PHOTO: DARCY PETERKA

Off-Policy Evaluation for Taxi Driver Repositioning Policies

Team: Danyang Han, Anita Pinto, Elizabeth Yum
Mentors : Vineet Goyal, Zhiwei Qin

Introduction and Objective

The Problems

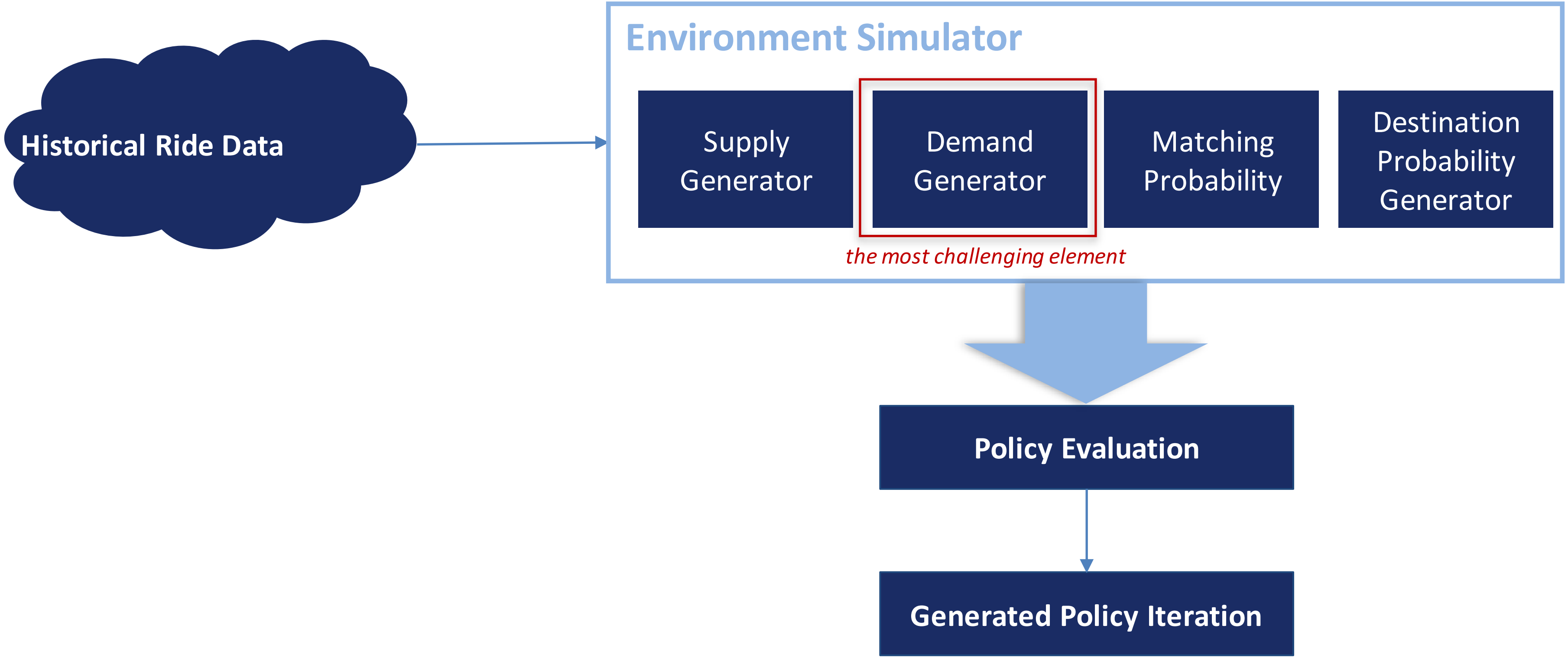
In large, highly competitive taxi markets, taxi drivers' incomes are highly dependent on their ability to be able to find their next passenger quickly and can vary widely depending on their working behaviors and preferences – when and where to start a shift and their repositioning behavior between rides

Many studies have been conducted on pattern recognition and reinforcement learning to minimize driver idle time and maximize driver income, but most efforts have been limited to evaluation of historical data

Objective

Tackle the problem of driver repositioning on ride-hailing platforms through off-policy evaluation by the creation of an environment simulator that leverages historical data to stochastically generate demand, passenger/driver matching and destination probability to be used in off-policy evaluation

Our Methodology



Effectively Modeling Real World Behavior

Source: New York City, Taxi & Limousine Company (TLC) Trip Record Data, June 2013, Yellow (Medallion) Taxis

Discretizing the State Space

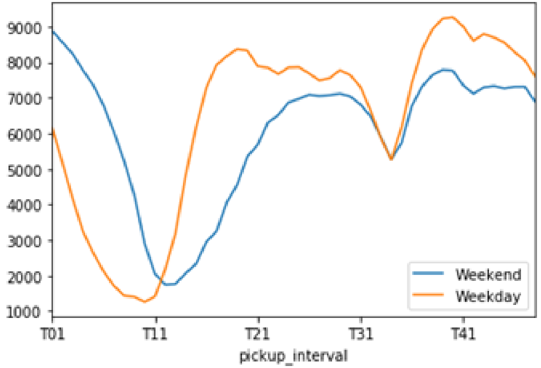
- Focused solely on transformed taxi data that tagged pickups and drop-offs to the 263 NYC Taxi Zones (see map below)
- For certain computations, discretized time into 48 30-minute intervals throughout the day

Map of All 263 NYC Taxi Zones

Effectively Modeling Real World Behavior

- Historical observed demand and driver behavior exhibited differing patterns depending on the day of week, which necessitated further splitting the state space and computations by weekend and weekday
- The majority of NYC TLC drivers work either one of two shifts – a morning shift or an evening shift of approximately 11 hours

Average Number of Active Taxis by Time Interval



Source: NYC Open Data

Supply Generation

Two Approaches Were Initially Considered…

Modeling a Single Driver's Behavior

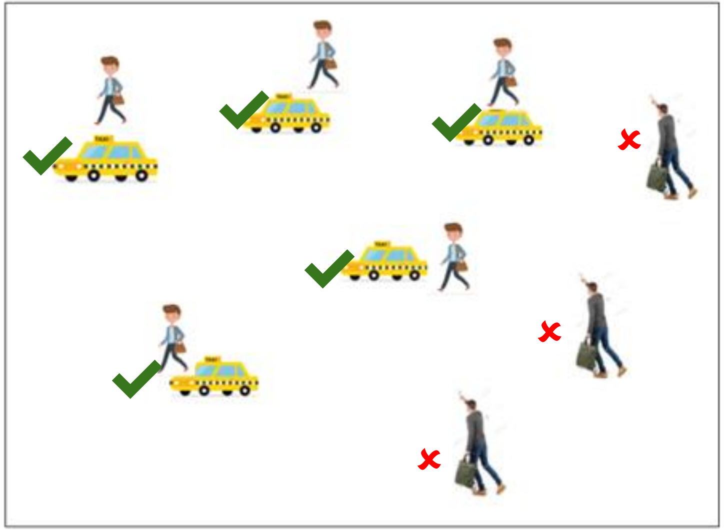
- A single agent learning within a broader environment that is simulated by historical data
- Pros: Computationally simple
- Cons: Assumption that only a single agent is learning

Our Chosen Approach for This Iteration

A Multi-Agent Model

- Multiple agents learning and following a specified policy
- Pros: The ability to dynamically change the environment as multiple agents learn at once
- Cons: Computational complexity

Arrival Rate → Demand Generation



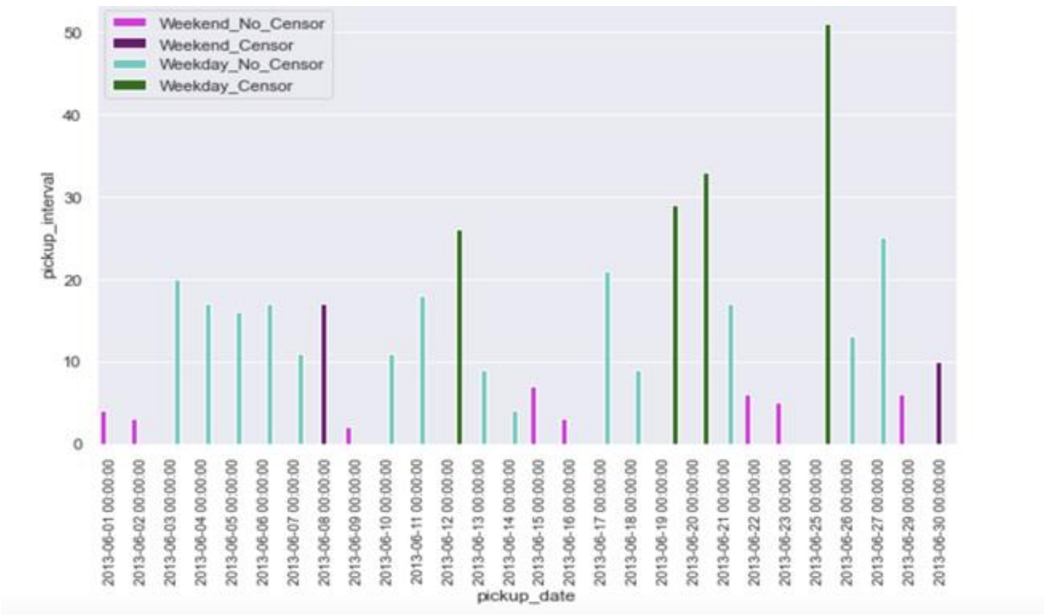
Calculating the uncensored demand rate

- Historical trips only gives us information on successful rides, i.e the observed demand.
- How do we estimate the lost demand ? The passengers who arrive at a pick up zone and never get a ride ?

Arrival Rate - Censoring Indicator

- 24 hours into 30 min time intervals i.e 48 time intervals per day
- **Passenger Arrival Rate** would be Average passengers arriving at the zone per time interval → Observed Demand
- **Few considerations:**
    - Price surge?
    - Inter pick up intervals - Time difference between 2 pick up intervals → longer intervals indicate longer wait for taxis and possible lost demand.
    - Difference in Pickup - Drop

Pick up Intervals Zone 36 (Brooklyn Bushwick North) at 12 Midnight



Calculation of Uncensored Demand

Generally, arrival rate, the λ 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 (δ_1, δ_2, .. δ_{30}) Where $\delta = 0$ means no censoring, δ=1 means censoring present

Our goal is to find the parameter λ given pairs of data (x_1 δ_1) … (x_{30},δ_{30})

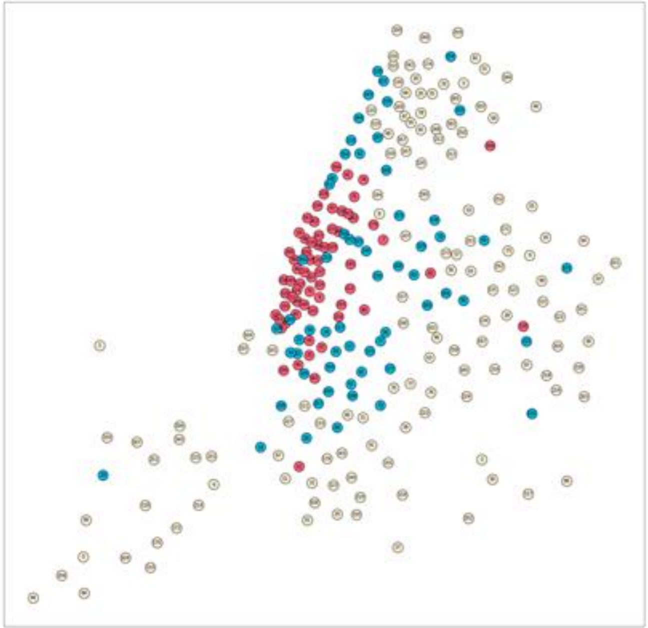
$$\lambda_{ML} = argmax(\sum_{i=1}^{n} (1-\delta_i) \, log(f(x_i,\lambda) + \delta_i log(1-F(x_i,\lambda) \,)$$

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

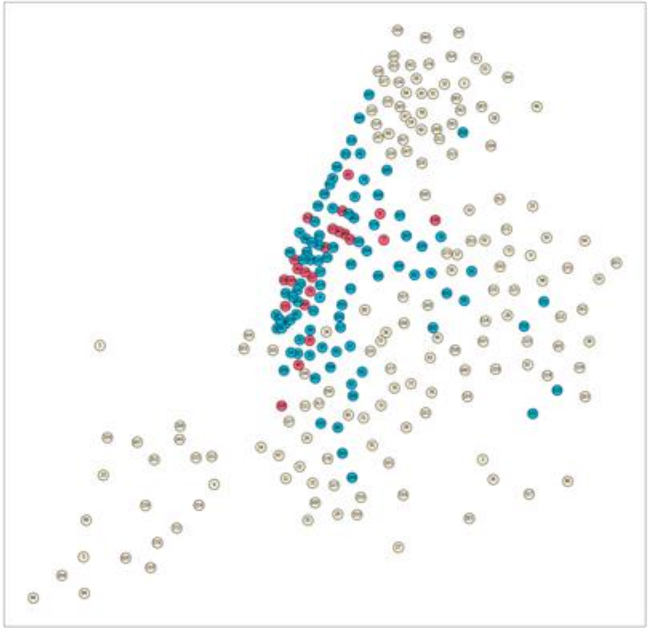
$$F(x_i,\lambda) = \sum_{i=1}^{n} \frac{e^{-\lambda} * \lambda^{-x_i}}{x_i!}$$

Censoring Indicator at Different Times

Censoring Indicator by Zone During Rush Hour



Censoring Indicator by Zone During Off-Peak

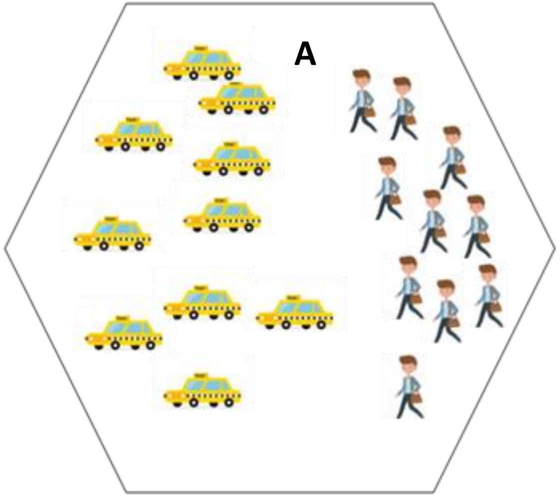


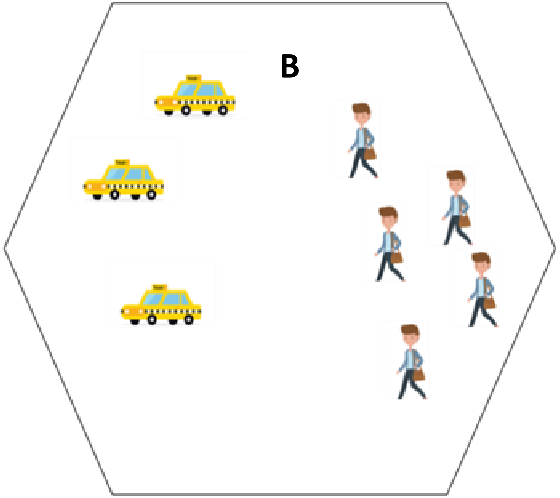
Red → Arrivals with Censoring Present, Blue → Arrivals with no censoring

Matching Probability? What does this mean?

How do we evaluate a policy?
Does the repositioning decision matter?
Do you always go to a zone with a higher arrival rate?
Does this depend only on arrival rate?

Matching Probability

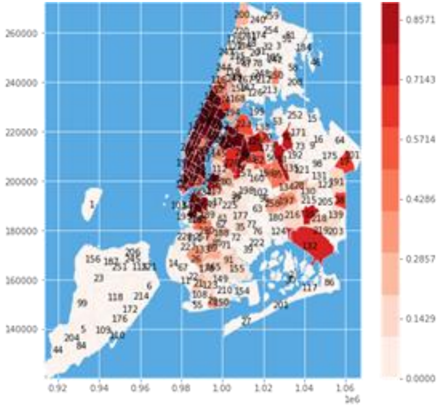




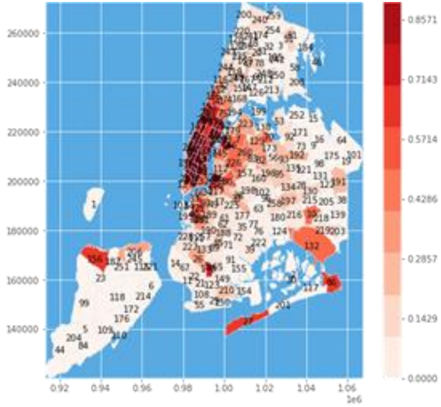
Matching Probability depends on the following product of probabilities.
Prob of Arrival Rate → 1- Prob of no arrivals (Poisson distribution)
Prob of no vacant taxi → Prob of no drop off in that zone in that time interval (Poisson distribution)
Prob of a zone to nearest arrival → using softmax on arrival rates difference from neighboring zones

Matching Probability

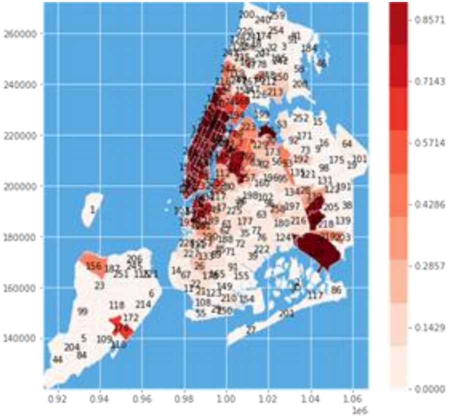
5:00:00 AM - 5:30:00 AM matching probability heatmap weekday



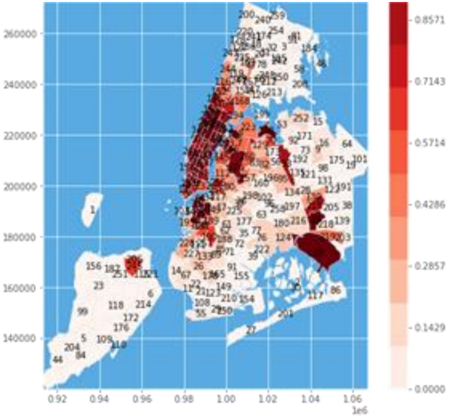
5:00:00 AM - 5:30:00 AM matching probability heatmap weekend



6:00:00 PM - 6:30:00 PM matching probability heatmap weekday



6:00:00 PM - 6:30:00 PM matching probability heatmap weekend



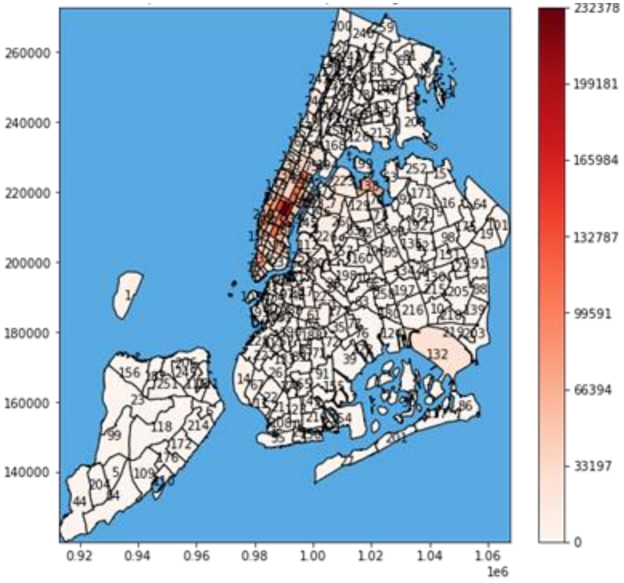
Columbia University Data Science Institute

Destination Probability

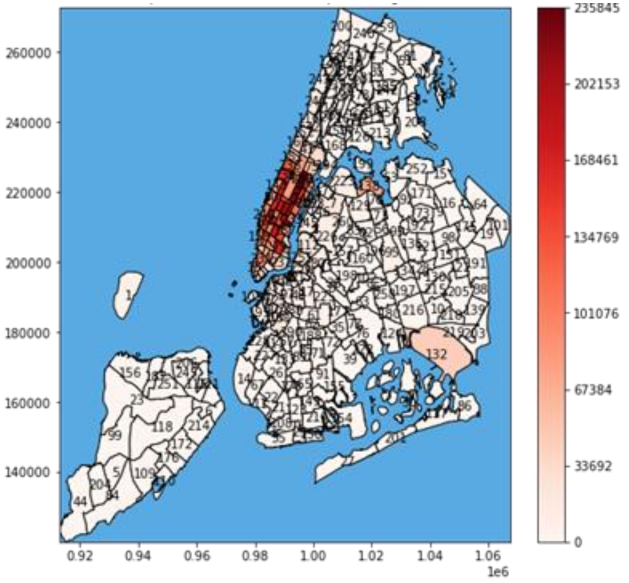
Estimating Destination Probability

- Reviewed historical data by shift and weekend/weekday to create a Markov matrix of historical rider demand based on their point of origin
- However, in instances where demand was censored, historical destinations observed may not accurately reflect total destinations demanded - plans to extend this further

Top destinations - heatmap morning shift



Top destinations - heatmap evening shift



Policy Evaluation

Hot Spot Tracing (Deterministic)

- Always reposition to the adjacent zone with the highest historical demand

Hot Spot Tracing (Probabilistic)

- Use softmax to normalize demands of adjacent zones into a probability distribution
- Reposition according to probability distribution

Random Reposition Policy :

- Reposition randomly to an adjacent zone.

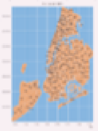
COLUMBIA UNIVERSITY
Data Science Institute

Simulator and Environment

Simulator

Taxi

- Current Time
- Current Location
- Total Fare
- Total Cost

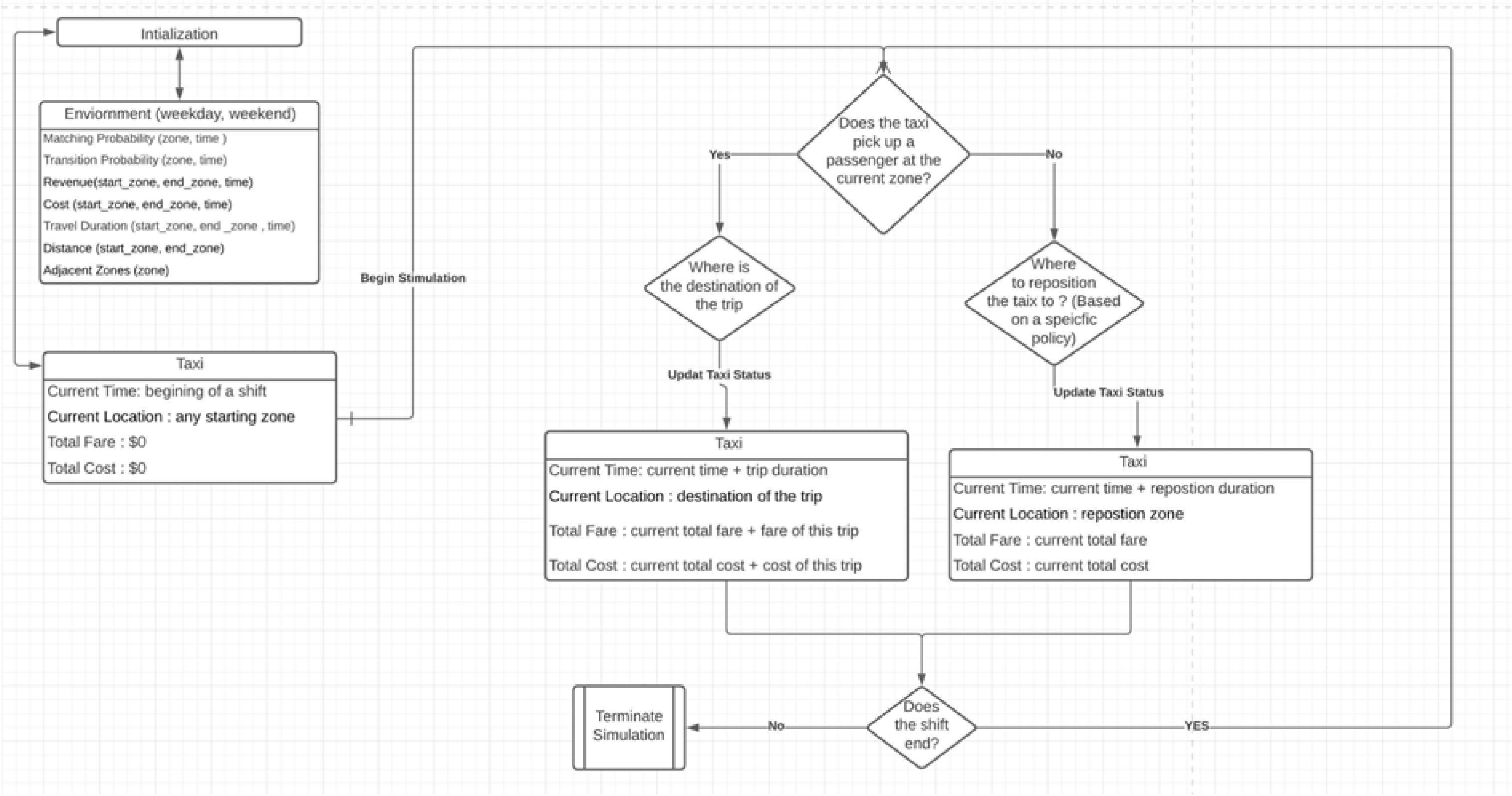


Enviornment (weekday, weekend)

- Matching Probability (zone, time)
- Transition Probability (zone, time)
- Revenue(start_zone, end_zone, time)
- Cost (start_zone, end_zone, time)
- Travel Duration (start_zone, end _zone , time)
- Distance (start_zone, end_zone)
- Adjacent Zones (zone)

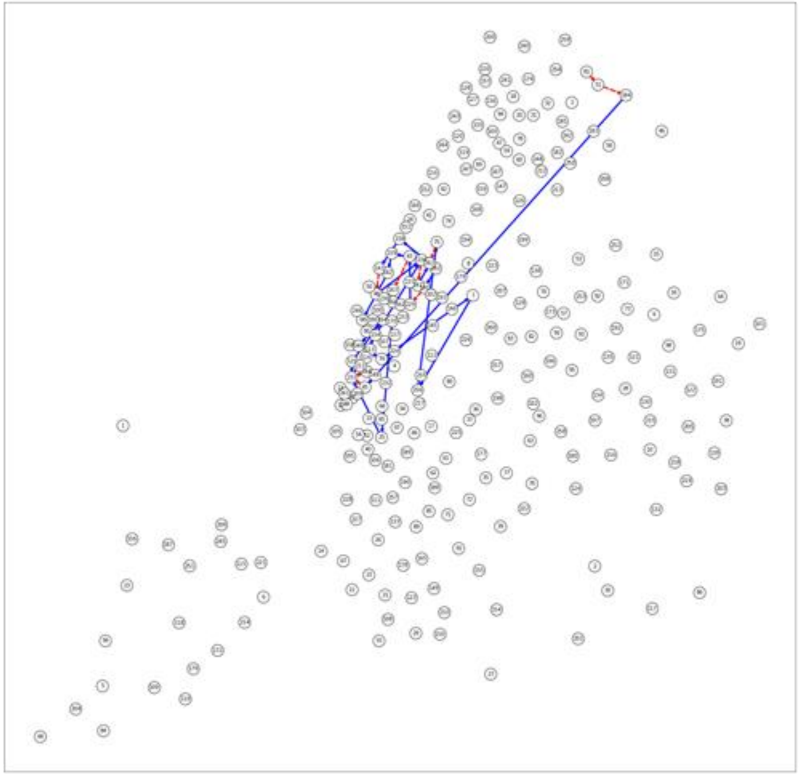
COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Simulator and Environment

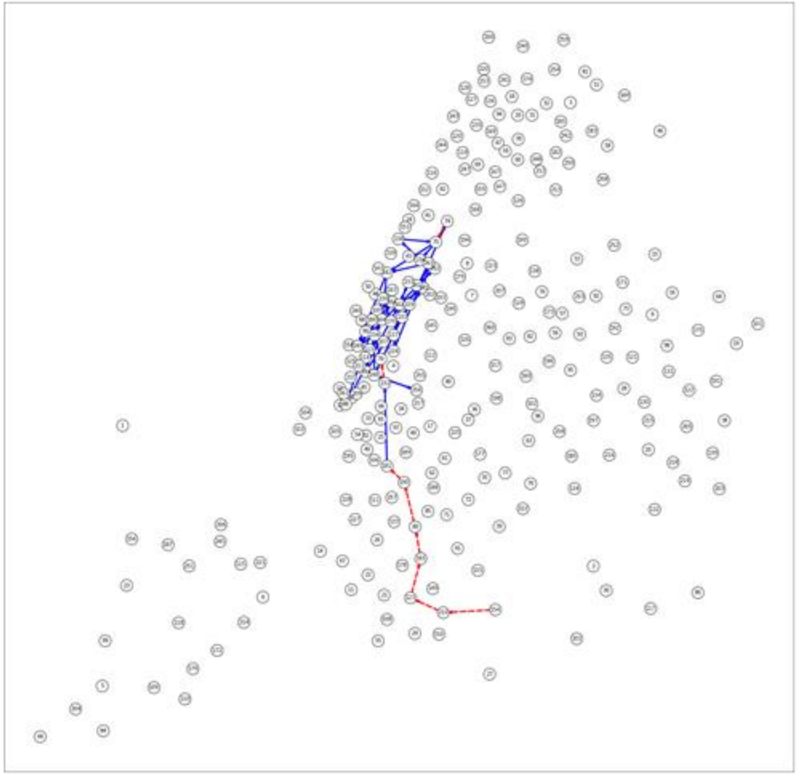


Driver Trajectory Examples in the simulator

Driver Trajectory starting point from Zone: 81 on Weekend Shift B



Driver Trajectory starting point from Zone: 154 on Weekday Shift A



Policy Evaluation - Monte Carlo Based GPI

State Space -> 263 (Spatial Taxi Zones) * 48 Time Intervals (Temporal)

Initialize:

v(s) = v(z,t) = state value functions

Returns(s) -> empty list for all returns

N(s) -> 0 -> State Hits

Repeat forever:

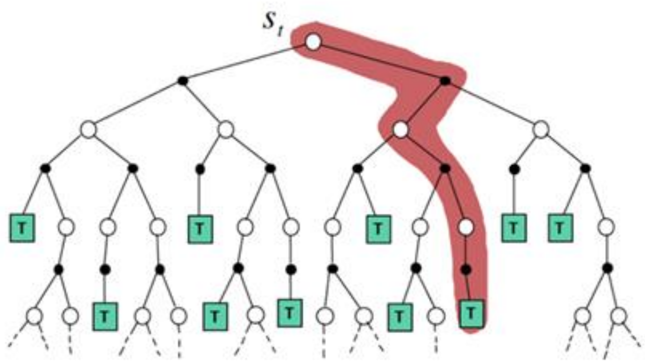
Generate an episode using policy π

Calculate G(s) -> R(t+1)+R(t+2) + …. R(T)

Append G(s) to Returns(s)

N(s) = N(s) + 1

V(s) = V(s) + 1/N(s) * (G(s) - V(s)) (Incremental Mean)



MC backup

[Silver, lecture]

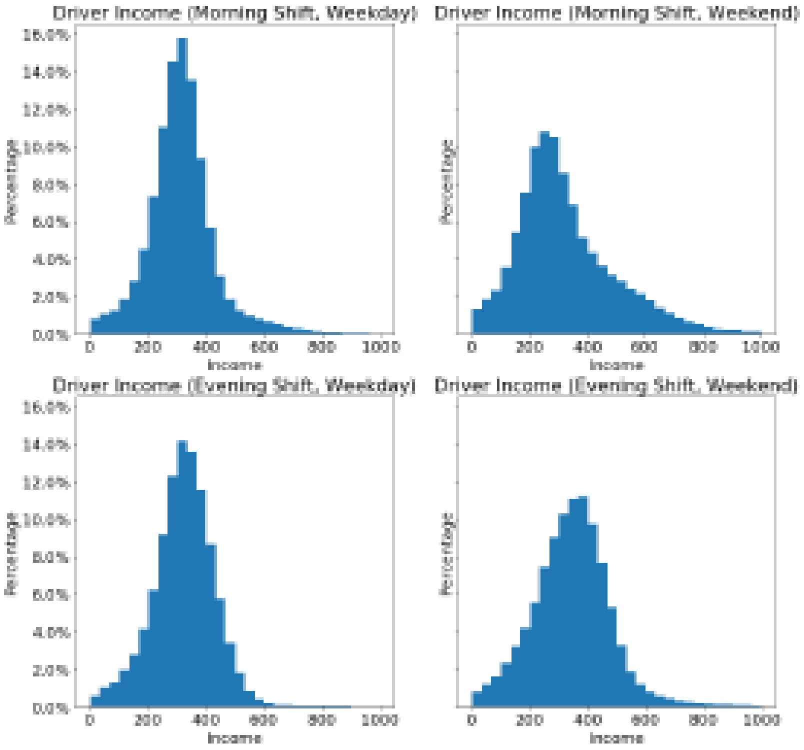
Simulation Stats: 12 types

3 Policies were evaluated *** Weekday/Weekend * Shift A / Shift B (3*2*2 = 12)**

Type1: Run randomly each simulation for 6000 iterations

Type2: Brute force, run each zone 40x times iteratively ~= 263*40 = 10520 *12 ~= **120K runs**

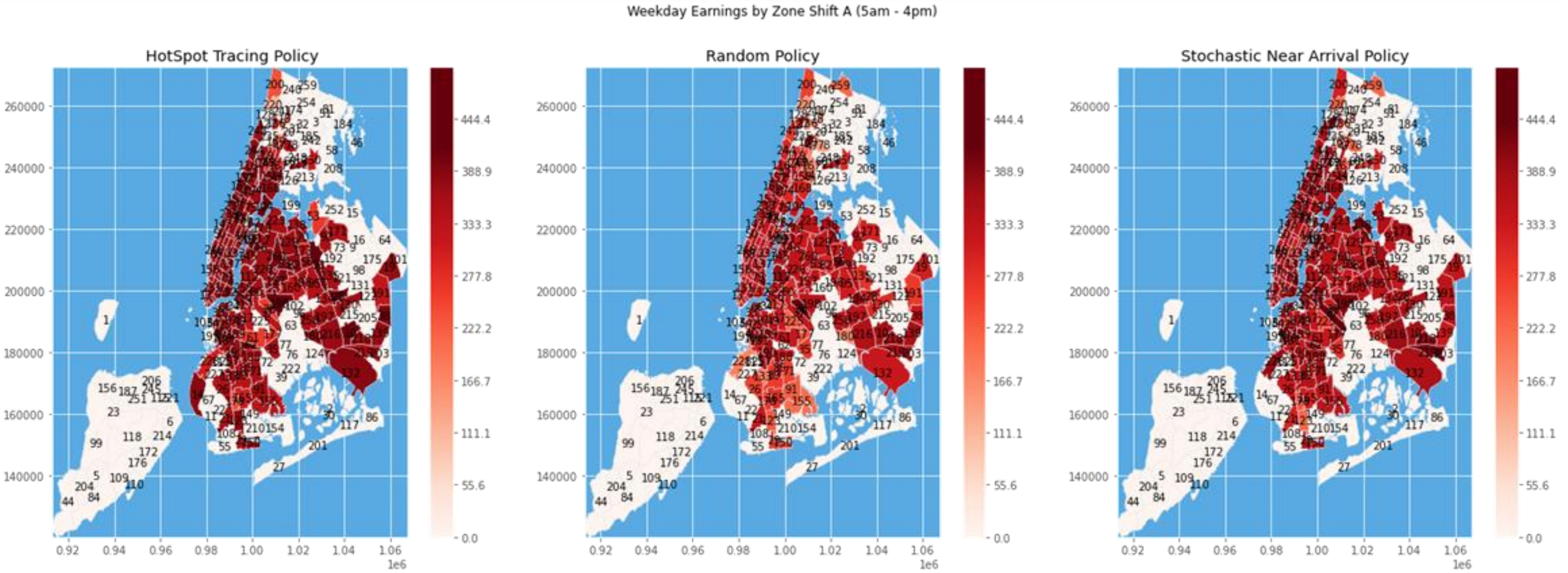
Simulation Runs - Earnings vs Historical Earnings



Earnings Histogram from Historical Trip Data

Earnings Histogram from Simulator Runs

Off-Policy Evaluation Results



Final Notes

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

&

A word of thanks…

Q&A