Bus & Rail Metrics Methodology, Data Sources, and Validation White Paper

Version 1.1
Updated January 2021

Table of Contents

Introduction

This technical document describes the data sources and methodology employed by StreetLight Data to develop our transit travel pattern Metrics for bus and rail modes, which are available in the StreetLight InSight® Multimode subscription – in addition to validation work to build confidence in the sources and data processing. Deriving transit metrics from Big Data can help transportation planners and transit agencies with many use cases, including ridership estimation by analyzing ridership across transit networks for any given month, along with diving deeper into bus and rail rider demographics that can aid Title VI requirements. This document is updated regularly – please check our website or reach out to your StreetLight contact to ensure you have the most up-to-date version of this document.

Locational Data Sources and Probe Technologies

Data Sources for Multimode Metrics

StreetLight's Multimode Metrics are currently derived from several types of data, predominantly:

1. Location-Based Services (LBS) data
2. Well-validated bus and rail ridership counts

As the mobile data supply landscape has evolved and matured over time, we have determined that LBS data sources are currently the best suited to analyze bus and rail trips. In particular, we've evaluated and ruled out cellular tower data for active modes due to poor spatial precision and ping rate. Similarly, most navigation-GPS data comes from connected cars, making it useless for transit-related analytics. For a detailed review and comparison of available data sources, please see StreetLight's "Methodology and Data Sources" white paper.

LOCATION-BASED SERVICES (LBS) DATA

LBS data can be processed into travel patterns at a comprehensive scale. Its high spatial precision and regular ping rate allow for capturing trips as well as activity patterns (i.e., probable home and work locations), trip purpose, and traveler demographics, and allow for mode inference to differentiate active mode from vehicular travel (more on this in sections below). This makes it an ideal alternative to data derived from cellular towers, which also has a large sample size but unfortunately lacks spatial precision and pings infrequently.

Our LBS data suppliers usually provide pieces of software called SDKs (software developer's kit) to makers of mobile apps to facilitate Location-Based Services. These smartphone apps include: couponing, weather, tourism, productivity, locating nearby services (i.e., restaurant/ bank/gas station), and many more apps, all of which utilize their users' locations in the physical world in order to provide value to the users.

The apps collect anonymous user locations when they are operating in the foreground. In addition, these apps may collect anonymous user locations when operating in the background. This “background” data collection occurs when the device is moving. LBS software collects data with WiFi proximity, a-GPS (Assisted GPS), and several other technologies. In addition, locations may be collected when devices are without cell coverage or in airplane mode. The data that StreetLight uses has better than 20m spatial precision, with an average ping interval of 200 secs/ping. Note: Some of our LBS data suppliers build their own LBS collection code directly into the app, instead of using a common SDK, but the outcome is largely the same.

Well-Validated Bus and Rail Ridership Counts

We use a set of well-validated rail and bus passenger counts to validate our Metrics. Unlike other modes of travel, where permanent loop counters can be installed along roads, sidewalks, and bike lanes to count trips, transit data cannot rely on these same sources for validation. Because LBS data corresponds to passenger trips, and not to transit vehicles, we cannot use traditional bus or rail timetables or records to validate our trips. Instead, passenger counts are required for direct comparison. Passenger counts are often recorded at the station level, specifically capturing boarding and/or alighting metrics. Some passenger count data can also be supplied at the route level as well. We attempt to use only ridership counts that we have deemed reasonable, providing consistent results across stations and across time. We understand there is debate around the accuracy of ridership count technology, which will be addressed in later sections.

Given the difficulties in obtaining timely and comprehensive passenger counts from transit agencies, StreetLight is limited in the types of comparisons that can be conducted. We’re always looking for additional transit data sources for validation, be it boardings and alightings, Origin-Destination details, or route-level metrics. StreetLight is open to collaboration with any agency or organization willing to share transit data for further validation. This type of collaboration provides valuable feedback, critical to improving our Metrics for customers in the future.

Additional Sources for Development of Algorithms

Additional sources of data were used during algorithmic development, though these sources do not directly provide data into the Metrics that StreetLight provides to customers.

1. OpenStreetMap layers.
2. GPS-enabled travel diaries with detailed, user-entered mode tags.
3. Traditional surveys about bus and rail trips.
4. GPS data procured from transit agencies.

The following section gives details on each data source used in the development of these StreetLight Metrics.

OpenStreetMap Layers

Critical to the development of bus and rail trips is the use of OpenStreetMap (OSM) to identify bus and rail infrastructure. Specifically, we rely on OSM rail stations and rail line designations for the creation of rail trips, using proximity of LBS device pings to the OSM rail layers as a primary indication of rail activity. For bus trips, we look for whether OSM has identified particular roads as bus routes. We then factor this into algorithms that determine the probability of LBS device pings belonging to a specific mode.

GPS-Enabled Travel Diaries with Detailed, User-Entered Mode Tags

Historically, many GPS-enhanced travel surveys have been performed. In some, users carried around GPS devices and then, via an app or online tool, tagged each trip with information such as the mode, trip end point, trip purpose, etc. While this data is private as it describes personal activities, the data from several such surveys are housed at the Transportation Secure Data Center, hosted by NREL (https://www.nrel.gov/transportation/secure-transportation-data/). By applying to this group, and working within the databases' digital firewalls, we were able to use this data as calibration/ground truth data to develop machine-learning-derived training algorithms that allow us to recognize bus trips or (most difficult) transitioning between modes in our general LBS data. We also sourced bike and walk trips from our own proprietary sources of trip diary data, which were used to improve bus trip precision.

Traditional Surveys About Rail and Bus Trips

As we developed our algorithms, we used general characteristics of bus and rail modes to help guide us. For example, we used distribution of rail and bus trip distance published by the National Household Travel Survey. Say a local survey found that the average bus trip length in a city is four miles. We have two versions of our algorithm being tested. For version A, the average length of all bus trips in that city is five miles; for version B, the average is four miles. We will then favor version B.

GPS Data Procured from Transit Agencies

To better understand expected patterns of bus and rail trips, GPS points from a variety of tagged reference data sources were also used. Automatic Vehicle Locators (AVLs) provided GPS pings from bus and rail lines in service, sourced from Metro Transit (Minnesota – bus), SFMTA (San Francisco – bus and rail), MBTA (Boston – bus), and Capital Metro (Austin – bus and rail). These data sources did not provide ridership data specifically, but informed us about the expected movement characteristics of buses and trains, and we assume riders would inherit similar patterns in our LBS data.

Development and Implementation of Our Mode-Tagging Algorithms and Metrics

Based on our deep knowledge of our own Big Data resources, and knowledge of travel patterns, we knew that we needed a multi-faceted algorithm to measure bus and rail trips and distinguish them from our existing modes. For example, using average trip speed alone would not be adequate to decipher a bus from a car. Academic research on this topic concurs.[1,2]

However, we felt that much of the academic literature was not directly useful to us, because it assumes that incoming data will be very refined (for example, pinging every second). We know that large, affordable samples of locational data are messier and less granular. Hence, we needed to use an approach that accommodated the characteristics of our LBS data. We also felt that it was not reasonable to segment our data into trips, and then infer mode, because different modes have unique ways of ending a trip. The end of an underground rail trip, as someone disembarks a train and walks through a station, looks very different than the end of a car trip, when someone goes from 35 mph to 0 mph very quickly, and then walks a few feet to their front door.

Therefore, we decided that the best way to approach mode inference was to:

- Isolate rail trips based on a series of heuristic rules, given the fact that rail trips often occur independent of the road network.
- With the remaining LBS data, assign mode probabilities to every "ping." We chose to model a device at rest, when it continues to ping, as a "stationary" mode.
- Group pings into trips, with mode-adaptive criteria for breaking trips.
- Assign the most probable mode to each trip, given ping probabilities combined with overall characteristics of the entire trip.

It should also be noted that we will continue to evaluate and update our algorithm in order to provide the highest-quality Metrics possible. Our latest iteration applies to bus and rail trips from the months of April, May, September, and October of 2019 and 2020.

Step 1 – Identify Rail Trips

Unlike other modes, which utilize a machine-learning algorithm to classify pings by mode, rail trips are developed using a strategy that relies on a series of methods and rules.

First, rail stations and railways are identified within the OSM layer. Rail lines in OSM may include commuter or regional rail, metro or heavy rail (subway or elevated), light rail, and streetcar. It's also possible that freight lines will be included in this category; however, due to limited passenger ridership on freight lines, we don't expect any significant sample of devices

[1] Dabiri and Heaslip. "Inferring transportation modes from GPS trajectories using a convolutional neural network." Transportation Research Part C 86 (2018) 360-371.
[2] Wu, Yang, and Jing. "Travel mode detection based on GPS raw data collected by smartphones: a systematic review of the existing methodologies." Information 2016, 7, 67; doi:10.3390/info7040067.

from this category. Once rail stations and lines are identified, we look for LBS pings that occur within close proximity to rail infrastructure. We then look for attributes of those pings that increase the likelihood of it actually being a rail trip, such as speed and acceleration. Next, groups of pings are assessed in aggregate to determine whether they retain rail-like qualities consistently, or whether there might be another mode present, such as a false-positive car point passing by the rail line.

StreetLight uses a “sequence-of-linked-pings” approach where the linkage is based on identifying the likely rail pings ordered by their timestamp. The “sequence-of-linked-pings” grows into what is eventually determined to be a rail trip. Once a series of rail points are identified, we define the likely start and end points of the trip.

Determining the accurate boundary of a rail trip can be challenging. In actuality, rail trips should always start and end at rail stations. However, because devices ping at different rates, it’s not guaranteed that devices will ping within the exact boundaries of a rail station. They may ping a few minutes after passengers have disembarked a train, pinging as they walk onto nearby city streets, for example.

As a result of these challenges, we choose to end trips when a series of potential conditions are met. At a high level, we aim to start trips when we start seeing LBS pings along the rail line and we aim to stop them once we stop seeing LBS pings along the rail line. More specifically, we look for whether a device’s speeds might indicate any change in mode, like emitting a series of slower-speed walk-like pings or pings at higher speeds that do not occur in close proximity to rail lines, indicating a potential vehicular mode of travel. Rail trips will also end when a device is stationary for 20 minutes or more. As a result, most rail trips will continue if a passenger transfers trains, or is stalled at a station, as long as that lack of movement is within 20 minutes’ time. This ensures that we capture end-to-end trips without terminating them too early due to stops at stations. This is especially relevant for longer-distance train trips, like Amtrak.

Once rail trips are created, each undergoes a series of additional quality checks to ensure validity. For example, if a trip both starts and ends within a short distance of a rail station, with no other pings occurring away from rail lines, we assume that the trip is good. Focusing on the starts and ends of rail trips near stations allows us to capture rail trips that occur underground, even without a cell or GPS signal capturing the trip in motion. We also perform additional checks to ensure rail and car trips are not conflated, given the fact that many rail lines run parallel to roads. Once all rail trips are created, the remaining LBS pings continue on to the following steps for further mode classification.

Step 2 – Identify Bus Trips Through a Mode Classification Model

IDENTIFY WALK AND STATIONARY TRIPS

After identifying rail trips, we implement similar custom strategies to identify trips that may be stationary or walking. We are able to model a person at rest (or stationary) and then, when a new trip is started again, we can change the travel mode to “walk,” for example. This technique makes the starting and ending of motion harmonized as a simple mode change.

As a result, after extracting rail, walk, and stationary trips from our data sample, we’re left with pings that we presume belong to cars, bicyclists, or people riding in buses. Cars, bicycles, and buses often share the same roads and thus go through a specialized mode classification model to decipher modes from one another.

Assign Mode Probabilities to Every Ping

For this step we used machine-learning techniques to assign mode probabilities to every remaining ping. This process utilizes a random forest (RF) classifier. This was chosen because it would yield the most accurate results while still yielding an algorithm that can be efficiently implemented in our product at scale.

Training Data

The first part of a successful machine-learning project is to develop a very clean set of “training” or calibration data. This data must contain pings (single points with a location and time) that comprise many trips, to which the user has confirmed that they were riding in a bus at this time. To do this, we used a combination of NREL’s Transportation Secure Data Center, GPS data from transit agencies, data derived from bus lines in service, and vehicular data from navigation-GPS devices. The addition of GPS pings from bus lines and personal cars allowed us to model a wider breadth of bus and car trips. Data was reviewed to ensure appropriate tags, and that labeled points were part of active trips. For sources with high ping frequency (less than 1 minute between pings), individual points within trips were also subsampled, so that the time in between points mirrored that of our LBS data.

Model Features and Model Training

We experimented with over 50 features. Some are “interior” to the pinging geospatial data, including time, distance, speed, acceleration, jerk, circuity, and angular velocity for each ping, as well as for its preceding and subsequent pings, day of week, hour of day, etc. Others are “exterior” or “contextual,” including road classification, weather, and density of commercial activity nearby. Critical to deciphering buses from other modes, we also considered additional geospatial features, such as the presence of bus routes and bike lanes, proximity to parks, and road network density.

Features can be thought of as attributes or explanatory variables in a model. In the end, we used a subset of the features that were most impactful in the training of the random forest algorithm. Also, we did not want to use too many features to avoid over-fitting, and instead allowed the model to be more adaptable.

To train the machine-learning model, we used a classic 80-20 split, training the model on 80% of the training data and testing it against the remaining 20% which we’d held back from training. We used a technique called “bagging” to try this iteratively with 100 decision trees in the random forest to arrive at an “out-of-bag score” that we could use as a measure of the quality of predictions from the model. Furthermore, we also monitored and made changes to improve the precision/recall scores and F-scores for each mode (car, bus, bicycle) that the model was supposed to be able to predict.

Our model is constantly improving. Some versions of the model score higher but use much deeper trees (which have a higher computational cost to run). Our selected version balances high scores with maintaining cost and efficiency for our customers at scale.

The modes predicted by the RF classifier are reflective only of the immediate vicinity of the ping. There can be a lot of noise and variation from ping to ping in the span of an entire trip. A vehicle driving down a congested road may behave similarly to a slower-moving bus.

Hence, we took a probabilistic approach to consume the results of the classifier. When we run our algorithm, pings are not assigned a single mode. Instead, they are assigned a mode probability distribution (0 – 1). For example, Table 1 lists a time sequence of pings where a person transitions from riding a bicycle to riding in a bus in about 10 minutes.

Individual Mode Probabilities				
Ping Timestamp	Bus	Car	Bike	Prediction
5/28/2020 10:51	0.040	0.021	0.939	Bike
5/28/2020 10:53	0.095	0.033	0.872	Bike
5/28/2020 10:55	0.231	0.017	0.752	Bike
5/28/2020 10:57	0.345	0.154	0.501	Bike
5/28/2020 11:03	0.465	0.320	0.215	Bus
5/28/2020 11:04	0.668	0.165	0.167	Bus
5/28/2020 11:06	0.693	0.100	0.207	Bus
5/28/2020 11:07	0.873	0.037	0.090	Bus
5/28/2020 11:09	0.913	0.036	0.051	Bus
5/28/2020 11:11	0.888	0.063	0.049	Bus

Table 1: Time sequence of pings for a mode transition. This demonstrates the probabilistic approach used to classify individual points.

Group Pings into Trips, and Assign Modes to Trips

Similar to the approach described for rail trips, a “sequence-of-linked-pings” approach is used where linkage is based on identifying the same mode or a similar mode of the pings ordered by their timestamp. The “sequence-of-linked-pings” grows into what is eventually determined to be a trip.

In a nutshell, we first determine the trip boundary where one trip ends and another begins, then decipher the ending trip's probabilistic travel modes by calculating it from the individual pings in the trip. Given the probabilistic nature of our approach, we consider a primary and a secondary mode for each ping with an associated normalized probability.

We read the pings ordered by their timestamp, predict the travel mode of each ping using the machine-learning model-based classifier and send the stream of mode-tagged pings to the process that creates trips. When we encounter enough consecutive pings with a different-enough set of modes, we conclude that the mode has changed. We then end the current trip and start the next trip with the new mode.

As a final check, we verify the correctness and feasibility of each trip for the mode assigned to it. If we find a trip that appears to be missing its end or beginning – for example, a bus trip that stops in the middle of the highway – we eliminate it. If we find a trip that appears to have erroneous data – for example, goes from San Diego to Addis Ababa, Ethiopia and back in four seconds – we eliminate it.

Step 3 – Bus and Rail Trip Locking

A trip from an LBS device is a series of connected pings. If the traveler turns a corner but the device is pinging only every 10 seconds, then that intersection might be "missed" when all the device's pings are connected to form a trip. For bus and rail trips, StreetLight utilizes network information from OpenStreetMap (OSM), including route types, speed limits, and directionality to "lock" the trip to the network. This "locking" process ensures that the complete route of the train or bus is represented, even though discrepancies in ping frequency may occasionally occur.

Rail trips will be locked only to rail lines, and will never lock to the vehicular road network, though there may be some locations where rail and vehicular OSM lines share the same road (think streetcars). As with any mode, it's possible that the distribution of pings within a trip may result in full trips, or portions of trips being unlocked. This may be more likely to impact rail trips that occur underground due to loss of GPS signal. As we mentioned above, we found no difference in the reliability between aboveground and underground rail metrics.

Bus trips can lock to any existing portion of the vehicular road network, even if OSM has not flagged a road as a bus route, because our research has shown that OSM's bus route labeling can be incomplete. In locations where bus routes are not specified in OSM, we don't want to risk excluding viable bus trips. In addition, we understand that buses can come in many forms (local public buses, commuter buses, private charter buses, tour buses, regional buses etc.), and that these buses will not always use traditional city bus lines.

Step 4 – Contextualization: Demographics and Trip Purpose Assignment

If a device that creates LBS data regularly pings on a block that contains residences, and those pings often occur overnight, there's a high probability that the device's owner lives on that block. This allows us to infer a probable home location to that device. By using daytime and nighttime ping frequency and land-use context, we avoid assigning a device to "live" at a place where the owner may work a night shift (an airport, for example) or goes on a vacation just for a few days a month. We reset each device's home location each month to accommodate people who move residences.

The trips that have the likely home location as a trip start or trip end are to be considered home-based. To assign workplaces, we look for where a device most frequently spends daytime hours (10am-4pm). We do not use land use for this assignment, because work occurs in all land uses. Note that we may allow people to work from the same place as their home. We do allow people to "work" at places like college campuses; thus, students going to school may be classified as working at the school. We will also miss people who work at irregular locations, like plumbers. In addition, we append demographic information at the Census Block, Census Block Group, or Dissemination Area (Canada) level. Our demographic data sources for the U.S. are the 2010 Census and American Community Surveys. In Canada, our source is Manifold Data.

Step 5 – Quality Assurance

Quality Assurance is done at multiple levels:

- **Classification of individual pings:** For the testing of the classifier, we used typical techniques used in testing machine-learning algorithms. We trained the model on 80% of the training data and tested it against the 20% we had held back and not exposed to the model. This was discussed in the sections above.
- **Unit-level testing for trips:** For testing the creation and breaking of the trips, we hand-picked 100+ unit test trips against which we verified the trip boundaries, the overall trip travel mode, and the mode of the individual pings as determined by the system.
- **Meso-level testing for trips:** We looked at certain regions, primarily metropolitan areas, to make sure trip distributions looked reasonable. Focusing on areas with well-known behavior – like bus and bike lanes, etc. – to verify that our trips mirror the real world.
- **System-level testing for trips:** We performed a number of tests on the overall trips generated in each month of data, including visual checks, statistical checks, spatial and spatiotemporal checks, and real-world data comparisons.

Step 6 – Normalization and Expansion

For bus and rail trips, StreetLight uses a set of ridership metrics provided by transit agencies to measure the change in trip activity each month. Using a set of bus and rail station polygons, we quantify the number of bus and rail trips that start at each station and compare those to the agency-reported values, and use this ratio to normalize appropriately. For example, if the agency reports that there are 400 rail passengers that disembark at a station in a day, and the StreetLight sample contains 40 rail passengers, the expansion ratio will be 10.

Due to limited availability of agency data, and the high frequency of monthly updates in StreetLight's data processing pipeline, we determined that creating monthly expansion ratios based on agency counts may be an accurate process, but it is risky due to the uncertainty of timely, available truth data. Therefore, we calculate expansion ratios for each mode in a seed month in 2019, then use variation in our LBS vehicular penetration rates to adjust scaling values month over month. As a result, the StreetLight Index for bus and rail data is normalized to adjust for change in our sample size. Given the limited availability of transit agency counts and the consistent availability of vehicular counts, this allows us the flexibility to add new months of data and normalize our sample with an efficient turnaround for customers. This process assumes that the variation in our LBS penetration rate is consistent across modes. We performed a thorough review of monthly changes in our sample and ensured that our normalization process allowed for comparisons across time.

Due to varied normalization ratios across road types and geographies, the StreetLight Index for bus and rail data is not yet "expanded" to estimate the actual volume of bus and rail passenger trips. Downloads are available with the Metrics output as StreetLight Index. For customers who are interested in deriving ridership estimates from StreetLight Metrics using calibration, they can do so on their own after downloading Metrics and working with the StreetLight Index provided for each zone.

Step 7 – Aggregate in Response to Queries

Whenever a user runs a query via StreetLight InSight®, our platform automatically pulls the relevant trips from the data repository and aggregates the results. For example, if a user wants to know the share of trips from origin zone A to destination zone B vs. destination zone C from September 2019, they can specify these analysis settings in StreetLight InSight®. Trips that originated in origin zone A and ended in either destination zone B or destination zone C during September 2019 will be pulled from the data repositories, aggregated appropriately, and organized into the desired Metrics. For additional information, see our Support Center.

Only trips that are bus and rail (as selected by the app user) will be queried. In setting up an analysis in StreetLight InSight®, users are able to specify the desired mode when selecting the "Mode of Travel" as a first step in the "Create Analysis" process.

StreetLight is constantly improving its bus and rail Metrics, so please share your feedback with us. We will share future improvements in methodology and validation in updates of this white paper, so check our website regularly for new information.

Rail Validation

Understanding existing transit behavior is critical to transportation and planning efforts. StreetLight has developed algorithms and machine-learning techniques that utilize Location-Based Services data to identify rail trips across the United States and Canada. This validation section focuses on comparisons between StreetLight's Rail Metrics, published travel survey metrics, and permanent counter locations in Boston, Chicago, and the San Francisco Bay Area.

We've taken two approaches to validate our transit Metrics. First, we compared our aggregated trip characteristics to information published by household travel surveys (NHTS) and transit databases (NTD) to evaluate whether our Metrics are within the general range of expected trip characteristics for each mode. Second, we compared our trip volumes at the station level to published ridership counts provided by transit agencies.

We attempt to use only transit ridership counts that we have deemed reasonably accurate. We do not use modeled passenger counts for validation, as they have additional sources of error. Our goal in this validation section is to demonstrate that our transit Metrics can be used in place of surveys or short-term passenger counts, and be used to augment additional transit data sets.

Comparing Our Rail Results to Travel Surveys and Transit Databases – NHTS and NTD

We analyzed our aggregated rail trip attributes and compared them to trip metrics published in the National Household Travel Survey 2017 (NHTS)[3] and the National Transit Database (NTD)[4] with transit reports from 2018. These two sources are fundamentally different but allow for valuable comparisons of trip attributes.

NHTS surveyed ~129,000 households across the U.S. and captured roughly 4,400 rail trips made by passengers. Meanwhile, NTD collects transit data directly from transit agencies through uniform reporting, which allows for the analysis of agency-based financial and operating information. There were 9 transit agencies from 2018 that exclusively reported details about their rail trips in 2018. To create a comparison metric set, StreetLight analyzed roughly 20.6 million rail trips across the continental U.S. that occurred in April, May, September, and October of 2019 and 2020 (for more on how we infer and create rail trips, please see the separate Methodology section).

We compared key average characteristics of rail trips from our data set to the NHTS and NTD reported values. We do not set an "exact match" as the goal. All three data sets are samples, and all three thus have different strengths and sources of error. Where discrepancies occurred, we believe they are explainable by known differences in collection methods. It should be noted that NHTS reports rail in two categories, "Amtrak/Commuter Rail" and "Subway/Elevated/Light Rail/Streetcar." Table 2 shows StreetLight average Rail Trip Attribute Metrics relative to the published NHTS and NTD numbers. Note that in the case of NTD, duration metrics were not reported; thus, they were inferred from reported distance and speed values.

[3] https://nhts.ornl.gov/
[4] https://insights.transitcenter.org/

Source	Avg Duration (min)	Avg Distance (mi)	Avg Speed (mi/hour)
NHTS (Amtrak/Commuter Rail)	99.08	43.08	26.09
NHTS (Subway/Elevated/Light Rail/Streetcar)	52.41	9.49	10.86
NTD	36.54 (inferred)	15.16	24.89
StreetLight Rail	30.79	10.25	18.01

Table 2: Comparison of average trip duration, distance, and speed from NHTS, NTD, and StreetLight trips.

As shown in Table 2, we found the StreetLight trip duration, distance, and speed averages were fairly close to the NTD reported values. NHTS values differed from both the NTD reported data and StreetLight's Metrics, specifically in the duration category. Because NHTS trips are self-reported from memory, we suspect that the NHTS duration metrics may be inflated, potentially including wait-times, transfers, and other transitional modes in the reported rail trip. NTD and StreetLight Rail (to varying extents) will isolate the rail portion of the trip from the rest of the trip that involves access and egress modes, such as walk or park-and-ride.

It's important to note the difference in average trip metrics reported in the two NHTS rail categories. Characteristics of rail trips can differ substantially based on the different types of rail included. Ultimately, we believe that StreetLight's rail trips skew more toward the "subway/elevated, etc." category, as these are a higher percentage of trips nationally. The NTD sample did not include any values reported by Amtrak.

NHTS also provided trip distance distributions that can be used for further comparison. In the following example, we've included trip distance comparisons to the "Subway/Elevated/Light Rail/Streetcar" category. Figure 1 shows the distribution of rail trips by length.

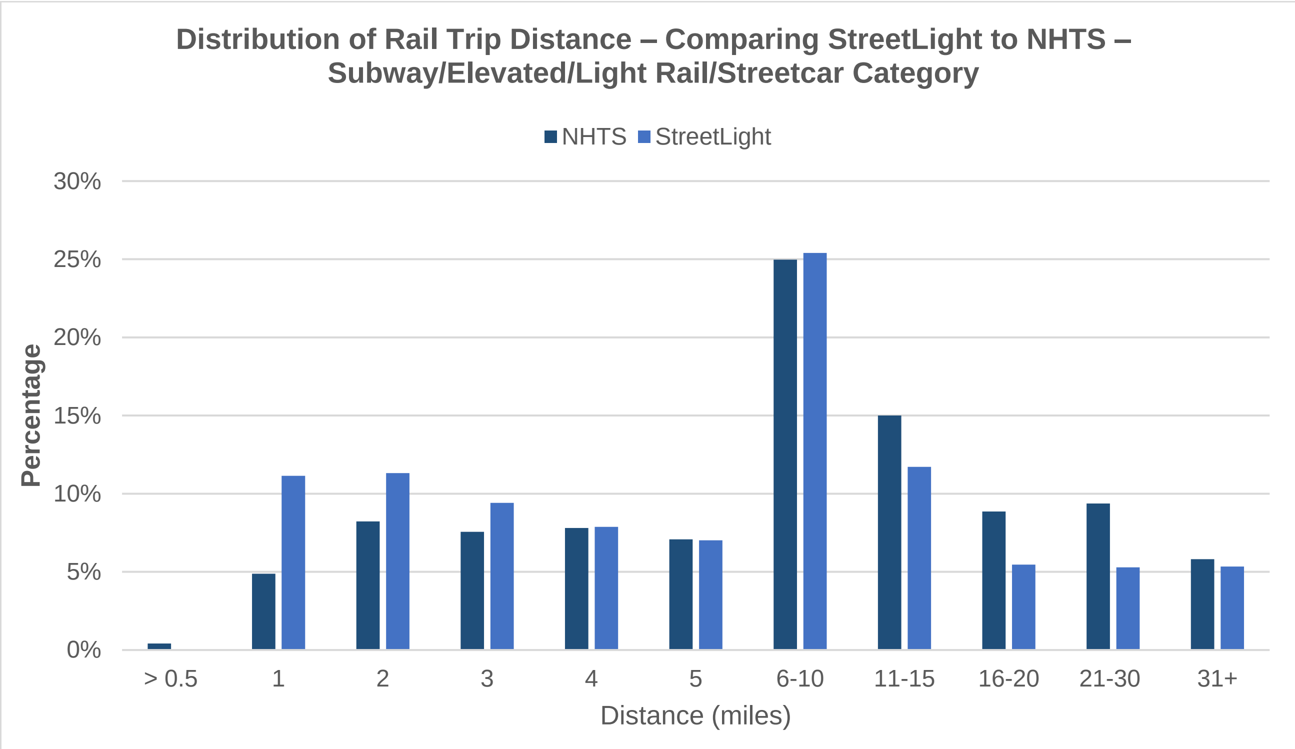


Figure 1: Histogram of rail trip length distribution for StreetLight and the NHTS.

StreetLight's distribution matches very closely with the NHTS survey, with a high proportion of trips between 6 miles and 10 miles in length. StreetLight does capture slightly more short-distance rail trips than the survey, although distributions are similar in pattern.

Comparing Our Results to Rail Ridership Counts

To validate data at a more granular level, we compared StreetLight rail trips at known transit stations to agency passenger counts across a number of cities nationwide.

Rail Data Source Review

For this validation, we aimed to compare agency data containing boarding and alighting passenger counts with monthly data from 2019 and 2020 to be able to validate as many of StreetLight's available months as possible. Because many sources of ridership data are published annually, or at the route level as opposed to the station level, this left us with a limited set of potential sources for comparison. We needed passenger counts that published their data on a daily or monthly basis to meet our strict criteria for temporal coverage. To assess performance across a diverse range of stations, our goal was to obtain passenger count data from a range of transit networks across the country, as well as a range of types of rail (elevated, subway, at-grade, etc.).

The first subset of station-level ridership data was obtained from Bay Area Rapid Transit (BART), comprising commuter rail data in the San Francisco Bay Area.[5] BART's ridership data set was comprised of 50 stations, both above and below ground. Due to BART's turnstile system that requires passengers to tag in and out at stations, data could be reported as total trip starts and stops by station, as well as station Origin-Destination patterns. Data was available as monthly averages by weekday and weekend for all months in 2019 through November 2020.

The next agency we were able to obtain station-level ridership data from was the Chicago Transit Authority (CTA).[6] Rail data provided by CTA encompasses the Chicago "L," a rapid-transit system composed mostly of elevated rail lines, with a small portion of underground stations. It's one of the busiest rail mass-transit systems in the country, encompassing 148 stations in Chicago and the surrounding suburbs. CTA's system allows for counting of passengers as they pass through entry turnstiles at each station. It does not report passenger counts exiting at each station. Data was made available at the daily level for all months in 2019 through September 2020.

We also obtained station-level transit agency ridership data from Massachusetts Bay Transit Authority.[7] MBTA (known colloquially as "the T") is a rapid-transit rail system composed of trains that travel both above and below ground in the greater Boston area. Data obtained by StreetLight encompasses 65 stations that report passenger counts as turnstile entries. It does not report passenger counts exiting at each station. Data was made available at the daily level for all months in 2019 through November 2020.

In evaluating the station counts for comparison, it's important to consider the range of average daily trip counts recorded across the agencies. Agency data was aggregated to capture average daily trip starts by station across April, May, September, and October 2019 and 2020. In Figure 2, all agencies' boarding frequency distributions are displayed with the same scale (x-axis) to ensure an easy visual comparison across the sources. Note that the y-axis range for number of stations varies across the three transit systems. Figure 2 contains distributions from 2019 months only, given the variability of transit ridership in 2020 due to COVID-19 restrictions.

[5] https://www.bart.gov/about/reports/ridership
[6] https://data.cityofchicago.org/Transportation/CTA-List-of-CTA-Datasets/pnau-cf66
[7] https://massdot.app.box.com/s/21j0q5di9ewzl0abt6kdh5x8j8ok9964

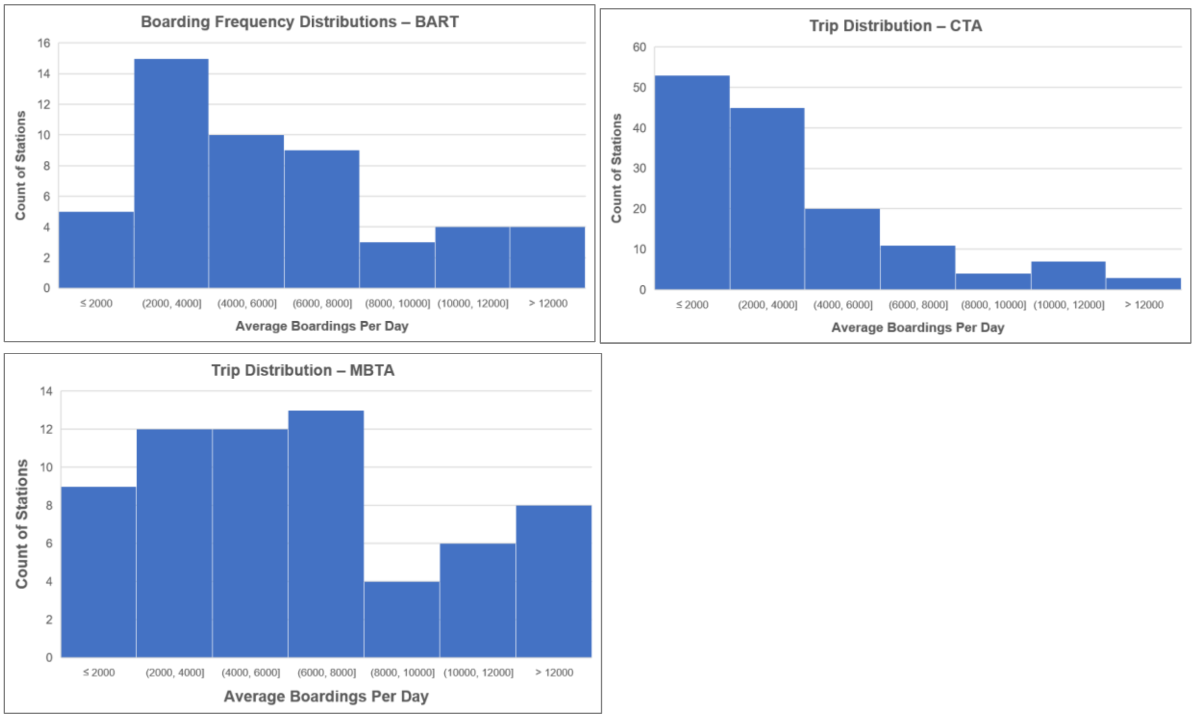


Figure 2: Histograms of station boardings for BART (top left graph), CTA (top right graph), and MBTA (bottom left graph) averaged across four months in 2019 (2019 months were used to show "pre-COVID" conditions). BART and CTA have station totals clustered below 4,000 daily boardings, while MBTA totals are more evenly distributed across defined bins.

INITIAL COMPARISON TO STREETLIGHT RAIL METRICS

After the source data was cleaned and aggregated, we compared the monthly boarding counts at stations with the origin counts on the matching rail station zones obtained from the StreetLight InSight® platform. Our goal was to determine the correlation between the two sets of counts.

To create zones in StreetLight InSight® that mirrored the transit agency's data, we first obtained station points from GTFS through Transitland's API.[8] Each point represents the latitude and longitude of the rail station. To translate this into a zone in StreetLight, we ingested the points as a shapefile into a GIS platform, and buffered the points so that each station was represented by a larger polygon. The size of each buffer zone depended on the density of stations in the given rail network. If buffered stations overlapped too much, the StreetLight InSight® platform risked double-counting a portion of rail trips. But if buffered stations were too small, it risked capturing too few rail trips as they end. We know that rail trips may not always end exactly within the bounds of the station platform, and we also know that station platforms can vary in size and length, and thus a standard circular buffer may not always capture the entirety of the station. To balance these challenges, we decided to implement maximum buffers of 500m when possible, but in rail networks where stations overlap, we reduced the size of the buffer to 150m at

[8] https://www.transit.land/documentation/an-open-project/contributor-agreement.html

minimum. Examples of buffered stations can be seen in Figure 3 below. BART stations were buffered 250m to avoid overlap, while CTA and MBTA stations were buffered 150m with minimal overlap.

Station Points for BART and MBTA with Buffers

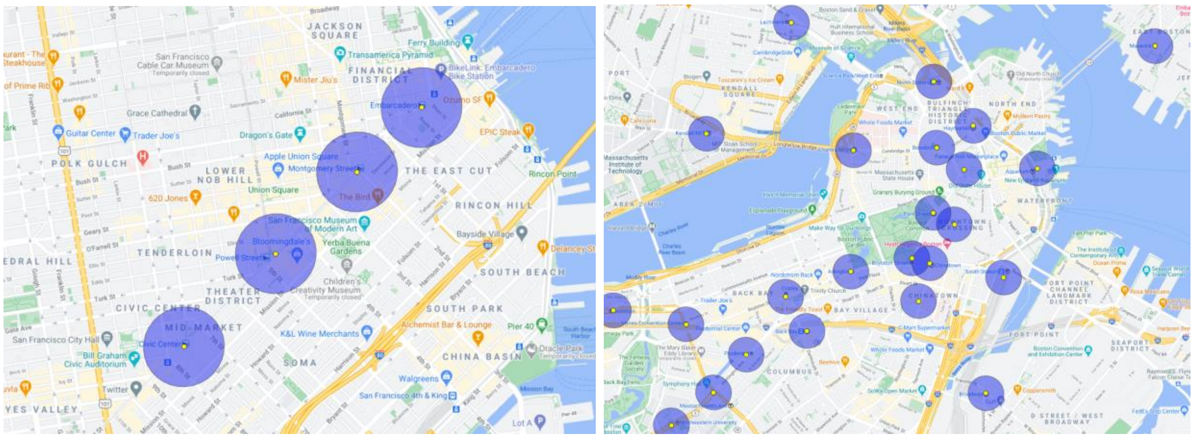


Figure 3: Station points with buffers. BART stations in the San Francisco Bay Area were assigned 250m buffers in order to prevent overlap (left). MBTA stations in Boston were assigned the minimum 150m buffer, resulting in some minimal station overlap (right).

Once the stations are all converted to polygons and exported as shapefiles, they can be uploaded directly to the StreetLight InSight® platform. A zone set was created for each rail station network, and all zones were marked “non-pass-through” in order to capture the starts and ends of rail trips. For more information on creating zones in StreetLight InSight®, see our Support Center. Figure 4 below illustrates the station polygons as zone sets in StreetLight InSight®.

Station Zones in StreetLight InSight® for BART, CTA, and MBTA

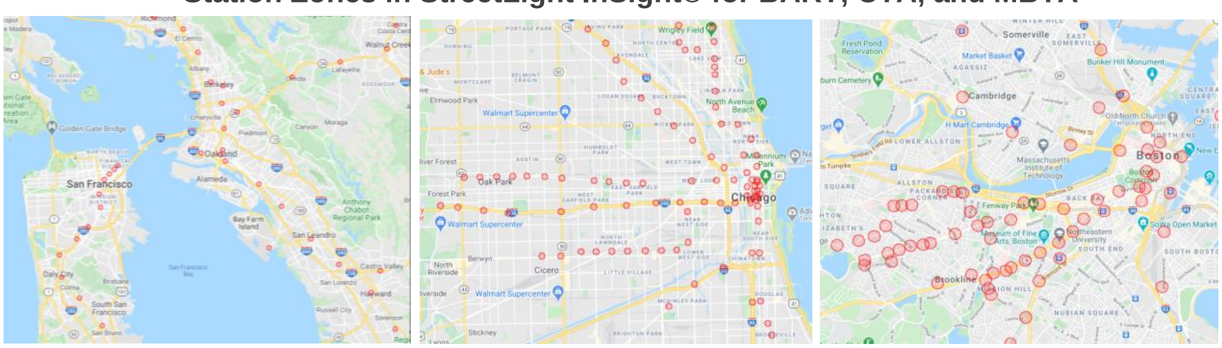


Figure 4: Station zones as uploaded in StreetLight InSight®. BART stations in the San Francisco Bay Area (left), CTA stations in Chicago (center), MBTA rail stations in Boston (right).

Note that this is not the only approach to creating station zones. Customers might also upload the station polygons themselves, rather than as latitude/longitude points. In this case, we would still recommend buffering the station polygons, for reasons stated earlier. Customers also have the option to manually draw station zones within the StreetLight InSight® application. Although this is a manual process, hand-drawn zones would allow for customers to utilize our recommended maximum buffer size of 500 meters, while modifying the geometry to prevent zone overlap between stations and account for unusual station geometries, which optimizes capture of station-specific trips.

Before diving into deeper validation, we wanted to highlight some potential sources of error involved when comparing agency counts to StreetLight Rail Metrics. Agency-reported passenger boardings for rail are less prone to error relative to other sensor technologies, making them a very reliable source for comparison. This is because passengers enter train stations or platforms via a turnstile, as opposed to tagging on themselves, which has the potential to introduce significant error because there is less accountability and many passengers don't tap-on/tap-off. While there's always the potential for broken turnstiles or passengers that illegally enter the system, we expect this to be an insignificant source of error. As mentioned earlier, StreetLight's data available for rail station matching has an additional source of potential error, accurately ending rail trips within close proximity to a rail station, which is why we recommend buffering stations before analysis. Thus, we suspect our rail station matching validation to be weaker than analyzing aggregated rail patterns. Finally, it's possible that the station points obtained from GTFS may not be centered perfectly with the rail station, thus causing buffered zones to miss portions of the track or the station that may capture additional trips.

To get a sense of our general performance across stations, it's important to consider the penetration rates recorded at each station. Penetration rates are calculated as the sum of the average daily StreetLight sample trips divided by the sum of the average daily station passenger count. A relatively consistent penetration rate is key to consistent results in a validation. Penetration rates are also key to helping us understand which stations are outliers, indicating some other source of error may be at play. We expect different rates of penetration across each transit network due to regional differences in penetration rates, as well as the differing buffer sizes used to create station zones. Figure 5 illustrates the variation in penetration rates across the three available sources.

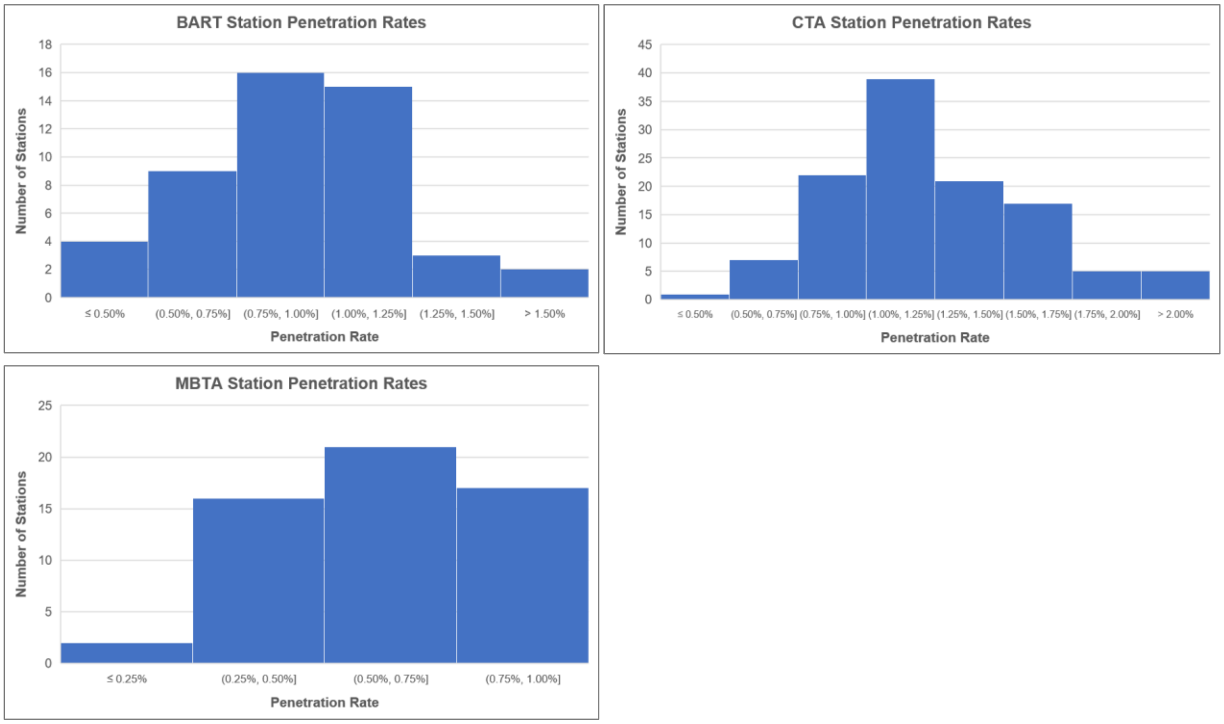


Figure 5: Distribution of average trip start penetration rates at rail stations. BART penetration rates are clustered around 1% (top left graph), CTA penetration rates tend to be above 1% (top right graph), and MBTA penetration rates tend to be below 1% (bottom left graph). Penetration rate is the ratio of StreetLight rail trips to reference rail trips.

Overall, penetration rates vary across agencies but still fall within expected ranges for our LBS data source. As mentioned earlier, the size of the rail station polygon analyzed in StreetLight InSight® may be one factor influencing the different penetration rates across agencies. Geographic location is another factor; we've seen our LBS penetration rates tend to be higher in the Midwest and Southeast relative to East Coast and West Coast regions, which might be contributing to CTA's higher values. Additionally, we expect slight differences in penetration rates when stations are above or below ground due to weaker signals in underground stations. We found these differences were not dramatic, and did not impact the validation process, but it should be noted that MBTA has the highest portion of below-ground stations relative to other agencies, which may be another contributing factor to those slightly lower penetration rates.

Combined Correlation Results

To evaluate correlation between agency-reported rail data and StreetLight Rail Metrics, we evaluated each agency separately. The benefit of evaluating each agency on its own is that it enables us to confirm the quality of metrics across a variety of regions and types of rail networks.

First, we evaluated the correlation between StreetLight Index values and the sum of average daily trips occurring at each rail station. Agency metrics were aggregated as daily averages across eight months when available: April, May, September, and October 2019 and 2020.

StreetLight Metrics were reported as average daily Index values generated from LBS data aggregated across the same calendar months. For each station, we compared the average daily agency count of passenger boardings to the StreetLight Index value representing trips starting at the same locations.

Stations underwent some additional outlier detection. In this case, stations with penetration rates two standard deviations outside the mean penetration rate were considered outliers. In most cases, this removed a small portion of stations with higher and lower than expected penetration rates. For example, in the case of BART, one zone out of 47 was determined to be an outlier with a penetration rate above 5% occurring at a station with low passenger counts due to construction. In the case of CTA, six zones out of 123 were considered outliers. Zones were checked to evaluate the cause of abnormal penetration. One correlate of higher penetration rates at individual stations was overlapping station zones, described in the zone setup process. Due to the occasional overlap of station buffers, some rail trips have the potential to be double-counted. This had the highest likelihood with CTA stations, due to the high density of the rail network in downtown Chicago.

For BART stations, the results showed high correlation, with an R^2 value of 0.90. This means that StreetLight's Metrics explain 90% of the variation in the rail station counts from the selected months in 2019 and 2020.

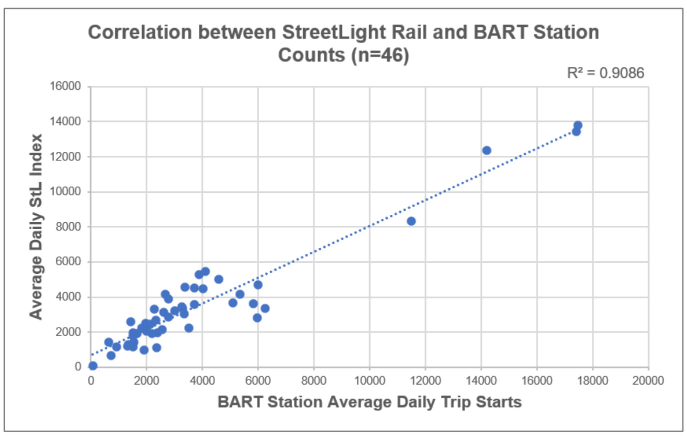


Figure 6: Correlation between average daily rail trips reported by BART and average daily rail StreetLight Index. Data was aggregated across eight months in 2019 and 2020, with an R^2 value of 0.90.

CTA showed similarly strong results, with an R^2 value of 0.81. Due to variation in penetration rates at some select underground stations, these results included above-ground CTA stations only. Unlike other sources, CTA had seven months available for comparison, rather than eight.

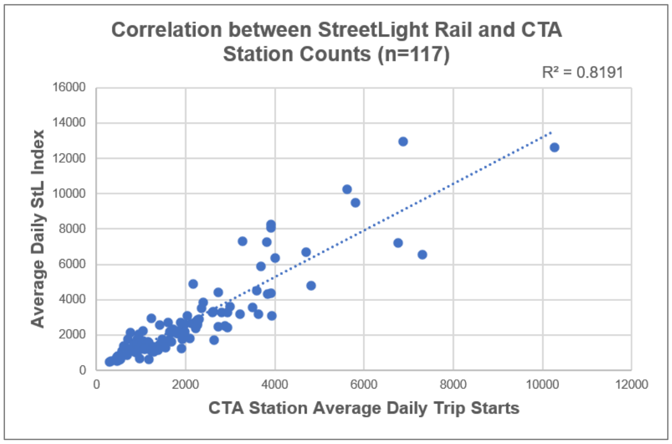


Figure 7: Correlation between average daily rail trips reported by CTA and average daily rail StreetLight Index for above-ground stations. Data was aggregated across seven months in 2019 and 2020, with an R^2 value of 0.81.

Finally, MBTA showed promising results, with an R^2 value of 0.77.

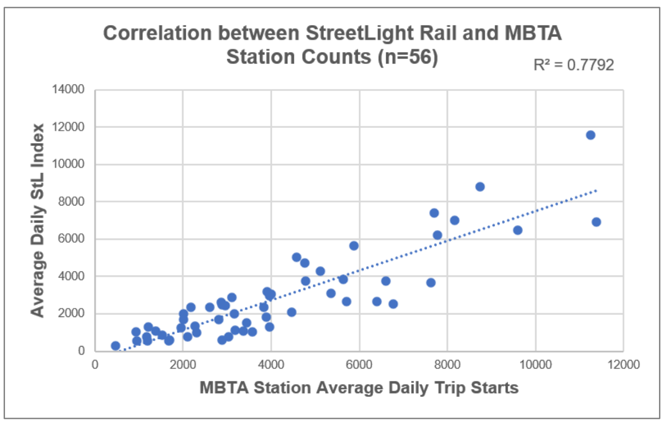


Figure 8: Correlation between average daily rail trips reported by MBTA and average daily rail StreetLight Index. Data was aggregated across eight months in 2019 and 2020, with an R^2 value of 0.77.

Overall, given the limited availability of transit data, as well as the differences in transit networks across the country, we are very pleased with these results. We were not surprised that BART returned the strongest correlation results across the three agencies. We expect this is likely due to the fact that BART is a relatively simple transit network relative to CTA and MBTA. With fewer intersecting lines, BART stations tend to be more spread out, thus subject to cleaner locking, as well as starting and ending of StreetLight trips. These qualities, as well as others discussed throughout this paper, should be considerations for customers as they explore and validate rail data.

Time Trends

Another important component of rail data is the ability to compare StreetLight Metrics across time in order to assess patterns and trends. StreetLight offers the ability to perform hourly, daily, and monthly comparisons of its Metrics. First, we'll evaluate monthly trends by comparing StreetLight Metrics to published sources, specifically monthly trends published by the Federal Transit Administration (FTA) and the American Public Transportation Association (APTA).

The National Transit Database (NTD)[9] maintained by the FTA publishes "Unlinked Passenger Trips" reported monthly by mode and type of service. The values below are aggregated across all agencies reporting rail trips to the FTA. According to the FTA, some agencies may report a 100% count of their unlinked passenger trips, while others may not have the resources or technology available for a 100% count and thus report a sample of trips.

APTA has partnered with the Transit App[10] to estimate and track ridership estimates over time. They publish weekly ridership estimates nationally, as well as for specific agencies. Estimated values are statistically modeled based on measures of Transit App usage, and therefore do not represent actual reported ridership counts. Below we've compared the total StreetLight Index for rail trips (U.S. and Canada) to APTA's National Estimates. It should be noted that APTA's National Estimates may include both bus and rail transit usage. This is because the Transit App includes usage of both modes but does not model them separately, instead modeling a single combined transit value. Regardless, we felt these numbers provided a useful point of comparison.

In Table 3 below, we compare the year-over-year change (2019 to 2020) by month for the StreetLight Index aggregated across all StreetLight trips in the U.S. and Canada, to the ridership samples and estimates published by the NTD and APTA. When validating, we looked for changes over time to be directionally accurate when compared to other sources, as well as within similar magnitude.

YOY Change by Month	National Transit Database	APTA	StreetLight Index (Rail)
April	-89%	-80%	-80%
May	-88%	-75%	-78%
Sept	-70%	-63%	-56%
Oct	-69%	-66%	-58%

Table 3: Comparison of year-over-year changes by month across the NTD, APTA, and the StreetLight Index. Changes over time match closely across the three sources.

Overall, the StreetLight Index for rail trips consistently falls within 15% of the trends reported by the two reference sources. April and May in particular, which saw the most dramatic decline in rail trips due to COVID-19 restrictions, match very closely across the three sources. This gives

[9] https://www.transit.dot.gov/ntd/data-product/monthly-module-adjusted-data-release
[10] https://transitapp.com/APTA

us confidence that StreetLight Rail Metrics are consistent with nationally reported year-over-year trends.

In addition to evaluating national trends, we can assess system-wide trends using monthly data provided by the agencies referenced earlier. For BART, CTA, and MBTA, we calculated average daily trips by month based on their reported data. In StreetLight InSight®, we ran monthly analyses using the same station zones described earlier in the paper. Below we show what year-over-year trends look like when comparing April 2019 to April 2020 across the three agencies.

Agency	Agency's reported YOY % change for April	StL Index reported YOY % change for April
BART	-94%	-92%
CTA	-88%	-85%
MBTA	-92%	-90%

Table 4: Agency reported year-over-year changes compared to StreetLight Index year-over-year changes for April. Trends are evaluated across three agencies – BART, CTA, and MBTA.

As illustrated in Table 4 above, the StreetLight Index is able to capture the dramatic drop in rail trips across the two years due to COVID-19 impacts.

Results can also be visualized across months to assess broader trends. As a reminder, StreetLight currently provides eight seasonal months of rail data, rather than a full contiguous year. We will be adding more data months to the platform in the future – make sure to check our Support Center for updates. In the figures below, the average daily ridership is compared to the StreetLight Index for the average daily trip boardings by station. In the following graph, the StreetLight Index has been adjusted to BART's system-wide counts for April 2019 in order to provide a simpler comparison of metrics across time. The StreetLight Index is a normalized value and not meant to be an estimate of real-world rail trips.

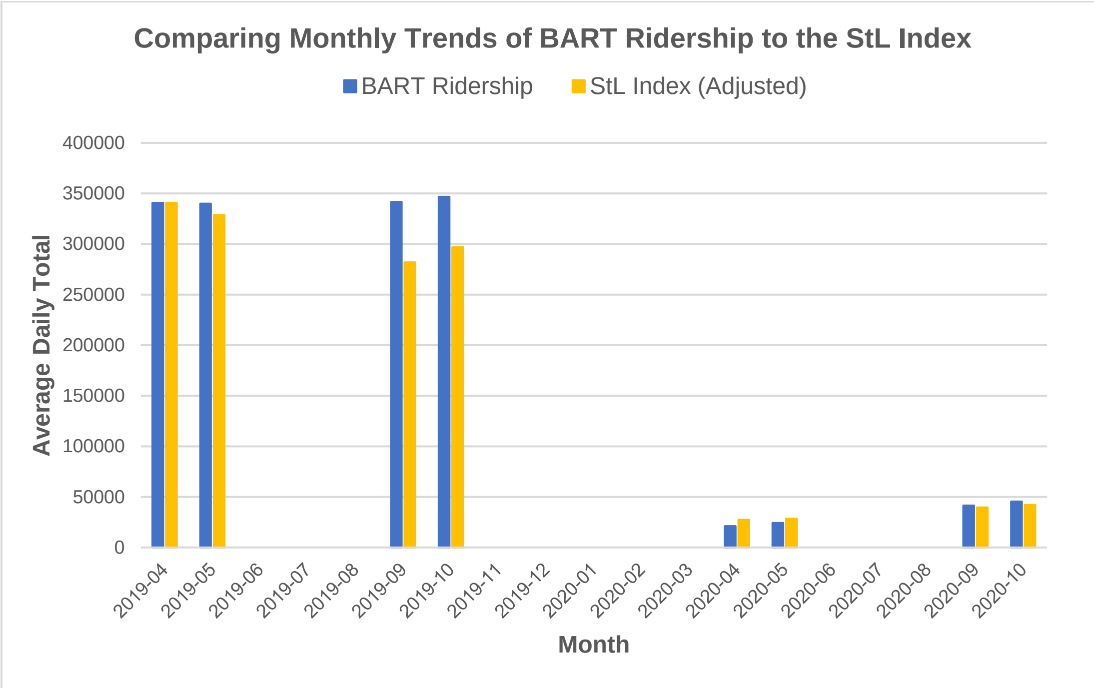


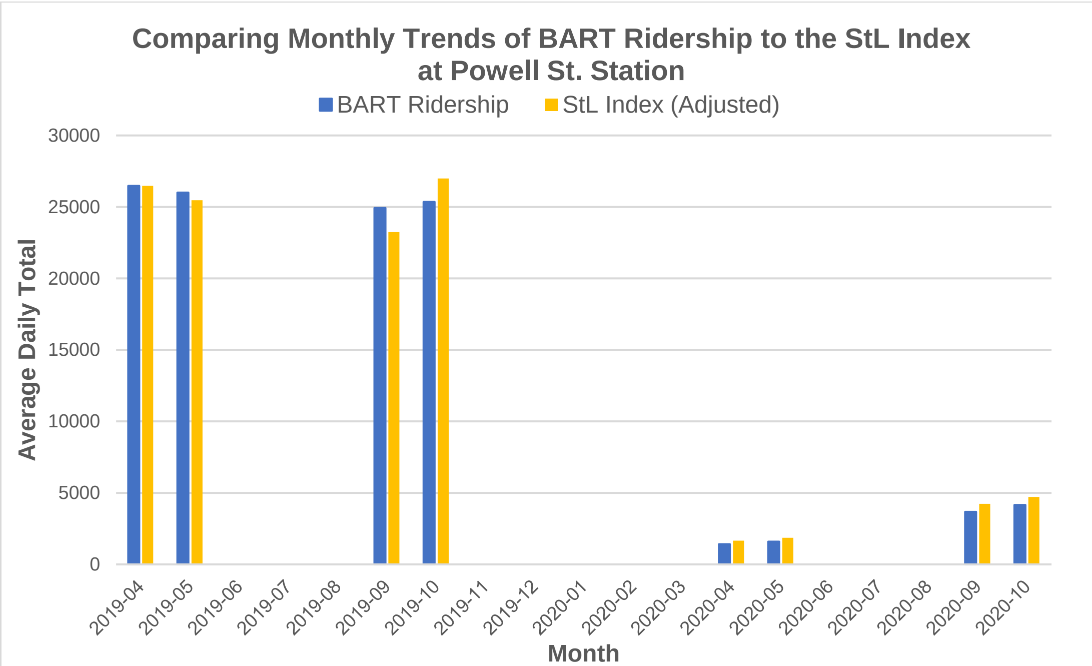
Figure 9: Monthly BART trends comparisons of agency-reported trips to StreetLight Metrics (StreetLight Metrics represent a normalized value, not a passenger volume estimate).

Generally, we're pleased to see trends match directionally across months. The most significant change occurs between 2019 and 2020 with a dramatic decline in trips across all sources, echoing what we know about COVID-19 and transit metrics. In some cases, specifically with CTA and MBTA, there appears to be an increase in ridership between spring 2020 months and fall 2020 months, which the StreetLight Index is able to capture.

In addition to assessing system-level time trends, changes at individual stations can also be evaluated. The examples below draw from BART ridership at two unique stations. The first, Powell St. Station, is a busy underground station in downtown San Francisco. It caters to commuters, as well as shoppers and tourists. The following station is located in the Bay Area suburb of Walnut Creek. It's an above-ground station heavily used for commuting. In both cases, year-over-year trends for April are nearly identical, and ridership totals closely mirror the trends in the StreetLight Index. In Figure 10, the StreetLight Index has been adjusted to the system-wide counts for April 2019 in order to provide a simpler comparison of metrics across time.

Station	Agency's reported YOY % change April	StreetLight Index reported YOY % change April
Powell St. Station (urban station)	-95%	-94%
Walnut Creek Station (suburban station)	-96%	-93%

Table 5: BART reported year-over-year changes compared to StreetLight Index year-over-year changes in April for Powell St. and Walnut Creek Stations.



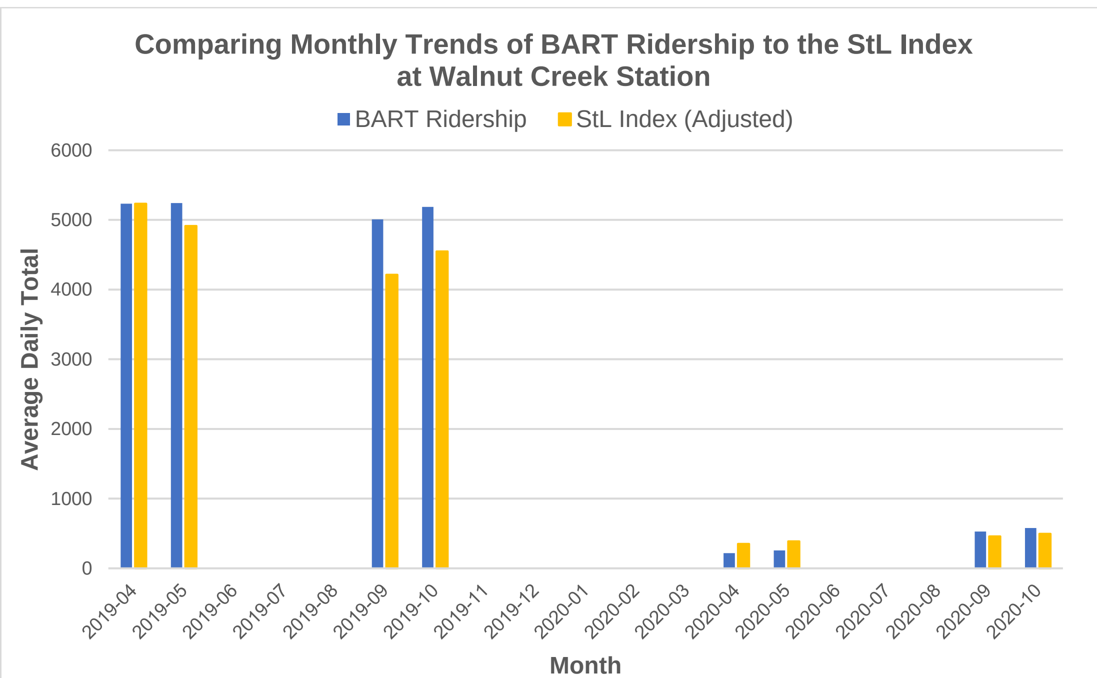
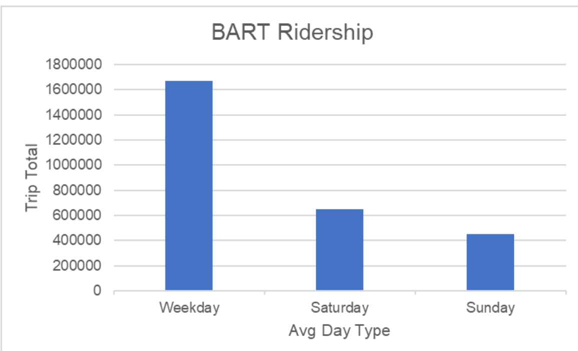


Figure 10: Monthly comparisons of agency-reported passenger counts to StreetLight Metrics for unique BART stations. Monthly trends for Powell St. Station (top graph) and monthly trends for Walnut Creek Station (bottom graph).

As a result of the monthly validations performed at the national, system-wide, and station levels, we recommend the use of the StreetLight Index as a tool for evaluating large changes across time. Large changes include significant seasonal increases or decreases of more than 15%, or in the case of 2020, the dramatic decrease in transit activity due to COVID-19. The StreetLight Index is meant to be a relative value for comparison, and requires calibration in order to represent an estimated count. As with many of our metrics, more subtle monthly changes may be difficult to capture, especially when analyzing locations with lower or more variable sample sizes.

Time trends can also be assessed at a more granular level in StreetLight InSight®. BART provides their station boarding data as daily averages across Weekdays, Saturdays, and Sundays. For this example, we've selected a month from 2019 in order to better assess trends prior to COVID-19 impacts. Figure 11 below compares average daily boarding values reported by BART to those derived from StreetLight. The goal is to have the patterns match across the two sources.



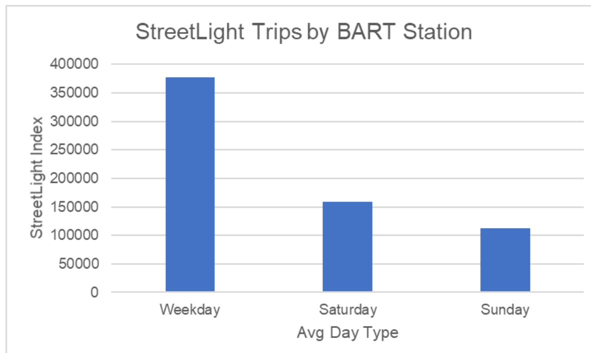


Figure 11: Average daily BART ridership on Weekdays, Saturdays, and Sundays in October 2019 (left). Average daily StreetLight Index values at BART stations on Weekdays, Saturdays, and Sundays in October 2019 (right). Day-of-week trends are very similar across the two sources.

BART ridership data shows a 61% decrease in trips between an average weekday and an average Sunday. StreetLight shows a similar 58% decline. Meanwhile, BART ridership data shows a 31% decrease in trips between Saturday and Sunday, while StreetLight shows a 29% decline. Overall, we're pleased to see these results match so strongly, validating StreetLight's ability to capture traffic trends over the course of an average week.

Although BART did not provide its transit data with hourly granularity, we wanted to assess the hourly curve derived from StreetLight's Metrics to ensure that hourly trends were reasonable. As seen in Figure 12 below, the hourly trend of the StreetLight Index shows clear morning and afternoon peaks, mirroring expectations around BART ridership patterns.

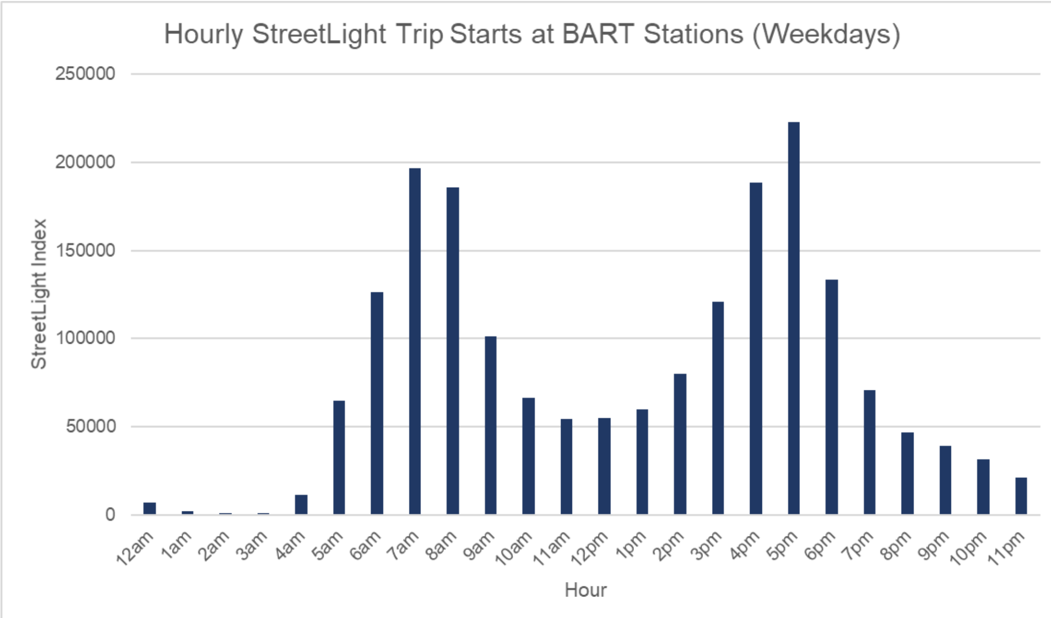


Figure 12: Hourly variation of the StreetLight Index for BART stations on average weekdays. StreetLight Metrics show distinct morning and afternoon peaks. Reference hour-of-day trip counts for BART alightings were not available for comparison.

Bus Validation

Understanding existing behavior is critical to transportation and planning efforts. StreetLight has developed algorithms and machine-learning techniques that utilize Location-Based Services data in order to identify bus trips across the United States and Canada. This validation section focuses on comparisons between StreetLight's Bus Metrics, published travel survey metrics, and bus ridership data in Los Angeles, California, and Cleveland, Ohio.

We've taken two approaches to validate our transit Metrics. First, we compare our aggregated trip characteristics to information published by household travel surveys (NHTS) and transit databases (NTD) to evaluate whether our Metrics are within the general range of expected trip characteristics for each mode. Second, we compare our trips at the bus-stop level to published ridership counts provided by transit agencies.

We use only transit ridership counts that we have deemed reasonably accurate, though the accuracy of bus passenger counts may be variable depending on the technology, which can differ across agencies. We do not use expanded counts or otherwise modeled passenger counts for validation, as they have additional sources of error. Our goal in this validation section is to demonstrate that our transit Metrics can be used in place of surveys or short-term passenger counts.

Comparing Our Bus Results to Travel Surveys and Transit Databases – NHTS and NTD

We analyzed our aggregated bus trip attributes and compared them to trip metrics published in the National Household Travel Survey 2017 (NHTS)[11] and the National Transit Database (NTD)[12] with transit reports from 2018. These two sources are fundamentally different but allow for valuable comparisons of trip attributes.

NHTS surveyed ~129,000 households across the U.S. and captured roughly 7,000 bus trips made by passengers. Meanwhile, NTD collects transit data directly from transit agencies through uniform reporting that allows for the analysis of agency-based financial and operating information. There were 46 transit agencies from 2018 that exclusively reported details about their bus trips in 2018. To create a comparison metric set, StreetLight analyzed roughly 81.3 million bus trips across the continental U.S. that occurred in April, May, September, and October of 2019 and 2020 (for more on how we infer and create bus trips, please see the separate Methodology section).

We compared key average characteristics of bus trips from our data set to the NHTS and NTD reported values. We do not set an "exact match" as the goal. All three data sets are samples, and all three thus have different strengths and sources of error. Where discrepancies occurred, we believe they are explainable by known differences in collection methods. It should be noted that NHTS reports buses in three categories: "Public or Commuter" and "Private/Charter/Tour/Shuttle" and "City-to-City." StreetLight does not currently separate these different types of buses. Table 6 shows StreetLight average bus Trip Attribute Metrics relative to the published NHTS and NTD numbers. Note that in the case of NTD, duration metrics were not reported; thus, they were inferred from reported distance and speed values.

Source	Avg Duration (min)	Avg Distance (mi)	Avg Speed (mi/hour)
NHTS (Public/Commuter)	49.92	7.15	8.59
NHTS (Private/Charter/Tour/Shuttle)	51.24	24.15	28.28
NHTS (City-to-City)	86.32	37.41	26.00
NTD	23.97 (inferred)	5.17	12.94
StreetLight Bus Metrics	24.46	4.10	11.31

Table 6: Comparison of average trip duration, distance, and speed from across NHTS, NHD, and StreetLight Bus Metrics.

As shown in Table 6, we found the StreetLight trip duration, distance, and speed averages were very close to the NTD reported values. NHTS values differed from both the NTD reported data and StreetLight's Metrics, specifically in the duration category. Because NHTS trips are self-

[11] https://nhts.ornl.gov/
[12] https://insights.transitcenter.org/

reported from memory, our suspicion is that the NHTS duration metrics may be over-inflated, potentially including wait-times, transfers, and other transitional modes in the reported bus trip. NTD and StreetLight Bus (to varying extents) will isolate the bus portion of the trip from any potential chained multimode tour.

It's important to note the difference in average trip metrics reported in the three NHTS bus categories. Characteristics of bus trips can differ substantially based on the different types of bus included. By design, StreetLight's bus trips skew toward the "public/commuter" category, because public and commuter bus trips will make up the vast majority of bus trips occurring across the country. StreetLight may include private buses, but we expect them to make up a much smaller portion of the real share of bus trips in a year, and hence of our sample. Additionally, our current bus trip design was focused on capturing typical trips within a city, and thus we expect to under-sample City-to-City trips, such as via Greyhound and Megabus. The NTD sample includes only agency-reported data, and thus covers only public and commuter buses.

NHTS also provided trip distance distributions that can be used for further comparison. In the following example, we've included trip distance comparisons to the "Public/Commuter" category. Figure 13 shows the distribution of bus trips by length.

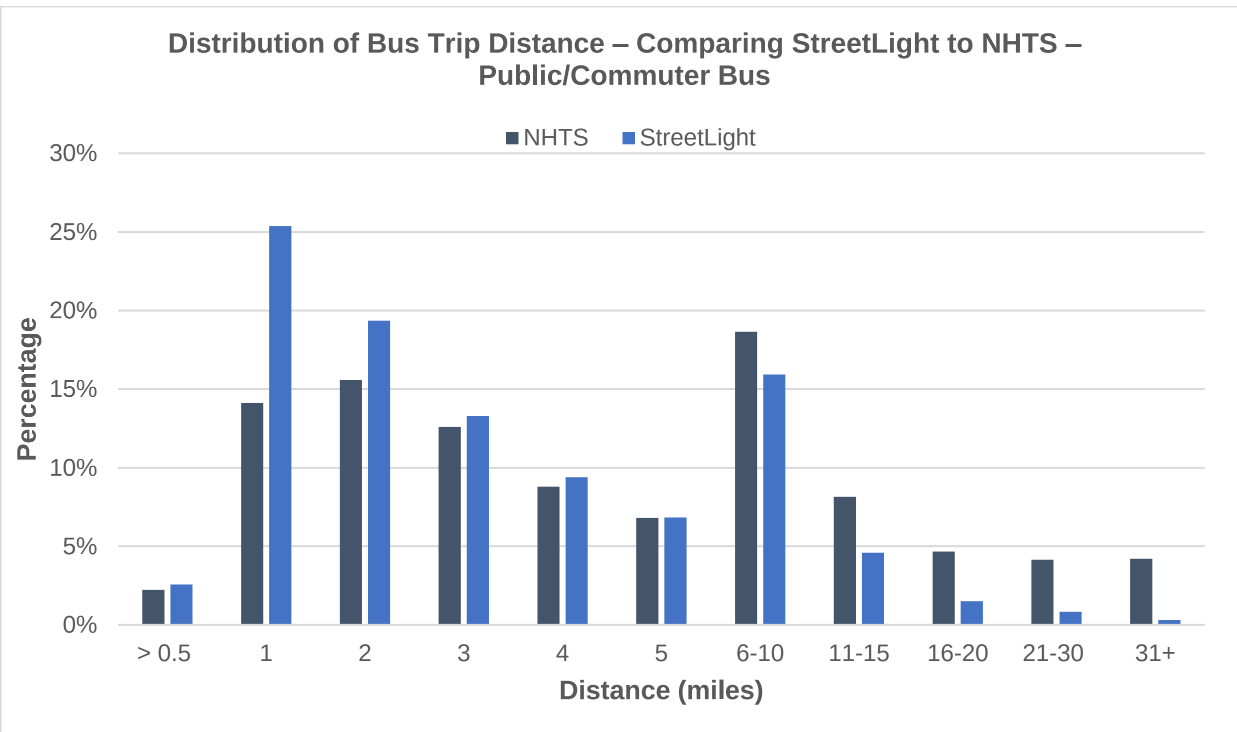


Figure 13: Comparison of bus trip length distribution between StreetLight and the NHTS.

StreetLight's distribution matches fairly closely with the NHTS survey, with a high proportion of trips in the 1-to-2-mile categories, then another peak between 6 miles and 10 miles in length. StreetLight does appear to capture more short-distance bus trips than the survey. Reasons for this may be that survey respondents are more likely to under-report short trips. Additionally, longer-distance commuter bus trips are more likely to take highways, as opposed to local roads, and also to make fewer stops. Our bus trips may under-sample trips along highways, as their characteristics make it harder to distinguish them from cars in our mode classification algorithm. This is something we'll continue to evaluate in future updates to this validation work, though for now, our strong parity with the NTD data source provides us with confidence in our bus Metrics.

Comparing Our Results to Bus Ridership Counts

To validate data at a more granular level, we compared StreetLight bus trips at known bus stations to agency passenger counts across a number of cities nationwide.

Bus Data Source Review

For this validation, we aimed to analyze agency data containing boarding and alighting passenger counts with monthly data from 2019 and 2020 to be able to validate as many of StreetLight's available months as possible. Because many sources of ridership data are published annually, or at the route level as opposed to the station level, this left us with a limited set of potential sources for comparison. We needed passenger counts that published their data on a daily or monthly basis in order to meet our strict criteria for temporal coverage. To assess performance across a diverse range of locations, our goal was to obtain passenger count data from a range of bus networks across the country.

The first subset of bus stop-level ridership data was obtained from Los Angeles Metropolitan Transit Authority (LAMTA). This data was provided to us directly by the agency. LAMTA's ridership data set included 13,677 bus stops in Los Angeles County comprised primarily of local bus lines. Data could be reported as passenger boardings and alightings by bus stop, and was available as monthly averages by weekday and weekend for October 2019.

The next agency from which we were able to obtain stop-level ridership data was Central Ohio Transit Authority (COTA),[13] which services the greater Columbus region. COTA's ridership data set was comprised of 3,076 bus stops and was reported as passenger boardings and alightings at the bus stop level. Data was available as monthly averages by weekday and weekend for all months in 2019 through August 2020.

To count passenger boardings and alightings, both agencies utilized Automatic Passenger Counting (APC) systems to quantitatively monitor passenger use on their bus networks. APCs are most commonly sensors mounted in bus doorways that count passengers as they are boarding and exiting a vehicle while information on bus stop location and time is also being recorded.

[13] https://discovery.smartcolumbusos.com/dataset/central_ohio_transit_authority/601d7cf2_82d9_4a66_8f37_fff5392ab617

In evaluating the station counts for comparison, it's important to consider the range of average daily trip counts recorded across the agencies. Agency data was aggregated to capture average daily trip starts by station in October 2019 for LAMTA. For COTA, data was aggregated across 6 available months, April, May, September, and October 2019 as well as April and May 2020. In Figure 14, all agencies' trip distributions are displayed with the same scale (x-axis) to ensure an easy visual comparison across the sources. The figure contains distributions from 2019 months only, given the variability of transit ridership in 2020 due to COVID-19 restrictions.

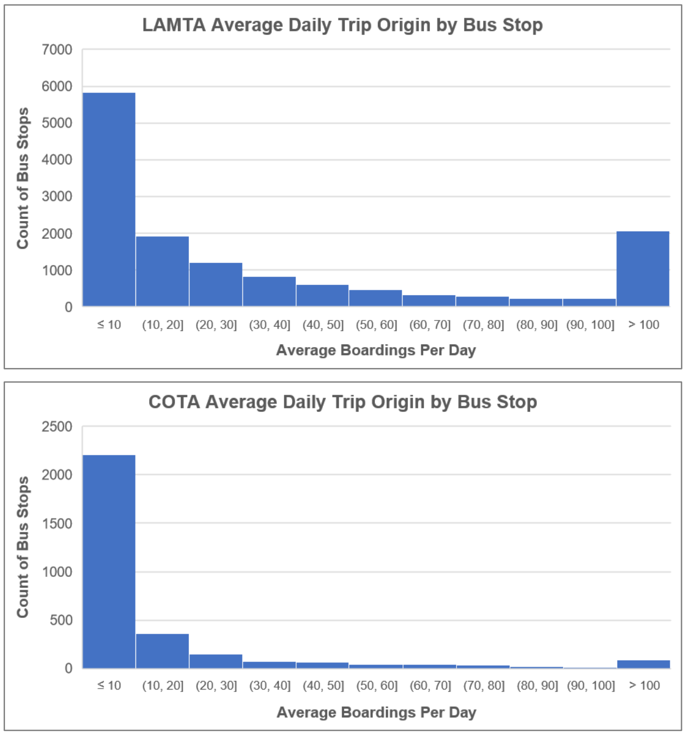


Figure 14: Histogram of bus stop boardings for LAMTA and COTA across available 2019 months. The majority of bus stops in both networks record fewer than 10 trips a day.

What's notable in the figure above is that both agencies have a very high proportion of bus stops with fewer than 10 passenger trips per day. This will impact our ability to analyze these stations both for sampling and privacy constraints, noted later in the validation portion of this report.

Initial Comparison to StreetLight Bus Metrics

After the source data was cleaned and aggregated, we compared the monthly boarding counts at stations with the origin StreetLight Index value on the matching bus station zones obtained from the StreetLight InSight® platform. Our goal was to determine the correlation between the two sets of counts.

To create zones in StreetLight InSight® that mirrored the transit agency's data, we first obtained station points from GTFS through Transitland's API.[14] Each point represents the latitude and longitude of the bus station. As noted in the station trip distribution charts in the previous section, many bus stops have very few recorded trips. This poses a challenge for a variety of reasons. First, StreetLight's trip penetration rates for individual analyses varied within a contained range (depending on mode, location, and other factors); thus, the likelihood of capturing a single daily trip at some of these bus stop locations decreases, contributing to significant variability in our validation results. Second, many bus stops are in very close proximity to one another (think of an eastbound bus stop on one side of the street and a westbound bus stop on the other side). Though pings for LBS devices have fairly high precision, differentiating a trip from one side of the street vs. another is outside the bounds of spatial precision for many devices.

As a result, we determined that bus stop points will need to be aggregated via a clustering method. We do not currently require a bus trip to start or end at exactly the bus station. A rider's pings may not always perfectly align with a bus station, and a trip may include a small distance spent walking to arrive at a stop. Buffering zones for bus stops, and clustering stops that are close together, both ensure that bus trips were properly attributed to each location. Clustering bus stops also increases sample size, which additionally improves the signal. For this analysis, we chose to cluster bus stops that occur within 100m of each other. Once assigned to clusters, we buffered each cluster by an additional 50 meters and dissolved the buffered stops into a single polygon based on their assigned cluster ID. All steps in this process were completed using a GIS tool. The result is a polygon shapefile of grouped bus stops, as shown in Figure 15 below.

Clustered and Buffered Bus Stops for LAMTA

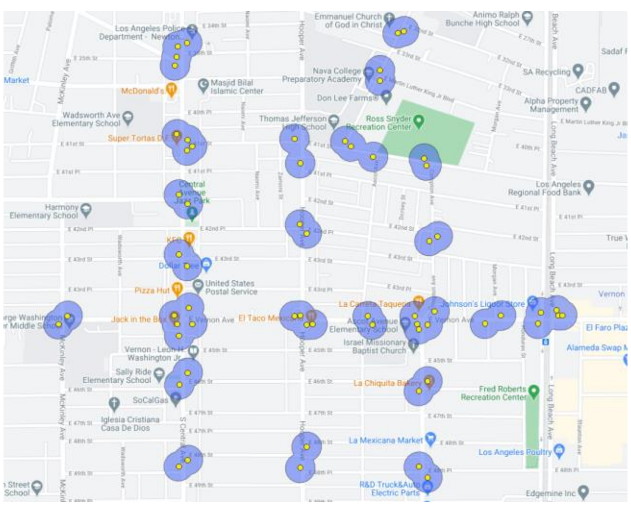


Figure 15: Clustered and buffered bus stops for LAMTA. Bus stops on opposite sides of the street, or bus stops situated in the same intersection, are grouped together for purposes of comparing to agency data.

[14] https://www.transit.land/documentation/an-open-project/contributor-agreement.html

Once the clustered bus stops are all converted to polygons and exported as shapefiles, they can be uploaded directly to the StreetLight InSight® platform. A zone set was created for each bus stop cluster and all zones were "non-pass-through" in order to capture the starts and ends of bus trips. For more information on creating zones in StreetLight InSight®, see our [Support Center](). Snapshots of the final zone sets in StreetLight InSight® are shown in Figure 16 below.

Station Zones for LAMTA and COTA in StreetLight InSight®

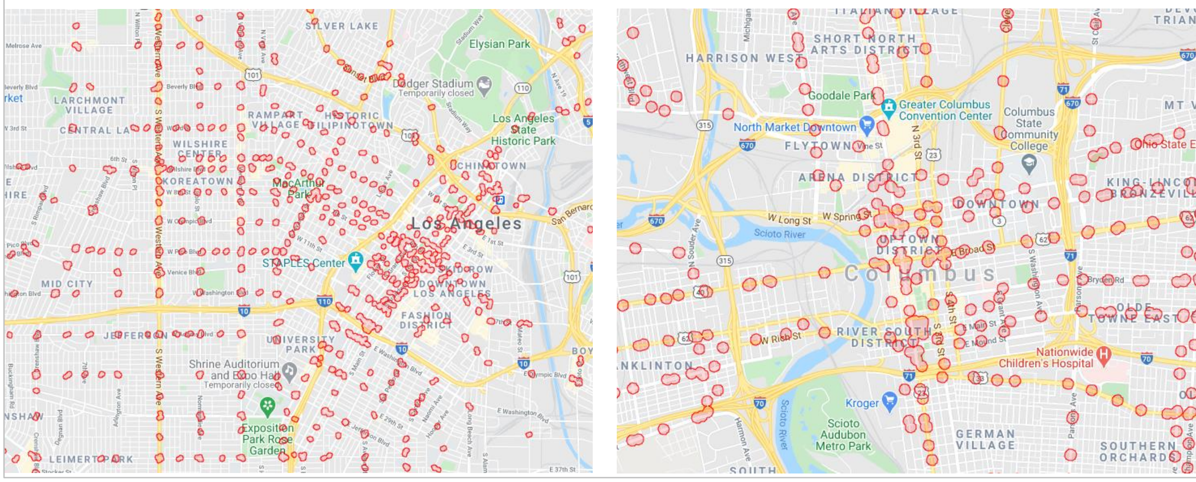
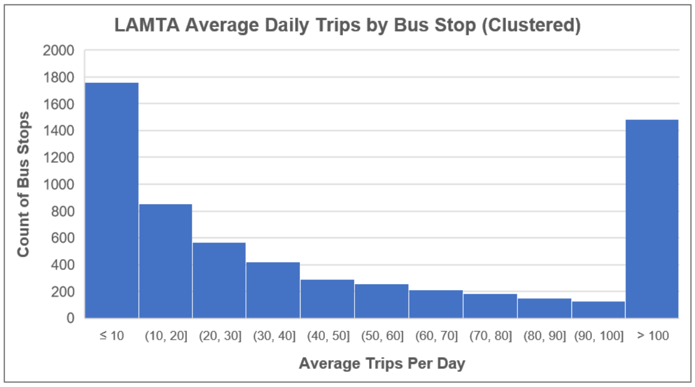


Figure 16: Station zones as uploaded in StreetLight InSight®. Clustered LAMTA bus stops near downtown Los Angeles (left map) and COTA bus stops in Columbus (right map).

This process results in 6,361 bus stop cluster zones for LAMTA bus stops (from 13,677), and 1,684 bus stop cluster for COTA bus stops (from 3,076). After re-evaluating the agency's boarding counts by cluster, we see an increase in locations with more than 100 daily bus trips, as shown in Figure 17. As with the previous histograms, these metrics are shown with available 2019 months only.



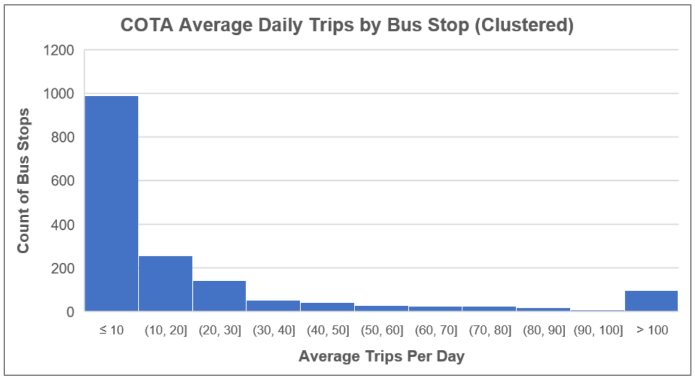


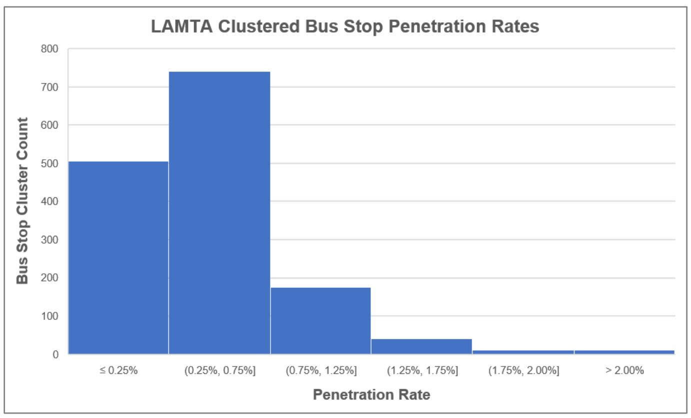
Figure 17: Histogram of bus stop trip starts for LAMTA (top graph) and COTA (bottom graph) for available 2019 months. The majority of clustered bus stops in both networks still record fewer than 10 trips a day, but a higher proportion of locations record more than 100 trips per day.

Note that customers also have the option to manually draw station zones within the StreetLight InSight® application.

Before diving into deeper validation, we wanted to highlight some potential sources of error for bus passenger counts. Like any counter technology, APC sensors are subject to malfunction or other technical issues. Specifically, APCs are known to struggle counting passengers in overcrowded buses. National Transit Database reporting policy guidelines recommend validating APCs against manual samples. In some cases, APC data should be processed to correct for anomalies or adjusted with factors.[15] As mentioned earlier, StreetLight's data available for bus stop matching has an additional source of error, accurately ending bus trips within close proximity to a bus stop, which is why we recommend buffering and clustering bus stops for analysis. Given the breadth of bus networks, especially in large urban areas with lots of frequent stops and intersecting bus routes, this can be challenging. Thus, accuracy will be improved wherever it is possible to use clustering, such as was done in this study. Finally, it's possible that the station points obtained from GTFS may not be centered perfectly relative to the real-world bus stop, especially with larger bus depots, thus causing buffered zones to miss portions of stations that may capture additional trips.

[15] https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/ntd/133146/2019-ntd-reporting-policy-manual-v1-2.pdf

To get a sense of our general performance across bus stops, it’s important to consider the penetration rates recorded at each bus stop cluster. Penetration rates are calculated as the sum of the average daily StreetLight sample divided by the sum of the average daily bus stop passenger boardings. A relatively consistent penetration rate is key to consistent results in a validation. Penetration rates are also key to helping us understand which bus stop clusters are outliers, indicating some other source of error may be at play. For the following penetration rate evaluation, we’ve limited the agency clusters to only those with more than 100 trips per day to ensure a viable sample size for analysis. This does not mean that no station with under 100 trips/day can be analyzed in StreetLight InSight® – just that we expect results to be more variable for such small stations. In Figure 18, penetration rates for different bus networks are aggregated across all available months in 2019 and 2020, and are displayed with the same scale (x-axis) to ensure an easy visual comparison across the sources.



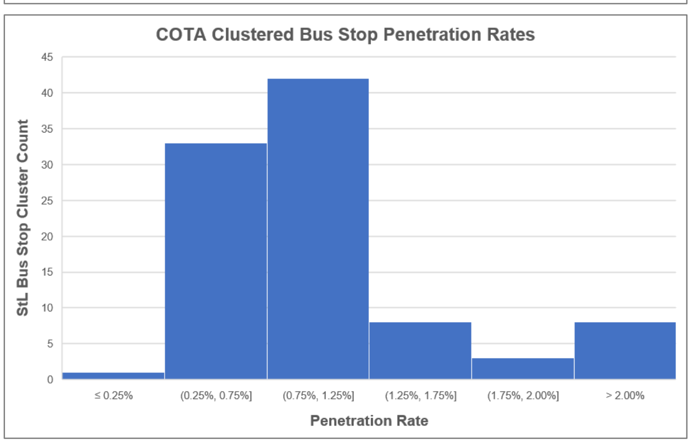


Figure 18: Distribution of average trip start penetration rates at bus stop clusters. LAMTA penetration rates are clustered around 0.5% (top graph); COTA penetration rates tend to be above 0.75% (bottom graph). Penetration rate refers to the ratio of StreetLight bus trips to reference passenger count.

Overall, penetration rates differ across agencies but still fall within expected ranges for our LBS data source. As mentioned earlier, the size and volume of the bus station clusters analyzed in StreetLight InSight® influences the different penetration rates across agencies. Geographic location is another factor: we’ve seen our LBS penetration rates tend to be higher in the

Midwest and Southeast relative to East Coast and West Coast regions, which might be contributing to COTA's slightly higher values.

COMBINED CORRELATION RESULTS

To evaluate correlation between agency-reported bus data and StreetLight Metrics, we evaluated each agency separately. The benefit of validating each agency on its own is that it enables us to confirm the quality of metrics across a variety of regions. First, we evaluated the correlation between StreetLight Index values and the sum of average daily trips occurring at each bus stop cluster. Agency metrics were aggregated as daily averages across April, May, September, and October 2019 and April and May 2020 for COTA; LAMTA had only October 2019 available. StreetLight Metrics were reported as average daily Index values generated from LBS data aggregated across the same calendar months. For each station, we compared the average daily agency count of passenger boardings to the StreetLight Index value representing bus trips starting at the same locations.

As noted in the previous discussion around penetration rates, we chose to compare only bus stop clusters with agency counts above 100 trips per day on average. Bus stop clusters underwent some additional outlier detection. In this case, bus stop clusters with penetration rates 2 standard deviations outside the mean penetration rate were considered outliers. In most cases, this removed a small portion of stations with higher-than-expected penetration rates.

COTA correlation results were strong, with an R^2 value of 0.90. Results in Figure 19 are displayed on a logarithmic scale for optimal visualization.

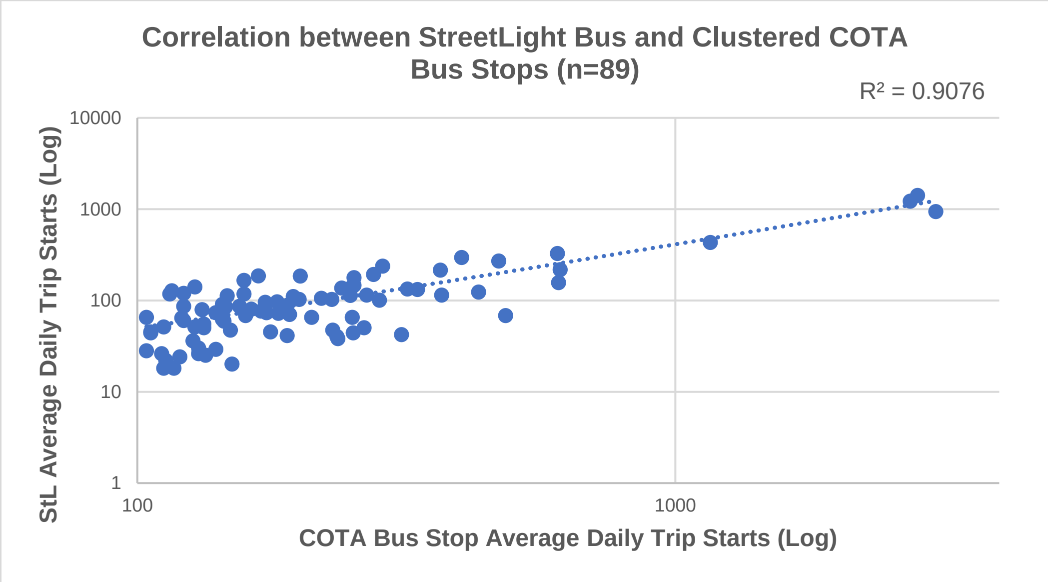


Figure 19: Correlation between average daily rail trips reported by COTA and average daily rail trips reported by StreetLight. Data was aggregated across six months in 2019 and 2020, with an R^2 value of 0.90. Data is logged on the y and x axis to better visualize values due to the high range and skew of data toward low counts.

For LAMTA bus stops, the results showed high correlation, with an R^2 value of 0.62. This means that StreetLight's Metrics "explain" 62% of the variation in the bus stop counts from October 2019. Results in Figure 20 are displayed on a logarithmic scale for optimal visualization.

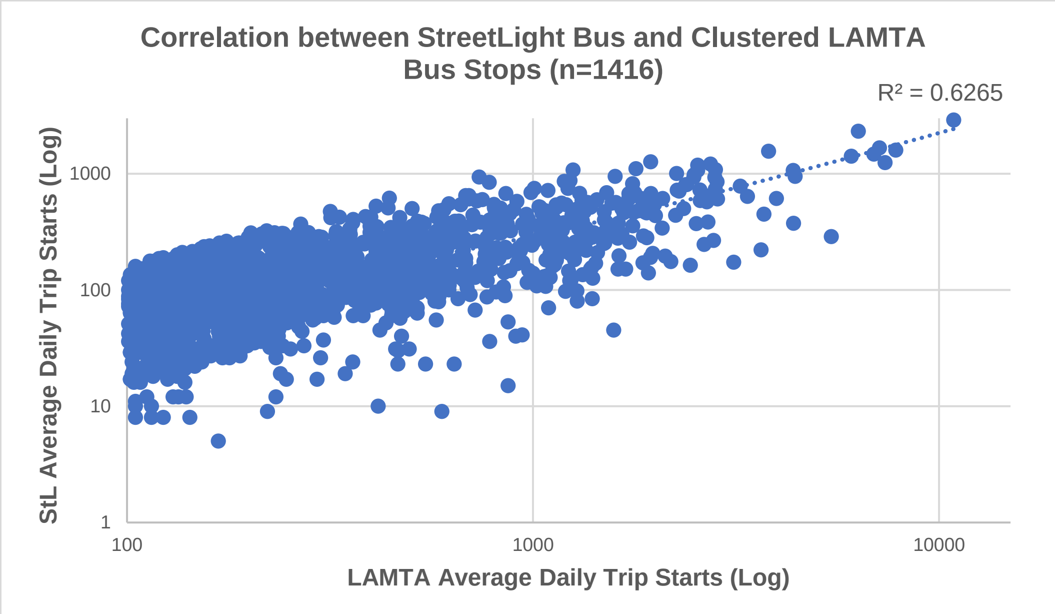


Figure 20: Correlation between average daily bus trips reported by LAMTA and average daily bus trips reported by StreetLight. Data was aggregated across October 2019, with an R^2 value of 0.62. Data is logged on the y and x axis to better visualize values due to the high range and skew of data toward low counts.

After deeper investigation, we have some explanations for the increased variation of the LAMTA results. First, we'll highlight the fact that LAMTA made only one month of data available for comparison, relative to other agencies where we were able to compare six months of available data. Second is that single GTFS-obtained latitude and longitude points used to derive bus stop locations may not always be in correct locations. In some cases, they were not centered appropriately relative to the actual bus stop, and in others they represented larger bus depots or areas where a 150m buffer may not accurately capture the extent of where buses load passengers. Third, due to the vast and complex road network in Los Angeles County, we also noted a number of stations that closely overlapped other infrastructure where mode confusion might be higher. These locations included bus stops very close to rail lines or complex highway interchanges with bus stops underneath. Ultimately, it may be that these locations require additional manual tuning of polygon zones in order to best capture bus trips, especially for differentiating local stops and large depots. More aggressive data cleaning or outlier detection will likely improve correlation results even further.

Overall, given the limited availability of transit data, as well as the high volume of low-count bus stations available for comparison, we are very pleased with these results. Further research will go into optimal bus stop or bus line configurations, and in future white papers we hope to address these topics more directly.

TIME TRENDS

Another important component of bus data is the ability to compare StreetLight Metrics across time in order to assess patterns and trends. StreetLight offers the ability to perform hourly, daily, and monthly comparisons of its Metrics. First, we'll evaluate monthly trends by comparing StreetLight Metrics to published sources, specifically monthly trends published by the Federal Transit Administration (FTA) and the American Public Transportation Association (APTA).

The National Transit Database (NTD)[16] maintained by the FTA publishes "Unlinked Passenger Trips" reported monthly by mode and type of service. The values below are aggregated across all agencies reporting bus trips to the FTA. According to the FTA, some agencies may report a 100% count of their unlinked passenger trips, while others may not have the resources or technology available for a 100% count and thus report a sample of trips.

APTA has partnered with the Transit App[17] to estimate and track ridership estimates over time. They publish weekly ridership estimates for individual agencies as well as national metrics. Estimated values are statistically modeled based on measures of Transit App usage, and therefore do not represent actual reported ridership counts. Below we've compared the total StreetLight Index for bus trips (U.S. and Canada) to APTA's National Estimates. It should be noted that APTA's national estimates may include both bus and rail transit usage. Regardless, we felt these numbers provided a useful point of comparison.

In the table below, we compare the year-over-year change by month for the StreetLight Index aggregated across all StreetLight trips in the U.S. and Canada, to the ridership samples and estimates published by the National Transit Database and APTA. When validating, we look for changes over time to be directionally accurate when compared to other sources, as well as within similar magnitude.

YOY Change by Month	National Transit Database	APTA	StreetLight Index (Bus)
April	-72%	-80%	-66%
May	-67%	-75%	-57%
Sept	-54%	-63%	-43%
Oct	-56%	-66%	-45%

Table 7: Comparison of year-over-year changes by month across the National Transit Database, APTA, and the StreetLight Index. Changes over time match closely across the three sources.

[16] https://www.transit.dot.gov/ntd/data-product/monthly-module-adjusted-data-release
[17] https://transitapp.com/APTA

Overall, the StreetLight Index for bus trips is directionally accurate and consistently falls within 20% of the trends reported by the two referent sources. This gives us confidence that StreetLight Bus Metrics are consistent with nationally reported trends.

In addition to evaluating national trends, we can assess system-wide trends using monthly data provided by the agencies referenced earlier. For system-wide trends, we have limited agency data available for comparison. With COTA, we can utilize reported monthly data through May 2020. But LAMTA's limited monthly data does not allow for time trend comparisons. In StreetLight InSight®, we ran monthly analyses with COTA bus stop clusters using the same station zones described earlier in the paper.

In an effort to expand our number of available sources for time trend comparisons, we also referenced agency-based metrics provided by the National Transit Database. We randomly selected agencies in mid-sized cities with bus ridership reported through Fall 2020. These cities included Buffalo, New York, Wichita, Kansas, and Savannah, Georgia. While we did not have station polygons to draw from, we created large area zones in StreetLight InSight® that encapsulated the bus routes in each selected metropolitan region. We then evaluated the bus trip starts that occurred within the region by month via a series of monthly Zone Activity analyses. While not as precise as the station-based methods available for COTA, these regional metrics still provide a useful reference point for understanding the performance of the StreetLight Index over time. Below we show what year-over-year trends look like when comparing April 2019 to April 2020 across the available sources.

Agency	Agency's reported YOY % change for April	StreetLight Index reported YOY % change for April
COTA (station-based)	-47%	-67%
Buffalo, NY (region-based)	-53%	-62%
Wichita, KS (region-based)	-63%	-52%
Savannah, GA (region-based)	-80%	-64%

Table 8: Agency reported year-over-year changes compared to StreetLight Index year-over-year changes for April. Trends are evaluated across Columbus, Ohio, (COTA) with a station based-approach, and Buffalo, Wichita, and Savannah with a region-based approach.

As illustrated in the table above, the StreetLight Index is able to capture the dramatic drop in bus trips across the two years due to COVID-19 impacts, though the magnitude of the directional change may vary.

Results can also be visualized across months, as a broader assessment of time trends. As a reminder, StreetLight currently provides eight seasonal months of bus data, rather than a full contiguous year, and in the case of COTA, only six months are available for direct comparison to StreetLight Metrics. In the figures below, the average daily ridership is compared to the StreetLight Index for the average daily trip boardings by bus stop cluster. In Figure 21, the StreetLight Index has been adjusted to COTA's system-wide counts for April 2019 in order to provide a simpler comparison of metrics across time. As a reminder, the StreetLight Index should be used for relative comparisons and is not meant to be an estimate of real-world counts.

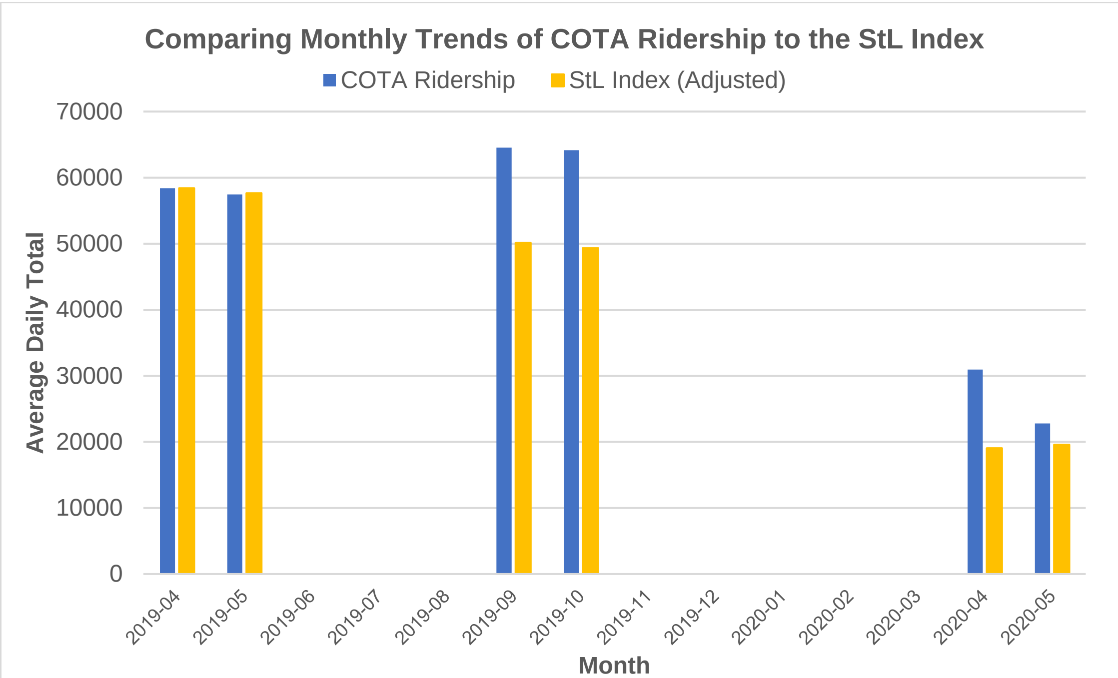


Figure 21: Monthly comparisons of COTA's agency-reported trips to StreetLight Metrics.

The figure above indicates that our StreetLight Index is able to capture the drop in bus trips between 2019 and 2020. However, small month-to-month variations in passenger count are captured less robustly.

As a result of the monthly validations performed at the national, system-wide, and station levels, we feel confident in the use of the StreetLight Index as a tool for evaluating large changes across time at the macro level. Large changes include significant seasonal increases or decreases of more than 15%, or in the case of 2020, the dramatic decrease in transit activity due to COVID-19. As with many of our metrics, more subtle monthly changes from typical seasonality may be difficult to capture, especially when analyzing locations with lower sample sizes (few passengers on an average day) or more variable sample sizes. Relative to StreetLight's Rail Metrics, StreetLight's Bus Metrics may be more prone to noisy fluctuations across time due to localized issues, which can range from confusion between bus and car modes to variation in the assignment of trip ends to bus stops. This should be a consideration for customers evaluating bus time trends, especially in finer-grained analyses, like those that are station or systemwide.

Due to the fact that few agencies provide stop-level data at the monthly level, there is still room for further validation and assessment of our bus metrics over time. We urge customers to use caution when comparing Bus Metrics over time and encourage calibrating to external sources whenever possible. We expect daily and hourly trends to be consistent, but we hope to further explore this in future validations, as more comparison data becomes available.

About StreetLight Data

StreetLight Data, Inc. pioneered the use of Big Data analytics to help transportation professionals solve their biggest problems. Applying proprietary machine-learning algorithms to over four trillion spatial data points over time, StreetLight measures multimodal travel patterns and makes them available on-demand via the world's first SaaS platform for mobility, StreetLight InSight®. From identifying sources of congestion to optimizing new infrastructure to planning for autonomous vehicles, StreetLight powers more than 6,000 global projects every month.


STREETLIGHT
DATA