Ride Substitution Using Electric Bike Sharing: Feasibility, Cost, and Carbon Analysis

JOHN WAMBURU, University of Massachusetts Amherst, USA
STEPHEN LEE, University of Pittsburgh, USA
MOHAMMAD H. HAJIESMAILI, University of Massachusetts Amherst, USA
DAVID IRWIN, University of Massachusetts Amherst, USA
PRASHANT SHENOY, University of Massachusetts Amherst, USA

While ride-sharing has emerged as a popular form of transportation in urban areas due to its on-demand convenience, it has become a major contributor to carbon emissions, with recent studies suggesting it is 47% more carbon-intensive than personal car trips. In this paper, we examine the feasibility, costs, and carbon benefits of using electric bike-sharing—a low carbon form of ride-sharing—as a potential substitute for shorter ride-sharing trips, with the overall goal of greening the ride-sharing ecosystem. Using public datasets from New York City, our analysis shows that nearly half of the taxi and rideshare trips in New York are shorts trips of less than 3.5km, and that biking is actually *faster than using a car* for ultra-short trips of 2km or less. We analyze the cost and carbon benefits of different levels of ride substitution under various scenarios. We find that the additional bikes required to satisfy increased demand from ride substitution increases sub-linearly and results in 6.6% carbon emission reduction for 10% taxi ride substitution. Moreover, this reduction can be achieved through a hybrid mix that requires only a quarter of the bikes to be electric bikes, which reduces system costs. We also find that expanding bike-share systems to new areas that lack bike-share coverage requires additional investments due to the need for new bike stations and bike capacity to satisfy demand but also provides substantial carbon emission reductions. Finally, frequent station repositioning can reduce the number of bikes needed in the system by up to a third for a minimal increase in carbon emissions of 2% from the trucks required to perform repositioning, providing an interesting tradeoff between capital costs and carbon emissions.

CCS Concepts: • **Applied computing → Transportation**; • **Human-centered computing → Empirical studies in ubiquitous and mobile computing**.

Additional Key Words and Phrases: Carbon Emission, Optimization, Electric Bikes, Ride Substitution

ACM Reference Format:
John Wamburu, Stephen Lee, Mohammad H. Hajiesmaili, David Irwin, and Prashant Shenoy. 2021. Ride Substitution Using Electric Bike Sharing: Feasibility, Cost, and Carbon Analysis. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 38 (March 2021), 28 pages. https://doi.org/10.1145/3448081

1 INTRODUCTION

Urban transportation has seen significant changes over the past decade due to the emergence of the sharing economy and the ubiquity of smartphones. Ride sharing services, such as Uber, Lyft, Grab, and Didi, have

Authors' addresses: John Wamburu, jwamburu@cs.umass.edu, University of Massachusetts Amherst, USA; Stephen Lee, stephen.lee@pitt.edu, University of Pittsburgh, USA; Mohammad H. Hajiesmaili, hajiesmaili@cs.umass.edu, University of Massachusetts Amherst, USA; David Irwin, irwin@ecs.umass.edu, University of Massachusetts Amherst, USA; Prashant Shenoy, shenoy@cs.umass.edu, University of Massachusetts Amherst, USA.

2474-9567/2021/3-ART38 $15.00
https://doi.org/10.1145/3448081

become immensely successful due to their promise of personal on-demand transportation at any time [25]. Early proponents of ride sharing suggested that these services would reduce our reliance on privately-owned cars, reduce traffic congestion, and reduce carbon emissions, with early studies estimating that at least five private vehicles would be replaced for each shared car and there would likely be carbon emission reductions if shared cars were newer vehicles [4]. However, the success of these services has resulted in an increase in traffic and more congestion on roads—a type of rebound effect [23]. For example, in New York City, a recent study has shown that ride sharing constitutes more than 50% of traffic on the roads [9, 41]. Another study on the climate impact of ride sharing has estimated that a typical ride sharing trip is less efficient than a personal car trip, mainly due to the "dead" miles travelled by a ride-share car between consecutive hired rides, and that this generates 47% higher carbon emissions than an equivalent private car ride [5, 8]. The study also showed that the greater convenience of ride sharing has steered passengers away from public transit options. From a carbon emissions standpoint, the study argues for greater use of travel modes that produce fewer, or zero, carbon emissions such as biking, walking, or mass transit [8].

Motivated by the need to green ride sharing, we address a key question: what are the carbon benefits and costs of encouraging more bike sharing as a substitute for shorter ride sharing trips, and what is the feasibility of such an approach? Bike sharing programs have become popular in major urban areas in recent years, with cities deploying tens or hundreds of thousands bikes within the urban core. Bike sharing provides many benefits—it is pollution free, provides wellness benefits through exercise, and can reduce our reliance on other forms of transportation, especially for shorter rides within urban areas [10]. However, bike sharing has traditionally been considered to be *complementary* to car-based ride sharing, rather than an *alternative*. There are many challenges in using bikes as a substitute for car rides. Biking can become challenging in rainy or cold weather and in the dark or may be infeasible for transporting cargo items. The pedaling effort involved in biking discourages its use for longer rides and for those requiring uphill terrain or steep slopes. The latter limitation can be addressed by using electric bikes, or e-bikes, in bike sharing programs. E-bikes, which are already popular in many countries, such as China, provide pedal assist through a battery-powered motor that makes biking easier for longer or uphill rides [47]. Although some bike sharing systems are beginning to adopt e-bikes, they incur higher deployment and maintenance costs than regular bikes. We argue that a hybrid mix of regular and e-bikes can be more effective than using either type of bike in a bike share program. Even with the use of some, or all, e-bikes in a bike sharing system, a car ride can be potentially replaced by a bike ride only when (i) the distance makes it amenable to biking, (ii) a bike can be easily picked up and dropped off at the start and end of the car trip, and (iii) factors such as weather and time of day are favorable. Analyzing the feasibility of such ride substitution using real-world car and bike data is a key goal of our work.

In addition to these user-centric challenges, there are several provider-centric challenges as well. Since bike share programs were not explicitly designed as an alternative to ride sharing, they will face capacity constraints if riders substitute car rides with bike rides even in modest numbers. The capacity of a bike share system will need to be significantly enhanced to handle this higher demand. Further, the bike share system will need to expand to new areas that lack coverage, especially if those areas see a significant volume of ride pickups or drop-offs. A detailed analysis of these capacity enhancement costs and the carbon benefits they provide is another goal of our work.

Motivated by these questions, we conduct a data-driven study to analyze the benefits of adding e-bikes to bike sharing programs with the broader goal of reducing carbon emissions and greening the ride sharing ecosystem. Our emphasis on green carbon-aware design of hybrid bike share systems and carbon analysis of using biking for short rides differentiates this study from prior work. To understand feasibility and benefits of using electric bikes within bike share systems, our paper seeks to address the following questions.

(1) What are the characteristics of typical bike share and ride share trips within an urban area? Based on the distances traversed by ride share trips, what fraction are amenable to substitution by bike rides? What percentage of ride share trips have a bike share station near the pickup and dropoff location?
(2) What is the amount of capacity enhancement needed to accommodate a certain percentage (say 5 or 10%) of car trips using bikes? What is the optimal mix of regular and electric bikes needed to accommodate those trips, and what are the carbon benefits of such ride substitution?
(3) How should coverage of bike share systems be intelligently extended to parts of the city that lack bike coverage so as to maximize the use of bikes as an alternative to short ride share trips? What are the carbon benefits and costs of such expansion?
(4) How can the capital and operational costs of these hybrid bike share systems be optimized using factors such as repositioning and what are their cost vs. carbon tradeoffs?

We adopt a data-driven analysis approach to address these questions by leveraging public datasets on bike sharing and ride sharing (taxi, Uber, Lyft, Juno, Via) for New York City. Our algorithms and data-driven analysis for addressing these questions yield the following contributions.

- First, we characterize and compare bike sharing and car trips in New York City and show that 50% of car rides are less than 3.6km and 69% are within 200 meters of a bike station. Our results imply that many shorter car rides could be easily substituted by bike rides if the user is motivated to do so.
- Second, we use optimization-driven data analysis to determine the capacity enhancements needed to substitute a modest fraction of car rides with bike rides. Our results show a 6.6% reduction of carbon emissions with 10% ride substitution, and although this would triple demand in the bike share system, it would require only a 2x increase in the number of bikes needed due to slack in the system from skewed spatial demand. A hybrid design that assumes electric bikes for longer bike rides of more than 5km requires only a quarter of these bikes to be electric, saving on the overall cost of the system. We also show that our results broadly hold under different strategies for ride substitution.
- Third, we use traffic estimates from ride sharing in urban areas that lack bike sharing coverage to determine popular pickup and dropoff locations and compute the additional carbon benefits of expansion into these areas. Our results show that adding bike coverage to new regions requires higher capacity enhancements for ride substitution than that for regions with current coverage. Specifically, we show that adding coverage to parts of New York City that lack bike sharing coverage would require a doubling of the number of bike stations and 50% more bikes, but the greening benefits of adding such coverage are significant, yielding up to 5000MT reduction in carbon emissions.
- Finally, we use our optimization-driven methods to characterize the tradeoff between capital costs of expansion and the operational costs of bike repositioning. We show that reducing the frequency of repositioning by 2 hours (e.g., from 2 to 4 hours) can reduce the cost of repositioning and the overall carbon emission by 58% while increasing the capital cost (number of bikes required) by only 29%.

Overall, our analysis shows that the carbon benefits of substituting even a modest fraction of ride share trips using electric and regular bikes are substantial and can help green the overall ride sharing ecosystem.

2 BACKGROUND

In this section, we provide background on ride and bike sharing.

Shared On-demand Mobility. The sharing economy has had a significant impact on personal transportation and has led to ride sharing becoming common, or even the preferred, form of transport in urban areas [27]. The ability to get a ride on demand using a smartphone "anywhere" has made them a popular alternative to traditional taxi rides and private car ownership. A side effect of this convenience has also caused them to supplant more carbon-efficient public transit options. Researchers have analyzed the climate impact of ride sharing and found ride sharing trips to generate 47% more carbon emissions than a private car trip, mainly due to the "wasted" driving between two hailed trips [5, 8]. This observation motivates our study on whether other types of ride sharing, such as bike sharing, which have a low carbon footprint could be used for short rides, while providing similar convenience such as on-demand pick-up/drop-off and wide availability.

Bike Sharing. Bike sharing programs are now commonplace in many urban areas. Biking is considered to be a "fun" mode of transportation that is ideal for last-mile short distance rides at a much lower cost than ride-sharing services. Moreover, bikes have dedicated bike lanes in many urban areas and are less susceptible to traffic congestion as a result, allowing for relatively quick short distance rides. In addition to being a clean and a zero-carbon transportation mode, biking is seen as a form of exercise and promotes an active lifestyle with health benefits. Bike sharing programs have led to a significant increase in the popularity of biking within urban areas in recent years. Bike sharing comes in two flavors: docked programs, where bikes are picked up and dropped off from bike stations, and dockless programs, where bikes can be dropped off anywhere [43]. Our work currently focuses on docked programs due to the need to charge electric-bikes at bike stations; however many of our insights carry over to dockless bike sharing programs as well.

There has been significant research attention on the design of bike sharing systems over the years. A range of problems which are discussed in more detail in Section 7 such as optimizing station placement, capacity planning, rebalancing algorithms, and bike placement methods have been studied. A key difference with this body of prior work is our focus on sustainable carbon-aware design and on using bike sharing as a viable alternative for shorter car rides from a carbon standpoint.

Electric Bikes. An electric bike provides pedal assist to its rider using an inbuilt motor and battery, which makes biking nearly effortless and attractive for longer rides or rides on uphill roads. Their attractiveness for handling more challenging rides and reducing the biking effort makes them a key design element for encouraging substitution of short car rides with bikes. Electric bikes have long been a popular form of transportation in countries like China [32]. Some bike share program such as in Riverside [40] and Raleigh [38] in the USA and Guildford [22] in UK use only electric bikes. Moreover, New York's Citibike program has plans to add up to 4000 electric bikes to its fleet [15]. While an all-electric bike system requires higher capital and operational cost, e.g., electric bikes are more expensive to acquire, and a charging infrastructure must be put in place for operation, a hybrid system is cheaper to build and maintain as only a subset of bikes and infrastructure will require electric capability. Given the higher deployment and maintenance costs of electric bikes, a goal of our work is to highlight the benefits of hybrid systems that use a combination of regular and electric bikes as a cheaper alternative to an all electric-bike system while retaining the key advantage of electric bikes for enabling ride substitution.

Given this background, the overall goal of our work is to analyze the feasibility of substituting shorter but more carbon-intensive car rides using bike sharing and to understand the carbon benefits of such ride substitution as well as the costs of handling a higher bike sharing demand resulting from such substitution.

3 HOW SUBSTITUTABLE ARE RIDE SHARE TRIPS BY BIKE TRIPS?

In this section, we present a feasibility analysis that examines the degree to which ride sharing trips are substitutable by bike trips.

3.1 Overview of Datasets

Our analysis is based on two public datasets from New York City. These datasets are described below and summarized in Table 1.

Table 1. Ride Sharing Datasets

	CitiBike 2019	TLC 2019	TLC 2016
Vehicle Type	Bikes	Taxi and FHV	Taxi
Number of vehicles	17,954 bikes/941 station	14K taxi, 78K FHV	22K taxi
Number of trips	18.98 million	354 million (101 taxi, 253 FHV)	156.77 million
Trip-level records	yes	yes	yes
Pickup/dropoff coordinates	GPS	Geo-fenced Zones	GPS
Duration	Jul 2018 - Jun 2019	Jul 2018 - Jun 2019	Jul 2015 - Jun 2016

CitiBike Dataset. CitiBike is the official bike share system for New York City and has released extensive data about its operations. For the purposes of our work, we use a 12 month period (July 2018 - June 2019) that comprises of 18.98 million trips. This is the most recent 12 months for which data is available. During this period, the CitiBike system had 17,954 bikes[1] and 941 bike stations. The data contains a trip's duration, its start and end time, the ID of start and end stations, the station names, the start and end GPS locations (latitude and longitude), the type of user, age and gender, and the bike ID.

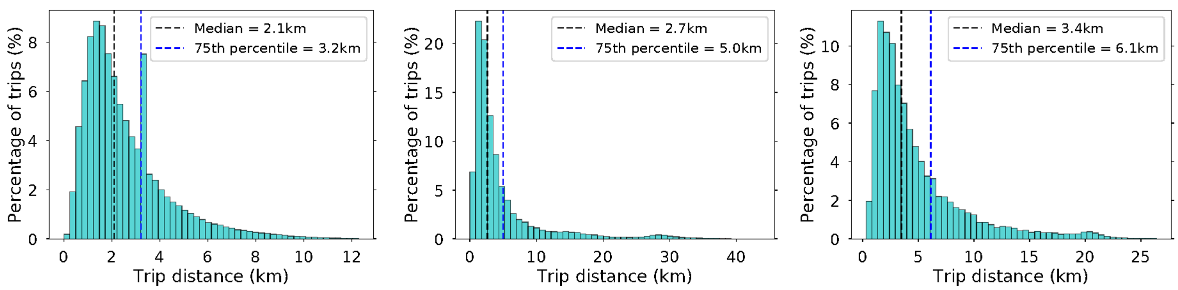
New York TLC Dataset. The New York City Taxi and Limousine Commission (TLC) [1] provides comprehensive statistics for taxi trips as well as ride sharing vehicles (referred to as For-Hire-Vehicles (FHV)) such as Uber, Lyft, Juno, Via and Limousine series. While many years of data are available, the nature of what information is released on a trip-level record has changed over the years. For the July 2018 - June 2019 period, which we refer to as the TLC 2019 dataset, there were a total of 354 million rides, of which 101 million come from taxis and 253 million come from FHVs, with Uber and Lyft accounting for a large majority (> 90%) of the FHV rides. Trip-level records are available for all of these 354 million rides, which include the pickup and drop-off date and time, pickup and drop-off location, trip distance, number of passengers, and itemized fare. The latter three fields are optional fields and only reported for taxi rides and unavailable for FHV rides. For the TLC 2019 data, pickup and drop-off locations are reported in the form of coarse-grain taxi zones, rather than exact GPS coordinates to preserve privacy. Each zone represents a geofenced neighborhood. Since some of our analysis requires precise pickup and drop-off location coordinates, we also use TLC data from July 2015 to June 2016 (TLC 2016 dataset), the most recent year for which precise GPS pickup and drop-off coordinates are available for taxi rides. The TLC 2016 dataset comprises of 156 million taxi rides as shown in Table 1. We note that as the popularity of FHV services like Uber and Lyft increased, the number of taxi rides dropped from 156 million in 2016 to 101 million in 2019, but the total volume of rides increased significantly, with FHV rides increasing to 2.5 times of taxi rides.

3.2 Feasibility Based on Distance

We begin our feasibility analysis by computing distributions of trip distances for bike, taxi, and FHV car sharing for the above datasets.

We analyze trip distances either by using the actual trip distance whenever reported in the trip records, or by computing the shortest road distance between the pickup and dropoff locations. For taxi rides, the actual trip distance is available in trip records. For bike rides, we use the GPS coordinates of the pickup and dropoff bike stations to estimate the bike trip distance. For the FHV rides, the trip distance is not available in trip-level records. Further, the pickup and drop-off locations are reported in terms of coarse-grain geo-fenced zones. To estimate the trip distance from coarse-grain pickup and dropoff zones, we compute the centroid of each pickup and dropoff zone and compute the shortest road distance between the two centroids. For trips that originate and

[1]The vast majority of CitiBikes are regular, non-electric bikes. CitiBike began a pilot deployment of electric bikes in 2019, but these constitute a small portion of the total bikes during this period.



(a) Bikes (b) Taxis (c) FHVs

Fig. 1. Distribution of trip distances for bikes, taxis, and for-hire (ride-share) vehicles. The median ride is short, with median trip distances ranging from 2.1-3.4km

end within the same zone, we assume the trip started at the zone's centroid and ended at the zone boundary and use half the zone length as the distance.

Figure 1 depicts our results. As shown in Figure 1a, most bike rides are short, with the median distance of a bike trip at 2.1km (for our analysis, we use median, instead of mean, since all our trip distributions have a long tail, which can skew the mean). The figure also shows that 75% of the trips, represented by the 75-th percentile of the distribution, are 3.2km or shorter. At the same time, we find that the distribution has a long tail with a small number of trips as long as 12.5km.

Figure 1b depicts the distribution of taxi trips from the TLC dataset. Like our bike results, we find that most taxi trips are short—the median taxi trip in 2019 was 2.7km long; this is virtually unchanged from the 2016 TLC data (not shown here for brevity) where the median taxi trip was 2.8km long. The 75-th percentile of the distribution shows that three-quarters of the trips are 5km or shorter, with the distribution exhibiting a long tail.

Figure 1c depicts the distribution of FHV ride distances. The results show that the median FHV ride is 3.4km, which is slightly longer than the median taxi ride. Note that many FHV rides start and stop within the same zone, and due to lack of exact pickup and dropoff coordinates for FHV rides, we approximated all of these rides to be half the zone distance, which may bias our result somewhat. Nevertheless, the majority of FHV rides are short (<3.5km), which is similar to our observations for taxi trips; the 75-th percentile of FHV trip distance is 6.1km.

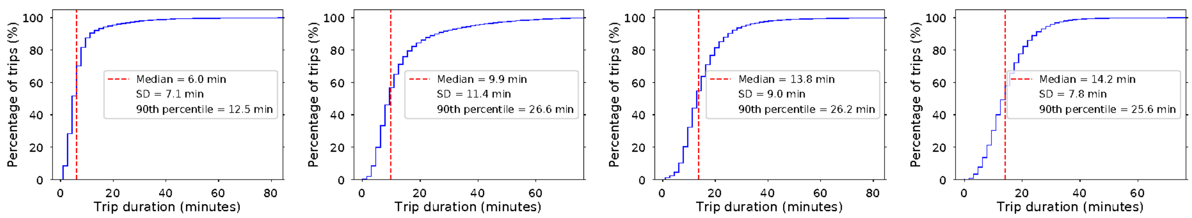
Key Takeaways:

- The median car trip is quite short (2.7km for taxis and 3.4km for FHVs), and this is comparable to the median bike trip (2.1km). Further, three quarters of all car trips, represented by the 75-th percentile of the distribution, are less than 5 and 6.1km, for taxis and FHVs, which represent bikeable distances.
- Around 43% of the bike rides are medium-distance trips (between 2 and 5km) long, which indicates that current bike riders are amenable to biking longer distances even without electric bikes. This implies that some of medium-distance car rides are also be amenable to substitution, in addition to shorter rides. Electric bikes have the potential to increase the feasibility of such substitutions.

The observations above indicate that, from a distance perspective, there are a substantial fraction of car rides could be potentially replaced by bike rides. In practice, however, there might be some practical challenges beyond the physical distance that are further discussed in Section 3.4.

3.3 Feasibility Based on Convenience

Next, we examine the convenience of ride substitution by examining two key user convenience metrics: end-to-end ride duration and distance to the nearest bike station.



(a) Bike short trips (b) FHV short trips (c) Bike medium trips (d) FHV medium trips

Fig. 2. Cumulative distribution functions of trips duration. Short bike rides of 2km or less are faster than FHV rides. FHV trips have a greater variance due to variable traffic conditions

Convenience Based on Trip Duration. The ride duration is a key convenience metric since users will have less motivation to substitute a bike ride by a car ride if bike trips take significantly longer than car trips. To analyze the feasibility of ride substitution based on trip durations, we compute the CDFs of trip durations of short distance (less than 2km) bike and FHV rides. We also compute the CDFs of trip durations for medium distance (between 2 and 5km) rides. Figure 2 depicts our results.

As shown, the median short bike trip takes 6 minutes (Figure 2a), while the median short FHV trip is 9.9 minutes long (Figure 2b). This reveals an interesting result—although bicyles are slower than cars, for short rides, *biking is faster than a car ride*—likely since dedicated bike lanes in cities such as New York are less congested than regular roads, so the bikes do not get caught in congestion.

Importantly, Figure 2b shows that distribution of short FHV rides has a long tail and high variance, which means that many short car rides can have longer durations during periods of traffic congestion (the 90-th percentile is 26.8min and the standard deviation is 11, even when traveling less than 2km). In contrast, Figure 2a shows that the bike trip duration distribution has a relatively short tail, with the 90-th percentile at 12.5min and a standard deviation of 7.1. Similarly, Figures 2c and 2d plot the CDF of the duration of medium distance (between 2 and 5km) trips. The result shows, that similar to short trips, the median duration of medium trips is shorter for bikes (13.8 minutes) as compared to FHV (14.2 minutes), though the difference is not as substantial as the short trips.

The key takeaway is that for very short rides of 2km or less, biking is 40% faster than an equivalent FHV ride due traffic congestion seen by cars on city roads and the availability of dedicated bike lanes.

Accessibility of Bike Stations. While the above analysis considered trip durations in terms of travel times, a bike trip has two components that impact trip durations: the time to walk to a bike station and the actual travel time on the bike. Users will have less motivation to use a bike if the nearest pickup or dropoff bike stations are inconveniently located, which increases the walking distance.

To analyze accessibility of bike stations, for each car ride, we consider their pickup and dropoff coordinates and estimate the distance to the nearest bike station from those locations. Note that this analysis requires precise (GPS) pickup and dropoff coordinates, and hence, we use the 2016 TLC dataset for this analysis. For each trip, we compute the distance to the nearest bike station from the pickup and dropoff GPS coordinates and consider the greater of the two values as the nearest station for that trip.

Figure 3 shows the distribution of the distance to the nearest bike station for each trip. As can be seen, 18% of rides are within 100m of a bike station, 45% of the rides are within 150m and 69% are within 200m of a bike station, where 200m distance takes less than 5 minutes to walk. *This implies that about 70% of riders are within a short walking distance of a bike station, making ride substitution convenient.*

Together, Figures 2 and 3 indicate that even if we were to add a 5 minute pickup and dropoff overhead to each bike trip (equivalent to a 200m walk to/from a station), the end-to-end duration of bike trips for short rides is still

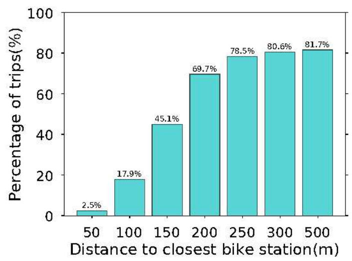


Fig. 3. Distribution of pickup/dropoff locations to the closest bike stations

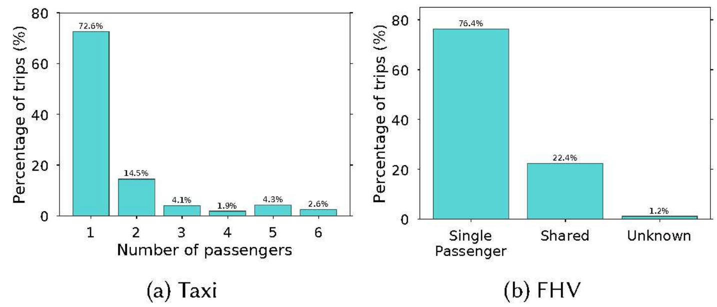


Fig. 4. Distribution of number of passengers in each taxi and FHV car trip. A quarter of the trips involve more than one passenger

comparable to the FHV trips. If we were to add pickup wait times to FHV trip durations, end-to-end duration for FHV rides would increase as well, making short bike trips more attractive from a time standpoint.

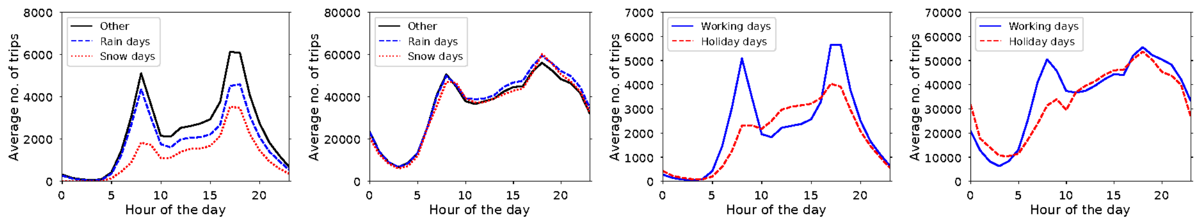
3.4 Challenges in Ride Substitution

Despite the results above, there are many scenarios where ride substitution may not be feasible even for short trips. First, if a taxi or FHV trip has multiple passengers, we assume that ride substitution is infeasible, i.e., only single passenger trips can be substituted using bike trips. The TLC data includes the number of passengers for each trip, hence in Figure 4, we plot the distribution of passengers per ride for taxi and FHV. The results show that for both taxi and FHV trips, more than 70% of trips are single passenger, which implies that 30% are multi-passenger trips and are not substitutable.

Second, the weather plays an important role in bike usage. Figure 6a depicts the demand for bike trips by the month of the year. As can be seen, bike usage drops by half in winter months in New York and with an even greater drop during January, which is the coldest month of winter. In addition, an analysis of weather data for New York (see Table 2) shows that it rained for 170 days of the year, and there were 158 days with an average temperature of less than 10°C. To quantify the effect of weather on trip demand, we analyze the number of trips during snow, rain, and other weather conditions. We group days as either snow, rain, or 'other' as follows; for example, if snow (or rain) falls during a day, we label the day as snow (or rain) day. However, if both rain and snowfall exist on the same day, we label the day as a snow day. All other days are labeled 'other' and capture other weather conditions during which biking is more convenient than during rain or snow. We then compute the cumulative number of trips every hour for each day and then compute the hourly average across all days of the year.

Figure 5a shows the distribution of the number of bike trips for different weather conditions. Not surprisingly, rainy or snowy days see lower bike share usage than sunny days. Figure 5b depicts a similar distribution for taxi trips. The morning peak for taxi trips shows higher demand for taxi rides during sunny days than rainy and snowy days. However, we observe a slightly higher demand for taxi trips in the evening hours during rain and snow than sunny days, which indicates that people generally prefer to avoid unfavorable weather after work hours. Since our optimization focuses on the upper bound of trip demand (the number of bikes is as many as is required to satisfy peak demand), the reduction in demand during rain and snow conditions is not likely to increase the number of bikes required to satisfy demand. We, therefore, do not consider these conditions for the rest of the paper.

Third, social factors such as holidays also play a role in bike usage. To quantify their effect on trip demand, we compute the average hourly number of trips during holidays and regular working days. We consider holiday



(a) (b) (c) (d)

Fig. 5. (5a) Distribution of bike trip demand by time of day during rainy, snowy and sunny (other), and (5b) distribution of taxi trip demand during rainy, sunny and snow days, and (5c) distribution of bike trip demand by time of day during working days and holiday days, and (5d) distribution of taxi trip demand during working days and holiday days.

days to be Federal Holidays only and all other days as working days. Figure 5c shows the distribution of bike trip demand during holidays and working days, while Figure 5d shows the same distribution for taxi trip demand. In both cases, we find that the average hourly demand for both bike and taxi trips is lower during holidays than during working days. Our analysis in the following sections considers all days; those with favorable weather and those without, and days with holidays. However, since the above analysis shows that demand drops during holidays and bad weather days, such factors do not need to be considered separately since lower demand will not increase bike capacity needed to handle peak demand.

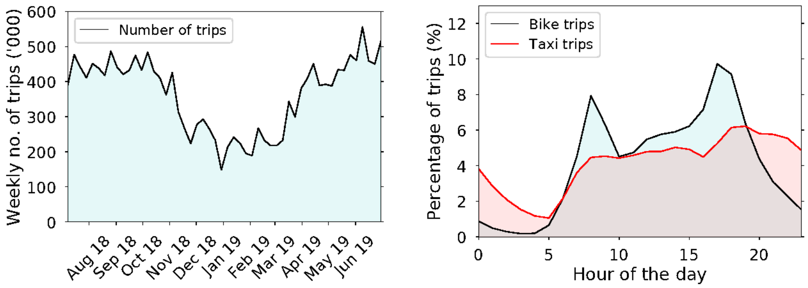
Summary and Key Takeaways. In summary, our data-driven analysis shows that the median taxi and FHV ride is quite short (2.7 km and 3.4km) and around three quarters of these trips are less than 5km, which represent distances that could be traveled on a bike. Importantly, for very short rides of 2km or less, biking is actually faster than taking a ride share trip. Further, nearly 70% of the car rides start and stop within 200m of a bike station. Even some of the medium distance car trips between 2 and 5km appear to be substitutable by bikes, since 43% of bike rides involve commuting these longer distances already, and future availability of electric bikes can further reduce the biking effort involved. The above data-driven observations show that a large fraction of taxi and FHV rides are feasible for substitution with bike rides.

In practice, however, a much smaller fraction of the rides may be amenable to such substitution due to several challenges involved in bicycling. For our subsequent analysis, we focus on the costs and benefits of substituting only a modest fraction of car rides (e.g., 1% to 10% of the total rides) using bike rides—especially since such a modest amount of ride substitution may actually be feasible with the introduction of electric bikes into a bike sharing system and appropriate user incentives.

Lastly, while our analysis has focused on New York City, our insights are broadly applicable to other dense urban areas. For instance, other studies have shown that a non-trivial fraction of the FHV and Uber rides in downtown urban districts of many cities are short [23] and share similar characteristics with our New York City analysis. Traffic congestion in cities is a global problem, and our observation that biking is faster than a car ride for ultra-short rides will likely hold for many other cities as well.

4 COSTS AND CARBON BENEFITS OF ELECTRIC BIKE-BASED HYBRID BIKE SHARE SYSTEMS

Having analyzed the feasibility of substituting shorter car rides with bike rides, we next analyze the cost and carbon benefits of modest levels of ride substitution. Specifically, we analyze the number of additional bikes needed to handle the higher demand due to ride substitution, the costs of using a hybrid mix of electric and regular bikes to meet this higher demand, and the carbon emission reductions resulting from such ride substitutions. We first describe our research methodology and then our results.



(a) Demand of bike trips by month

(b) Demand of bike and taxi trips by hour

Fig. 7. Trip demand by (a) month of the year and (b) hour of the day.

Table 2. Summary of NYC weather 2018-2019

Weather condition	No. of days
Rain	170
Average temp < 10°C	158
Average temp ≥ 10°C	207

4.1 Research Methodology

Our methodology for analyzing the additional capacity and the mix needed in the bike share system to absorb the higher demand from ride substitution is based on data-driven optimization. We discuss our data-driven methodology and our optimization algorithm in turn.

Methodology for Estimating Ride Substitution Demand. We begin with the CitiBike dataset described in Table 1 since it represents the baseline (existing) demand for bikes in the bike share system. We then consider various scenarios, each representing a different fraction of ride substitution of car rides by bike rides. For each scenario, we augment the bike dataset with car rides that have been substituted by bike rides—by sampling the TLC dataset and adding those sampled trips to the bike dataset. We use three sampling strategies to estimate ride substitution demand. Our baseline strategy is described here. In addition, we consider two other strategies in Section 4.2.1 to understand their impact on our results. To illustrate, given a certain fraction of ride substitution, say 1%, we sample the TLC dataset to extract 1% of the total trips (here, 1% of the 354 million rides) and add them to the bike dataset. This augmented bike dataset represents the enhanced demand for bike rides as a result of that level of ride substitution. Note that sampling of the TLC dataset must be done carefully to ensure only feasible rides are extracted from the overall dataset. Consequently, our sampling method only considers rides with the following constraints when sampling.

(1) Only rides less than a threshold distance are considered, which eliminates the long tail of the distribution representing long taxi and FHV rides. That is, we exclude all trips longer than 10km from the sampling and strongly bias the sampling towards short and medium distance car rides (<5km) in the TLC dataset, since these trips are the most likely candidates for substitution.
(2) Only rides that start and stop within a certain threshold walking distance (e.g. 200m) from a bike station are considered, which ensures only "convenient" rides are considered for sampling and substitution.
(3) Rather than uniformly sampling the TLC dataset across time, we bias the sampling to follow the temporal distribution of bike rides. As shown in Figure 6b, car rides occur at all times of the day—even when it is dark—while bike rides are more common during daylight hours and exhibit peaks that are correlated with the morning and evening commute hours. We assume that ride substitution is more likely when bikers prefer to bike, and hence we sample the TLC dataset based on the temporal distribution of bike rides (i.e. sampling of car rides from the TLC data follows the temporal distribution of bike rides shown in Figure 6b). As a result, the enhanced bike data set follows the same temporal distribution as the original Citibike dataset.

We use the process above to create multiple enhanced bike datasets, each representing a different level of ride substitution (e.g. 1%, 2%, 3%, 5%, 10%). This yields enhanced datasets with the total number of trips in the enhanced datasets ranging from 18 to 54 million (see Table 4 in appendix for details about the enhanced datasets).

Finally, since the above sampling process needs to ensure substituted car rides start and stop within a threshold walking distance of a bike station, we need precise GPS coordinates of pickup and dropoff locations for trips in the TLC data. Recall that the 2019 TLC data only provides coarse-grain zones for location coordinates. Hence, we use a combination of the 2019 and 2016 TLC data for constructing the enhanced bike datasets. In Section 3, we showed the median distance of the 2016 and 2019 taxi trips was similar (2.8 and 2.7km). Further, our analysis (omitted for brevity) shows that both 2019 and 2016 TLC datasets exhibit similar spatial and temporal distributions. Hence, we use the 2019 TLC data to determine the volume of trips that need to be sampled for a certain level of ride substitution and then sample rides from the 2016 data to construct an enhanced bike dataset with complete trip level records having the desired substitution volume.

Optimization Problem. We now cast an optimization problem with the objective of minimizing the number of regular and electric bikes needed to satisfy the demand represented by the above enhanced bike dataset. We divide the trips in the dataset into short, long, and medium categories by using two distance thresholds: $T_{\texttt{short}}$ and $T_{\texttt{long}}$. Short trips are assumed to have distances less than $T_{\texttt{short}}$, long trips are assumed to have distances greater than $T_{\texttt{long}}$, and medium trips lie between the two. We assume that the demand for short and long trips is handled by regular and electric bikes, respectively. Demand for medium trips can be handled by either regular or electric bikes, based on the preference of the rider. Our optimization formulation finds the allocation of medium trips between regular and electric bikes such that the overall number of bikes is minimized (in addition to computing the optimal number of regular and electric bikes needed to handle the demand from short and long trips, respectively).

We define our optimization model as follows. Let $\mathcal{S} = \{1, \dots, n\}$ denote the set of bike stations in a bike sharing system, each indexed by i. We assume a time-slotted model where in each slot t and station i, we have incoming trips denoted by $I_i(t)$. We use mean bike trip duration as the value of t. Trips that start and end in a particular time slot are considered independently, i.e., trips that start in one time slice and end in another are considered as starting in the first time slot and ending in the other, respectively. We further divide the incoming trips $I_i(t)$ into $I_i^{\texttt{L}}(t)$, $I_i^{\texttt{M}}(t)$, and $I_i^{\texttt{S}}(t)$ to indicate the number of incoming *long, medium, and short* incoming trips to station i at time t, respectively, and we have

$$I_i(t) = I_i^{\texttt{L}}(t) + I_i^{\texttt{S}}(t) + I_i^{\texttt{M}}(t), \quad \forall i, \forall t. \tag{1}$$

Next, let $O_i(t)$ denote the number of outgoing trips from station i at time t. Similarly, outgoing trips can be further divided into long $O_i^{\texttt{L}}(t)$, medium $O_i^{\texttt{M}}(t)$, and short $O_i^{\texttt{S}}(t)$ outgoing trips, so, we have

$$O_i(t) = O_i^{\texttt{L}}(t) + O_i^{\texttt{S}}(t) + O_i^{\texttt{M}}(t), \quad \forall i, \forall t. \tag{2}$$

As the optimization variables, let $x_i(t)$ and $y_i(t)$ denote the number of e-bikes and regular bikes available at station i and time t. Then, the flow conservation constraint indicates that the outgoing flows from station i at time t should be less than or equal to the incoming flow and available bikes at the station, i.e.,

$$I_i(t) + x_i(t) + y_i(t) \geq O_i(t), \quad \forall i, \forall t. \tag{3}$$

Note that bikes from incoming trips can also be used to satisfy outgoing trips. To optimize the bike usage for the medium trips, for station i at time t, let us define ${I_i}^{\texttt{M,e}}(t)$ and ${O_i}^{\texttt{M,e}}(t)$ as additional optimization variables that determine the e-bike incoming and outgoing medium trips respectively, and ${I_i}^{\texttt{M,r}}(t)$ and ${O_i}^{\texttt{M,r}}(t)$ denote medium trips satisfied by regular bikes. Now, the flow conservation constraint could be further divided to separately enforce the dedicated e-bike and regular flows, i.e.,

$$I_i^{\mathsf{L}}(t) + I_i^{\mathsf{M,e}}(t) + x_i(t) \geq O_i^{\mathsf{L}}(t) + O_i^{\mathsf{M,e}}(t), \quad \forall i, \forall t, \tag{4}$$

$$I_i^{\mathsf{S}}(t) + I_i^{\mathsf{M,r}}(t) + y_i(t), \geq O_i^{\mathsf{S}}(t) + O_i^{\mathsf{M,r}}(t), \quad \forall i, \forall t. \tag{5}$$

Note that with the formulation of Equations (4) and (5) formulated, Equation (3) becomes redundant, and is stated for the sake of better illustration of the flow conservation constraints. We further emphasize while $I_i^{\mathsf{M}}(t)$ and $O_i^{\mathsf{M}}(t)$ are the inputs to the optimization problem, the allocation of medium trips to regular and e-bikes is performed by the optimization problem, and we have

$$I_i^{\mathsf{M}}(t) = I_i^{\mathsf{M,e}}(t) + I_i^{\mathsf{M,r}}(t), \quad \forall i, \forall t, \tag{6}$$

$$O_i^{\mathsf{M}}(t) = O_i^{\mathsf{M,e}}(t) + O_i^{\mathsf{M,r}}(t), \quad \forall i, \forall t. \tag{7}$$

The next constraint determines the evolution of available bikes across time slots, i.e.,

$$x_i(t+1) + y_i(t+1) = x_i(t) + y_i(t) + I_i(t) - O_i(t), \quad \forall i, \forall t. \tag{8}$$

That is, the overall bikes available of station i at the next slot $t + 1$ is equal to the aggregation of incoming and available bikes subtracted by the outgoing bikes at t. Similarly, the bike evolution constraints could be further partitioned to determine the number of e-bikes and regular bikes available, i.e.,

$$x_i(t+1) = I_i^{\mathsf{L}}(t) + I_i^{\mathsf{M,e}}(t) + x_i(t) - O_i^{\mathsf{L}}(t) - O_i^{\mathsf{M,e}}(t), \quad \forall i, \forall t, \tag{9}$$

$$y_i(t+1) = I_i^{\mathsf{S}}(t) + I_i^{\mathsf{M,r}}(t) + y_i(t) - O_i^{\mathsf{S}}(t) - O_i^{\mathsf{M,e}}(t), \quad \forall i, \forall t. \tag{10}$$

Equation (9) (resp. Equation (10)) states that the e-bikes (resp. regular bikes) available at $t + 1$ is equal to the aggregation of the incoming long (resp. short) trips, medium trips dedicated to e-bikes (resp. regular bikes), and the available e-bikes (resp. regular), subtracted by the outgoing long (resp. short) and medium trips handled by e-bikes (resp. regular). The last set of constraints simply enforce the availability of bikes in each slot.

$$x_i(t) \geq 0 \quad \forall i, \forall t, \tag{11}$$

$$y_i(t) \geq 0 \quad \forall i, \forall t. \tag{12}$$

Finally, let $x_i(1)$ and $y_i(1)$ denote the number of bikes available at the beginning in each station, then the optimization objective is to minimize the number of regular and electric bikes at the first slot, while respecting the constraints. Hence, the optimization problem that determines the optimal mix between the regular and electric bikes (called `OptMix`, hereafter) could be formally formulated as

$$\begin{aligned} [\texttt{OptMix}] \quad \min \quad & \sum_{i=1}^{n} x_i(1) + y_i(1) \\ \text{s.t.,} \quad & \text{Equations (1) – (12)}, \\ \text{vars.,} \quad & x_i(t), y_i(t), I_i^{\mathsf{M,r}}(t), I_i^{\mathsf{M,e}}(t), O_i^{\mathsf{M,r}}(t), O_i^{\mathsf{M,e}}(t), \quad \forall i, \forall t. \end{aligned}$$

We use our enhanced bike datasets, each representing a certain percentage of ride substitution demand, as the input to this optimization problem; since each dataset represents one year of data, our approach assumes that demand estimates for the entire year are available in advance to solve this optimization problem. Our optimization approach yields the minimum possible number of bikes to fully cover the trips in the dataset.

Note that the actual number of bikes deployed in the system needs to be *greater* than this minimum (lower bound) solution to reduce the so-called blocking probability—where a user arrives at a bike station and finds the station empty. Bike share systems usually *overprovision* the number of bikes well above the minimum levels needed to match estimated demand in order to reduce the blocking probability of turning away bikers, especially since real-world demand will not exactly follow this estimated demand and will exhibit real-time stochastic variations.

In addition, bike-sharing systems often rebalance (or reposition, used interchangeably) bikes twice a day to cater to demand surge or depleted stations [36], as discussed in detail in Section 6. Hence with repositioning, the number of available bikes in each station changes and the bike evolution constraints (Equations (9) and (10)) will need to be modified to account for arrivals, departures, *and* repositioning within the system. To capture the effects of repositioning, we can run the `OptMix` problem in each repositioning period separately. Specifically, let $\mathcal{P} = \{1, \ldots, P\}$, be the set of repositioning periods, each indexed by p. To find the overall optimal mixture of regular and electric bikes, we solve P instances of `OptMix`, separately. In addition, let $x_i^p(1)$ and $y_i^p(1), i \in \mathcal{S}$ be the optimal number of e-bikes and regular bikes for the p-th instance of `OptMix` that takes inputs from the p-th repositioning period. In addition, $\texttt{OptMix}(p) = x_i^p(1) + y_i^p(1)$ is the optimal heterogeneous mix of bikes that will satisfy all bike trip demand within repositioning period p. Given the optimal solution of all instances, and to compute the global optimal mix of bikes, we take the maximum across all instances, i.e.,

$$x_i^\star(1) = \max_{p \in \mathcal{P}} x_i^p(1), \quad y_i^\star(1) = \max_{p \in \mathcal{P}} y_i^p(1), \quad \texttt{OptMix}^\star = x_i^\star(1) + y_i^\star(1), \tag{13}$$

where $x_i^\star(1)$ and $y_i^\star(1)$ are the maximum number of e-bikes and regular bikes required across all repositioning periods, and $\texttt{OptMix}^\star$ is the global optimum of the `OptMix` problem across all repositioning periods, and we have $\texttt{OptMix}^\star \geq \max_{p \in \mathcal{P}} \texttt{OptMix}(p)$.

4.2 Cost Analysis

Given the above optimization approach, we use our enhanced bike dataset to compute the total minimum number of bikes needed to handle a certain level of ride substitution, as well as the mix of regular and electric bikes. Unless specified otherwise, we assume $T_{\texttt{short}}$ = 2km and $T_{\texttt{long}}$ = 5km, which implies that very short rides of 2km or less are satisfied using regular bikes, rides longer than 5km are satisfied using electric bikes and all medium rides of 2 to 5km can be satisfied using either type of bike based on user preference. Our analysis below assumes a rebalancing frequency of 8 hours (twice a day) - Section 6 analyzes the impact of other rebalancing periods. Since the CitiBike system sees peak demand during the morning and evening rush hours (see Figure 6b), we assume that bikes are repositioned once after the morning rush to fill in depleted stations, and once more after the evening rush hours [36]. We run our optimization algorithm for various levels of ride substitution ranging from 1% to 10% using the enhanced datasets, and compute the total number of bikes needed to satisfy the corresponding demand as well as the relative proportion of electric and regular bikes. Figure 8b depicts our results.

Note that the original CitiBike dataset has around 18 million trips. Ride substitution of 5% (i.e. 5% of 354 million TLC rides) adds around 17.7 million additional bike trips and doubles the aggregate bike trip demand. Ride substitution of 10% adds 35.4 million additional bike trips to the dataset, and triples the demand. As shown in Figure 8b, the total number of bikes grows sub-linearly with increasing demand from increased ride substitution. The graph yields the following observations.

First, we observe that for 5% ride substitution—which doubles the total number of bike trips—the total number of bikes increases to 13,348 (from 9,199), roughly a 52% increase. For 10% ride substitution—which triples the total number of trips—the total number of bikes increases to 19,065, a 117% increase.

This sub-linear increase in bike share capacity to handle a linear increase in demand is an interesting result and attractive from a greening perspective. The primary reason for this sub-linear increase is due to the skewed nature of demand that causes some bike stations to see high demand while others have slack. As overall demand increases, heavily utilized bike stations need additional bikes to handle the extra demand, but less utilized stations can absorb some of the extra demand using available slack rather than requiring extra bikes. Similarly, any increase in demand during off-peak hours typically does not require a proportionate increase in bikes, since there

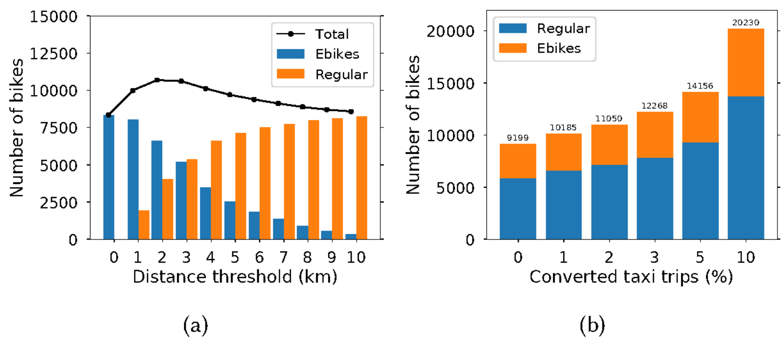


Fig. 8. (a) Distribution of e-bikes and regular bikes at varying distance thresholds. (b) Bike capacity enhancement to service varying taxi trip substitution

is surplus slack available in the system during off-peak periods. As a result of these spatial and temporal skews in usage, a linear increase in demand requires a sub-linear increase in bike capacity. Note that our optimization computes the minimum (lower bound) on the number of bikes and yet benefits from slack in the system. In a real deployment with over-provisioning over this lower bound, the actual slack is even higher. Note that while this sub-linear increase may not be the main driver for expanding an existing bike fleet, our main focus is on ride substitution and each bike trip added results in a net positive i.e. there are lower emissions, better space management due to reduced parking space and less traffic on roads.

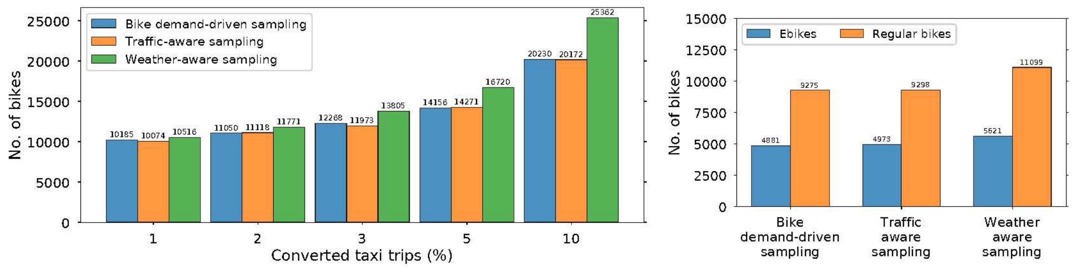
Second, we observe that the minimum number of bikes needed to handle the baseline demand (depicted as 0% ride substitution) is 9,199. The current CitiBike system (Table 1) has 17,954 bikes, which implies that the system is already over-provisioned by a factor of 2 above the minimum required capacity computed by our optimization. This is in order to minimize blocking probability of turning away users due to unavailable bikes.

A final observation about Figure 8b is that around 26% of the total bikes are electric bikes - a significant saving to the system over using all electric bikes which tend to be more expensive to buy and maintain.

Next, we vary the distance threshold $T_{\texttt{long}}$ that represents the cutoff distance above which all rides are satisfied using electric bikes. To simplify the analysis, we set $T_{\texttt{short}} = T_{\texttt{long}}$ for each experiment run, which yields a single threshold above which electric bikes are used. We vary this threshold from 0 to 10km in steps and run our optimization algorithm to compute the bike capacity in each case.

Figure 8a depicts our results. As shown, a distance threshold of 0km implies an all-electric bike system, while that of 10km implies using mostly regular bikes (with a small number of electric bikes for satisfying ultra long bike rides that exceed 10km). The black curve depicts the sum of the regular and electric bikes needed for each distance threshold value. As can be seen, the total number of bikes needed is a bit higher for intermediate values of the threshold $T_{\texttt{long}}$ than at the extremes. This is because the bike share system uses a mix of bikes for intermediate $T_{\texttt{long}}$ values. In this case, it is not sufficient to have a non-empty bike station when a user arrives to check out a bike - the correct type of bike needs to be available as well (e.g. a non-empty bike station that only has non-electric bikes available still causes blocking if the user wants to use an electric bike for a longer ride). Since we cannot substitute one type of bike with another, the total bike capacity needed for intermediate values is around 30% higher. At the two extremes, most of the demand is serviced by a single type of bike reducing the need for such "over-provisioning".

The figure shows that for $T_{\texttt{long}}$ = 3km (roughly equal to the median trip distance), the number of regular and electric bikes needed is equal. For $T_{\texttt{long}}$ = 5km, the percentage of electric bikes needed drops to 26%. This result



(a) (b)

Fig. 9. (9a) Number of bikes resulting from different sampling strategies at varying substitution levels, and (9b) number of ebikes and regular bikes resulting from different sampling methods at 5% taxi trip substitution.

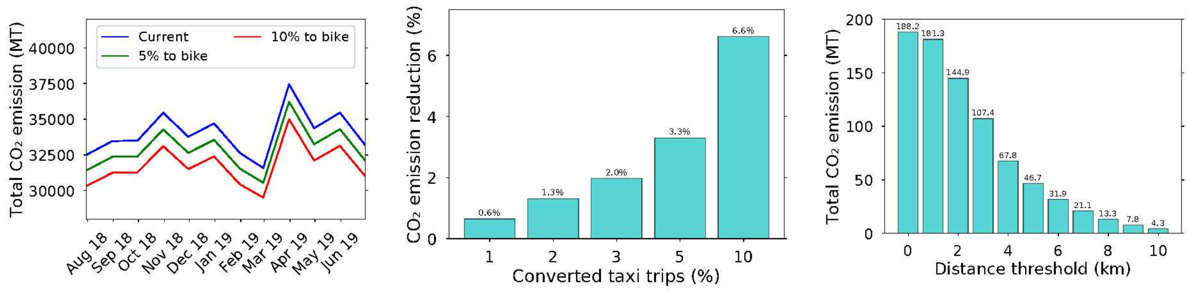
indicates that a hybrid system can yield significant cost savings over an all-electric bike system by requiring between a quarter and a half of the total number of bikes to be electric bikes.

4.2.1 Impact of Other Sampling Strategies. Our results above are based on our baseline strategy driven by bike demand. Next, we consider two other sampling strategies for ride substitution to understand how other substitution strategies impact out results.

(1) **Traffic-aware sampling.** For this strategy, we compute the hourly PMF of taxi trip demand and sample trips to substitute based on this PMF since during peak traffic hours, more people are likely to switch to bike trips to avoid traffic congestion.
(2) **Weather-aware sampling.** For this strategy, we gather weather data at hourly granularity (weather data at this frequency was not previously used in the paper), and perform sampling of taxi trips to be substituted as follows. During each hour, we check the weather observed during that hour and if rain or snow, we sample taxi trips taken during that hour at zero probability (i.e. we do not sample trips during rain or snow at all). For all other hours, we follow the original bike trip distribution to determine sampling probability.

We run our optimization on the generated set of trips based on these sampling strategies and compare the resulting number of bikes with the original sampling strategy driven by the hourly distribution bike trip demand. Figure 9a depicts the results of this analysis. At low taxi substitution levels, the impact of different sampling strategies on the number of bikes is fairly small (3.2 and 6.5% for 1 and 2% substitution between bike demand driven sampling and weather-aware sampling respectively). As the level of substitution increases, the impact of different sampling strategies also increases, with the maximum impact being 25% for 10% ride substitution between bike demand driven sampling and weather-aware sampling respectively. As shown in Figure 9b, the impact on the number of electric vs regular bikes is also small (15.2% and 19.7% at 5% substitution for regular and electric bikes respectively). Usually, bike share systems are over-provisioned to reduce the probability of users missing a bike when they need one e.g. the CitiBike bike share system is already over-provisioned by 2× the number of bikes required to satisfy demand. Since the number of bikes is within 11.8% of each other on average for all sampling strategies considered, the resulting number of bikes will not differ by much regardless of the sampling strategy used.

Key Takeaways. First, a linear increase in ride substitution results in a sub-linear increase in bike capacity, due to available slack in the bike system. When the number of bike trips is doubled, it results in a 52% increase in bike capacity. Second, although the overall number of bikes may be higher in a hybrid bike system compared to an all-electric bike system, it can still yield cost savings as the overall number of e-bikes drops to 26% of total



(a) (b) Distance Threshold=5km (c)

Fig. 10. (10a) Monthly emissions from taxis, FHVs and electric bikes under various ride substitution scenarios. (10b) Overall carbon emission reduction under different ride substitution scenarios. (10c) Overall carbon emission in the current bike share system under varying distance threshold.

capacity. Lastly, various sampling strategies yield similar results (within 11.8% of one another on average) for bike capacity needed to handle ride substitution.

4.3 Carbon Analysis

Next, we analyze the carbon emission reductions from substituting car rides with bike rides. Our methodology for estimating the carbon emissions from cars and bikes is as follows.

To compute the emissions from a car trip, we assume the gasoline consumption rate of taxis and FHVs is equivalent to the USA average consumption rate for passenger vehicles, which is 22.3 miles per gallon (9.48 km/litre) [3]. The amount of gasoline consumed for a specific trip can be converted into equivalent CO_2 emissions as follows.

$$\text{CO}_2 \text{ emissions per trip} = \frac{\text{CO}_2 \text{ emissions per litre} \times \text{total trip length}}{\text{km per litre}}. \tag{14}$$

When a car trip is substituted with a non-electric bike, we assume that CO_2 emissions are zero for that bike trip. When a car trip is substituted with an electric bike, we need to estimate the small amount of emission from charging the battery. To compute the carbon emission, we assume the average e-bike energy consumption of 7.9 kWh/km [20], which captures a broad range of energy consumption scenarios from when a user starts an e-bike trip, to reaching maximum speed and coasting such that the motor is no longer turning. We then convert the electricity consumed during a bike ride into equivalent CO_2 emission as follows.

$$\text{CO}_2 \text{ emissions per trip} = \frac{\text{CO}_2 \text{ emissions per kWh} \times \text{total trip length}}{\text{km per kWh}}. \tag{15}$$

Finally, we use the Greenhouse Gas Equivalencies Calculator CO_2 emission factors of electricity and gasoline [2]: (1) Electricity: 0.486 MT/MWh, and (2) Gasoline: 0.00234 MT/litre.

Figure 10a depicts the monthly carbon emissions from taxis and FHVs for all trips during that month. The blue curve labelled "Current" depicts the existing CO_2 emission footprint with no ride substitution. Since the trips see a seasonal drop during winter months, the emission footprint is also lower in the winter than in other seasons. Overall, the CO_2 emission ranges between 32,000 and 35,000 MT per month over a 12 month period. The figure shows that there is a noticeable drop in CO_2 emission under a 5% and 10% ride substitution scenario.

Figure 10b depicts the percentage reduction in CO_2 emission under different ride substitution scenarios. As shown, 5% ride substitution yields a 3.3% reduction in CO_2 emission, while a 10% ride substitution yields 6.6% reduction in CO_2 emission. Even a small 2% ride substitution yields 1.3% savings in CO_2 emission. These results show the greening potential of ride substitution.

Interestingly, the reduction, while good at an absolute level, is sub-linear with respect to the percentage of ride substitution. This is due to the nature of sampling used to determine trips for ride substitution. If we had uniformly sampled the TLC dataset, a certain percentage of trips would yield a corresponding reduction in miles driven and an equivalent reduction in CO_2 emission. However, since the sampling is biased towards short and medium distance car trips less than 5km, these shorter rides yield a lower reduction in total miles driven (by virtual of being short), and a sub-linear reduction in CO_2 emission.

There are two reasons why these estimates are likely to be conservative. First, a reduction in trips reduces traffic congestion and brings additional benefits to other vehicles (e.g. faster moving traffic, less idling at traffic lights), which reduces carbon emissions. Second, as noted in Section 1, taxis and FHVs consume dead miles driving between hailed trips, which increases CO_2 emission by 47% as per a study [8]. Our analysis does not account for these dead miles, which will subsantially increase the carbon emissions reported in Figure 10b.

Finally, Figure 10c shows CO_2 emissions from the bike share system. The distance threshold results in different number of electric bikes, and hence different levels of emission. As shown, even for a threshold of zero which yields an all-electric bike system, the total emissions for a year are 188MT, which is negligible compared to the 32,000MT of taxi and FHVs. The result highlights the benefit of using a zero or low carbon form of ride sharing for greening the overall ride share ecosystem.

Key Takeaways. The overall emissions represented as the sum of taxi and e-bike emissions see a reduction of 7% under 10% ride substitution. Under this scenario, e-bikes account for only 0.05% of the emissions, indicating substantial carbon benefits from greening ride shares.

5 EXPANDING A BIKE SHARE SYSTEM

Having explored how existing taxi trips can be substituted with a combination of regular and electric bikes, we now study how to effectively expand the bike share system and analyze its impact on the overall carbon emission of the ride-sharing ecosystem. We first present our methodology for our analysis, followed by our results.

5.1 Research Methodology

We begin with the TLC dataset since it represents ride-share trips that can potentially represent bike demand in the expanded system. Since not all taxi rides can be substituted as bike rides, we sample the dataset with the following constraints:

- We only sub-sample the taxi rides with pickup and drop-off locations outside the existing bike share system. In other words, we exclude the taxi trips that are within the 300m range of the current bike share system.
- We exclude the long trips that are likely not substitutable by bike trips and only consider ride trips less than a threshold distance (e.g., 12km).

We construct our expanded bike-share dataset using the above methodology. Table 3 describes the characteristics of the expanded bike-share system dataset and adds 3.89 million additional trips to the CitiBike dataset.

Next, we use the additional trips as input to our data-driven methodology for expansion. Our methodology for intelligent expansion of the bike share system includes four major steps. *First*, we identify geographical locations with high ride-sharing activity that are outside the current bike coverage (Section 5.1.1). Second, we estimate the number of stations required in these locations. *Third,* we devise an algorithm to determine the minimum number of bike stations that ensures maximum coverage of the area. (Section 5.1.3). *Finally,* we determine the optimal number of regular and electric bikes in the expanded stations by solving the `OptMix` problem (Section 5.1.4).

5.1.1 Clustering the Geographical Areas outside the Current Coverage. The main goal is to determine the geographical regions with high ride-sharing demand, as they may have a higher potential of conversion if bikes are more accessible. To do so, we first cluster the taxi trips to find the areas with high ride traffic. Our algorithm is built upon DBSCAN [17], a density-based clustering algorithm that groups points based on proximity. At a high level, DBSCAN only clusters points at high-density regions. Points that lie in low-density regions far from other clusters do not belong to any group. DBSCAN uses two main parameters to estimate density-based clusters: *first,* an application-specific distance threshold ϵ which is a radius within which points should be in order for them to considered neighbors and within a cluster. The *second* parameter is `minPts`, a threshold for the minimum number of neighbors for a point to form a cluster. We now explain how DBSCAN constructs a cluster. The points that have more than `minPts` neighbors within the radius ϵ are considered as the *core points*. All points within the radius ϵ of a core point are *directly density-reachable* points. If a directly density-reachable point is again a core point, all its neighbors are progressively added to the cluster, and are said to be *density-reachable*. All these points together are said to be *density-connected*, and form a cluster in the data. DBSCAN therefore alienates dense areas from areas of lower density, and ϵ could be interpreted as the gap between two clusters. The final output from DBSCAN is spatially labeled clusters representing geographical regions with high ride density.

To apply DBSCAN to our dataset, we need to determine the values of `minPts` and ϵ. Our parameter choices are as follows. We set ϵ = 300m, as the threshold distance between points to ensure that a bike station is available within walking distance from a trip's pickup location. We then set the `minPts` parameter to 10,000, which is at least 2 times the average number of trips per bike station in CiteBike share dataset in the one year period of analysis. Finally, we use ELKI [42], an open-source software library containing the implementation of DBSCAN to cluster the potential areas for expanding the bike share system.

5.1.2 Determining the Number of Bike Stations for Expansion. Next, we estimate the number of stations needed in each cluster. In our approach, we assume the coverage density provided by stations in the expanded bike-share system will be similar to the existing CitiBike system. Hence, we use the geographical information, the ride demand, and the number of stations in the existing system to estimate the coverage density of future stations. Specifically, we calculate the cluster density d, which represents the number of stations per unit area, of the CitiBike dataset. We then estimate the number of stations k_j in cluster $j \in \mathcal{K}$ by dividing the total area of the cluster by the density d.

5.1.3 Station Placement Algorithm. The third step is to develop a bike station placement algorithm with the geographical clusters $\mathcal{K}$ and k_j as the number of stations for cluster $j \in \mathcal{K}$ as the input. The objective is to place k_j bike stations in cluster j such that the total number of trips that could be covered by at least one station is maximized.

We use the following algorithm for the bike station placement. Since the coverage area of a cluster might be an arbitrary polygon, we first convert the corresponding polygon of the cluster into a square by first computing the bounding rectangle, i.e., the smallest rectangular polygon that contains the polygon, and then deriving a square of length l where l is the length of the longer side of the rectangular polygon, and the rectangular polygon is contained in the square. Then, the station placement mechanism works as follows: compute $n = \lfloor \sqrt{k_j} \rfloor$, and split the square into $n \times n$ cells. In each cell, place a station randomly (in practice the placement within each cell might come with the several practical constraints, and our approach just indicates the area in which the station should be placed). Due to the rounding process, we might have some remaining stations to be placed. In this case, we distribute them across the whole square region at random. Finally, we discard all placements which do not cover a single trip to remove the extra points brought about by transforming the original polygon into a square.

5.1.4 Optimal Number of Regular and Electric Bikes. The last step is to determine the number of bikes for each station. Towards this, we first use a preprocessing analysis on the taxi trips and assign them to the final set of

Table 3. Number of stations, trips, bikes and carbon reduction in the expanded system.

	Current System	Expanded System
No. of stations	941	1761
No. of trips (millions)	18.98	22.86
No. of e-bikes	3,337	4,361
No. of regular bikes	5,862	7,949
Total bikes	9199	12310
Total annual e-bike CO_2 emission	47 MT	100 MT
Total annual taxi CO_2 emission	183,648 MT	168,835 MT

bike stations and calculate the incoming and outgoing flows of the stations, and then run the `OptMix` problem in Section 4 to determine the optimal mix between regular and electric bikes to meet the demand.

5.2 Cost Analysis

We now evaluate the performance of our expanding strategy for the existing bike share system to cover new areas. Similar to the previous section, we consider short rides of 2km or less use regular bikes, rides longer than 5km use e-bikes, and medium rides between 2 to 5km are satisfied using either e-bikes or regular bikes. The expanded bike sharing system adds an additional 3.89 million trips, which is around 20% of the original CitiBike dataset, in neighborhoods that have no bike stations. We use the expanded bike sharing dataset to compute the number of stations and bikes required to satisfy the demand.

We begin by highlighting the regions with ride-share traffic for one month (see Figure 14a in the appendix). The figure depicts trips originating from areas outside the coverage of the current bike system. The green markers show the snapshot of the taxi trips' pickup points, and the red markers are the existing bike stations. It reveals that most stations are centered around Manhattan and parts of Brooklyn. Our goal is to expand the bike share system's coverage into other areas indicated by the green markers.

Figure 14b (included in appendix) depicts our density-based clustering results. The current CitiBike system shows that most stations are centered around Manhattan and parts of Brooklyn. Our algorithm identified 31 high-density clusters outside these areas, and the overall trips within each cluster ranged from 10,000 to 2.5 million trips in size. The black markers depict points that lie in low-density regions, distant from other clusters, and thus, not classified into any cluster. In total, the number of trips that do not belong to any cluster is 1.2 million (about 13% of the uncovered region). We exclude these points in our analysis.

As seen in the figure, newer stations are placed in neighborhoods such as Upper Manhattan, Bronx, Queens, among others, that have higher ride-sharing traffic. The newer stations identified by trip clusters are placed in neighborhoods such as Upper Manhattan, Bronx, Queens, among others, that have higher ride-sharing traffic. These neighborhoods have an average number of 4,766 bike trips per station and lie close to the boundaries of the existing bike system. Since they can enhance the connectivity of the existing system, stations in these neighborhoods could be likely candidate locations for expansion. We observe that stations are also located in neighborhoods such as Coney Island Beach, which is distant from the existing CitiBike stations. Note that our expanded bike-share dataset contains taxi trips that are feasible for bike substitution (i.e., trip distance less than 12 km). Since the average number of bike trips per station in these geographically spread out regions is close to the other places, these locations show bike-share potential, even though they are distant from other neighborhoods.

A second key observation is that the number of new bike stations required is close to 88% of the existing capacity. Although this number is substantial, the expanded bike-share system also has high coverage. If we assume the new stations can provide service at a radius of 300m, it has a coverage area of 113 km^2, which is similar to the coverage provided by the existing system. Overall, our results show that 820 stations are added to

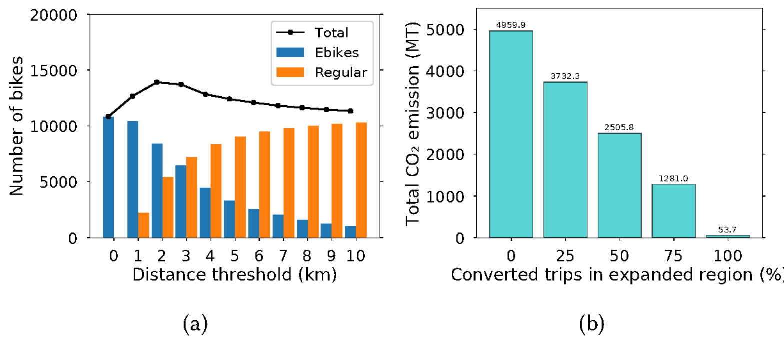


Fig. 11. (11a) Number of bikes for the expanded bike system with different distance threshold. (11b) Overall carbon emissions for different bike expansion scenarios.

the existing bike-share system. Table 3 shows the overall statistics of the expanded bike station, which is the sum of current and the new system.

We also see a noticeable increase in the number of bikes required to meet demand in the expanded system. Previously, we discussed a 5% ride substitution (i.e, 17.7 million trips) increased the overall bike capacity by 52%. However, in the expanded bike share system, which has 3.89 million additional trips, the overall bike capacity increased by 14.5%. Note that the utilization, that is, the number of trips per bike is 1,772 in the expanded bike-share system, which is about 950 less trips than the ride substitution scenario. This is because there is enough slack available in the existing system to compensate in the ride substitution scenario. But, when expanding to new regions, since there are no existing bikes, we need more bikes to meet the demand.

In the next experiment, we analyze how the distance threshold affects the mix of bikes in the expanded bike-share system. We note that a distance threshold of 0km implies an all-electric bike-share system while that of 10km represents a bike-share system with mostly regular bikes. Figure 11a depicts our results. As shown in Figure 11a, the black curve depicts the overall number of bikes needed to satisfy the demand for each distance threshold value. It reveals that the intermediate threshold value requires a higher number of bikes than at the extremes. This is because the correct type of bike needs to be available in the station to meet different trip types, e.g., long trips will need e-bikes. We also observe that an all-electric bike-share system will require 10,842 e-bikes. However, with the distance threshold of 3km, the number of e-bikes required is 6,494. This indicates that we can achieve significant cost savings if we consider a hybrid bike system in the expanded bike-share system.

5.3 Implications of Our Results

We note that the cost of expansion into previously uncovered areas, while higher than expanding an existing system, is reasonable since adding 3,111 bikes and 820 stations gives the same geographical coverage as the existing system. We also emphasize that the main focus of our study is ride substitution, and for every bike trip we add, a car trip is reduced. The benefits of this are multifaceted, and far outweigh the cost of expansion. For the end user, the ability to travel the same distance at lower cost and sometimes faster speed than a car ride is realized. At the same time, there is reduction in parking space, road traffic and CO_2 emissions. Lastly, increased bike usage results in increased profit for the bike share provider.

Key Takeaways. Expanding a bike share system to new areas incurs a higher investment cost, in terms of number of new bikes and bike stations needed. Our expanded bike system provides service to an area that similar to the existing system, and adds an additional 820 stations and 3,111 bikes. Some newer stations are located close to the existing bike stations, and could potentially enhance the connectivity of the current system.

5.4 Carbon Analysis

We now analyze the emissions from expanding the bike-share system. We evaluate the overall emissions from taxis and e-bikes in the expanded dataset for various levels of ride substitution ranging from 0% to 100%. Figure 11b depicts our results. We note that the expanded bike dataset has an additional 3.88 million trips and, as seen in the figure, responsible for 4960MT carbon emissions. As we increase the ride substitution levels, the carbon emissions also reduce significantly due to near zero emissions from bikes. The results also reveal that there is a significant benefit to expanding bike-share systems even if a fraction of the taxi rides are substituted. For example, the emissions are reduced by a quarter if 25% of car trips, (i.e., 0.95 million trips) are replaced by bikes. The emissions are proportionately reduced for different level of ride substitution. Lastly, it is worth noting that our emission estimates are conservative, as we do not account for emissions due to the deadheading of FHVs. As noted previously, deadheading accounts for 47% of the miles logged in FHVs. Hence, we expect to see a much higher overall reduction if we see an uptake of bikes in the expanded region.

Key Takeaways. There is significant potential greening benefit from expanding the existing bike station. Carbon emissions drop from 4960 MT to 3700MT, when we consider 25% ride substitution in our expanded dataset. To put this in perspective, this is equivalent to a 141,780 gallons reduction in of gasoline consumption [2].

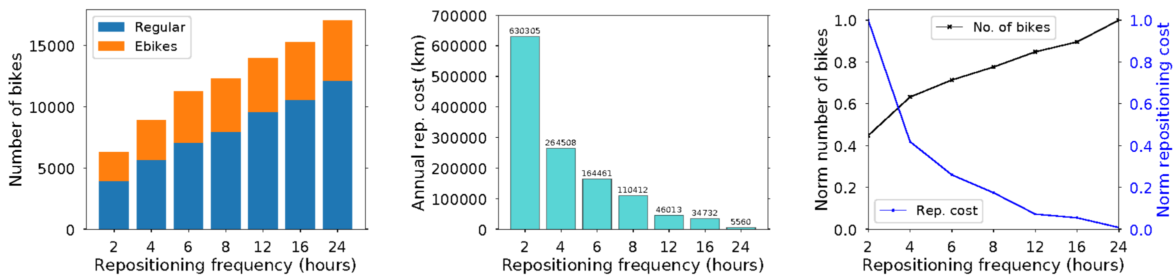
6 COST AND CARBON BENEFITS OF BIKE REPOSITIONING

Asymmetric distribution of user demand creates an imbalance in the bike system, leading to depleted bikes in one station and overcapacity in another. This imbalance pattern may occur in residential and commercial areas in the morning where riders may ride bikes to work, causing residential areas to be devoid of bikes. To solve this imbalance problem, trucks or bike trailers are often used to *reposition* bikes within the system. There have been many studies that focus on mitigating imbalance in bike stations [11, 18, 21, 30, 34, 36, 44]. Here, we focus on understanding the cost of repositioning bikes from the emissions standpoint. Moving bikes involves trucks or trailers that produce emissions. Hence, we seek to answer the following key research questions: (i) What is the cost of repositioning bikes, i.e., the distance traveled by trucks to move bikes, and how does it impact bike capacity if we reposition frequently? (ii) How much carbon do trucks emit in repositioning bikes? (iii) How does repositioning impact the overall carbon emission of the rideshare ecosystem?

6.1 Research Methodology

In Section 4, we formulated the `OptMix` problem to compute the optimal mix of bikes needed at the start of each rebalancing period (i.e., *initial state*) at each station. At the end of each run, we get a *final state*, which represents the number of bikes available at each station after the simulation. Recall that we implicitly assume bike repositioning between each rebalancing period, where the number of bikes changes from the final state to the initial state of the next period. Here, we describe our approach to calculate the cost of repositioning bikes within stations.

We define the cost of repositioning as the distance traveled by trucks to move bikes between stations. It is calculated by determining the distance traveled in moving the bikes in the final state to resemble the initial state of the next rebalancing period. Our algorithmic approach to minimize this cost is as follows. We first calculate the *surplus* and *deficit* stations by taking the difference between the current final state $S_p^{\texttt{final}}$ and the initial state $S_{p+1}^{\texttt{initial}}$ of the system, where p and $p+1$ denote the current and next rebalancing period, respectively. A station



(a) Number of bikes (b) Repositioning cost (km) (c) Comparing the normalized costs

Fig. 12. Cost analysis of bike repositioning.

has surplus bikes if the final state has more bikes than it is the initial state at the next period i.e., $S_p^{\mathtt{final}} > S_{p+1}^{\mathtt{initial}}$. Similarly, when $S_p^{\mathtt{final}} < S_{p+1}^{\mathtt{initial}}$, the station is said to be in deficit.

We then use a simple greedy-based approach to relocate bikes from surplus to deficit stations and assume no constraints on the number the available trucks for moving bikes. Although in practice, most real-world bike-sharing systems deploy around five vehicles for repositioning [36], here, we are only focus on determining the cost of moving the bikes from one station to another.

Our greedy algorithm works as follows. We create a sorted list of *surplus stations* based on the number of surplus bikes in descending order. We then iterate over this sorted list, and for each station, distribute the surplus bikes to the closest deficit station. We repeat the above process until one of the following conditions is met: (i) there is no station with a surplus of bikes or (ii) no other stations have a deficit of bikes.

The above conditions ensure that there is no room left for distributing bikes among the stations. We also enforce a capacity constraint on the number of bikes that repositioning vehicles can carry. If the number of bikes exceeds the capacity, the algorithm uses an additional vehicle to move the bikes. We calculate the overall cost of repositioning by summing all the distances traveled by the vehicles to rebalance the system.

We note that the repositioning cost may differ depending on when bikes are repositioned. This is because the number of bikes at stations is at constant flux throughout the day; thus, the repositioning cost will also vary accordingly. For example, there might be more repositioning required when it is done right before a peak period compared to a non-peak period. To compensate for this discrepancy, for each repositioning frequency, we shift the start period by 1 to 10 hours and report the average of all runs as the final cost of repositioning.

Lastly, we note that the proposed greedy algorithm could be extended to consider more sophisticated settings. For example, we could impose a limit on the number of vehicles used for repositioning or assume a more dynamic scenario, where trucks can move bikes to multiple locations at a time. Our approach here is the first step to provide a cost versus carbon tradeoff analysis of repositioning.

6.2 Cost Analysis

We now evaluate the cost of repositioning and the impact of repositioning on the bike capacity. We assume vehicles can carry a maximum of ten bikes at a time. We run our experiments on different repositioning frequency, which represents how often bikes are rebalanced among stations.

We first analyze the impact of repositioning on the overall bike capacity. Figure 12a shows the number of bikes needed to meet the demand with varying repositioning frequency. We observe that the number of bikes decreases as repositioning frequency increases. This is because when we rebalance more frequently, it reduces the probability of having empty bike stations. Any bike demand is promptly met due to frequent movement of bikes by repositioning vehicles. Conversely, if we do not rebalance bikes frequently, we will require more bikes to meet demand surges or asymmetric usage behavior. As shown in the figure, when stations are rebalanced every two hours, the overall bikes needed to meet all demand is 6,324. As we increase the frequency to the bare minimum of once per day, the number of bikes in the system increases three-fold. Thus, a system that rebalances frequently will need a much smaller bike capacity to operate. Of course, in practice, it is much difficult to predict the demand, which is why most bike-sharing systems are over-provisioned to ensure availability. This also reduces the need to rebalance frequently.

Next, we examine the cost of repositioning in terms of the total distance traveled by the repositioning vehicles. Figure 12b depicts the results. The figure shows that a decrease in repositioning frequency decreases the cost of repositioning significantly. The primary reason for this pattern can be attributed to asymmetric user demand. Daily commuters may pick bikes up in the morning to commute to work and return to the same station in the evening. Similarly, a commuter may pick bikes up and ride it to a depleted station. As a result of this, some stations may get replenished, reducing the need for trucks to balance the deficit. This *auto balancing* effect may reduce the movement of bikes within stations. In particular, our results show that the total annual repositioning cost of bikes is >600,000km, if we rebalance every 2 hours, and decreases by 99% if we rebalance every 24 hours.

Interestingly, the figure also indicates a significant drop in repositioning cost when the frequency increases from 2 to 4 hours. Since the vehicles travel less distance under the 4-hour scenario, it suggests that there is no need for trucks to rebalance frequently, as the system partially *auto balances* itself — presumably due to riders returning bikes to the same station. However, we see diminishing returns in repositioning cost as we increase the repositioning frequency beyond 12 hours — as a result of the self-balancing demand pattern discussed earlier.

Lastly, there is a clear tradeoff between the cost of repositioning and the number of bikes as the repositioning frequency changes. More precisely, repositioning bikes frequently reduces the overall bikes in the system — a capital expenditure (CapEx) benefit. On the downside, however, it increases the operational expenditure (OpEx) associated with repositioning. To analyze the CapEX vs. OpEx tradeoff, we normalize the data with their maximum and report the results in Figure 12c. As seen, the normalized number of bikes, depicted by the black line, increases when the system rebalances less often. However, we see that the normalized repositioning cost decreases as we increase the repositioning frequency because bike trailers travel less distance.

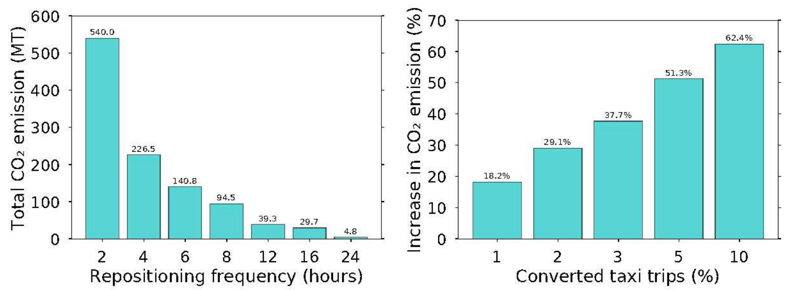
Key Takeaways. Repositioning bikes less frequently increases the overall number of bikes needed in the system. For example, repositioning bikes once per day can result in a three-fold increase compared to repositioning every 2 hours. However, the cost of repositioning is high if done frequently. Vehicles may travel an estimated 600, 000 miles annually if the repositioning frequency is 2 hours. This will diminish the green benefit of bike-share systems. We can reduce the repositioning cost by 73.9% to 82.4% if we reposition every 6-8 hours, and this does not increase the overall number of bikes adversely.

6.3 Carbon Analysis

We evaluate the impact of repositioning frequency on carbon emissions generated by repositioning vehicles. In our analysis, we assume the repositioning vehicles' mileage is similar to the national average fuel economy of trucks of 6.5 miles per gallon [6] and use the methodology in Section 4 to calculate the carbon emissions.

We begin by analyzing the emissions from repositioning on the CitiBike dataset. Figure 13a depicts the results. The results show a 58% drop, i.e., from 540MT to 226MT, in carbon emissions when we vary repositioning frequency from 2 to 4-hour periods. This rapid drop observed in the total distance traveled by the vehicles as reported in Figure 12b is due to the *auto-balance* effect explained in Section 6.2. At the extreme, we see a drop in emissions to 4.8MT, when the repositioning frequency is 24 hours.

Another key observation is that the annual emission footprint of e-bikes only is 47MT (Table 3), whereas that from repositioning vehicles under the 8-hour period is 94.5MT. This indicates that emissions from repositioning



(a) (b)

Fig. 13. (13a) Total emissions from repositioning operations. (13b) Increase in repositioning emissions with increase in converted taxi trips.

can overshadow the e-bikes' emission footprint. However, this number is still small (0.05%) compared to emissions from taxi rides. Even if we sum the emissions from e-bikes and repositioning vechiles, they represent less than 0.1% of the emissions from taxi rides.

Finally, in Figure 13b, we report the increase in carbon emissions under varying ride substitution scenarios ranging from 1% to 10%. In this experiment, we use the fixed repositioning frequency of 8 hours. Our results show that a 10% ride substitution increases the repositioning emissions by 62% (i.e., from 94MT to 154 MT) compared to the no-ride substitution scenario. However, we note that this additional emission is less than 0.1% of the emission reduction due to transitioning from taxi to bike, as depicted in Figure 10b.

Key Takeaways. The annual carbon emissions from repositioning vehicles can be twice as much as emissions from riding e-bikes. However, they still represent a small emission footprint compared to taxis.

7 RELATED WORK

The growth of bike-sharing systems has motivated significant research interest in analyzing usage patterns [7, 29, 31, 39], optimizing infrastructure [12, 13], and improving service availability [19]. Studies analyzing bike infrastructure at a city-scale provide insights into reducing congestion, pollution, improved public health, and land use [35]. This, in turn, enables policymakers to invest and transition into sustainable transportation. Separately, there have been multiple surveys that indicate e-bikes displace other modes of transport and are often used for long-distance commute [14, 37, 39]. The analysis in [37] indicates that users find e-bikes more fun to use, thereby increasing the trips made on bikes. While these studies have looked at other aspects of bike-sharing systems, our work is the first to quantify the sustainability of bike-share systems by analyzing the conversion of taxi rides to bike rides and factoring in the emissions brought about by electricity production.

Analysis of usage patterns and trip history is especially useful in urban planning that enables the placement of bike stations [12, 13, 48]. For example, prior work uses demand predictions to recommend locations for planning future stations with high demand potential [12]. Extending bike-sharing programs at city-scale requires extensive planning and thus propose the use of data-driven approaches to keep cost low. Data-driven approaches typically rely on characterizing factors (e.g., neighborhood density, point of interest popularity) that impact bike demand mobility. Locations that show high traffic demand indicate the potential for future bike stations [13]. However, our work's main goal is analyzing carbon reduction by substitution for-pay car rides and therefore uses FHV car trips to predict demand. Our work has also looked at different strategies of trip substitution, e.g., weather and traffic-aware substitution. Our work has also explored the expansion of existing bike-share systems in a multimodal hybrid manner, i.e., by using different distributions to drive trip substitution and using trip characteristics to inform the split of electric and regular bikes.

Rebalancing bikes has also been extensively studied [11, 18, 21, 30, 34, 36, 44–46]. Bike deployment models focus on improving the availability of bikes in stations while reducing the bikes shuffled across stations. Modeling demand surge periods, depleted bike stations, hotspots for bike pick up and drop off can play a key role in rebalancing [36, 44]. Since there is an operating cost associated with rebalancing bikes, most approaches optimize for truck routes with the number of trucks and time of day as additional constraints. For example, in [36], the authors limit the number of trucks for rebalancing and consider different rebalancing strategies during rush hour and night time [36] since traffic during rush hour is vastly different during this period. Differently, [21, 34] propose online re-positioning strategies that consider truck routes and future expected demand. In [46], a data-driven model is proposed to predict the safe rebalancing range for bike stations. In comparison, while prior work focuses on reducing the cost of repositioning bikes by optimizing routes, our work focuses on quantifying the carbon footprint of repositioning activity and analyzing the tradeoff between capital cost and frequency of repositioning. Our work also presents an optimal way of bike-share expansion by finding clusters of high FHV activity with high potential for trip substitution, thereby identifying optimal carbon reduction opportunities by converting car trips to bike trips.

Mobility patterns from taxi cabs [19, 24, 33], and extrinsic sources such as mobile, social networks, demographics, have also been used for modeling and analyzing bike systems [16, 26, 28]. These studies mainly focus on designing and optimizing bike infrastructure systems. However, our approach provides a novel analysis of the carbon footprint in designing hybrid bike-sharing systems.

8 CONCLUSIONS

To green the ride sharing ecosystem, we conducted a data-driven approach to study the feasibility, costs, and carbon benefits of using hybrid electric bike share system as an alternative to ride sharing systems. From the feasibility perspective, our analysis using publicly available datasets showed that a large fraction of car rides is feasible for substitution with bike rides. We also showed that various sampling strategies yield similar results (within 11.8% of one another on average) for bike capacity needed to handle ride substitution. Our analysis showed that only 10% ride substitution reduces the carbon emissions of the entire ride share system by up to 7%, while e-bikes account for only 0.05% of the emissions. We also proposed a data-driven clustering approach to expand the bike coverage area and analyzed the benefits of such expansion from the emissions' perspective. Finally, we studied the cost vs. emissions tradeoff of frequent repositioning of the system to prevent empty or full stations, and our results showed that the annual carbon emissions from repositioning vehicles can be twice as much as emissions from riding e-bikes. However, they are still negligible compared to the emissions from cars.

While our results were specific to New York City, many of our observations apply more broadly to other cities. Biking using hybrid electric bike share systems can be a viable form of transport for short rides in many congested cities, and it can provide a reduction in carbon emissions from cars. As part of future work, we plan to study how the use of electric cars in ride sharing schemes can be used in conjunction with electric bike sharing to further green the ride sharing ecosystem.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their insightful comments. This research was supported by NSF grants 2020888, 1908298, 1836752, and 1645952.

REFERENCES

[1] [n.d.]. About TLC - TLC. https://www1.nyc.gov/site/tlc/about/about-tlc.page. (Accessed on 11/23/2019).

[2] [n.d.]. Greenhouse Gas Equivalencies Calculator. https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator. (Accessed on 10/5/2020).

[3] [n.d.]. Greenhouse Gas Equivalencies Calculator | Energy and the Environment | US EPA. https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator. (Accessed on 04/25/2020).
[4] [n.d.]. How Green is the Sharing Economy? https://knowledge.wharton.upenn.edu/article/how-green-is-the-sharing-economy/. (Accessed on 05/05/2020).
[5] [n.d.]. Ride-hailing increases emissions, contributes to climate pollution - study. https://www.reuters.com/article/us-uber-emissions/ride-hailing-increases-emissions-contributes-to-climate-pollution-study-idUSKBN20J27K. (Accessed on 05/05/2020).
[6] [n.d.]. vm1.pdf. https://www.fhwa.dot.gov/policyinformation/statistics/2016/pdf/vm1.pdf. (Accessed on 05/05/2020).
[7] Kang An, Xiaohong Chen, Feifei Xin, Bin Lin, and Longyu Wei. 2013. Travel characteristics of e-bike users: Survey and analysis in Shanghai. *Procedia-Social and Behavioral Sciences* 96 (2013), 1828–1838.
[8] Don Anair, Jeremy Martin, Maria Cecilia Pinto de Moura, and Joshua Goldman. [n.d.]. Ride-Hailing's Climate Risks. https://www.ucsusa.org/resources/ride-hailing-climate-risks. (Accessed on 05/05/2020).
[9] Laura Bliss. 2019. How Much Traffic Do Uber and Lyft Cause? https://www.citylab.com/transportation/2019/08/uber-lyft-traffic-congestion-ride-hailing-cities-drivers-vmt/595393/.
[10] Craig Bullock, Finbarr Brereton, and Sive Bailey. 2017. The economic contribution of public bike-share to the sustainability and efficient functioning of cities. *Sustainable cities and society* 28 (2017), 76–87.
[11] Rossana Cavagnini, Luca Bertazzi, Francesca Maggioni, and Mike Hewitt. 2018. A two-stage stochastic optimization model for the bike sharing allocation and rebalancing problem.
[12] Longbiao Chen, Daqing Zhang, Gang Pan, Xiaojuan Ma, Dingqi Yang, Kostadin Kushlev, Wangsheng Zhang, and Shijian Li. 2015. Bike sharing station placement leveraging heterogeneous urban open data. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 571–575.
[13] Longbiao Chen, Daqing Zhang, Leye Wang, Dingqi Yang, Xiaojuan Ma, Shijian Li, Zhaohui Wu, Gang Pan, Thi-Mai-Trang Nguyen, and Jérémie Jakubowicz. 2016. Dynamic cluster-based over-demand prediction in bike sharing systems. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 841–852.
[14] Christopher Cherry and Robert Cervero. 2007. Use characteristics and mode choice behavior of electric bike users in China. *Transport policy* 14, 3 (2007), 247–257.
[15] citi 2019. Citi Bike Electric Fleet Will Grow to 4,000 — With $2 Fee. https://nyc.streetsblog.org/2019/02/28/e-xciting-but-also-e-xpensive-citi-bike-electric-fleet-will-grow-to-4000-with-2-fee/.
[16] Lisette Espín Noboa, Florian Lemmerich, Philipp Singer, and Markus Strohmaier. 2016. Discovering and characterizing mobility patterns in urban spaces: A study of manhattan taxi data. In *Proceedings of the 25th International Conference Companion on World Wide Web*. 537–542.
[17] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *Kdd*, Vol. 96. 226–231.
[18] Daniel Freund, Shane G Henderson, and David B Shmoys. 2017. Minimizing multimodular functions and allocating capacity in bike-sharing systems. In *International Conference on Integer Programming and Combinatorial Optimization*. Springer, 186–198.
[19] Raghu Ganti, Mudhakar Srivatsa, Anand Ranganathan, and Jiawei Han. 2013. Inferring human mobility patterns from taxicab location traces. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 459–468.
[20] Lukas Gebhard, Lukasz Golab, Srinivasan Keshav, and Hermann de Meer. 2016. Range prediction for electric bicycles. In *Proceedings of the Seventh International Conference on Future Energy Systems*. 1–11.
[21] Supriyo Ghosh, Pradeep Varakantham, Yossiri Adulyasak, and Patrick Jaillet. 2017. Dynamic repositioning to reduce lost demand in bike sharing systems. *Journal of Artificial Intelligence Research* 58 (2017), 387–430.
[22] guildford [n.d.]. You can now rent a 'Boris bike' in Guildford from the University of Surrey. https://www.getsurrey.co.uk/news/surrey-news/you-can-now-rent-boris-14982984. (Accessed on 05/05/2020).
[23] Alejandro Henao. 2017. *Impacts of Ridesourcing-Lyft and Uber-on Transportation Including VMT, Mode Replacement, Parking, and Travel Behavior*. University of Colorado at Denver.
[24] Mohammad Asadul Hoque, Xiaoyan Hong, and Brandon Dixon. 2012. Analysis of mobility patterns for urban taxi cabs. In *2012 international conference on computing, networking and communications (ICNC)*. IEEE, 756–760.
[25] Shan Jiang, Le Chen, Alan Mislove, and Christo Wilson. 2018. On ridesharing competition and accessibility: Evidence from uber, lyft, and taxi. In *Proceedings of the 2018 World Wide Web Conference*. 863–872.
[26] Chaogui Kang, Stanislav Sobolevsky, Yu Liu, and Carlo Ratti. 2013. Exploring human movements in Singapore: a comparative analysis based on mobile phone and taxicab usages. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*. 1–8.
[27] Farshad Kooti, Mihajlo Grbovic, Luca Maria Aiello, Nemanja Djuric, Vladan Radosavljevic, and Kristina Lerman. 2017. Analyzing Uber's ride-sharing economy. In *Proceedings of the 26th International Conference on World Wide Web Companion*. 574–582.
[28] Clayton Lane. 2005. PhillyCarShare: First-year social and mobility impacts of carsharing in Philadelphia, Pennsylvania. *Transportation Research Record* 1927, 1 (2005), 158–166.

[29] Yexin Li and Yu Zheng. 2019. Citywide bike usage prediction in a bike-sharing system. *IEEE Transactions on Knowledge and Data Engineering* (2019).
[30] Yexin Li, Yu Zheng, and Qiang Yang. 2018. Dynamic bike reposition: A spatio-temporal reinforcement learning approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1724–1733.
[31] Yexin Li, Yu Zheng, Huichu Zhang, and Lei Chen. 2015. Traffic prediction in a bike-sharing system. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 1–10.
[32] L. Liang. 2018. The National, China rides into a bike-sharing future. https://www.thenational.ae/business/technology/china-rides-into-a-bike-sharing-future-1.700338.
[33] Yu Liu, Chaogui Kang, Song Gao, Yu Xiao, and Yuan Tian. 2012. Understanding intra-urban trip patterns from taxi trajectory data. *Journal of geographical systems* 14, 4 (2012), 463–483.
[34] Meghna Lowalekar, Pradeep Varakantham, Supriyo Ghosh, Sanjay Dominik Jena, and Patrick Jaillet. 2017. Online repositioning in bike sharing systems. In *Twenty-seventh international conference on automated planning and scheduling*.
[35] Peter Midgley. 2011. Bicycle-sharing schemes: enhancing sustainable mobility in urban areas. *United Nations, Department of Economic and Social Affairs* 8 (2011), 1–12.
[36] Eoin O'Mahony and David B Shmoys. 2015. Data analysis and optimization for (citi) bike sharing. In *Twenty-ninth AAAI conference on artificial intelligence*.
[37] Natalie Popovich, Elizabeth Gordon, Zhenying Shao, Yan Xing, Yunshi Wang, and Susan Handy. 2014. Experiences of electric bicycle users in the Sacramento, California area. *Travel Behaviour and Society* 1, 2 (2014), 37–44.
[38] raleigh [n.d.]. Citrix Cycle, a Bikeshare Program for Raleigh, is Now Rolling! https://www.visitraleigh.com/plan-a-trip/visitraleigh-insider-blog/post/citrix-cycle-a-bikeshare-program-for-raleigh-nc-is-now-rolling/. (Accessed on 05/05/2020).
[39] Ivan Rios, Lukasz Golab, and S Keshav. 2016. Analyzing the usage patterns of electric bicycles. In *Proceedings of the Workshop on Electric Vehicle Systems, Data, and Applications*. 1–6.
[40] riverside [n.d.]. City of Riverside Launches Bike Riverside: The Citys First Bike Share Program. https://riversideca.gov/press/city-riverside-launches-bike-riverside-citys-first-bike-share-program-0. (Accessed on 05/05/2020).
[41] Bruce Schaller. 2017. *Unsustainable? The Growth of App-Based Ride Services and Traffic, Travel and the Future of New York City*. Technical Report. Schaller Consulting, New York, NY.
[42] Erich Schubert and Arthur Zimek. 2019. ELKI: A large open-source library for data analysis-ELKI Release 0.7. 5" Heidelberg". *arXiv preprint arXiv:1902.03616* (2019).
[43] Yu Shen, Xiaohu Zhang, and Jinhua Zhao. 2018. Understanding the usage of dockless bike sharing in Singapore. *International Journal of Sustainable Transportation* 12, 9 (2018), 686–700.
[44] Jia Shu, Mabel C Chou, Qizhang Liu, Chung-Piaw Teo, and I-Lin Wang. 2013. Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems. *Operations Research* 61, 6 (2013), 1346–1359.
[45] Shuai Wang, Tian He, Desheng Zhang, Yunhuai Liu, and Sang H. Son. 2019. Towards efficient sharing: A usage balancing mechanism for bike sharing systems. In *The World Wide Web Conference*. 2011–2021.
[46] Shuai Wang, Tian He, Desheng Zhang, Yuanchao Shu, Yunhuai Liu, Yu Gu, Cong Liu, Haengju Lee, and Sang H Son. 2018. BRAVO: Improving the rebalancing operation in bike sharing with rebalancing range prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–22.
[47] Jonathan Weinert, Chaktan Ma, and Christopher Cherry. 2007. The transition to electric bikes in China: history and key reasons for rapid growth. *Transportation* 34, 3 (2007), 301–318.
[48] Jiawei Zhang, Xiao Pan, Moyin Li, and Philip S Yu. 2016. Bicycle-sharing systems expansion: station re-deployment through crowd planning. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 1–10.

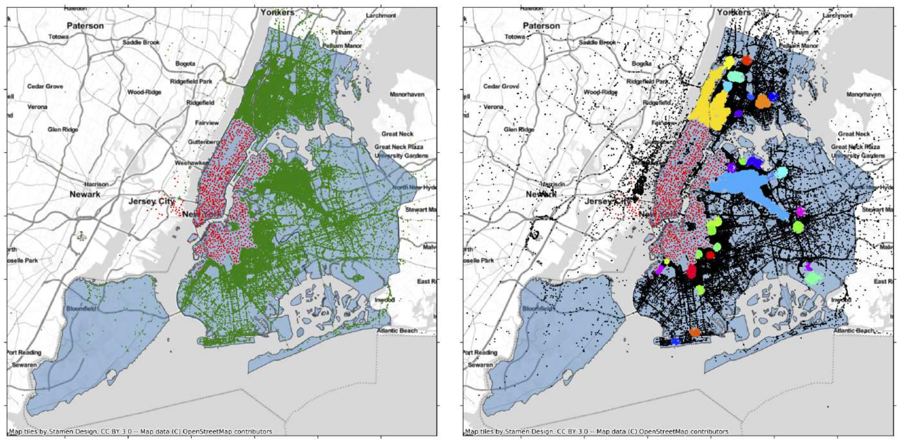
9 APPENDIX

9.1 Dataset Characteristics

Table 4 depicts the characteristics of our enhanced bike datasets. The original CitiBike dataset consists of around 18 million trips (see Table 4). We use the process described in Section 4.1 to sample taxi trips and create multiple enhanced bike datasets, each representing a different level of ride substitution (e.g. 1%, 2%, 3%, 5%, 10%). This yields enhanced datasets with the total number of trips ranging from 18 to 54 million, which triples the original bike trip demand.

Table 4. Summary of synthetic datasets from taxi trip conversion.

	Number of trips (millions)
Original bike share system	18.98
With 1% converted taxi trips	22.52
With 2% converted taxi trips	26.06
With 3% converted taxi trips	29.6
With 5% converted taxi trips	36.68
With 10% converted taxi trips	54.38
Expanded bike-share system (Sec. 5)	22.86



(a) (b)

Fig. 14. (14a) Taxi trips (green markers) outside the 300m radius of any bike station (red markers). (14b) Trip clusters discovered by DBSCAN (multiple colors). Black markers are trips in low-density regions and do not belong to any cluster.

9.2 Bike Share System Expansion Analysis

Figure 14a depicts trips originating from areas outside the coverage of the current bike system. The green markers show the snapshot of the taxi trips' pickup points, and the red markers are the existing bike stations. The figure reveals that most stations are centered around Manhattan and parts of Brooklyn. Our goal is to expand the bike share system's coverage into other areas indicated by the green markers.

Figure 14b, depicts the results of our density-based clustering algorithm built upon DBSCAN. Our algorithm identified 31 high-density clusters, and the overall trips within each cluster ranged from 10,000 to 2.5 million trips in size. The black markers depict points that lie in low-density regions, distant from other clusters, and thus, not classified into any cluster. The total number of trips that do not belong to any cluster is 1.2 million (about 13% of the uncovered region).

As can be seen in the figure, newer stations are placed in neighborhoods such as Upper Manhattan, Bronx, Queens, among others, that have higher ride-sharing traffic. These neighborhoods have an average number of 4,766 bike trips per station and lie close to the boundaries of the existing bike system. Since they can enhance the connectivity of the existing system, stations in these neighborhoods could be likely candidate locations for expansion.