# Probabilistic Allocation in Urban Bike-Sharing Systems

Andrej Basica, Roshen Chatwal, Ejike Ike, Marko Isakovic

# Introduction

### Problem

If you've ever tried to rent a Blue Bike, you might notice some stations are completely empty or full

This problem stems from **suboptimal** bike allocations across stations during the day, failing to adapt to rider demand patterns.

### Approach & Goal

Our goal is to develop a reliable **model** that estimates the minimum number of bikes needed at each station to ensure rider satisfaction.

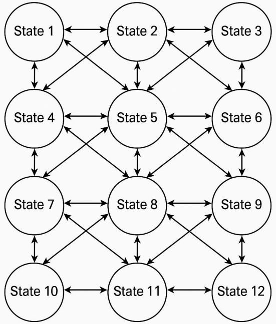
We used a combination of **simulations** and **analytical modeling** to predict station net changes and bike needs over the day.

### Data

We collected **real-world trip data** from the BlueBikes system, focusing on **12 Harvard-area stations**.

This dataset enabled us to track trip patterns, compute transition probabilities, and validate our bike allocation models.

# Model Setup




Each station is modeled as a state in a **Markov chain**

A trip from one station to another is a **state transition**, governed by **transition probabilities**

$$Q = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{12,1} & p_{12,2} & \cdots & p_{12,12} \end{bmatrix}$$

***pij*** represents the probability that a bike leaving station ***i*** will go to station ***j***

We got **P** using historical BlueBikes trip data, filtering only trips between the 12 Harvard stations

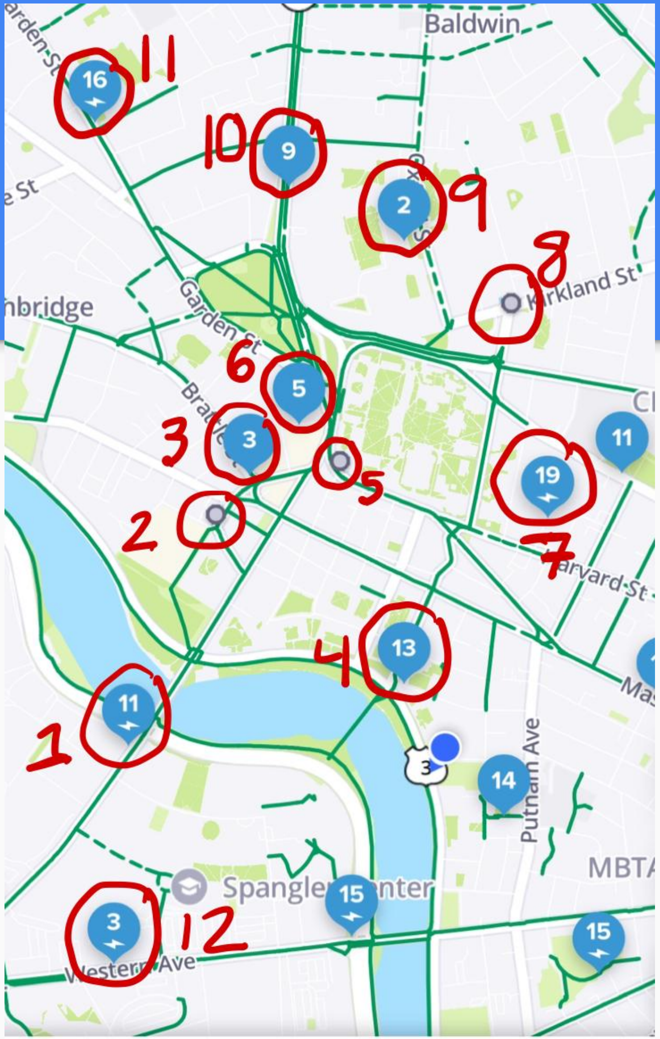
Each row was normalized so that the probabilities from one station sum to 1

Simulated bike flows 100 times using our probability model to capture daily station activity

By smartly allocating bikes to **minimize shortages**, we boost rental **revenue** and create a more reliable BlueBikes system for users.

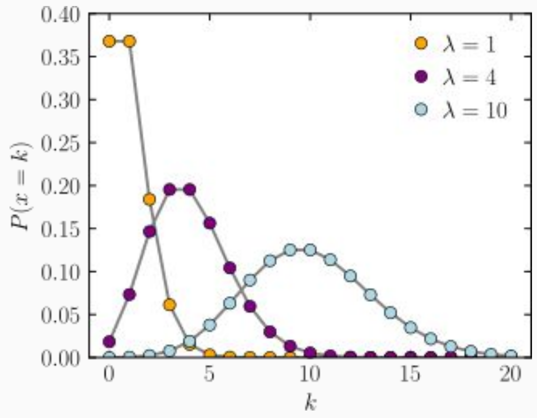
# 12 Selected Stations

1. Harvard Stadium: N. Harvard St at Soldiers Field Rd
2. Harvard Kennedy School at Bennett St / Eliot St
3. Harvard Square at Brattle St / Eliot St
4. Harvard University River Houses at DeWolfe St / Cowperthwaite St
5. Harvard Square at Mass Ave/ Dunster
6. Church St
7. Verizon Innovation Hub 10 Ware Street
8. Harvard University Gund Hall at Quincy St / Kirkland St
9. Harvard University / SEAS Cruft-Pierce Halls at 29 Oxford St (this one is right next to class)
10. Harvard Law School at Mass Ave / Jarvis St
11. Harvard University Radcliffe Quadrangle at Shepard St / Garden St
12. Innovation Lab - 125 Western Ave at Batten Way



# Bike Demand: A Poisson Process

- **Observations:** Bike rentals are rare events with large potential crowds — ideal conditions for a Poisson model
- **Timestep:** Chose $\Delta t$ = 10 minutes, matching the average ride duration between Harvard stations
- **Estimation:** Calculated each station's $\lambda_1, \lambda_2, \ldots, \lambda_{12}$ by counting April rides and dividing by 4320 timesteps in the month
- **Modeling:** Assume $A_i \sim \mathrm{Pois}(\lambda_i)$ to simulate the number of renters arriving at each station per timestep



# Net Change in Bikes: A Markov Simulation

**Algorithm 1** Bike Sharing Simulation Pseudocode (100 iterations of the below)

Initialize array $s0_{1\times 12}$ where $s[i]$ is the starting number of bikes at station $i$
Initialize array $x_{1\times 12}$ of zeros, where $x[i]$ is the # of bikes inflowing into station $i$ by the end of the current timestep
Initialize arrays $Z_{144\times 12}$ of zeros to track net changes in the state variable, successful rentals, and failed rentals at each station over 24 hours
**for** each timestep $\Delta t$ in 24 hours **do**
    **for** each station $i$ **do**
        Randomly draw the number of potential customers arriving at station $i$ over the interval using the distribution of $A_i$. Store this as $n_{t_i}$.
        Let $r_{t_i} \leftarrow \min(n_{t_i}, s_{t-\Delta t}[i])$ ▷ actual rentals possible
        Update: $s_t[i] \leftarrow s_{t-\Delta t}[i] - r_{t_i}$
        Track the net change in the # of bikes, the # of successful rentals ($r_{t_i}$), and the # of failed rentals ($n_{t_i} - r_{t_i}$) over the interval
        Map successful customers to destination stations using transition matrix row $Q[i]$, where destination counts are drawn from $\text{Mult}_{12}(r_{t_i}, \vec{p})$ with $\vec{p} = Q[i]$
        Update $x$ to reflect destination inflows of bikes from station $i$
    **end for**
    Update state: $s_t \leftarrow s_t + x$
    Reset $x$ to all zeros
**end for**

# Revenue Bridge

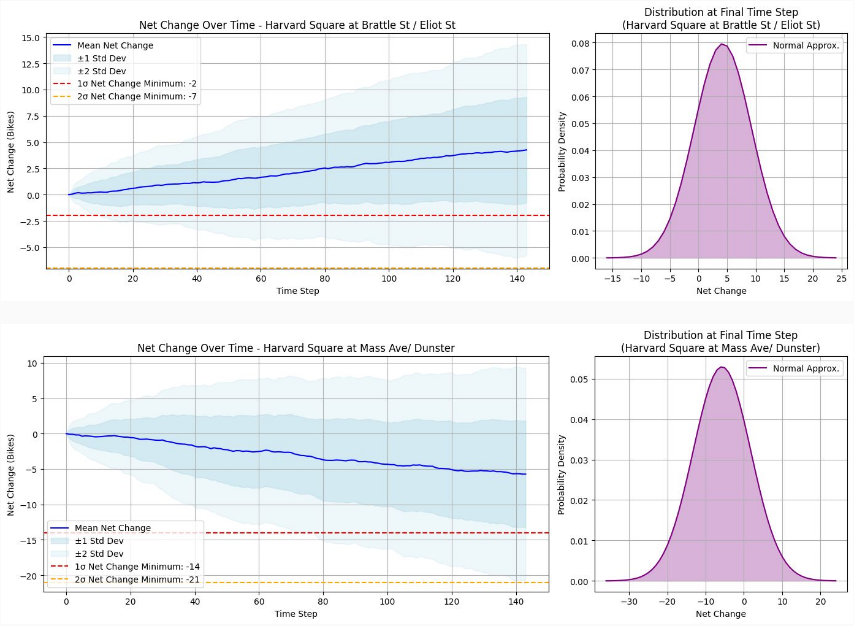
- Set rental price at **$5.50 per 10-minute ride** to translate rides to revenue
- Track rides with matrices: Zs for successful rentals and Zf for missed rentals
- Calculate Daily Revenue Earned and Daily Revenue Missed using these matrices
- Find the optimal initial bike allocation that maximizes earned revenue, minimizes missed revenue, and uses the fewest total bikes

$$\text{Daily Revenue Earned} = p \cdot \sum_{i=1}^{144} \sum_{j=1}^{12} Z_{s_{ij}}$$

$$\text{Daily Revenue Missed} = p \cdot \sum_{i=1}^{144} \sum_{j=1}^{12} Z_{c_{ij}}$$

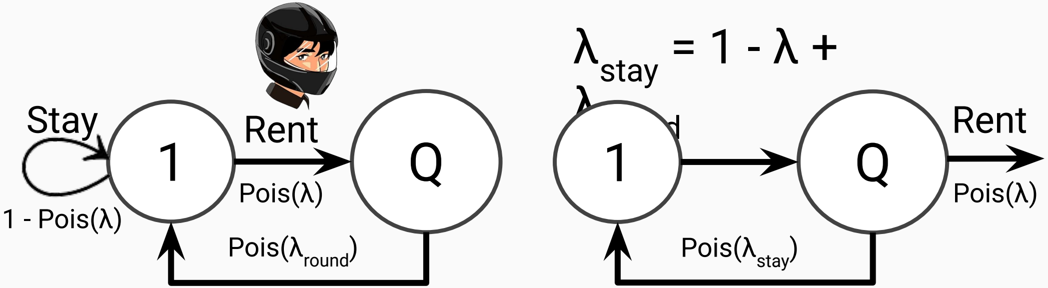
# Bike “Unconstrained” Model

- Every station starts with 100 bikes to avoid not counting missed opportunities within net changes
- Count the net changes and relevant standard deviations over 100 model simulations
- Having enough bikes to cover z = -2 across all time steps allows 97.7% of the possible rentals to happen
- Easier computationally but unrealistic, we don’t have unlimited bikes in reality



# Analytical Solution

- Redefine transition matrix Q' to capture the probability of staying in the same state
- Find Poisson arrival and departure rates: the difference between their random variables has a Skellam distribution
- For different values of our bike constraint B, select initial allocations of B bikes that greedily reduces **loss** the most




$$Q = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{12,1} & p_{12,2} & \cdots & p_{12,12} \end{bmatrix} \quad , \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \quad , \quad Q' = \Lambda Q + (I_{12} - \Lambda)$$

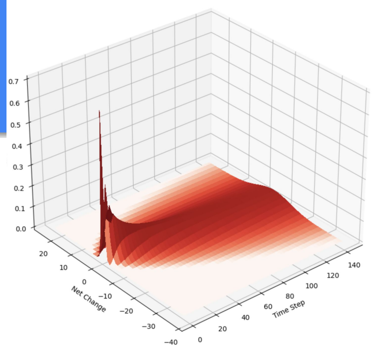
$$\mathrm{NF}_i = \mathrm{Imports}_i \text{ - } \mathrm{Exports}_i \sim \mathrm{Skellam}(\lambda_i^+, \lambda_i^-)$$

$$\mathrm{Loss}(B_i) = \sum_{t=1}^{T} \mathbb{P}\left(\mathrm{NX}_{i,t} < -B_i\right) = \sum_{t=1}^{T} F_{\mathrm{Skellam}(\lambda_i^+ t, \lambda_i^- t)}(-B_i - 1)$$

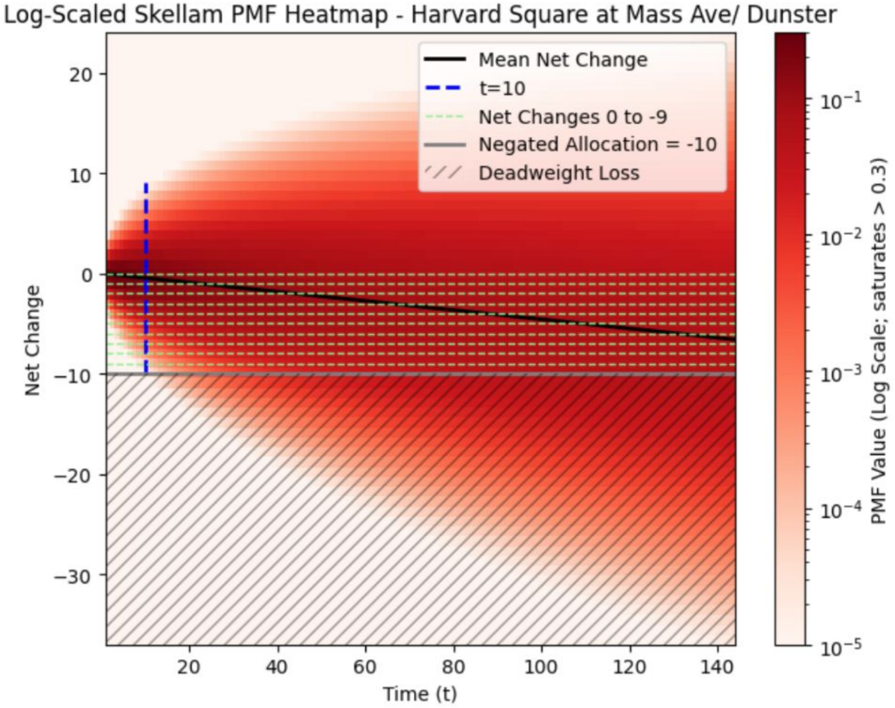
$$B = 74 : \begin{bmatrix} 5 \\ 4 \\ 7 \\ 3 \\ 12 \\ 5 \\ 6 \\ 5 \\ 8 \\ 9 \\ 5 \\ 5 \end{bmatrix} , B = 89 : \begin{bmatrix} 5 \\ 5 \\ 8 \\ 4 \\ 14 \\ 6 \\ 8 \\ 6 \\ 10 \\ 11 \\ 6 \\ 6 \end{bmatrix} , B = 104 : \begin{bmatrix} 6 \\ 6 \\ 9 \\ 5 \\ 16 \\ 7 \\ 9 \\ 7 \\ 12 \\ 13 \\ 7 \\ 7 \end{bmatrix} , B = 119 : \begin{bmatrix} 7 \\ 7 \\ 10 \\ 6 \\ 18 \\ 8 \\ 10 \\ 7 \\ 14 \\ 15 \\ 9 \\ 8 \end{bmatrix} , B = 133 : \begin{bmatrix} 8 \\ 8 \\ 11 \\ 7 \\ 20 \\ 9 \\ 11 \\ 8 \\ 16 \\ 16 \\ 10 \\ 9 \end{bmatrix}$$

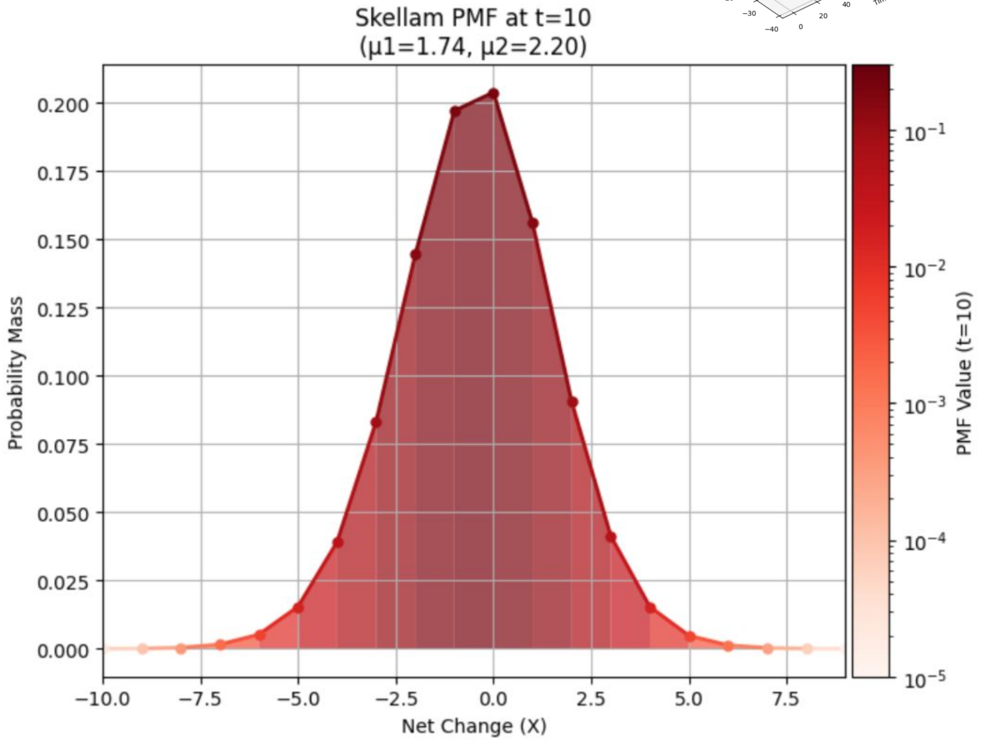
# Bike-Constrained Model

## Greedy Algorithm Visualization, Single Station



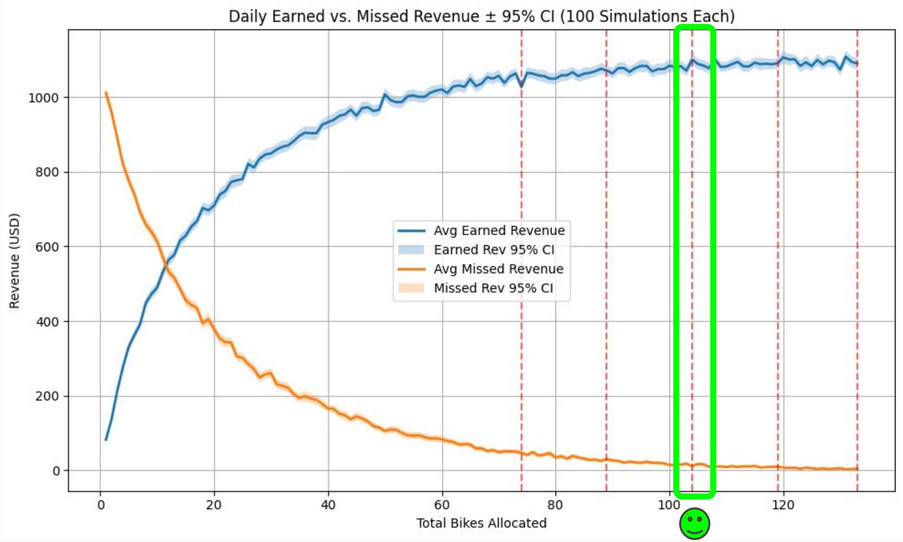

Station: Harvard Square at Mass Ave/ Dunster



# Simulated Revenue

- Feasible B-range bounded by the red vertical lines
- Inverse relationship between Avg Earned and Avg Missed Revenue
- Curves flatten near **B = 104**
  - Estimate for *feasible social optimum* quantity of bikes in circulation
  - Near optimum initial allocation on slide 9
- *Profit-maximizing quantity* requires additional info on marginal costs



# Discussion

- **Allocations:** Most popular departing stations at Harvard:
    1. Harvard Square at Mass Ave / Dunster
    2. Harvard Law School at Mass Ave / Jarvis St
    3. Harvard SEAS Cruft-Pierce Halls at 29 Oxford St
    - Allocation imbalance stable across different total bike values
- **Operations:**
    - B = 104 (social optimum)
    - Set Marginal Revenue = Marginal Cost (profit maximizing; unknown/TBD)
- **Shortcomings:**
    - Assumption of equal ride durations for all ride s
    - Closed system (only trips within 12 out of 480 Bluebikes stations around Boston)
- **Future Extensions:**
    - Road state to account for different times
    - More streamlined data analysis to allow for additional stations in the model