Curbside Technology Performance Assessment Report

Curbside Management Pilot

Bellevue, Washington

February 7, 2022

Contents

Tables

Figures

Abbreviations

API	application programming interface
City	City of Bellevue
CNS	cargo network service
Collaborative	Smart Cities Collaborative
CSV	comma-separated values
FN	false negative
FP	false positive
GPS	Global Positioning System
ID	identifier
MaaS	Mobility-as-a-Service
MUST	mobile unit for sensing traffic
N/A	not applicable
PDT	Pacific Daylight Time
STAR	Smart Transportation Application & Research
SUV	sport-utility vehicle
T4A	Transportation for America
TN	true negative
TNC	transportation network company
TP	true positive
UTC	Universal Time Coordinated
UW	University of Washington

Appendices

Prepared for the City of Bellevue by

Executive Summary

The rise and continued use of Mobility-as-a-Service (MaaS) and increased use of the roadway by a variety of modes, among other factors, have all resulted in an increased demand for curbside access and utilization. In 2017, Transportation for America (T4A) launched the Smart Cities Collaborative (Collaborative) to engage with public agencies and planners in pooling technical expertise and resources to advance smart mobility concepts. The purpose of the Collaborative is to share experiences about new mobility technologies and develop best practices that increase access, safety, and economic opportunity for all involved parties. In 2020, the City of Bellevue (City) became actively engaged with T4A and was selected to be one of three pilot cities to conduct a curbside pilot project. The City chose to procure and install several video-based curbside monitoring devices to test each system's respective accuracy, efficiency, and ease of use. This report summarizes the methodology, results, and conclusions from this pilot project.

The primary goal of this project was to evaluate the accuracy of video-based curbside monitoring technology solutions in a real-world environment. The fundamental building blocks of a useful curbside management system include the capabilities to correctly identify when a vehicle is present (i.e., curbside is occupied), identify when a vehicle is not present (i.e., curbside is vacant), and determine the duration for which a vehicle has occupied the space. In addition to the fundamental building blocks, other metrics such as activity type, vehicle type, and vehicle length can be useful in expanding the range of curbside data captured by vendor technology for further applications. The City was interested in identifying a scalable system that could detect high-volume curb areas accurately for future enforcement and fare payment functionality. Ideally, a system would be able to identify the presence or absence of a vehicle with 95 percent accuracy while also being able to estimate the duration for which the vehicle occupied the curbside 95 percent of the time. These two criteria form the basis for evaluation of five systems deployed by vendors as part of this project.

Five vendors deployed their respective systems within Bellevue's central business district along 106th Avenue NE between NE 4th Street and NE 6th Street. Three evaluation rounds were conducted: (1) on November 14, 2020, for 3 hours, (2) between December 14 and 18, 2020, and (3) between May 3 and May 7, 2021. In addition to obtaining data and information from each vendor, project staff viewed and analyzed normal video feeds to identify the ground truth during each of these periods.

Data from each vendor and the corresponding ground-truth information were assembled into an evaluation database following a set of standardized procedures. One of the critical components of developing the database was defining the matching criteria that were used to link vendor-reported vehicle occupancy events to ground-truth identified occupancy events. Because of inconsistencies in the various systems' internal clocks, this matching required the establishment of time "windows" around the ground-truth reported start time for the vehicle occupancy and the ground-truth reported end time for vehicle occupancy. Three such matching windows were explored: ±30 seconds, ±1 minute, and ±2 minutes. Additionally, to alleviate the potential impact of the size of the geographic zone and the lack of painted parking lines in the areas being monitored, matching was performed at both the stall level (e.g., a single parking space) and the zone level (i.e., multiple parking spaces). Ultimately, the analyses suggested that using a ±30-second time window at the stall level is feasible and appropriate.

Data analyses were conducted using a confusion matrix that enabled the estimation of five key performance measures: Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, and Accuracy. Comparison of Accuracy versus the 95 percent goal was conducted using T-tests[1]. Duration was examined as a function of the average difference in the duration of vehicle occupancy as well as the percentage of vehicle occupancy events where the vendor-reported duration was within ±30 seconds of the actual duration obtained from the ground-truth data. T-tests for statistically significant differences were used to identify results that were statistically significant.

Table ES-1 summarizes the key results for the five vendor systems under the least and most stringent matching criteria, which provides the upper and lower boundaries for accuracy, respectfully. As observed in the table, no vendor was successful in achieving the goal of 95 percent accuracy as the calculated accuracy for all five vendors was statistically lower than 95 percent. With respect to accuracy, there was no statistically significant difference in the performance between the systems of Vendors C, D, and E. All of the systems were able to achieve the targeted performance level with respect to duration under the most stringent matching criteria. Additionally, under this same criterion, the average difference between the vendors and ground-truth data were less than 1 minute.

[1] A T-test, or Student's T-test, is a type of statistical hypothesis test that can be used to compare differences between two means, compare differences between two proportions, or compare a mean to a hypothesized value. It is based upon the properties of the T-distribution.

Table ES-1. Evaluation performance measures for the most and least stringent matching criteria

Criterion	Most stringent matching criteria					Least stringent matching criteria				
	Vendor					Vendor				
	A	B	C	D	E	A	B	C	D	E
True positive rate (Sensitivity)	31.7%	13.0%	80.6%	70.1%	70.8%	63.5%	92.6%	91.7%	85.7%	90.3%
Positive Predictive Value	25.4%	5.0%	69.9%	70.1%	53.1%	50.8%	35.7%	79.5%	85.7%	67.7%
Negative Predictive Value	28.3%	6.9%	85.7%	67.6%	72.0%	72.5%	81.4%	94.5%	84.8%	93.1%
True negative rate (Specificity)	22.4%	2.6%	77.1%	67.6%	54.5%	61.0%	16.3%	86.0%	84.8%	75.2%
Youden Index	-0.46	-0.84	0.58	0.38	0.25	0.24	0.09	0.78	0.71	0.65
Accuracy	**26.6%**	**5.5%**	**78.5%**	**68.9%**	**61.4%**	**61.9%**	**41.8%**	**88.1%**	**85.3%**	**80.7%**
Average difference in vehicle occupancy duration (minutes)	-0.16	0.16	0.97	-0.26	-0.05	4.35	-5.66	7.41	2.47	7
Percentage of occupied vehicle events where duration was within ±30 seconds	93.9%	93.8%	96.6%	91.7%	98.1%	79.8%	26.3%	90.9%	73.8%	81.5%

In conclusion, none of the evaluated systems were able to meet the desired performance criteria with respect to accuracy. All were able to estimate duration to the desired level under certain conditions. However, three of the vendor systems were able to achieve accuracy in the 60–80 percent range under the most stringent matching criteria, which may enable use of these systems for some applications.

In future pilots examining this type of technology, care should be given to standardizing the structure and content of the data to be provided by each vendor so that they can be more readily matched to the ground truth. Additionally, provisions for synchronizing the internal clocks at the beginning and throughout the data collection period should be made.

1 Background

This section presents background information about the Bellevue curbside initiatives and the Collaborative.

1.1 Bellevue Curbside Initiatives

Demand for curb use has increased dramatically in urban areas, creating new competition for limited right-of-way. Beyond transit, bicyclists and pedestrians, parking, and delivery vehicles, new users have emerged in the form of rideshare pickup and drop-off and micro-mobility devices such as e-scooters. Recognizing this challenge, the City has spent several years advancing curbside pilot projects, observing curbside behaviors, and conducting a curb inventory. Monitoring curbside use to better understand demand is a critical first step in developing a strategy that fairly allocates usage of this public resource.

In 2019, the City began creating a dynamic curbside inventory in its downtown area. Data were collected and registered within the CurbLR data standard that helps describe curb regulations in a digital format. This inventory effort was conducted to provide a foundational common resource between the City and the traveling public. The initial inventory was published in June 2020.

Also in 2019, the City partnered with the University of Washington (UW) Urban Freight Lab and the Pacific Northwest National Laboratory on a Department of Energy–funded project to study the behavior of delivery vehicle drivers and develop a mobile application that predicts the availability of a curbside loading zone near their destination. The goal of the project was to reduce parking-seeking behaviors and illegal on-street freight parking. Through this ongoing partnership, the City installed sensors near 10 curbside loading zones, which enabled real-time assessment of occupancy status.

1.2 T4A Smart Cities Collaborative

> **Smart Cities Collaborative Mission**
>
> To maximize the transformative potential of emerging technologies and new mobility options to increase access, safety, and economic opportunity for all residents while reducing congestion, vehicle miles traveled, and greenhouse gas emissions.

In 2017, T4A launched the Smart Cities Collaborative to engage with public agencies and planners in pooling technical expertise and resources to advance smart mobility concepts. The purpose of the Collaborative is to share experiences about new mobility technologies and develop best practices that increase access, safety, and economic opportunity for all involved parties. For each annual session, the Collaborative selected a mobility challenge to focus on. In late 2019, T4A announced that the Collaborative would focus on curb management in the 2020 session. In December 2019, the City of Bellevue was selected by T4A as one of three cities to conduct a curb pilot program and act as one of the host cities for the Collaborative.

The original intent of the curb pilot was to make operational changes at the curbside with the goal of balancing curbside access among competing uses. When the COVID-19 pandemic disrupted travel patterns, the Collaborative pivoted and became a multiagency brain trust for curb-related rapid-response strategies, such as creating best practices for app-based food delivery curbside pickup zones, on-street dining areas, and “healthy street” networks. In addition to COVID-19-related rapid-response strategies, the City continued to pursue and examine technology solutions that are capable of accurately monitoring curb space utilization.

The T4A pilot project aligned well with the City’s vision to address the emerging challenges associated with growth in curbside activity volumes and complexity.

The City recruited the HDR team (HDR Engineering, Inc. and IDAX Data Solutions) to help collaborate on this pilot project. The team assisted Bellevue with collecting data, analyzing outputs, and compiling the final report. This document provides an overview of the vendors included in the pilot project, the methodology used for the evaluation, and a summary of the evaluation findings.

2 Project Goals and Objectives

The goal of this project was to identify, install, and evaluate the accuracy of video-based curbside monitoring technology solutions. More specifically, the primary goal of the technology performance evaluation was to weigh each vendor’s curbside detection accuracy against third-party “ground truth” video data to assess the performance of each vendor’s system.

The following performance goals were identified and communicated with the vendors prior to the start of the evaluation:

- Provide 95 percent time accuracy of curb arrival and departure events
- Provide 95 percent location accuracy of curb zone recordings

These performance goals were established based on the following reasons:

- Many curb activities occur in very short durations. This is especially true for activities related to pickup and drop-off activities. The ability to accurately capture data related to these events requires a high level of precision.
- Future applications such as automated curb enforcement or fare payment require a high level of accuracy to avoid dispute.
- Where there is a potential future application for automated fare payment, curb monitoring technology should provide similar levels of detection accuracy as electronic toll collection systems that charge usage fees to vehicles using tolled facilities.

A set of performance metrics were developed to measure the vendor’s system performance against these goals using the ground-truth data collected during the same evaluation periods. The data collection, evaluation methodology, and associated performance metrics are explained in more detail in Section 4 of this report.

3 Curbside Pilot Overview

This section presents an overview of the curbside pilot project, including site selection, assessment timeline, and participating vendors.

3.1 Site Selection

The technology for this project was installed along 106th Avenue NE between NE 4th Street and NE 6th Street. This area is in the central business district of downtown Bellevue. It is surrounded by commercial office buildings, hotels, restaurants, and shopping centers, and is near the major transit hub on 108th Avenue. Curbside usage in this area includes on-street parking, commercial loading zones, an electric vehicle charging zone, short-term passenger drop-off/pickup areas, bus stops, and employee shuttle stops.

The selected vendors deployed camera-based sensors in the pilot area. The sensor locations and their monitoring zones are shown in Figure 1 below.

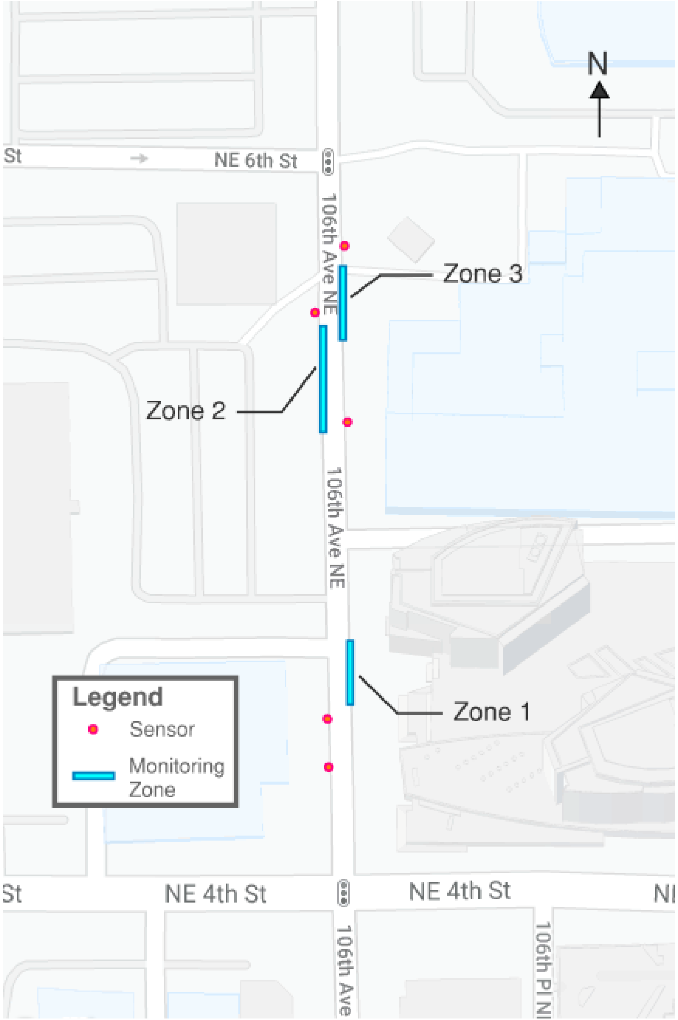


Figure 1. Project site map

3.2 Participating Vendors

This section describes the five participating vendors and the technologies they employed. **Error! Reference source not found.** summarizes the vendors' technology attributes and the zone(s) each vendor was responsible for monitoring.

Automotus

Automotus deployed two camera-based sensors, which transmitted video images to a web-based dashboard using cellular connectivity. Curb use statistics were automatically sent to the City via application programming interface (API).

University of Washington (STAR Lab)

UW deployed two camera-based mobile unit for sensing traffic (MUST) sensors with edge computing capability. Data were transmitted via cellular connectivity, without the ability to view the video feed. UW transmitted curb use statistics to the City manually each day.

Cleverciti

Cleverciti deployed one camera-based sensor, which included the ability to detect Global Positioning System (GPS) coordinates at four corners of a vehicle. While the Cleverciti sensor can use multiple communication protocols, cellular connectivity was used for this project. Video images were transmitted to a web-based dashboard using cellular connectivity, with curb use statistics automatically sent to the City via API.

Verizon

Verizon deployed two camera-based sensors for the final round of evaluation. Video and curb use data were sent directly to Verizon's in-house portal via cellular connectivity. Verizon transmitted curb use statistics to the City at the conclusion of the data collection period, with additional consolidation and cleaning performed by the City team.

Umojo

Umojo deployed two camera-based sensors for the final round of evaluation. Video images were transmitted to a web-based server using the city's fiber optic network, with curb use statistics automatically sent to the City via API.

Table 1. Summary of vendor technology attributes

Vendor name	Number of sensors	Technology	Video availability	Real-time data availability	Data format	Communication interface	Assigned zone(s)
Automotus	2	Camera-based	Video images available	Yes	Web-based dashboard, CSV pulled from API	Cellular data	2, 3
University of Washington	2	Camera-based with edge computing capability	N/A	No	Offline CSV	Cellular data	1, 2
Cleverciti	1	Camera-based with motion sensor detection	Video images available via offline XLSX	Yes	Web-based dashboard, CSV pulled from API	Cellular data	1
Verizon	2	Camera-based with edge computing capability	Video images available on vendor platform	No	Offline CSV	Cellular data	2, 3
Umojo	2	Camera-based	Video images available via Microsoft OneDrive	Yes	Web-based dashboard, CSV pulled from API	City fiber-optic network	2, 3

3.3 Assessment Timeline

The performance assessment initiative included three separate data evaluation periods spanning between November 2020 and May 2021. The first two assessments were conducted as a mutually beneficial exercise to allow vendors to correct errors and to allow City staff to reevaluate methodology and study approaches. The data listed in this report reflect the findings from the third and final assessment only.

3.3.1 Preliminary Assessment Rounds

The first round of preliminary evaluation was limited to 3 hours, from 10 a.m. to 1 p.m., on Saturday, November 14, 2020. The first evaluation period included three of the five vendors in this project. The project team conducted a preliminary assessment of the vendor performance for Automotus, Inc. and UW's Smart Transportation Application & Research (STAR) Lab. After reviewing and comparing the vendor data to the ground-truth data from November 14, 2020, the preliminary analysis identified areas for vendors to conform their reporting methodology with the ground truth methodology. The preliminary analysis results were shared with vendors for adjustments. Because of the discrepancies uncovered from the preliminary evaluation, a second round of analysis was required to adequately assess the technologies against the desired project goals.

The second round of data collection occurred from Monday, December 14 to Friday, December 18, 2020. In this round, Cleverciti joined Automotus and UW in participating in data collection and analysis. The analysis from the Round 2 evaluation resulted in additional insights to improve each vendor's accuracy performance. A narrowed set of data attributes were refined to resolve inconsistencies between each vendor's reporting mechanism and the ground-truth video data. Additionally, vendors were given the opportunity to recalibrate sensor settings to better capture the desired data output to meet the City's needs and expectations.

3.3.2 Final Assessment Round

For the final assessment, two additional vendors—Umojo and Verizon—were added to collect curbside data and increase sample size. Vendors captured curbside data on the following days: Monday, May 3–Friday, May 7, 2021

The ground-truth data were collected from Monday, May 3 to Friday, May 7, 2021, during the following two periods:

- 10 a.m.–1 p.m.
- 4 p.m.–6 p.m.

The data collection hours correlate to the peak usage of curbside activity in this area based on the ground-truth data collected during the second round of evaluation. A total of 25 hours of ground-truth data were collected to evaluate each vendor's performance.

Throughout the data collection and analysis process, the City and its consultant team actively coordinated with vendors to communicate data collection methodology, evaluation requirements, and data cleaning and formatting processes, and to resolve any concerns regarding analysis results.

4 Methodology

This section presents the methodology applied to the project, including data acquisition, data attributes, event matching criteria, and evaluation performance metrics. Methods described in this section are applicable primarily to the final round of evaluation.

4.1 Data Acquisition

This section describes the data acquisition process, including ground-truth data collection and vendor data acquisition and preparation.

4.1.1 Ground-Truth Data Collection

Ground-truth data were collected for the purpose of establishing a baseline data set to compare against vendor data. For the final round of evaluation, video cameras were mounted along 106th Avenue NE to capture traffic behavior footage at the curbside. This footage was then manually verified and compared to data outputs from vendor cameras.

Three main zones were identified and monitored for the final assessment. To set a foundation for comparison, individual stalls were assigned identifier (ID) tags within each monitored zone. No pavement markings delineate stalls along 106th Avenue NE, so ID stalls were denoted virtually and communicated to vendors. Figure 2 shows reference images that both the City team and vendor teams used to identify zone and stall IDs for their respective curbside data.

For the final evaluation, the City team installed four cameras in the study area. The cameras recorded the curbside zones, which were simultaneously monitored by vendors. The City team performed manual video review to extract curbside event data from the raw video footage and recorded the attributes for each of these vehicle occupancy events. The City team recorded event data in an Excel spreadsheet to allow side-by-side comparison to vendor event data. Only vehicle occupancy events as well as vehicle vacancy windows (explained in Section 4.1) within a specified time frame were considered. To provide consistency between the ground-truth and vendor data, the City team synchronized all data to Pacific Daylight Time (PDT) prior to the beginning of the evaluation period.

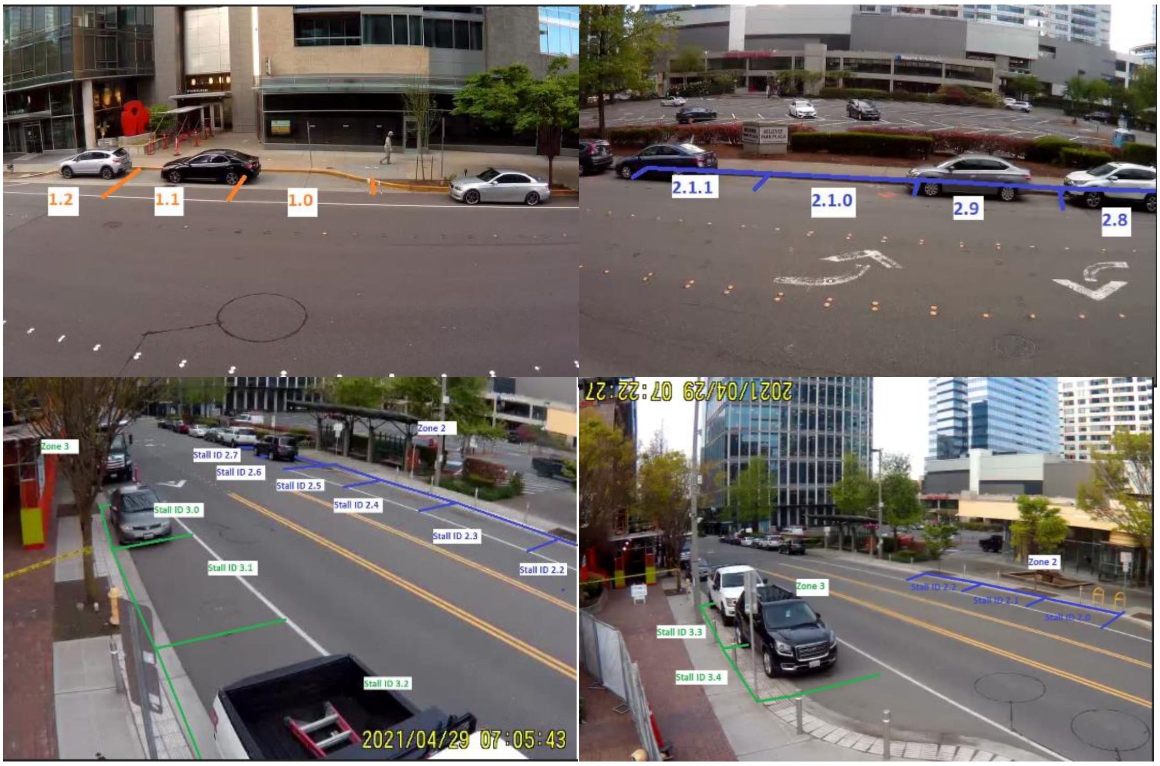


Figure 2. Curbside stall layout for Zones 1, 2, and 3

Note: The black vehicle shown in the upper-left image was recorded with a Primary Stall ID of 1.1 (full occupancy) and a Secondary Stall ID of 1.0 (partial occupancy) as specified in Table 3.

4.1.2 Vendor Data Acquisition and Preparation

Vendors prepared for the final round of evaluations by setting up and recalibrating their cameras and sensors. For this round, all vendors were asked to synchronize their sensors and systems to the Universal Time Coordinated (UTC) time zone. Vendors were then asked to report time in PDT to be consistent with the ground-truth data. Once the curbside data collection time frame was complete, vendors provided their data to the City team for comparison with the ground-truth data. The City established an expectation that data be provided by vendors via API given the desire for near real-time data and the long-term implications for curbside monitoring applications such as fare payment and parking enforcement. Three of the five vendors provided an API query that was used to pull the vendor data outputs from their respective online servers. Two of the five vendors were unable to provide an API and instead extracted data separately and submitted offline comma-separated values (CSV) files each day of the evaluation period. Data from the API queries and CSV files were prepared for evaluation using the following process:

1. Consolidate granular activity data into event data (if provided)
2. Verify and convert timestamps to PDT as needed
3. Remove unqualified vehicle occupancy events (as described in Section **Error! Reference source not found.**)
4. Translate GPS coordinates into Zone and Stall IDs (if provided)
5. Remove vehicle occupancy events that were outside of the study area

6. Adjust other attributes of interest to match ground-truth data format

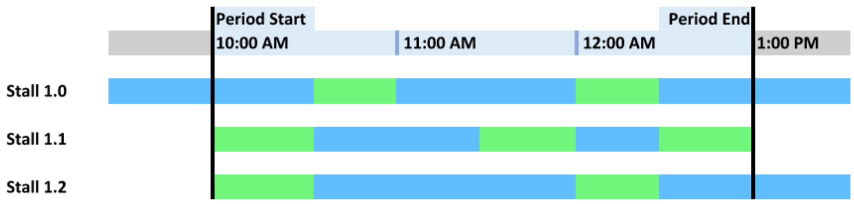
The general flow of the data cleaning and preparation process is described in the flow chart in Appendix A.

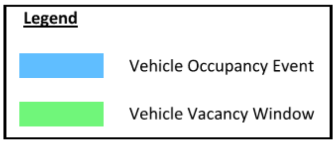
4.2 Data Attributes

The purpose of the ground-truth data collection was to assess the performance of vendors in collecting event data attributes described in the following sections. Vendors were provided with the attributes of interest and evaluation methodology prior to the evaluation period so they could calibrate their devices and adjust data formatting.

4.2.1 Vehicle Occupancy Events and Vehicle Vacancy Windows

The assessment of curbside use within the evaluation period required the establishment of vehicle occupancy events, or windows of time during which a vehicle was present at a curb. Ground-truth video was evaluated to determine the exact arrival time (Event Start) and departure time (Event End) of individual vehicles at the curb. The time bounds of curbside events also allowed the establishment of curbside vacancy windows during which the curbside was unoccupied. **Error! Reference source not found.** below illustrates how vehicle occupancy and vehicle vacancy windows were determined.





Count	Vehicle Occupancy Event	Vehicle Vacancy Window	Total
Stall 1.0	3	2	5
Stall 1.1	2	3	5
Stall 1.2	2	2	4
Total	7	7	14

Figure 3. Sample vehicle occupancy event vs. vehicle vacancy window diagram

Vendors, in turn, recorded curbside activity using the same approach to defining individual curbside events and submitted the data attributes reported for each unique vehicle occupancy event. Vendor vehicle occupancy events were evaluated against ground-truth data to compare reported time frames for vehicle occupancy and vacancy.

4.2.2 Locational Attributes

Vehicle occupancy events were also evaluated for the ability of each vendor's sensors to report a precise and accurate vehicle location. Locational accuracy was reported in the following two ways:

- **Zone ID:** A length of curb with logical start and end points (such as a curb cuts and pull-outs). Zones corresponded to those shown in Figure 1 above and each consisted of space suitable for 3 to 13 vehicles.
- **Stall ID:** A length of curb within a parking zone corresponding to the space required for a single vehicle. Because of the lack of pavement markings to define individual parking spaces, parking stalls were defined virtually. In many cases, vehicle occupancy events spanned two parking stalls, in which case both primary and secondary stall locations were recorded.

4.2.3 Other Attributes of Interest

While they were not evaluated as part of this project, some vendors provided additional attributes such as vehicle type, vehicle length, and park type. These were excluded from evaluation because of the subjectivity involved in classifying these attributes during the ground-truth video reduction process. Given the variability of this type of data and the capabilities across different vendors, these other attributes were not analyzed as part of this project. However, these attributes will likely be of future interest in assessing curb use, developing curb management policy, and monetizing curb use. presents a summary of attributes and describes each attribute in more detail.

Table 3 presents a summary of attributes and describes each attribute in more detail.

Table 3. Summary of attributes

Category	Attribute	Description	Notes
Time Attributes	Event Start	Timestamp when vehicle comes to a complete stop at the curb. Time will be recorded ***5 seconds*** after vehicle comes to a complete stop. [a]	Vehicles that stop in one stall for longer than 5 seconds and then move to a second stall are considered two events. However, if the vehicle briefly stops and shifts within the same stall, this would be considered only one event recorded once the vehicle stops for 5 consecutive seconds. This 5-second approach was established after the first rounds of validation to better capture accurate vehicle occupancy events and reduce noise in the data.
	Event End	Timestamp when vehicle begins to move. First instance of movement is recorded.	
Location Attributes	Zone ID	For reference only—general zone that ground-truth cameras are capturing. Zones include multiple stalls.	
	Primary Stall ID	Majority (or full) area where vehicle is occupying.	No hierarchical ordering between Primary and Secondary Stall ID when matching.
	Secondary Stall ID	Spillover area if vehicle is straddling multiple zones	Vendor will need to report if the Secondary Stall is occupied by the vehicle or not via API or CSV. We will compare both Primary and Secondary stalls with ground truth.
Other Attributes of Interest	Vehicle Type	Manual observation of vehicle type (car, passenger truck, sport-utility vehicle (SUV), delivery van, bus, freight truck). Also includes type-by-behavior, including cargo network services (CNSs) and transportation network companies (TNCs).	
	Park Type	Manual observation of park type (pull to curb, partially pulled to curb).	
	Vehicle Length	Manual observation.	

a. This methodology approach is refined from the first two rounds of evaluation development. The reason for this methodology change is to better capture “true” curbside events, and to filter out brief pauses that may result in false positives.

4.3 Event Qualification and Matching

The following rules were used in determining whether a vendor vehicle occupancy event was evaluated:

- All vehicle occupancy events with both an Event Start and Event End time occurring within the peak period were considered.
- Any vehicle occupancy event with an Event Start **OR** Event End time within this period was considered.
- All vehicle occupancy events in which both the Event Start and Event End time occurred outside the evaluation period were removed from consideration. This type of vehicle occupancy event is shown as an unqualified event in Figure 4.

Figure 4 graphically presents examples of qualified vehicle occupancy events as well as vehicle occupancy events that were removed from the analysis.

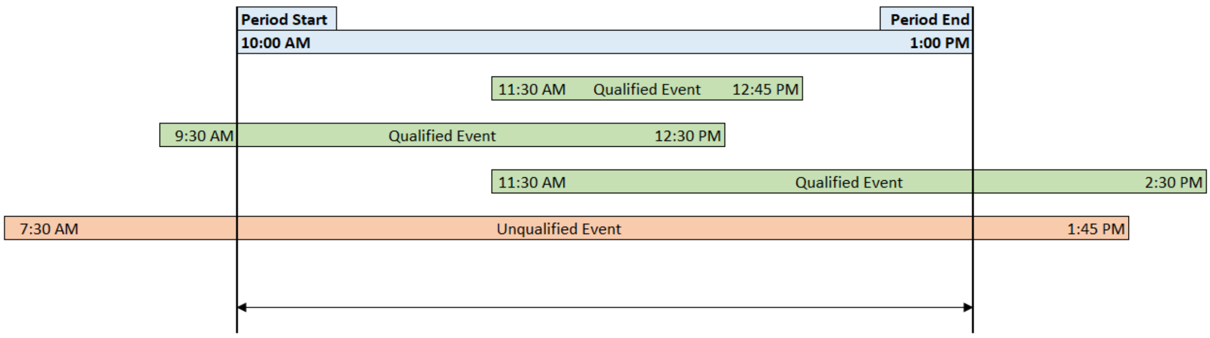


Figure 4. Qualified and unqualified vehicle occupancy event diagram

Each qualified vendor occupancy event was "matched" to a ground-truth vehicle occupancy event recorded by the City team. Events were matched on a one-to-one basis, meaning that each vendor-recorded vehicle occupancy event was associated with only one ground-truth vehicle occupancy event, and vice versa. The following criteria were used to determine and assess matched vehicle occupancy events:

- **Time:** The timestamps of vendor events and ground-truth events were compared. A maximum discrepancy of up to 2 minutes was allowed in matching Event Start or Event End timestamps to ground-truth data.
- **Location:** Vendor location data were compared to ground-truth video data. Zone ID was used as the minimum threshold in designating a match between vendor data and ground-truth data. Stall ID was also used as a more rigorous assessment of locational accuracy.

Error! Reference source not found. below illustrates the accuracy match types used for this analysis based on the time and location attributes ranging from the least to the most stringent criteria.

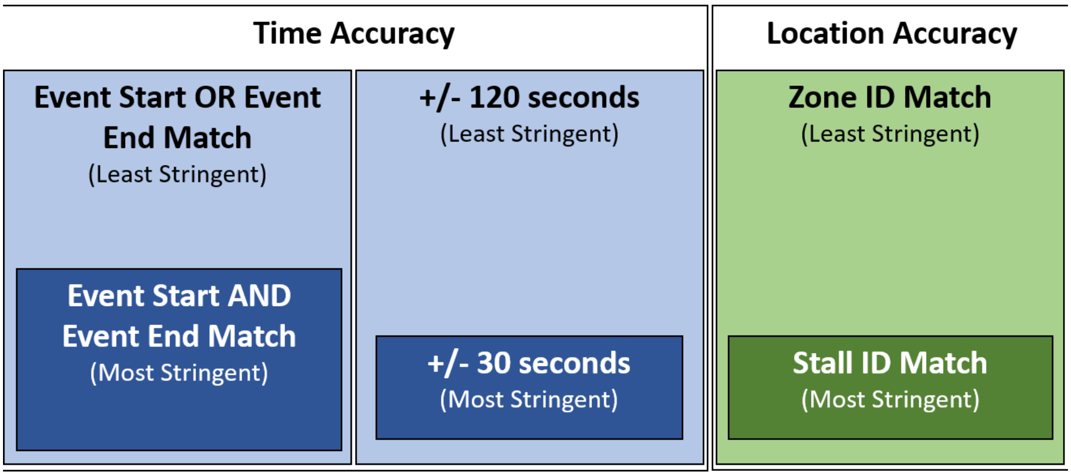


Figure 5. Overview of accuracy match types

4.4 Evaluation Performance Metrics

A set of performance metrics was developed to evaluate how well each vendor captured the curbside vehicle occupancy events and vehicle vacancy windows within its assigned zone(s). The description of each metric is summarized below.

Regardless of the differences in matching, the approach for evaluating each of the vendors' equipment was to observe a series of vehicle occupancy events and compare the reported status against the ground-truth data. Statistically, this is equivalent to a binary testing approach whereby both the observed outcome and the ground truth fall into one of only two categories corresponding to the reported presence or absence of a vehicle. The results can be readily described using a 2×2 confusion matrix as illustrated in Table 3. Correspondingly, six key performance measures are associated with this confusion matrix with the key metric being an assessment of the overall accuracy of the vendor's equipment.

Table 3. Confusion matrix and associated performance measures for evaluation

Confusion matrix		Observed ground truth		Measures
		Vehicle present	**Vehicle not present**	
Reported status by vendor	**Vehicle reported as present**	Number of true positives (TP)	Number of false positives (FP)	Positive Predictive Value $\frac{TP}{(TP+FP)}$
	No vehicle reported as being present	Number of false negatives (FN)	Number of true negatives (TN)	Negative Predictive Value $\frac{TN}{(FN+TN)}$
Measures		Sensitivity $\frac{TP}{(TP+FN)}$	Specificity $\frac{TN}{(FP+TN)}$	Accuracy $\frac{(TP+TN)}{(TP+FP+FN+TN)}$
		Youden Index $Sensitivity + Specificity - 1$		

Within the construct of the confusion matrix, the true and false positive/negative numbers along with the corresponding six performance measures can be interpreted as follows:

- **True positive (TP):** The number of true positives refers to the number of instances when the vendor correctly identified the presence of a vehicle that is also observed by the ground-truth data.
- **False positive (FP):** The number of false positives refers to the number of instances when the vendor reported the presence of a vehicle that was not reflected in the observed ground-truth data. A high number of false positives indicates an over-reporting issue from the vendor technology.
- **False negative (FN):** The number of false negatives refers to the number of instances when the vendor failed to report the presence of a vehicle that was recorded in the observed ground-truth data. A high number of false negatives indicates an under-reporting issue from the vendor technology.
- **True negative (TN):** The number of true negatives refers to the number of instances when the vendor correctly identified the curb space as being vacant among those times where the ground-truth data indicated that it was vacant.
- **Sensitivity:** Sensitivity refers to the proportion of times that the vendor's equipment correctly identified the presence of a vehicle among all of the instances where a vehicle was actually present. Sensitivity can also be referred to as the *true positive rate* as it is a measure of how well the vendor's equipment can identify true positive cases.
- **Specificity:** Specificity or the *true negative rate* is a metric associated with how well the vendor's equipment can correctly identify when a vehicle was not present among all of the times when a vehicle was actually not present. It is the proportion of times that the vendor's equipment matched the ground-truth data

with respect to no vehicle being present among all of the instances where a vehicle was not present.

- **Positive Predictive Value** is often referred to as *precision* and is defined by the proportion of events where a vehicle was actually present among all cases where the vendor reported a vehicle being present. Positive Predictive Value is an estimate of the probability that a vehicle is actually present when the vendor reports that a vehicle is present. A Positive Predictive Value of 1.0 would indicate that the vendor's equipment does not have any false positives.
- **Negative Predictive Value** is the percentage of cases where the curbside was actually empty among all cases where it was reported by the vendor's equipment as being empty. It is a measure of how many true negative events there were among all reported negative events and is an estimate of the probability that a vehicle is not present given that the vendor reports that a vehicle is not present.
- **Accuracy** is a measure of how well the vendor's equipment performed accounting for both identifying when a vehicle was present as well as when a vehicle was not present. It represents the proportion of times that the vendor's equipment correctly reported the status of the curbside with respect to the presence or absence of a vehicle.
- **Youden Index:** The Youden Index is an alternative measure to accuracy with respect to the overall performance of a vendor's system in correctly identifying vehicle occupancy events as well as correctly identifying curbside vacancy windows. This index is defined as follows:
    - Youden Index = Sensitivity + Specificity – 1.
    - According to W.J. Youden, "The index gives equal weight to false positive and false negative values, so all tests with the same value of the index give the same proportion of total misclassified results."[2] Although it is technically possible for the Youden Index to be negative, it is typically bound between 0 and 1. An index value of 0 would indicate that the system being tested does not have any ability to determine whether a vehicle is parked at the curbside. An index value of 1 would indicate that the vendor's equipment is perfect in both identifying vehicle occupancy events and determining when a vehicle is not present (i.e., no false positives or false negatives). As a rule of thumb, systems or tests with a Youden Index value of greater than 0.5 are considered to have positive predictability.

As all six of these performance measures are calculated using the same confusion matrix, they are closely related and interdependent. Ideally, a vendor's equipment would provide high Accuracy with high Positive Predictive Value, Negative Predictive Value, Sensitivity, Specificity, and Youden Index. In practice, there is often a tradeoff in performance between Predictive Value, Sensitivity, and Specificity.

---

[2] Youden, W.J. 1950. "Index for rating diagnostic tests". *Cancer*. **3**: 32–35. doi:10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3. PMID 15405679.

5 Results

This section presents the overall findings and results of the data analysis for the evaluation of the vendor's individual systems. However, as previously discussed, matching the vendor-reported events to the ground-truth events was a function of time windows around the start and end of each event as well as the location. As a result, the number of events and therefore the overall findings with respect to the six key performance measures varies as a function of the matching algorithm. In this section, we present the findings for only two sets of matching algorithms:

- **Most stringent matching criterion:** This criterion represents the algorithm or requirement that both the start and end times of the vendor-reported event needed to be within 30 seconds of the ground truth and they had to be assigned to the correct stall.
- **Least stringent matching criterion:** This criterion represents the algorithm or requirement where either the start time or end time of the vendor-reported event fell within 120 seconds (2 minutes) of the ground truth. Location has to match only within the overall zone assigned to the vendor (zones contain multiple stalls).

Providing results associated with these two matching cases provides "bookends" for the overall findings, which are included in Appendix B for completeness.

5.1 Performance in Detecting a Vehicle Event

A key consideration was whether any of the systems provided by the five vendors could achieve the City's objective of 95 percent accuracy with respect to the ability to detect whether a vehicle was using a specific curb space.

Figure 6 summarizes the calculated accuracy for each vendor under the two matching criteria. As observed in the figure, none of the vendors' systems achieved an overall accuracy of 95 percent. Statistically, all were significantly lower than 95 percent based upon simultaneous Bonferroni[3] adjusted T-tests with an overall error rate of 5 percent. Depending upon the matching algorithm selected, the estimated accuracy ranged from a low of 6 percent to a high of 88 percent.

[3] The Bonferroni adjustment consists of dividing the overall error rate desired across multiple tests by the number of tests to be conducted. It is a method to control the overall "family-wise" error rate to 5% by using a smaller "per-comparison" error rate as the error rate for each individual statistical test.

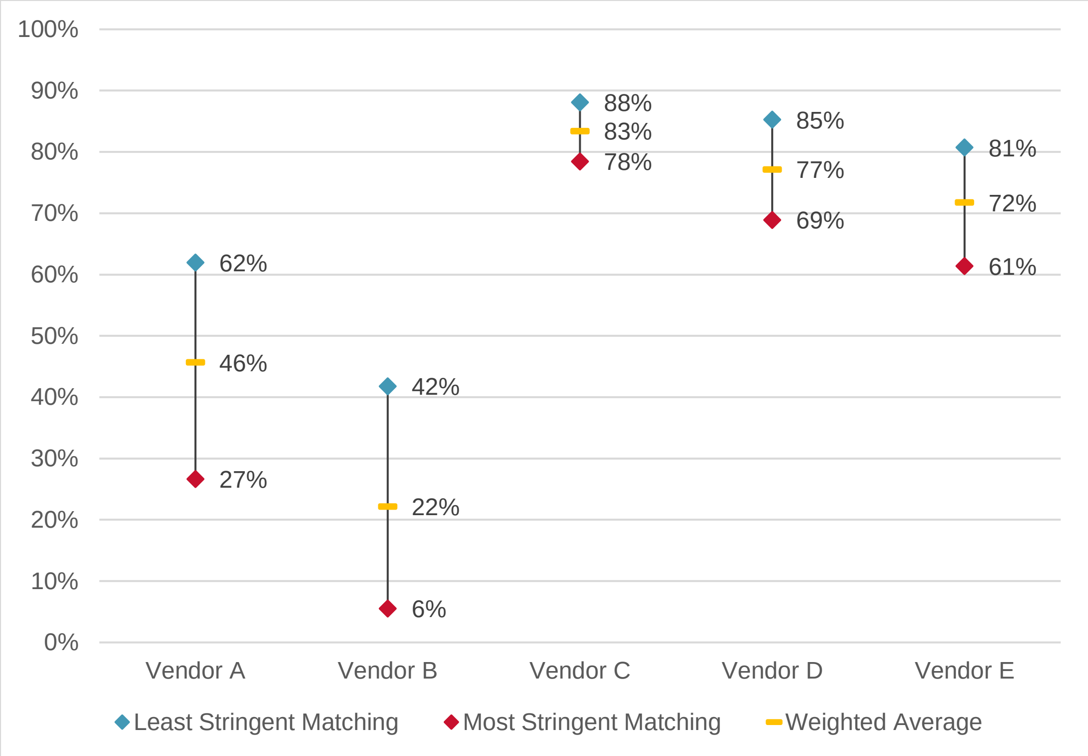


Figure 6. Accuracy for each vendor's system under the most and least stringent matching criteria

Pairwise, statistical tests comparing the estimated accuracy between each vendor indicates that Vendor B had the lowest accuracy, followed by Vendor A. There is no statistically significant difference in the accuracy of the Vendor C, Vendor D, and Vendor E systems though all three of these systems are significantly higher in estimated accuracy than those of Vendors A and B (see Table 5).

Table 5. Results of pairwise statistical tests of estimated accuracy between vendors' systems

Accuracy	Vendor B	Vendor A	Vendor E	Vendor D	Vendor C
Least stringent matching	41.8%	61.9%	80.7%	85.3%	88.1%
Most stringent matching	5.5%	26.6%	61.4%	68.9%	78.5%
Statistical significance	Statistically significantly lower than all other vendors	Statistically significantly lower than Vendors D, E, and C	No statistically significant difference		

Table 6 summarizes the Youden Index calculated for the five vendors. These results are very similar to those observed with the estimated accuracy. Vendors A and B's equipment does not have a high degree of predictability based on the Youden Index. Vendors C, D, and E's equipment all appear to have an ability to determine vehicle occupancy events and curbside vacancy windows.

Table 6. Calculated Youden's Index for the five vendors with the most and least stringent matching criteria

Youden's Index	Vendor A	Vendor B	Vendor C	Vendor D	Vendor E
Least stringent matching	0.24	0.09	0.78	0.71	0.65
Most stringent matching	0.00	0.00	0.58	0.38	0.25

Examining the remaining performance measures provides insight into the specific characteristics of each vendor's system that are driving the overall accuracy and Youden's Index results. Table 7 summarizes the calculations and findings for each of the six performance measures for the least and most stringent matching criteria.

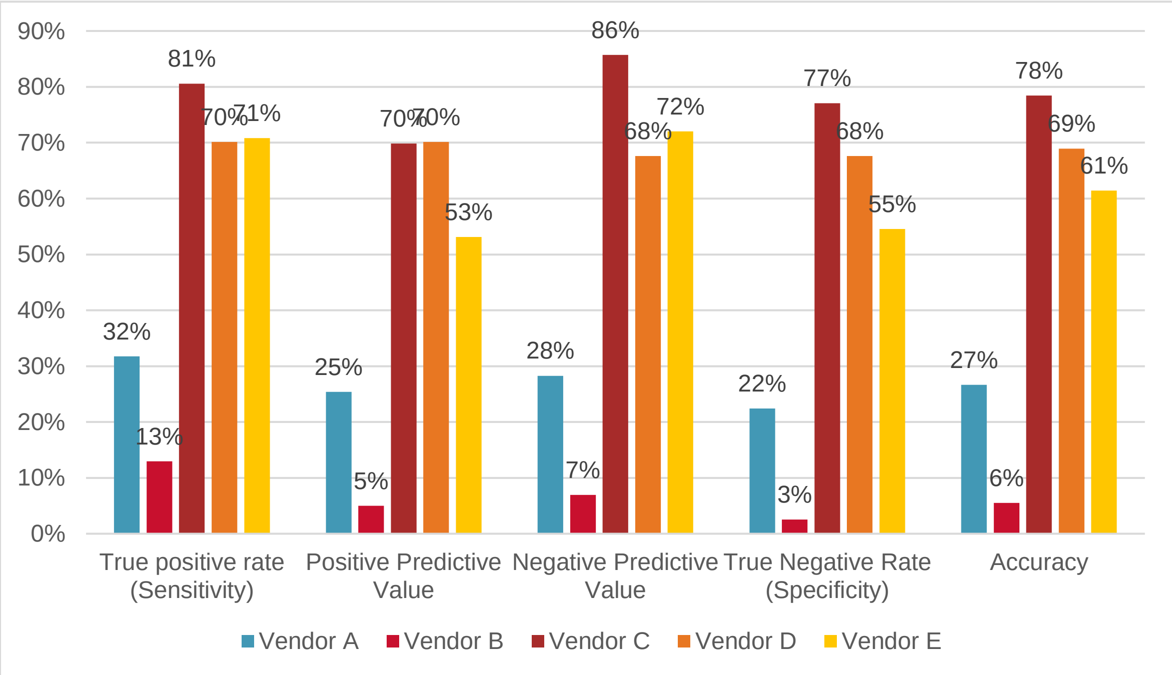


Figure 7 and

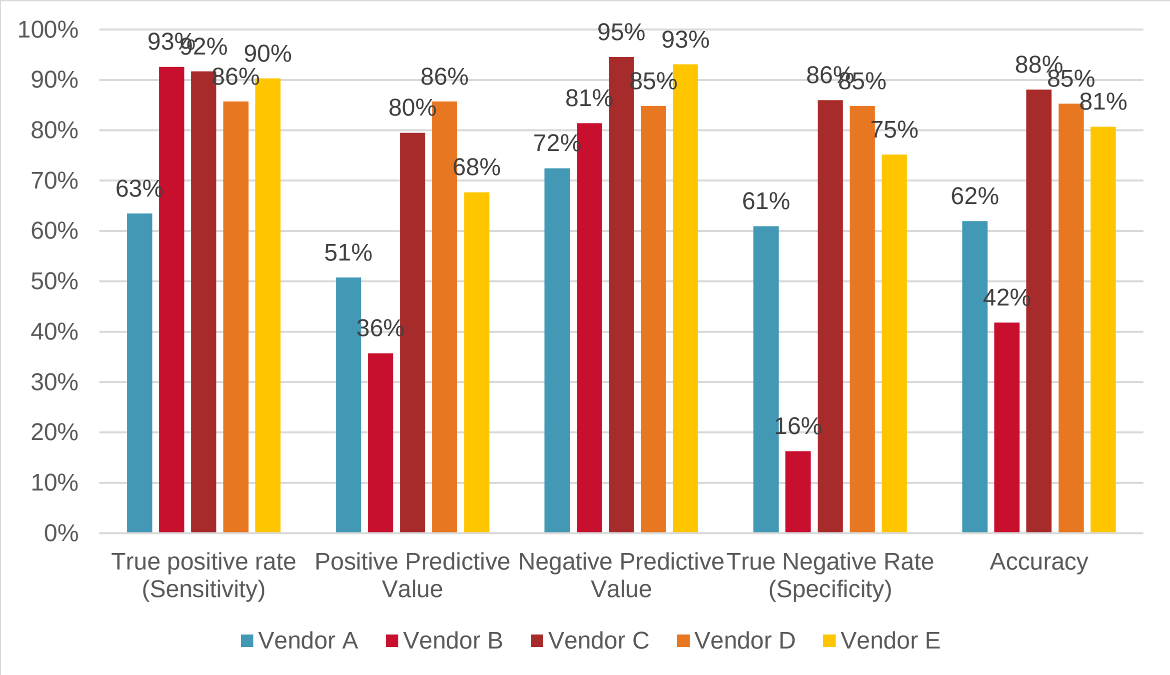


Figure 8 summarize the calculations graphically.

Table 7. Performance measure calculations for the most and least stringent matching criteria

Parameter	Most stringent matching criteria					Least stringent matching criteria				
	Vendor					Vendor				
	A	B	C	D	E	A	B	C	D	E
Number of true positive (TP)	33	14	58	108	51	66	100	66	132	65
Number of false positive (FP)	97	266	25	46	45	64	180	17	22	31
Number of false negative (FN)	71	94	14	46	21	38	8	6	22	7
Number of true negative (TN)	28	7	84	96	54	100	35	104	123	94
Calculations										
True positive rate (Sensitivity)	31.7%	13.0%	80.6%	70.1%	70.8%	63.5%	92.6%	91.7%	85.7%	90.3%
False positive rate	25.4%	5.0%	69.9%	70.1%	53.1%	50.8%	35.7%	79.5%	85.7%	67.7%
False negative rate	28.3%	6.9%	85.7%	67.6%	72.0%	72.5%	81.4%	94.5%	84.8%	93.1%
True negative rate (Specificity)	22.4%	2.6%	77.1%	67.6%	54.5%	61.0%	16.3%	86.0%	84.8%	75.2%

Parameter	Most stringent matching criteria					Least stringent matching criteria				
	Vendor					Vendor				
	A	B	C	D	E	A	B	C	D	E
Youden Index	0.00	0.00	0.58	0.38	0.25	0.24	0.09	0.78	0.71	0.65
Accuracy	26.6%	5.5%	78.5%	68.9%	61.4%	61.9%	41.8%	88.1%	85.3%	80.7%

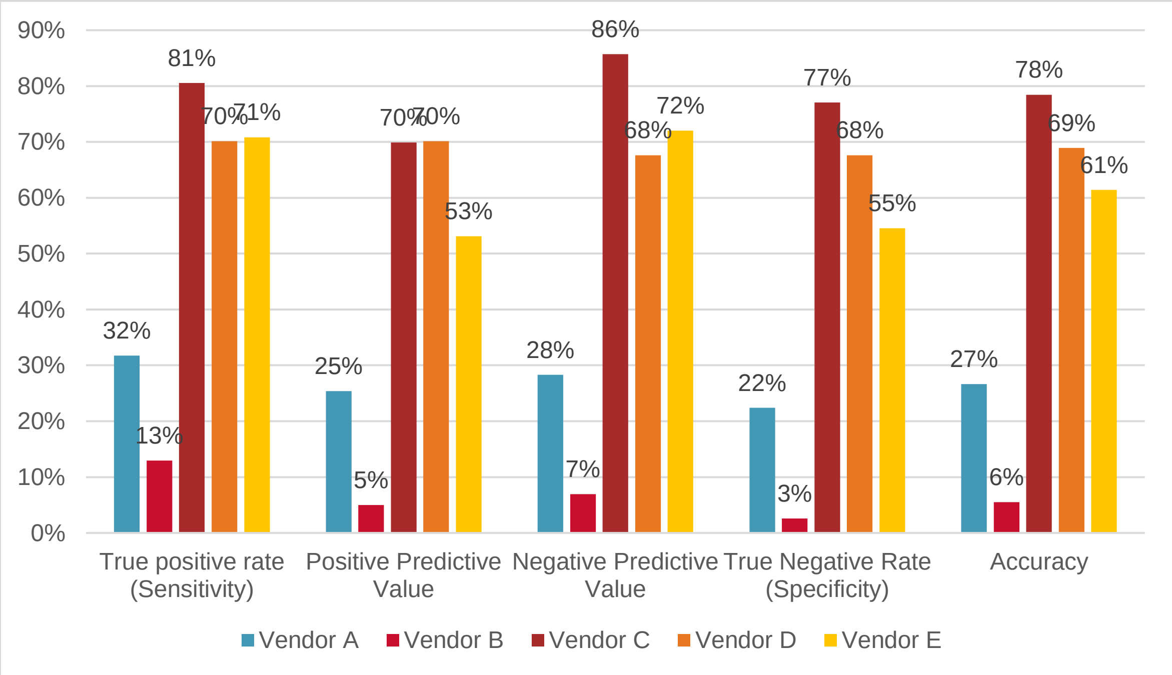


Figure 7. Performance measures for the most stringent matching criteria

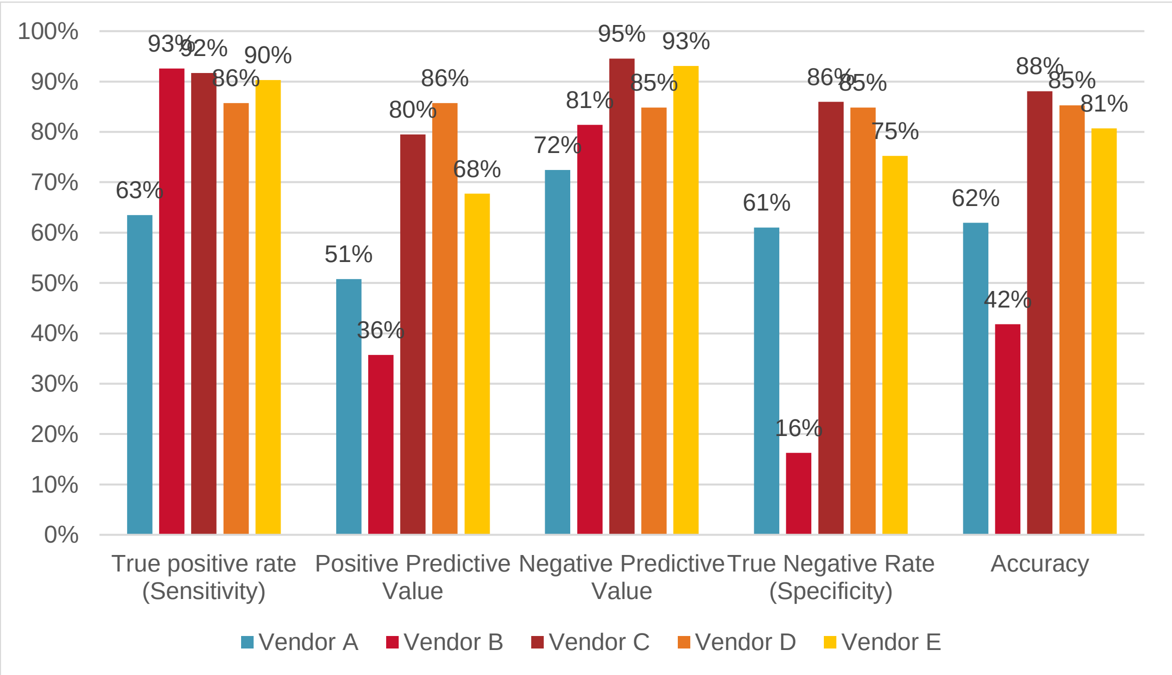


Figure 8. Performance measures for the least stringent matching criteria

The performance measures for Vendors C, D, and E are consistent across the five performance measures. In fact, the performance measures for these three vendors do not appreciably improve when relaxing the matching criteria. Higher Negative Predictive Values than Positive Predictive Values suggest that the systems for all three vendors perform better in identifying when a vehicle is not present compared to identifying when a vehicle is present (Positive Predictive Value). The closeness of the Sensitivity and Specificity for these same vendors, however, is not significantly different, suggesting that this difference, while noticeable in the results, may not be materially different.

Vendors A and B have noticeable and significant changes in performance as a result of the difference matching criteria going from least to most stringent. There is a very significant jump in both the Sensitivity and Positive Predictive Value, which also resulted in a dramatic improvement in the estimated accuracy between the two matching algorithms. Regardless, it is clear from the performance measures that the systems of these two vendors performed significantly below those of the other three across the board.

5.2 Performance in Estimating Vehicle Dwell Time (Duration)

A necessary prerequisite for a city to successfully implement dynamic fare payment or other policy related to active management of the curb space is to be able to accurately determine not only the presence of a vehicle but also the duration that a given vehicle remains curbside. This duration, commonly referred to as dwell time, was therefore a component of the evaluation.

Table 8 summarizes the average durations of matched events reported by the vendors and the corresponding ground truth. As observed in the table, there are cases where there are statistically significant differences between the average duration reported by the vendor and that of the ground truth. In particular, Vendors A, D, and E all (statistically) over-reported the length of time a vehicle was present at the curb on average based upon those matched cases under the most stringent matching criteria. However, while statistically different, the differences between the ground truth and the average durations are functionally very similar with the average difference being less than 1 minute under the most stringent matching criteria case.

The values between the least and most stringent criteria are significantly different as a direct result of the matching criteria. In the most stringent case, both the starting and ending times must be within 30 seconds for this to be a match by definition. As can be seen in the table, generally the durations were very close. In contrast, the least stringent matching criteria required only one of the two points (starting or ending) to be within 2 minutes of the ground truth. This means that the duration values are much less constrained—resulting in matches where the duration values are much different between the vendor-reported and the ground truth. The above phenomenon can easily be observed by comparing the extent of the points that are outside of the dotted line in Figure 9 versus Figure 10.

Table 9 presents an estimation of vehicle occupancy duration as a percentage of vehicle occupancy events.

Table 8. Summary of comparisons of the duration (in minutes) of vehicle occupancy events for the least and most stringent matching criteria

Vendor	Average duration (ground truth)	Average duration (vendor-reported)	Average difference	95% lower confidence limit on average difference	95% upper confidence limit on average difference	Statistically significant difference[a]
Least stringent matching criteria						
A	38.55	34.20	4.35	(16.07)	24.76	No
B	17.18	19.13	(1.95)	(10.67)	6.77	No
C	66.73	59.32	7.41	(6.06)	20.87	No
D	44.77	42.30	2.47	(0.67)	5.62	No
E	51.18	44.18	7.00	(8.46)	22.47	No
Most stringent matching criteria						
A	10.38	10.54	(0.16)	(0.22)	(0.10)	Yes
B	5.86	6.02	0.16	0.01	0.31	No
C	41.42	40.44	0.97	(1.07)	3.01	No
D	34.48	34.74	(0.26)	(0.29)	(0.23)	Yes
E	33.37	33.42	(0.05)	(0.08)	(0.02)	Yes

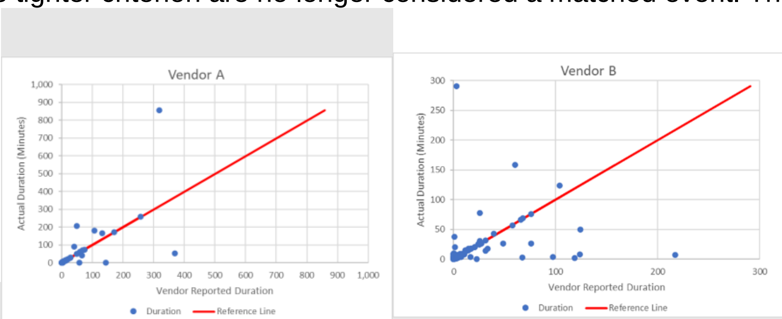
a. Statistical significance assessed using paired T-test with alpha of 5% (a paired T-test is a statistical method based upon the T-distribution that can be used to test whether the mean difference between pairs of measurements is zero).

Table 9. Estimation of vehicle occupancy duration as a percentage of vehicle occupancy events

Vendor	Actual number of vehicle occupancy events	Percentage of vehicle occupancy events with reported durations within 30 seconds of ground truth	95% lower confidence limit on percentage	95% upper confidence limit on percentage	Statistically significant from 95%[a]
Least stringent matching criteria					
A	66	75.8%	70.5%	81.0%	Yes (lower)
B	100	28.0%	19.2%	36.8%	Yes (lower)
C	66	90.9%	87.4%	94.4%	Yes (lower)
D	132	75.8%	72.0%	79.5%	Yes (lower)
E	65	81.5%	76.7%	86.4%	Yes (lower)
Most stringent matching criteria					
A	33	93.9%	89.8%	98.1%	No
B	14	92.9%	79.4%	100.0%	No
C	59	96.6%	94.3%	99.0%	No
D	108	91.7%	89.0%	94.3%	Yes (lower)
E	54	98.1%	96.3%	100.0%	Yes (higher)

a. Statistical significance assessed using T-test with alpha of 5%.

As with accuracy and other performance measures, the performance of the vendors' systems with respect to being able to quantify the duration of occupancy for a portion of the curb space varies as a function of the matching algorithm. As the matching criterion becomes more stringent, the vendor's equipment can be observed to match the actual occupancy duration more closely. By making the matching criterion more stringent, we can observe a closer match between the vendor and ground-truth occupancy duration. However, this comes at a cost of a lower accuracy number as events that do not fall within the tighter criterion are no longer considered a matched event. This can be clearly observed in



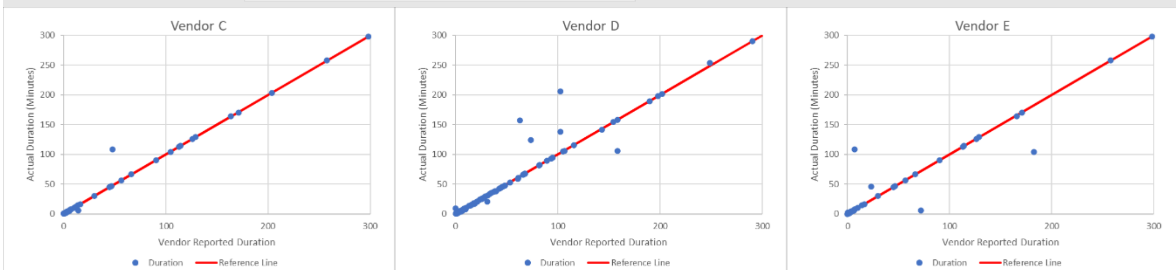


Figure 9 and Figure 10, which illustrate the vendor-reported versus actual ground-truth durations. Points in the graph that fall to one side of the reference line are indicative of over- or under-reporting by the vendor's equipment. Visually comparing the two sets of graphs reveals that there are many more points to either side of the reference line for all five vendors under the least stringent matching criteria compared to the most stringent matching criteria, but this is particularly noticeable with Vendors A, B, and D. This result can also be observed within the estimates presented in Table 9. The percentages shown in that table represent the percentage of dots that fall along the red reference line. For example, for Vendor A, 75 percent of the dots fall close to the red line while 25 percent of the dots do not. This result may indicate that even though relaxing the matching criteria improves the overall accuracy calculation as described in the previous section, it does so at the expense of adversely impacting the estimation of vehicle occupancy duration.

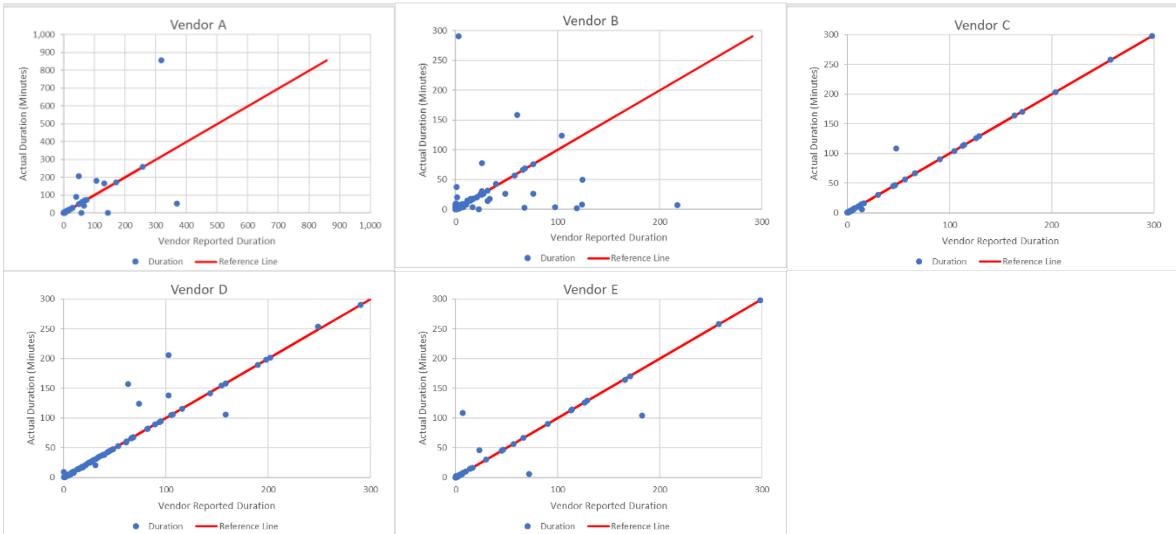


Figure 9. Comparison of reported vehicle occupancy duration versus actual ground-truth duration for the least stringent matching criteria

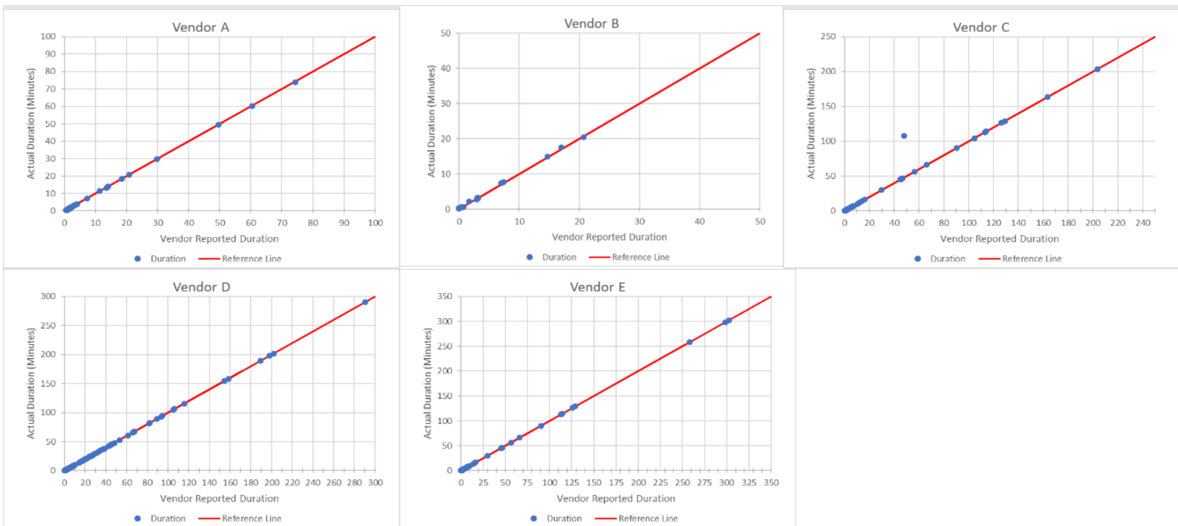


Figure 10. Comparison of reported vehicle occupancy duration versus actual ground-truth duration for the most stringent matching criteria

6 Conclusions and Recommendations

This project represents one of the very first experiments conducted to evaluate emerging technologies for identifying curbside activity. As with all such first-of-a-kind projects, this project was not without its challenges. In some cases, significant challenges and data cleaning processes were needed to correctly match the vehicle occupancy events reported by the vendors to those of the ground-truth system. It would be expected that over time, the reporting and output from the various systems would begin to converge and become easier to data cleaning processes. However, should this study be repeated elsewhere, it would be worthwhile to establish a clear data reporting structure prior to the start of the experiment that each vendor would be required to meet.

Extensive efforts were undertaken to ensure that the data provided by the various vendors could be matched to the ground-truth data. In the early rounds of the project, it was unclear whether the initial results were a function of the equipment inaccuracy or an inability to successfully match the reported occupancy events to the ground truth. This resulted in the development and application of several algorithms to use for matching these different sets of data. Following the conclusion of the study, however, the findings suggest that using a fairly stringent matching criterion that requires both the start and end times of the vehicle occupancy information to be within 30 seconds of the ground truth as well as the location being accurate to the stall level is an appropriate criterion. While the accuracy metrics for two vendors did improve with the relaxation of the matching criteria, their systems' ability to estimate the duration of occupancy correctly diminished. On the other hand, the matching criteria did not meaningfully impact the findings associated with the other three vendors, suggesting that the initial low accuracy was not a function of the matching criteria, but rather, reflective of equipment failure or subpar performance.

Based upon the findings presented in Section 5 and in the appendix, it is reasonable to conclude that none of the five tested systems met the desired levels of performance. More specifically, none of the five systems tested were able to achieve an overall accuracy of 95 percent under either the least or most stringent matching criterion. However, three vendors had systems that were able to achieve accuracy in the range of 60 to 80 percent under the most stringent matching criteria while simultaneously being able to, on average, estimate vehicle occupancy duration within 1 minute. The goal of the 95 percent accuracy for the most stringent case is specified initially with the potential of fare payment application in mind. Even though the results did not meet the expectation, they still offer a wealth of potential in serving as data collection tools and other application usages.

Lessons learned and corresponding recommendations for future studies include:

- Establishing a standardized data reporting structure and methodology for all vendors.
- Synchronizing clocks and time zone assumptions across systems prior to the start of any given trial and continuously verifying that this clock synchronization is maintained during the trial is critical to the success of matching vendor-reported vehicle occupancy events to ground-truth vehicle occupancy events.
- Some of the vendors' systems were observed to require more in-field calibration than others. Adequate time is required for each system to be correctly calibrated.

- Weather and other situational factors were not systematically included as factors with the experimental design of the study. However, anecdotal information provided by the vendors and field observation from the project team indicates that weather and other location-based conditions such as striping, sun reflection, etc. may impact these systems and should be included in the experimental design of future studies.
- This experiment did not seek to address the ability of the various systems to perform over an extended period (i.e., system durability). This should be a consideration for future research.

This page is intentionally left blank

Appendix A. Data Cleaning Flow Chart

Figure A-1 summarizes the process used to develop and clean the data set used for the analyses.

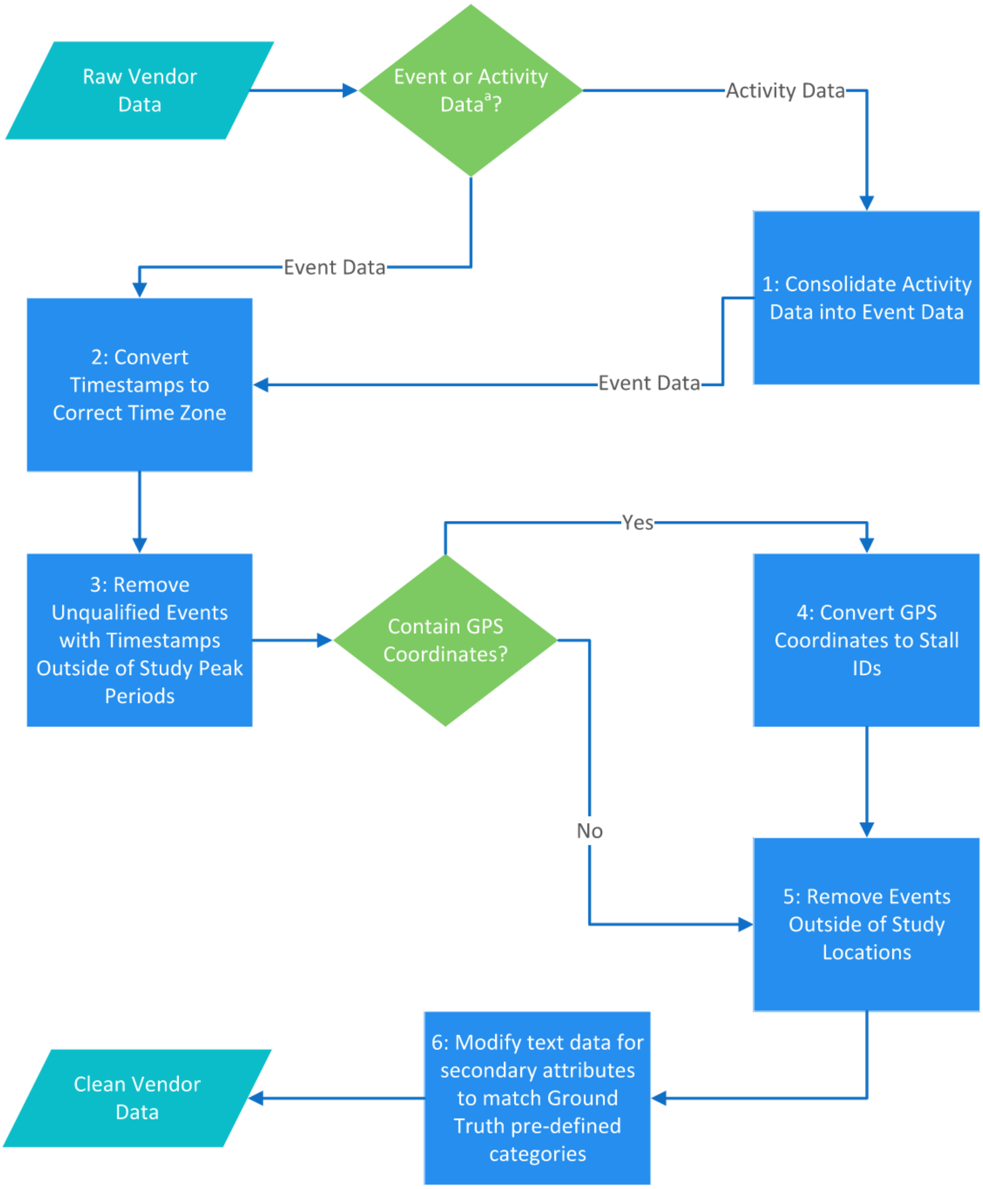


Figure A-1. Data cleaning flow chart

“Activity data” refers to data records that correspond to a vehicle being reported as entering or exiting a curbside location. An “event” in the diagram consists of the two activities represented by the (1) entry into the curbside space and (2) exit of the curbside space

This page is intentionally left blank

Appendix B. Detailed Vendor Performance Results

Table B-1. Vendor A detailed performance results

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	Vehicle's location matches to the zone	Vehicle's location matches to a specific stall
Time thresholds	±2 minutes	±30 seconds
Raw counts		
Number of ground truth vehicle occupancy events	104	104
Number of vendor-reported vehicle occupancy events	130	130
Number of matched vehicle occupancy events	66	33
Number of ground truth vacancy windows	158	158
Number of vendor-reported vacancy windows	133	133
Number of matched vacancy windows	100	28
Confusion matrix data elements		
Number of true positive (TP)	66	33
Number of false positive (FP)	64	97
Number of false negative (FN)	38	71
Number of true negative (TN)	100	28
Calculated performance measures		
True positive rate (Sensitivity)	63.5%	31.7%
Positive Predictive Value	50.8%	25.4%
Negative Predictive Value	72.5%	28.3%
True Negative Rate (Specificity)	61.0%	22.4%
Youden Index	0.24	0.00

Table B-1. Vendor A detailed performance results

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	Vehicle's location matches to the zone	Vehicle's location matches to a specific stall
Time thresholds	±2 minutes	±30 seconds
Accuracy	61.9%	26.6%
Duration measures		
Ground-truth average duration (minutes)	38.55	10.38
Vendor-reported average duration (minutes)	34.20	10.54
Average difference in duration (minutes)	4.35	-0.16
Percentage of vehicle occupancy within 30 seconds of ground truth	75.8%	93.9%

Table B-2. Vendor B detailed performance results

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	Vehicle's location matches to the zone	Vehicle's location matches to a specific stall
Time thresholds	±2 minutes	±30 seconds
Raw counts		
Number of ground truth vehicle occupancy events	108	108
Number of vendor-reported vehicle occupancy events	280	280
Number of matched vehicle occupancy events	100	14
Number of ground truth vacancy windows	124	124
Number of vendor-reported vacancy windows	73	73
Number of matched vacancy windows	35	7
Confusion matrix data elements		
Number of true positive (TP)	100	14
Number of false positive (FP)	180	266
Number of false negative (FN)	8	94
Number of true negative (TN)	35	7
Calculated performance measures		
True positive rate (Sensitivity)	92.6%	13.0%
Positive Predictive Value	35.7%	5.0%
Negative Predictive Value	81.4%	6.9%
True Negative Rate (Specificity)	16.3%	2.6%
Youden Index	0.09	0.00
Accuracy	41.8%	5.5%

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	Vehicle's location matches to the zone	Vehicle's location matches to a specific stall
Time thresholds	±2 minutes	±30 seconds
Duration measures		
Ground-truth average duration (minutes)	17.18	5.86
Vendor-reported average duration (minutes)	19.13	6.02
Average difference in duration (minutes)	(1.95)	0.16
Percentage of vehicle occupancy within 30 seconds of ground truth	28.0%	92.9%

Table B-3. Vendor C detailed performance results

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	Vehicle's location matches to the zone	Vehicle's location matches to a specific stall
Time thresholds	±2 minutes	±30 seconds
Raw counts		
Number of ground truth vehicle occupancy events	72	72
Number of vendor-reported vehicle occupancy events	83	83
Number of matched vehicle occupancy events	66	58
Number of ground truth vacancy windows	112	112
Number of vendor-reported vacancy windows	124	124
Number of matched vacancy windows	104	84
Confusion matrix data elements		
Number of true positive (TP)	66	58
Number of false positive (FP)	17	25
Number of false negative (FN)	6	14
Number of true negative (TN)	104	84
Calculated performance measures		
True positive rate (Sensitivity)	91.7%	80.6%
Positive Predictive Value	79.5%	69.9%
Negative Predictive Value	94.5%	85.7%
True Negative Rate (Specificity)	86.0%	77.1%
Youden Index	0.78	0.58
Accuracy	88.1%	78.5%

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	**Vehicle's location matches to the zone**	**Vehicle's location matches to a specific stall**
Time thresholds	**±2 minutes**	**±30 seconds**
Duration measures		
Ground-truth average duration (minutes)	66.73	41.42
Vendor-reported average duration (minutes)	59.32	40.44
Average difference in duration (minutes)	7.41	0.97
Percentage of vehicle occupancy within 30 seconds of ground truth	90.9%	96.6%

Table B-4. Vendor D detailed performance results

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	Vehicle's location matches to the zone	Vehicle's location matches to a specific stall
Time thresholds	±2 minutes	±30 seconds
Raw counts		
Number of ground truth vehicle occupancy events	154	154
Number of vendor-reported vehicle occupancy events	154	154
Number of matched vehicle occupancy events	132	108
Number of ground truth vacancy windows	131	131
Number of vendor-reported vacancy windows	160	160
Number of matched vacancy windows	123	96
Confusion matrix data elements		
Number of true positive (TP)	132	108
Number of false positive (FP)	22	46
Number of false negative (FN)	22	46
Number of true negative (TN)	123	96
Calculated performance measures		
True positive rate (Sensitivity)	85.7%	70.1%
Positive Predictive Value	85.7%	70.1%
Negative Predictive Value	84.8%	67.6%
True Negative Rate (Specificity)	84.8%	67.6%
Youden Index	0.71	0.38
Accuracy	85.3%	68.9%
Duration measures		

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	Vehicle's location matches to the zone	Vehicle's location matches to a specific stall
Time thresholds	±2 minutes	±30 seconds
Ground-truth average duration (minutes)	44.77	34.48
Vendor-reported average duration (minutes)	42.30	34.74
Average difference in duration (minutes)	2.47	-0.26
Percentage of vehicle occupancy within 30 seconds of ground truth	75.8%	91.7%

Table B-5. Vendor E detailed performance results

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	Vehicle's location matches to the zone	Vehicle's location matches to a specific stall
Time thresholds	±2 minutes	±30 seconds
Raw counts		
Number of ground truth vehicle occupancy events	72	72
Number of vendor-reported vehicle occupancy events	96	96
Number of matched vehicle occupancy events	65	51
Number of ground truth vacancy windows	112	112
Number of vendor-reported vacancy windows	137	137
Number of matched vacancy windows	94	54
Confusion matrix data elements		
Number of true positive (TP)	65	51
Number of false positive (FP)	31	45
Number of false negative (FN)	7	21
Number of true negative (TN)	94	54
Calculated performance measures		
True positive rate (Sensitivity)	90.3%	70.8%
Positive Predictive Value	67.7%	53.1%
Negative Predictive Value	93.1%	72.0%
True Negative Rate (Specificity)	75.2%	54.5%
Youden Index	0.65	0.25

Time criteria	Either the vehicle occupancy start or end time match within the time thresholds	Both the vehicle occupancy start and end time match within the time threshold
Location criteria	Vehicle's location matches to the zone	Vehicle's location matches to a specific stall
Time thresholds	±2 minutes	±30 seconds
Accuracy	80.7%	61.4%
Duration measures		
Ground-truth average duration (minutes)	51.18	33.37
Vendor-reported average duration (minutes)	44.18	33.42
Average difference in duration (minutes)	7.00	-0.05
Percentage of vehicle occupancy within 30 seconds of ground truth	81.5%	98.1%