# Recommendations on Electrifying Blue Bike Stations

Prepared by: Julie Compston, Jessie Finocchiaro, Katie Slyman (Boston College)

Last Updated: February 10, 2026

## Data

We used two datasets, both in the forms of CSVs downloaded from data.boston.gov.

1. Blue Bike ride data from October 2025.
   1. **Features**: Classic vs E-bike, Start time, End Time, Start Station, End Station, Lat/Lon Coordinates, Membership of Rider
   2. **Missing:** Charge information on E-bikes, Bike ID to track a specific bike’s path.
2. Station List as of November 17, 2025
   1. **Features:** Name, Lat/Lon Coordinates, Seasonal Status, Municipality, Number of Docks, Station ID
   2. **Missing:** Location of station on sidewalk vs street (seemed important in NYC)

## Initial Exploration of system usage

There is a heavy-tailed distribution on the total number of rides (classic or e-bike) that go to each station, and this follows similarly for just e-bike rides as well.

**A graph showing the end of a line

AI-generated content may be incorrect.A graph of a bike ride

AI-generated content may be incorrect.**

Some preliminary visualizations allude to a tension where high-traffic stations (larger radius of station) tend to have a lower ratio of e-bikes at the station (more blue coloration).

A map with many dots

AI-generated content may be incorrect.

We apply a “suburban” mask on each station, deeming it suburban if it is more than 4 kilometers from the “mean” station, representing the center of population. We observe that the most e-bike concentrated stations tend to be suburban. In general, we also find that 36% of e-bike rides involve a suburban station. 64% of rides are urban-to-urban, 23% are urban-to-suburban or the reverse, and 13% of rides are suburban-to-suburban.

A graph of a graph of a bicycle

AI-generated content may be incorrect.

However, in thinking about charging bikes, we considered that it was not just important to understand how many bikes were at a station, but how long each e-bike was staying at the station. Because we didn’t have access to a bike ID number associated with each ride, we looked at a “worst-case” (least latent time) estimate for each station, where the bike that was just returned was the “next bike out.” [This article](https://www.curbed.com/article/citi-bike-battery-ebike-charging-docks-maintenance-repair.html#:~:text=(A%20full%20charge%20takes%20about,hours%2C%20like%20an%20EV.)) about NYC’s electrification of stations seems to suggest that 4-6 hours will be necessary to fully charge a bike. Even the least visited stations are shy of this full charging block, but that might be because of how conservative our estimate is.

A graph of a number of bikes

AI-generated content may be incorrect.

## Station usage

An initial exploration of some of the busiest stations led us to many of the universities. For example, one of the most common start-end pairs for rides with from Commonwealth Ave at Agganis Way to Commonwealth Ave at Granby St and back—presumably taking students from the residential to academic parts of Boston University. All the 10 most frequent trips are less than 2km apart (even less than a mile), which we try to control for in further analysis, since these rides presumably do not consume as much electricity and are lower stakes if a bike runs out of battery power.

A screenshot of a computer

AI-generated content may be incorrect.

A graph of a graph of a graph

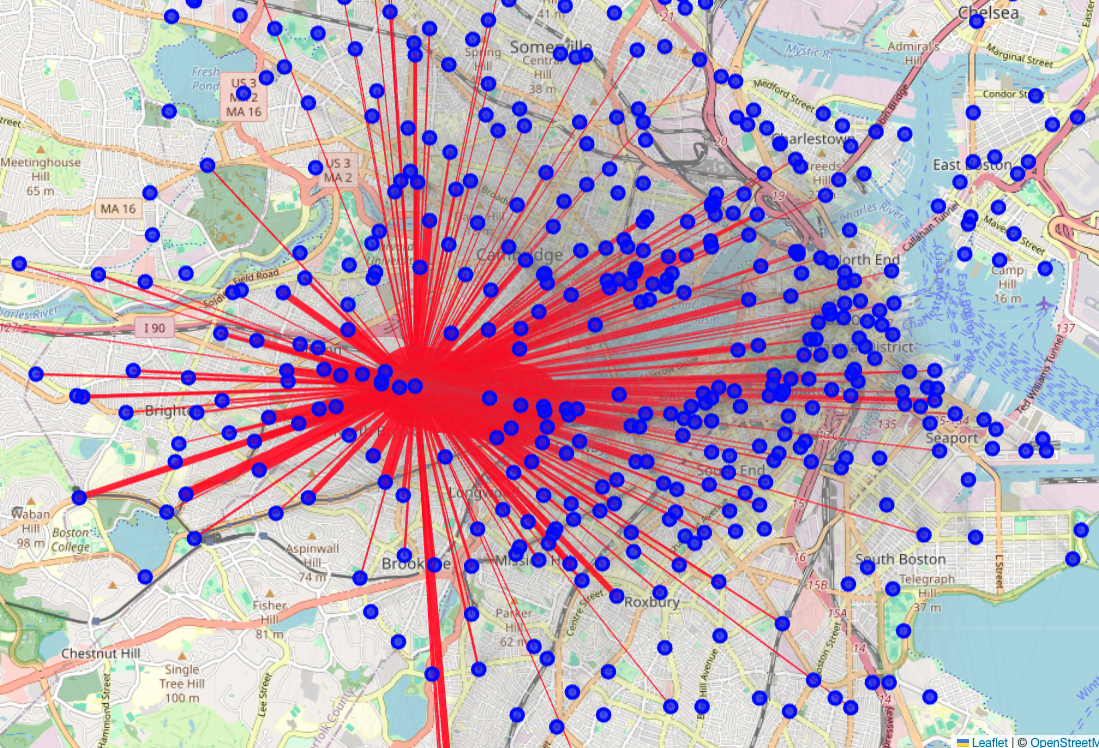
AI-generated content may be incorrect.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

AI-generated content may be incorrect.

A graph of a longwood and longwood

AI-generated content may be incorrect.



We also generated a graph of rides starting from give stations (above: Comm Ave at Agannis Way; below: Longwood Ave at Binney), weighted by how many rides end at each station. Longwood Ave at Binney is near the hospitals, and seems to be a receptor of many suburban rides, presumably handling commutes from many employees.

A map with many points

AI-generated content may be incorrect.

## Optimizing electrification

We consider a small suite of metrics to consider when looking at a set of stations to electrify.

### Utility metrics

A first naïve approach is to consider the most utilized stations and simply electrify those. However, as we observed in the initial exploration, these are often connected to each other, so electrifying a set of high-throughput, often-connected stations will likely have redundancies.

#### Charging Score

In observing the report that states e-bikes needs 4-6 hours for a full charge, we design a metric that rewards stations for timed use, where bikes are at least at the station long enough to get a significant partial charge.

The goal of this metric is to determine which stations balance high daytime usage with low nighttime usage/longer low usage blocks to allow for adequate charging windows. For example, if the busiest station in Boston has bikes leaving every 5 minutes throughout the day, we expect charging will not actually be that useful. Let

where is a time interval amount to bucket the data. Additionally, let be a threshold for the number of departures in the window considered to be efficient. For example, if 15 bikes are leaving a station every 30 minutes, then bikes charging for 2 minutes (on average) is not very helpful. With this, we say that the total charging time available at station in the day is:

where the indicator function is equal to 1 if the number of departures in the window is below the threshold , representing a sufficient “charging block.” We then scale by the total station utilization to obtain a charging score. This score is highest when both downtime and the number of total departures are high. Observe that abnormally high station utilization might skew results, so we do check this later.

With being 30 minutes and , we initially observe the following top-5 stations by charging score.

A table of numbers and a number of squares

AI-generated content may be incorrect.

In trying to control charging score to be less skewed by high utilization, we incorporate min-max normalization on the downtime fraction and number of departures, and a interquartile range-based score to penalize very high concentration of bike use at a particular time of the day.

A screenshot of a phone

AI-generated content may be incorrect.

#### Simulating a Markov Process

TODO

### Fairness metrics

#### Max-Min

One “worst case” notion of fairness is considering the farthest distance a customer would have to travel to access an electrified station. Since we do not have customer locations, we use station locations as a proxy. Suppose there are stations, and we can electrify up to of them. Throughout, we generally use . The location of station is denoted .

Max-min fairness is then defined by

Importantly, since there is a cluster of stations up in Salem, this metric inevitably requires a station in Salem be electrified. Including Salem, the tuple of stations (Seaport Blvd at Sleeper St, Coolidge Corner – Beacon St at Centre St, Salem State University – Bike Path at Loring Ave) best satisfies this notion of fairness, with a maximum distance of 10.53 kilometers needed to be traveled.

If we were to *exclude Salem* from this computation, then the tuple (Commonwealth Ave at Agganis Way, Forsyth St at Huntington Ave, Lewis Wharf at Atlantic Ave) becomes the most fair, with a maximum distance of 10.586 being the maximum distance necessary.

Observe that this notion of fairness leans heavily on one station to be the “closest”, so there are many station combinations that are “equally fair” according to this metric. When excluding Salem, it is most important to include the Commonwealth Ave at Agganis Way station.

### k-Medians

TODO