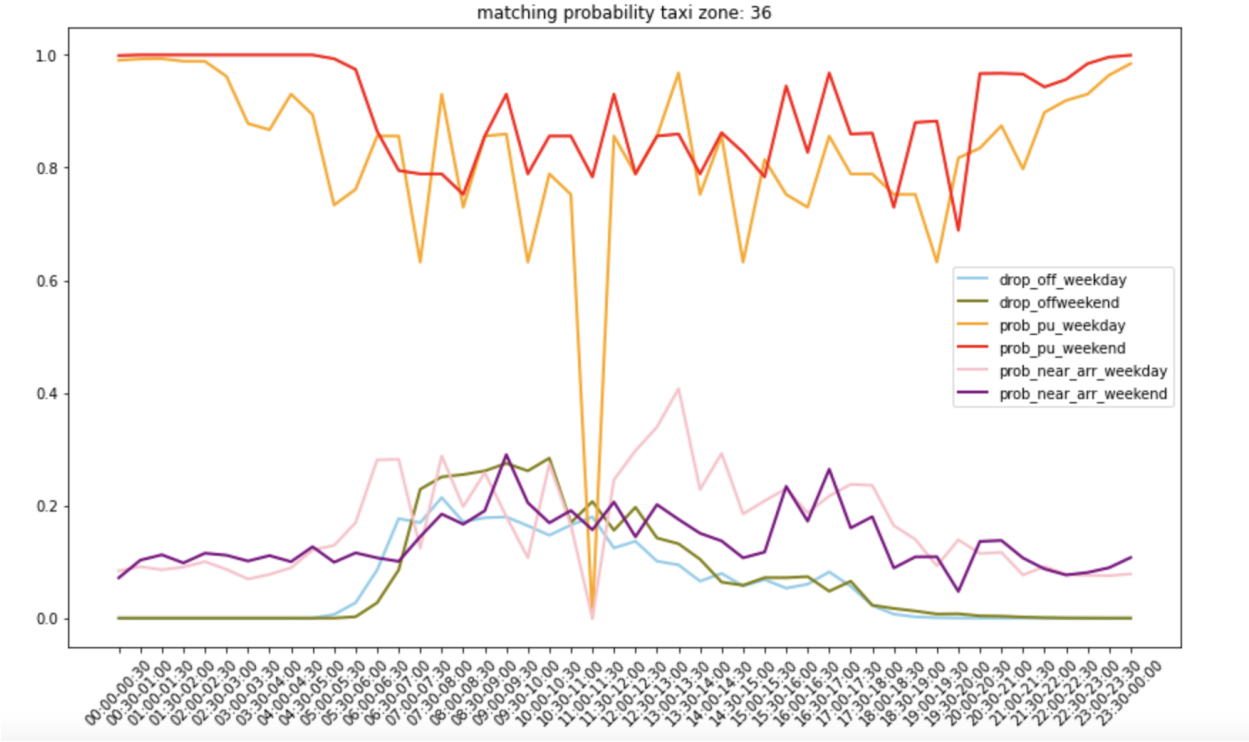
Per Yu et al (2019), one approach to the matching probability is to split up the probability of matching for a single taxi as the product of the following three probabilities. The three constituent probabilities and the way they could be adapted to our environment simulator are below.

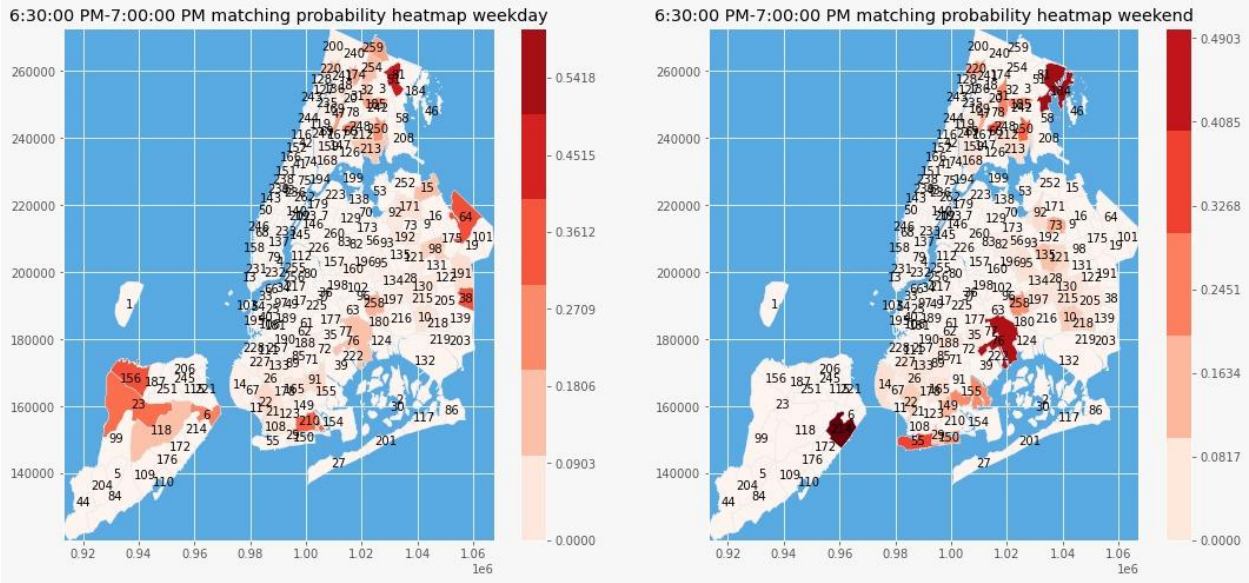
1. Probability of having at least one passenger within the pickup location
    a. The probability of at least one passenger = 1- Probability of no passengers
    b. To calculate arrival rates, we have used censor indicator to flag outliers in large pick-up intervals and for very busy zones, where the pickup-intervals are close to zero and no variance, we made use of the difference between pick up count and drop off count happening in that time interval for that time zone.
    c. These flags are specific for time interval and a taxi zone and weekday/weekend distribution analyzed separately.
    d. We use the censored indicators to estimate the censored demand / arrival rate
2. Probability that the earliest arrival during the time period happens at the pickup location
    a. Calculate $\lambda$□ which would be arrival rate at a pick up zone in $\Delta t$ ( 30 min )
    b. Calculate $\lambda_{N(j)}$ would be the arrival rate at respective adjacent nodes, use geoshapes for calculating adjacent zones
    c. The probability that an arrival from node h is earlier than all other nodes is

        $\lambda$□ / $\lambda_{N(j)}$

    d. Since $\lambda_{N(j)}$ would be fixed for a given time interval, this probability matching calculation would ensure to drive the policy to choose a pickup zone with a higher number of arrival rates.

3. Probability that the taxi in question is the nearest vacant taxi to the pickup location
    a. First calculate $\lambda_d$, the number of drop-offs around the taxi zone for that time interval $\Delta t$ ( 30 min )
    b. Use the adjacent zone calculator using geoshapes.
    c. Calculate $P_d$ being the probability of nearest taxi to a pick up zone, by calculating the probability of no taxis in that zone and its adjacent taxi zone.
    d. Let $\lambda_i \ldots \lambda_n$ be the drop off counts for adjacent zones i…,n, the probability of no taxis in that zone and its adjacent taxi zone for the next 10 minutes is calculated as follows:
        i. $P_d [X(S)=0] = \exp\{ -(\lambda_d+\lambda_i + \ldots+ \lambda_n)/\text{num of adj zones}\}$






**Figures above show the heat map of probability match distribution of different taxi zones at the time intervals mentioned.**

In addition, we also considered the alternative approach to the matching problem in which matching would occur globally among all available drivers rather than at the individual driver level. In this alternative example, the simulation would roll from time period to time period. Within each time period, the total number of available drivers would be tabulated. A transition matrix would then be applied to the available drivers based on historical driver behavior (where taxi's transitioned to after drop-off based on their next pickup). Then in the next time period, their probability of matching with a customer would be calculated based on a randomly generated demand number

from the demand generator and the available taxis in the zone after transitioning post-pickup. Please see Appendix C: Global Approach to Probability Matching, for a broader discussion of the assumptions and a toy example.

A comparison of the two approaches are summarized below:

| Driver Probability Matching | Global Probability Matching |
|---|---|
| • Emphasizes one driver's actions and assumes the environment is stationary | • Reviews the global landscape and allows for changing environment |
| • Can leverage matching probability from historical data | • Need to compute matching probability at each time interval |

As the Global Probability Matching approach could allow for more dynamic evolution of the environment, but is more computationally intensive, we will be pursuing a single agent approach for the first production environment.

## V. Destination Probability Generator

For the destination probability generator, we began with identifying where customers demanded to be driven to based on their pick-up zone from historical data.

As the base case within the simulator for a zone and time without demand censoring, if a driver picks up a passenger in a certain zone, the driver will be randomly assigned a destination zone based on historic demand from within that zone.

However, with censored demand, there is the added challenge that the destinations of censored demand is not available within the historical data. For these zones and times, a modified transition matrix will be used that takes into account the historical destinations of passengers in not just that zone, but also neighboring zones.

## VI. Next Steps

With the major components of the environment simulator built, the next step is to pull together elements of the environment simulator in an initial Policy Evaluation process. From there, we aim to begin a Generalized Policy Iteration process or Monte Carlo Rollout for Policy Improvement.

# Appendix

## Appendix A: References

[1] Xinlian Yu, Song Gao, Xianbiao Hu, Hyoshin Park, “A Markov decision process approach to vacant taxi routing with e-hailing”, Transportation Research Part B: Methodological, Volume 121, 2019, Pages 114-134, ISSN 0191-2615, https://doi.org/10.1016/j.trb.2018.12.013.

[2] Mersereau, Adam J. "Demand estimation from censored observations with inventory record inaccuracy." Manufacturing & Service Operations Management, vol. 17, no. 3, 2015, p. 335, ISSN: 1523-4614, https://doi-org.ezproxy.cul.columbia.edu/10.1287/msom.2015.0520

## Appendix B: Contributions

- Danyang Han: Main contributor on data transformation, processing data for arrival rates by time zone, testing proposed models for Poisson demand estimation, data visualization, and neighboring zones
- Anita Pinto: Main contributor on cleaning and preprocessing data, data visualization, formulating and testing models for probability matching calculation, inter-pickup time arrivals and lost demand estimation, and demand optimization.
- Elizabeth Yum: Main contributor on organizing materials for meetings, setting up logistics, managing progress, probability matching, transition matrices, collating and documentation of the results and reports.

## Appendix C: Average Start Times and End Times for Shifts in June 2013

| | Weekday | Weekend |
|---|---|---|
| **Morning Shift** | **7:03:18 start - 15:43:05 end** | **8:34:30 AM start - 16:24:12 end** |
| **Evening Shift** | **17:16:57 start - 06:19:23 end** | **17:29:32 start - 05:29:05 end** |

# Appendix D: Global Approach to Probability Matching

A simplistic description of the key assumptions and toy example in a two zone word (with Zone 3 and Zone 4) follows below.

## Assumptions:

- Each segment of 24 hours is broken down into blocks of time (not necessarily the same 30 minute blocks)
- Within each block of time, taxi drivers can either (1) drop-off a passenger, (2), pick-up a passenger, or (3) not be matched to a passenger. Taxi drivers may not both drop-off and pick-up a passenger within the same block of time.
- We assume that where taxis typically go after drop-off is not sensitive to the time of day/day of week (e.g., if a taxi drops off at zone 4, it has a 90% chance of staying in zone 4, and 10% chance of going to zone 3) for the entire day
- Taxis can migrate from one zone to another within the time period even after drop-off

## Inputs (and examples for a 2-zone world of Zone 3 and 4):

- Selecting a block of time to begin. Example: for illustration of how it would work at the *middle* of the day
    - Denoting Δt as our initial time slot (e.g., 8:0:00-8:09:59 AM if in 10 minute increments)
    - Subsequent time periods denoted as Δt+1 (8:10:00-0:19:59 AM), Δt+2 (8:20:00-0:29:59 AM), etc.
- Beginning # and location of "available" taxis (taxis that have dropped off, started their shift, or did not pick up a passenger in the prior shift and stayed in the area). Example:
    - Zone 3: At Δt, 20 taxis dropped off a passenger
    - Zone 4: At Δt, 30 taxis dropped off a passenger
- Adjusted demand parameters for each zone. Example:
    - ƛ_{zone 3, Δt+1} = 20
    - ƛ_{zone 4, Δt+1} = 40
- Transition probability matrix of ***after-drop-off***

| | **Zone 3** | **Zone 4** |
|---|---|---|
| **Zone 3** | **.8** | **.2** |
| **Zone 4** | **.1** | **.9** |

    - **Example available at: https://github.com/eyy4/Spr-2021-DiDi-RL/blob/master/notebooks/EDA_matching_probability.ipynb**

Out[35]:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.001487 | 0.000000 | 0.000000 | 0.000929 | 0.0 | 0.000000 | 0.000248 | 0.000000 | 0.000000 | 0.000062 | 0.0 | 0.000619 | 0.055380 | 0.000062 | 0.000000 | 0.0 | 0.000000 | 0 |
| 2 | 0.000000 | 0.571429 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0 |
| 3 | 0.000000 | 0.000000 | 0.020638 | 0.001876 | 0.0 | 0.000000 | 0.001876 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.001876 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0 |
| 4 | 0.000000 | 0.000000 | 0.000000 | 0.299248 | 0.0 | 0.000000 | 0.000209 | 0.000000 | 0.000000 | 0.000015 | 0.0 | 0.000060 | 0.001331 | 0.000000 | 0.000000 | 0.0 | 0.000060 | 0 |
| 5 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0 |
| 6 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.055556 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.021368 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0 |
| 7 | 0.000016 | 0.000000 | 0.000000 | 0.001458 | 0.0 | 0.000000 | 0.204451 | 0.000047 | 0.000000 | 0.000157 | 0.0 | 0.000016 | 0.000345 | 0.000031 | 0.000000 | 0.0 | 0.000063 | 0 |
| 8 | 0.000000 | 0.000000 | 0.000000 | 0.002439 | 0.0 | 0.000000 | 0.063415 | 0.051220 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0 |
| 9 | 0.000000 | 0.000000 | 0.000000 | 0.007788 | 0.0 | 0.000000 | 0.014019 | 0.000000 | 0.012461 | 0.001558 | 0.0 | 0.000000 | 0.001558 | 0.000000 | 0.003115 | 0.0 | 0.000000 | 0 |
| 10 | 0.000000 | 0.000000 | 0.000000 | 0.000339 | 0.0 | 0.000000 | 0.003051 | 0.000000 | 0.000000 | 0.049153 | 0.0 | 0.000000 | 0.000678 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0 |

- Transition probability matrix of ***after-pick-up*** - Example:

| | **Zone 3** | **Zone 4** |
|---|---|---|
| **Zone 3** | **.1** | **.9** |
| **Zone 4** | **1** | **0** |

## Sample Approach:

1. **Within Δt, estimate the total # available taxis in each zone**, i.e., # dropped off, started their shift, or did not pick up a passenger in the prior shift and stayed in the area
    a. Zone 3: 20 available taxis (Assumption)
    b. Zone 4: 30 available taxis (Assumption)
2. **Apply the transition matrix *after-drop-off* to determine total supply at the beginning of Δt+1**
    a. Zone 3: **19** = 20 taxis in zone 3 at time Δt * 0.8 + 30 taxis in zone 4 at time Δt * 0.1
    b. Zone 4: **31** = 20 taxis in zone 3 at time Δt * 0.2 + 30 taxis in zone 4 at time Δt * 0.9
3. **Randomly generate demand for Δt+1 based on ƛ_{zone 3, Δt+1}, ƛ_{zone 4, Δt+1}**
    a. Zone 3: poissrnd(ƛ_{zone 3, Δt+1}) = poissrnd(20) = 18
    b. Zone 4: poissrnd(ƛ_{zone 4, Δt+1}) = poissrnd(40) = 38
4. **Calculate P(matching) based on demand and supply for the period**
    a. Zone 3: min(# demand/# supply, 1) = 18/19
    b. Zone 4: 1
5. **Apply transition probability matrix of *after-pick-up* and repeat from step 1**
    a. During Δt+1:
        i. Zone 3:
            1. 18 taxis taken by customers (min(supply,demand)) -> apply after-pickup transition matrix and 16 are taken to zone 4, and 2 are taken elsewhere to zone 3
            2. 1 taxi is empty -> apply after drop-off transition matrix, so 20% chance of migrating to zone 4
        ii. Zone 4:

    1. 31 taxis are taken by customers(min(supply,demand)) -> apply after-pickup transition matrix and all 31 are taken to zone 3
6. **Repeat from step 1**...