BTS TRANSPORTATION PROBE DATA GUIDE: LOCATION BASED–SERVICES DATA

Location based–services (LBS) data comprise location sightings that are derived from the usage of location-enabled mobile applications (apps) on smartphones and/or cellular-enabled tablets. In other words, LBS data encompass the geographic information generated when mobile apps request position updates. These data are collected through software development kits (SDKs) embedded in mobile apps, creating an array of location pings that can be triggered by user interactions, background processes, or automated system functions. Unlike data sources that are purpose-built for transportation, LBS data are a byproduct of broader commercial mobile app functionality.

A fundamental limitation of LBS data is their opportunistic nature—location information is captured whenever apps require positioning for their core functionality, whether for navigation, advertising, social media check-ins, or weather services.[1] This behavior creates highly irregular spatiotemporal patterns that vary dramatically based on individual user behavior, device settings, and app-usage patterns.

The LBS data marketplace has experienced significant volatility due to evolving privacy regulations and platform (i.e., software ecosystem) restrictions.[2] Major changes to iOS and Android privacy controls—particularly Apple's iOS 14.5 update requiring explicit tracking consent [Apple 2021] and similar, albeit undocumented, Android developments—have substantially reduced data availability. Additionally, high-profile privacy concerns and regulatory scrutiny have led some major apps to discontinue reselling location data entirely based on each app, original equipment manufacturer (OEM), and carrier.

1. CAPABILITIES

LBS data offer the following distinct advantages for generating transportation statistics:

- Trip detection is possible through algorithms analyzing location sightings.
- Home locations can be inferred by clustering nighttime sightings.[3]
- Work locations can be inferred through daytime clustering on weekdays.

---

1 No available documentation indicates whether the location capture includes pings when the app is not engaged, either for capture by the device, original equipment manufacturer device, or carrier.

2 Isolating any particular regulation, either implemented or proposed, in combination with other events (media), that may have contributed to any specific business decision made by LBS vendors is difficult. The contract team observed that marketplace changes generally correlated with several similarly timed regulatory and platform update events.

3 This methodology is used by several vendors: "For home locations, we look at an entire calendar month, and identify the top five neighborhoods where a device pings during evening and night-time hours. Each of the five neighborhoods is assigned a probability weighting" [StreetLight 2025].

2. LIMITATIONS

Known limitations[4] of LBS data for transportation statistics include the following:

- Home location detection may be impacted by overnight work shifts.
- Work location detection may be impacted by overnight work shifts and school attendance.
- The ability to distinguish between transportation modes is limited.
- Trip purpose must be inferred.
- Sociodemographic inferences, such as the age and gender of device users in aggregate, have moderate to low confidence.
- Determining when an individual movement is taking place as part of group travel, such as trips taking place on buses with individual devices moving together, is difficult.
- Privacy protections and instability in the market make LBS data procurement volatile as LBS data generation and harvesting are dependent on an ecosystem of apps and must share that ecosystem with device OEMs and carriers.

3. VENDORS AND AGGREGATORS

LBS data are available from private vendors in their disaggregate and aggregate forms. While some companies sell both forms, most vendors are focused on one or the other.

Companies, such as Apple, Google, and Meta, collect location data via their various devices and offerings. However, no known instances of these data in a disaggregate form or outside platform offerings have been available within transportation.

For further information on vendors and aggregators, refer to *Location intelligence: the new frontier for data-driven success* [Jonker 2021].

4. MARKETS

Primary markets for LBS data include the following:

- Advertising via location-aware texts
- Retail and commercial intelligence for use in developing new locations
- Finance to understand activity at specific retail establishments

Transportation is a secondary market.

5. SCALE OF LBS DATA FOR TRANSPORTATION USES

LBS data are available for the contiguous United States at the national (the National Household Travel Survey),[5] state, and local levels (individually, from vendors). They can support nationwide analyses but require extensive data processing, data management, development,

---

[4] Knowledge on the limitations of LBS data is limited due to the following:
1. Research substantially lags the pace of technology companies to innovate.
2. Technology companies in this space, while referencing existing research, rely on innovation from other domains and not just transportation or mobility.

[5] Availability of data in rural areas, Alaska, and Hawaii, may present a challenge due to limited sample sizes.

and data science expertise, which are specialized skills not accessible to many transportation agencies.[6]

6. GENERAL DETAILS

The following sections detail background information on how LBS data are generated, why they exist as a data type in the first place, and the pieces of information represented in their component data points.

6.1. How Are Data Captured

The following sections detail elements of the technology underlying the generation, capture, and protection of LBS data points.

6.1.1. Trigger Mechanisms and Frequency

The following mechanisms generate LBS data points:

- LBS sightings are triggered when a user's mobile phone app requests location updates, which can happen for a variety of reasons:
  - Requires location information (maps, rideshare)
  - Runs spatially aware background processes even if the app does not have a primary spatial component (location-aware advertisements, fitness tracking)
  - Measures device motion (accelerometers)
  - Produces scheduled pings[7]
- Sightings are produced irregularly, depending on user behavior, device settings, app configurations, and operating system privacy settings.[8]
- Frequency ranges from seconds (navigation apps) to hours (weather apps).[9]
- Sightings are generally irregular, with navigation apps providing more regular data than other app types.
- Operating system background restrictions and battery-saving modes suppress background updates, making patterns sporadic for apps that send location sightings while not being actively used.

6.1.2. Location-Determination Technology

The location element of an LBS data point is generated per the following procedures:

---

[6] Disaggregate data, for a national scale, require cloud computing or a government or university secure data center. A number of computing resources within those systems can be leveraged for processing such large datasets. These resources can be bespoke user-developed systems within a cloud platform base (Amazon Web Services [AWS] Athena, Google Big Query), or the resources can be a pay-for-compute service within a cloud environment (Databricks). However, if an organization has not previously standardized a data pipeline, producing one involves considerable iteration to find the best mix of technologies for their needs and budget.

[7] Detailed primary research into how, when, and the frequency of location pings across different apps is limited and mostly tertiary.

[8] For example, location tracking is under privacy settings in iOS, but LBS can also be based on Wi-Fi. If Wi-Fi is off and location sharing is not allowed, then the user has a less degree of being located.

[9] Vendors and aggregators of these data typically will not disclose from which apps they are getting data due to nondisclosure agreements and relationships. In some cases, if asked, they may be willing to share the categories of the apps.

- Coordinates are primarily determined through Global Positioning System (GPS) signals received by mobile devices via a combination of technologies, including internal software, external hardware, and internal hardware.
- When GPS is unavailable or weak (e.g., indoors, urban canyons), the following fallback methods are used:
  - Cellular tower triangulation: The device's location is determined by measuring signal timing or strength from multiple cell towers and calculating where those signals intersect (typically hardware based).
  - Wi-Fi triangulation (signal strength matching): The device measures signal strength from multiple Wi-Fi access points and uses those measurements to calculate its position relative to the known locations of those access points (typically hardware based).
  - Wi-Fi receiver known locations: The device scans for nearby Wi-Fi networks and matches their unique identifiers (IDs) against databases that map those networks to specific geographic locations (typically hardware based).
  - Internet Protocol (IP) address-based location: The device's internet connection is traced back through IP address databases that associate ranges of IP addresses with geographic regions (can vary as to whether hardware or software based).

The precision of the resultant location sighting depends on the methods the phone must use to determine its position:

- GPS-based sightings under open sky typically achieve accuracy within 33 feet.
- Wi-Fi triangulation typically achieves accuracy within 20–164 feet.
- Cell tower–based estimates are much less precise (within 100–1,000 feet).

6.1.3. Data Transmission

LBS data points are transferred from a device per the following procedures:

- Data are transmitted from devices to servers[10] via Hypertext Transfer Protocol Secure- or Transport Security Layer (TLS)-encrypted connections.
- Transmission occurs when devices have internet connectivity (via cellular networks or Wi-Fi).[11]
- Some vendors implement data encryption beyond standard TLS, which is the successor to Secure Socket Layer, encryption utilized by internet sites.
- For stationary devices,[12] data are typically batched and sent periodically.
- For moving devices, transmission can be more frequent, especially for navigation apps.
- SDK-based collection typically uses app-defined transmission intervals as established by the app developers, but the specific SDK settings and options are not documented.

---

[10] Prior to the ubiquitous availability of cloud computing and storage, app owners would receive data on their servers and transmit the data to an aggregator or a vendor for storage on their servers. Currently, the contract team assumes that everything is moved through a cloud service.

[11] Transmission upon connection can include current location as well as location(s) while not connected based on device technology and settings (SDK technology and settings, app, etc.)

[12] For example, technologies like wildlife cameras send data over cellular networks and devices because they cannot connect to wired networks.

6.1.4. Identifiers and Privacy

The following are ways the owner of an LBS data–generating device is masked:

- Devices are tracked using the following advertising IDs:[13]
  - IDFA (Identifier for Advertisers) on iOS devices
  - AAID (Android Advertising ID) on Android devices
- These pseudonyms are alphanumeric strings of varying lengths that users can reset.
- User resets of advertising IDs, which limit how long an ID is observable, complicate longitudinal tracking.
- iOS requires explicit app-level permission for tracking since iOS 14.5 [Apple 2021].
- Android 12+ offers similar opt-out capabilities.

6.1.5. Observation Unit

The following circumstances characterize the unit of analysis for LBS data:

- The fundamental observation unit is an anonymized mobile device sighting, not a person.
- One person may carry multiple devices, creating oversampling. To control for this issue, data vendors and aggregators will identify the unique patterns in a process termed “deduplication,” or “de-duping.” This process is generally not exact and requires extensive computation if it is based on locations and not IDs.
- Shared devices create ambiguity in person-based attribution.
- Devices are not linked to specific vehicles, though movement patterns can suggest vehicular travel.

6.2. Why Does This Data Source Exist Originally and for Transportation

The following sections provide background on the origin of LBS as a data type.

6.2.1. Original Purpose

LBS data exist as a data type for the following reasons:

- LBS data primarily exist to support targeted advertising and app functionality.
- Location data enable more relevant, proximity-based advertising delivery. Retail apps use location for nearby store information (for example, the Best Buy and Walmart apps).
- The data support core functionality for navigation, local search, and LBS apps, such as the following:
  - Social apps use location for geo-tagging and “near me” features (for example, wayfinding, traffic, and some dating apps).
  - Weather apps use location to provide local forecasts

---

13 These IDs change over time, and research experience suggests that these IDs have a long depleting tail of how long an ID will persist. After, 1, 10, and 30 days, fewer and fewer IDs are similar. For example, with iOS, older models explicitly let users reset the Advertiser ID, and this ID can be reset on newer models by toggling to Apple Advertising and setting personalized ads to on or off.

6.2.2. Transportation Focus

LBS data are not collected with the transportation industry primarily in mind.

The following circumstances describe the relationship between transportation and LBS as a data type:

- Most LBS aggregators and vendors serve multiple market sectors beyond transportation.
- Documented transportation research using LBS data only started in 2016.
- Specialized transportation vendors apply additional processing to identify trips and movement patterns.

6.3. What Do the Data Capture

The following sections note the elements and attributes that come with LBS data points, differentiating those generated at their inception from those that must be estimated.

6.3.1. Core Attributes

Core attributes, or the data elements intrinsically available for all LBS data points, include the following:

- Timestamp (typically in Coordinated Universal Time [UTC] or local time)
- Latitude–longitude coordinates (World Geodetic System 1984 [WGS84] format)
- Hashed and anonymized device ID (typically of the advertising ID[14])
- Location accuracy estimate (in meters)

6.3.2. Non-Universal Attributes

Outside the core attributes, additional data elements may be available for an LBS data point when it is initially generated.

Non-universal attributes, such as the following, depend on many factors, including SDK configuration:[15]

- Device characteristics (OEM, operating system, model)
- Connection information (Wi-Fi, cellular)
- App ID (which app generated the ping)
- Battery level
- Location-detection method (GPS, Wi-Fi, cellular) and measure of positional accuracy
- Motion activity type (e.g., stationary, walking, running, automotive)

---

[14] Other forms of IDs exist, and these IDs may be unique to the app or even device. The advertiser ID is the most portable between technologies and platforms.

[15] Each SDK will have its own set of configurations and preferences.

6.3.3. Derived Attributes

Data elements that are not inherently generated with an individual LBS data point, and thus must be contextually derived or estimated, include the following:

- No direct speed, heading, or acceleration readings are provided. Given differences in technology, how often sightings are tracked, app settings, and other factors (including the need to compare attributes across datasets), these attributes must be inferred by analyzing sequences of positions over time.
- Elevation or altitude is sometimes available but often low quality as no measures indicate how well the device and technology are calibrated to measure these attributes.
- No native point-of-interest or land-use data are gathered (must be overlaid, or spatially joined, separately).
- No demographic data are directly collected (sometimes these attributes are inferred from location patterns, such as the demographics for the Census block group or tract of the inferred home location of the device).

6.3.4. Sample Characteristics

Details on the sample penetration of LBS data vis-à-vis the American public include the following:

- The sample typically comprises devices with greater than 10 daily sightings for 5–15 percent of the population [BTS 2022].
- Many devices are not used frequently or continuously, so they have few sightings.
- Land use (i.e., rural versus urban areas can have divergent demographic profiles) and regional biases can be observed. Demographic biases are difficult to ascertain for certain in the data (due to device anonymization and privacy restrictions) even though demographic biases related to age and gender are well measured and show that, as people age, their travel patterns change, and these patterns diverge between genders as well.
- Devices with location services enabled, coupled with region-specific laws, have opt-in bias.
- Sample correction methods often use census demographics, traffic counts, and/or other location features for weighting using machine learning (ML) techniques.

6.3.5. Data Quality

Documented concerns about the quality of LBS data include the following:

- Common issues, documented by data analysts and providers, include:
  - GPS drift in urban canyons
  - Imprecise indoor positioning
  - Spoofed locations from virtual private networks
  - Missing data during signal loss
  - Battery optimization interruptions
- Quality validation methods include:
  - Impossible speed filtering: measurement of timestamps and geodesic distance between sightings, which produce impossible travel speeds

- GPS jump detection, or the GPS canyon effect: measurement of the location of subsequent sightings in time and their distance apart and filtering out anomalies based on a user-determined, use-case-specific distance criteria
- Pattern consistency analysis: assessment of patterns, such as nighttime location, to understand if a device follows expected patterns or if the patterns look random (i.e., unusual activity in a residential area at night)

For more information on data quality, refer to *Understanding GPS and Mobile Phone Data for Origin–Destination Analysis* [Chen et al. 2017].

7. TEMPORAL AND SPATIAL SCALES OF THE DATA

The following sections characterize the historical, temporal, and spatial coverage, availability, and precision of LBS data.

7.1. Temporal Coverage

Details on the historical penetration and availability of LBS data in the United States include the following:

- Availability varies significantly by vendor and region (i.e., urban areas with more people).
- Historical archiving policies vary by aggregator and/or vendor.[16]
- Most vendors offer data from 2017 or 2018 onward.
- Quality and consistency improved in recent years due to a combination of many factors, including increases in smartphone penetration rates, proliferation of apps, improved technology, longer lasting battery life, and so on.
- Sample reductions occurred after 2021 due to app developers' privacy, regulatory, and risk-perception concerns, with some major apps no longer providing data. Many developers chose to depart the market ahead of these changes, although matching them to specific apps or user groups is difficult.

7.2. Temporal Resolution

Details on the precision of temporal information generated with an LBS data point include the following:

- Raw data include timestamps, typically with second precision.
- Time zones may be UTC or local time depending on the aggregator.
- Data delivery latency (time between creation of the sighting and receipt of the data by the data processor or vendor) is generally 24–72 hours.
- Real-time analysis is generally not available.
- Processing time for value-added products, like processed LBS products, is typically 2–6 weeks (where value-added can refer to any data enhancements that require analytical work, such as ML, map matching, data expansion, data wrangling, trip identification, pushing to production availability in a platform, etc.).

[16] Buyers can request information from vendors about the specific geographic areas and periods for which they have data available, and vendors will typically provide this coverage information when asked.

7.2.1. Spatial Coverage

Details on the spatial coverage and availability of LBS data in the United States include the following:

- Coverage correlates with population density and cellular network coverage.
- Geographic distribution is strongest in urban and suburban areas.
- Significant gaps exist in rural regions.
- Cross-border movement can create data discontinuities.[17]

7.2.2. Spatial Resolution

Details on the precision of spatial information generated with an LBS data point include the following:

- Raw data are collected as WGS84 latitude–longitude coordinates.
- For privacy protection, precise coordinates may be aggregated to Census or custom geographic units as well as roads.
- Spatial accuracy varies by collection method (GPS: less than 50 feet, Wi-Fi: 20–200 feet, Cell: 100–1,000 feet).
- Urban areas typically have higher point densities than rural areas.

8. USE CASES OF THE DATA

The following sections note documented use cases and applications of LBS data, highlighting those uses specific to transportation and transportation statistics.

8.1. Transportation Use Cases

The following applications are examples of how LBS data may be used in transportation:

- Origin–destination studies to understand travel patterns between regions
- Traffic operations and route analysis
- Regional activity patterns and mobility analysis
- Trip detection using algorithms that identify trip-end stops via spatial–temporal clustering
- Inference of common home or work locations
- Travel-time estimation
- Route choice and link estimation and analysis
- Evacuation monitoring during emergencies
- Tourism and visitor studies tracking nonresident movement

8.2. Nontransportation Use Cases

The following are examples of how LBS data may be used outside transportation:

[17] Cross-border data discontinuities occur not due to technical limitations, but because vendors typically acquire data within single countries (e.g., U.S.-only data purchases from aggregators). Combining data from multiple countries introduces processing complexities and edge cases in geofencing and analysis.

- Retail site selection and competitive analysis
- Billboard and advertising placement optimization
- Public space utilization studies
- Event attendance estimation

REFERENCES

Apple. 2021. "iOS 14.5 delivers Unlock iPhone with Apple Watch, more diverse Siri voice options, and new privacy controls," *Newsroom*. Cupertino, CA: Apple Inc. https://www.apple.com/newsroom/2021/04/ios-14-5-offers-unlock-iphone-with-apple-watch-diverse-siri-voices-and-more/. Last accessed November 17, 2025.

BTS. 2022. *BTS Daily Travel Metrics, Trips by Distance: Methodology and Validation*. Washington, DC: Bureau of Transportation Statistics. https://rosap.ntl.bts.gov/view/dot/67520. Last accessed September 18, 2025.

Chen, Cynthia, Xuegang Ban, Feilong Wang, et al. 2017. *Understanding GPS and Mobile Phone Data for Origin–Destination Analysis*. Washington, DC: Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/55805. Last accessed September 18, 2025.

Jonker, Shalie. 2021. *Location intelligence: the new frontier for data-driven success*. Mountain View, CA: Google Maps Platform. https://mapsplatform.google.com/resources/blog/location-intelligence-new-frontier-data-driven-success/. Last accessed September 18, 2025.

StreetLight. 2025. *Home and Work Locations Methodology*. San Francisco, CA: StreetLight.

About This Document

Title:	*BTS Transportation Probe Data Guide: Location-Based Services (LBS) Data*
Publication Date:	March 2026
Contributor(s):	BTS: Kyle Titlow, Mindy Liu, Ed Strocko, Rolf Schmitt, and Devon Lechtenberg. MacroSys: Stacey Bricka WRA: Byron Chigoy and Krishnan Viswanathan
DOI:	10.21949/hst8-0343