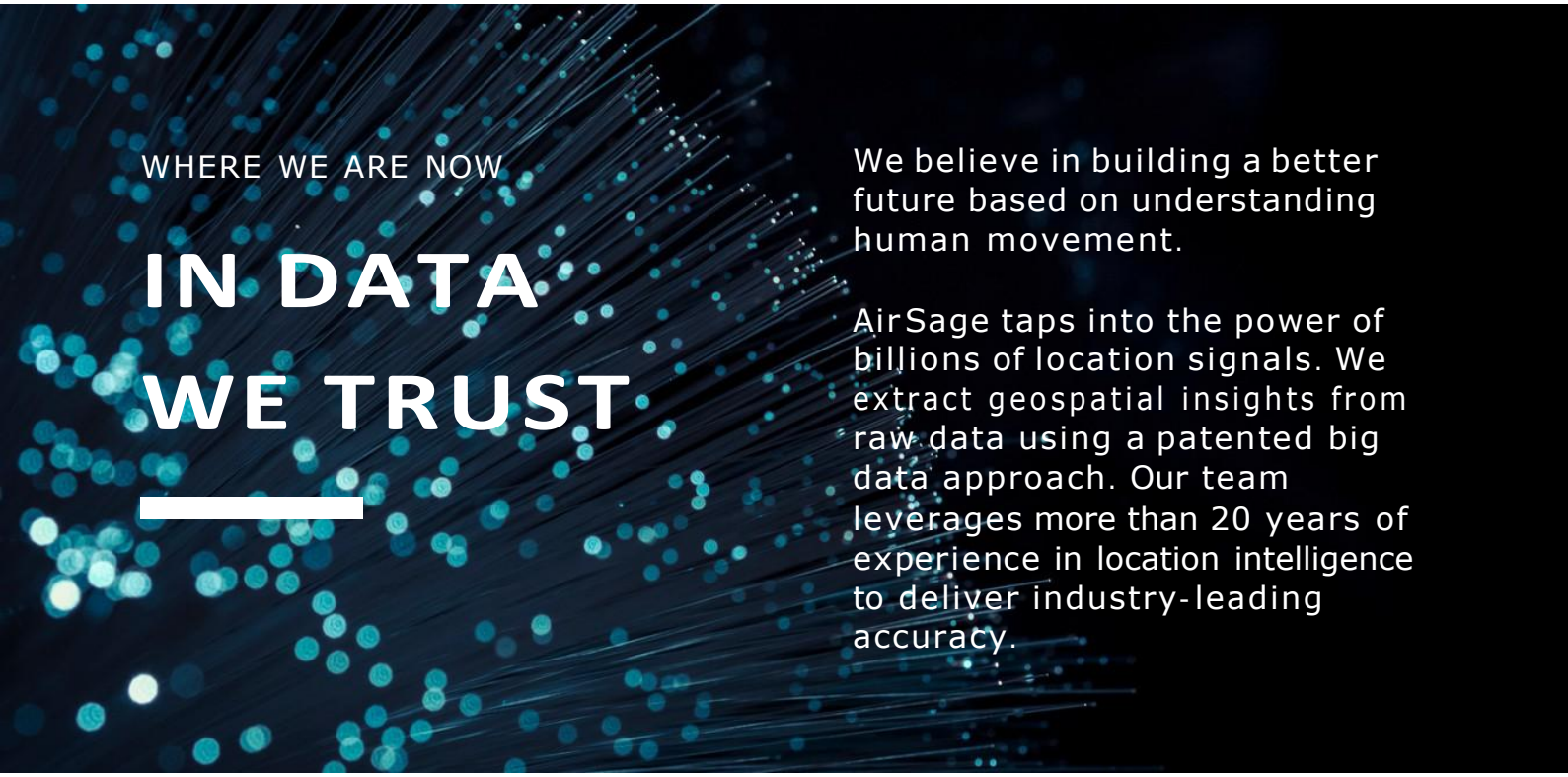
**CONTENT BY**
JONATHAN SILVERBERG, CTO AIRSAGE INC.


# MOBILE DEVICE DATA QUALITY OVERVIEW



WHERE WE ARE NOW

# IN DATA WE TRUST

We believe in building a better future based on understanding human movement.

AirSage taps into the power of billions of location signals. We extract geospatial insights from raw data using a patented big data approach. Our team leverages more than 20 years of experience in location intelligence to deliver industry-leading accuracy.

## Data Collection & Processing

AirSage's process to develop mobile device data, also known as Location Based Services (LBS) data, output requires various elements that occur in a sequence.

First, data must be sourced in a manner that meets applicable privacy requirements and regulations. Next, the dataset must be cleansed to make use of the most appropriate pieces of data - typically the device sightings with the highest level of location accuracy, obtained from highly visible and highly active devices.

The GPS signal locations from LBS data typically place a device within 5-15 meters (or better) accuracy, compared with the actual location. Following this, "Home" and "Work" zones are assigned to each of the devices, and the points that they generate (i.e., device pings) are classified into point types. After these steps, the travel patterns obtained from the sample devices are expanded to reflect estimates of the number of trips that would be made by the entire population.

The following pages describe specific details associated with steps that AirSage conducts during the collection and processing of the raw data.

# ACCURACY

AirSage's data accuracy is achieved based on several components: Data sources, Data processing, and Extrapolation.

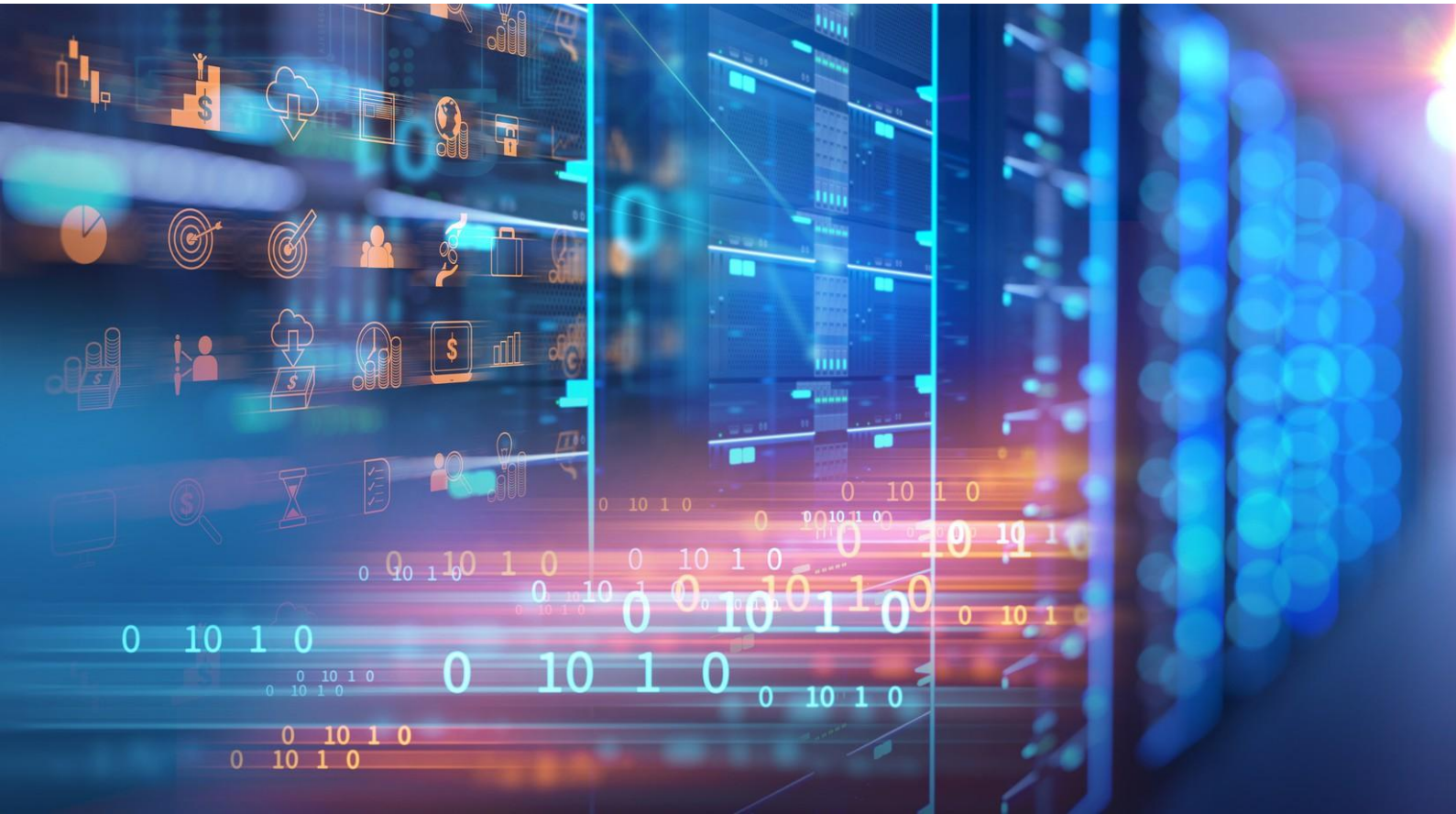
## Sourced Data

With more than a decade of experience with sourcing various types of anonymous location data (carrier data, connected vehicle data, fleet data, smartphone data, and more), AirSage has developed an expertise in sourcing the best available data and building an optimal data panel.

Nearly all mobile device data available in the open market for large scale sourcing has been evaluated and considered by AirSage to enter its data panel.

Each such candidate passes a thorough and efficient evaluation process that ultimately reveals its data volume, coverage, uniqueness, and multiple other quality metrics, all relevant for the AirSage analytics use cases.

Data that has been chosen to enter the panel goes through similar ongoing evaluation to ensure that the highest quality standards are also kept through time. Data feeds that fail to maintain such standards are removed from the panel.



*Diversity and Representativeness:* AirSage establishes its mobile device GPS data panel by sourcing data from data aggregators.

AirSage maintains relationships with various data aggregators that source data from individual app publishers, 1st party SDK providers, and Ad-Tech companies with processes installed on mobile devices.

These data points are obtained through "opt-in" services, wherein users consciously select whether to allow their device to share its location data with various apps installed on the device.

At any given time, AirSage's panel is composed of location data generated from hundreds of unique apps. Regardless of the source of data, AirSage verifies that no single app contributes to a significant portion of users of the full panel. Moreover, AirSage confirms that no single app category (e.g. News, Travel, Weather, Gaming, etc.) is more dominant than others, to avoid skewing the panel and increasing biases.

## Data Processing

AirSage evaluates discrete GPS sightings through an extensive cleansing process that refines the dataset to leverage only the most accurate and reliable sightings.

Further, AirSage looks for duplicate sightings (same device, same location within a few meters, and same minute in time) and collapses them to a single sighting using the sighting with the highest horizontal accuracy value. This is called sub-sampling, and is a part of AirSage's cleansing process.

Additionally, sightings that are too far apart spatially, in too short of a timeframe (i.e., device sightings at 12:00 PM in New York, New York and 12:15 PM in Los Angeles, California) are cleansed from the dataset. AirSage conducts this step to leverage only the highest quality devices.

*Home/Work Zone assignment:*
AirSage uses its proprietary algorithms to analyze GPS sightings generated by each sample device, over a period of time, to estimate home and work locations for the sample devices.

The assigned home location - effectively the common nighttime location - is where a device registers the highest number of sightings during the nighttime hours (9:00 pm – 6:00 am), while the assigned work location - effectively the common daytime location - is where a device registers the highest number of sightings during the daytime hours (10:00 am – 4:00 pm). To this end, AirSage's use of the term "work location" would also cover other regular daytime locations such as locations at schools/colleges/ universities.

*Point Type Classification:*
Based on AirSage's proprietary algorithms, each GPS sighting is classified as being stationary or transient at a **high reliability (above 99.9%) rate.** As it relates to AirSage's methodology, a Transient Point (TP) is in motion, and an End Point (EP) is stationary after a minimum time threshold is met.

Specifically, AirSage's standard definition of an EP is a point that has a resting location for at least 15 minutes. Thus, a dwell time of 15 minutes or longer is needed to meet AirSage's standard definition of an EP. This definition can be adjusted per client requirements, such as increasing the resting location time duration. End points are even further classified using the estimated home and work locations and refined into being a Home Point (HP), Work Point (WP), and Home-Work Point (HWP).

*Point Type Classification (continued):*
Using the assigned home and work location zones for each sample device, the sightings registered by a device within its assigned home zone are classified as an HP.

Similarly, device sightings registered within its assigned work zone are classified as a WP. Lastly, device sightings registered within the same assigned home and work zone are classified as an HWP. Device sightings at all other locations are classified into either stationary points (outside of home/ work zones) or moving points.

Device sightings classified as stationary points outside of home/ work are referred to EPs. Similarly, device sightings classified as moving points are referred to as TP. A device sighting is classified as an EP based on a rule set constructed as a function of estimated travel time between adjacent sightings, time difference, and forward moving angle between adjacent sightings in order of time. Device sightings that do not satisfy the above rule set are classified as TP.

Recent enhancements to these classification algorithms added distance to road network and buildings to further refine the points and grow AirSage's level of confidence in the determination of point type classification.

The aforementioned parameters were arrived at after empirical analysis of GPS sightings.

*Mode Imputation*:
AirSage can further process trip ends to include inferred mode of travel (air, rail, motorized, bicycle, and pedestrian) as an attribute. This mode imputation would be based on movement patterns. For instance, air travel is imputed for endpoints in one airport and subsequently appearing in another airport.

For rail, AirSage leverages a spatial dataset of railway line locations in the study area and observes movements along those corridors.

Pedestrian movements can be inferred by speed and distance traveled. For example, the whole trip typically has travel speeds consistently slower than any other mode.

Bike trips are also inferred based on speed and distance traveled, and the speed of a biker does not exceed a certain threshold throughout the device sightings associated with a single trip.

Remaining trips are assumed to be motorized. In most cases, we suggest pooling all motorized trips (i.e., automobile and bus) into a combined "motorized" classification.

## Extrapolation

One of the biggest challenges in the mobility data space is translating the results generated from a varying sample of mobile devices into insights about the full population. AirSage has developed the most efficient methodologies to do so.

The home location assigned to each device is used to compute weighted factors that allow AirSage's processes to expand the sample counts to estimate the total number of trips that would be made by the entire population. In order to do this effectively, weighted factors are computed at the census tract level geography.

These weights are calculated based on the number of devices (in AirSage's data panel) with an assigned home location in each census tract each calendar month.

Hence, the weights calculated are unique for each census tract and are subject to change on a monthly basis. The weighted factors are then applied to the trips made by each of the sample devices in order to expand those trips to represent the estimated movements in the study area for the entire population.

The final expanded trip estimates are then presented in the resulting trip matrix (or other) output.

## Data Validation

AirSage's data is routinely reviewed and analyzed for QA/QC purposes. It also has been reviewed by and substantiated by industry leaders at outside organizations such as University of Maryland and one of the largest global engineering and planning consultancies.

The quality and accuracy of AirSage output are considered best-in-class, likely a direct result of AirSage's rigorous methodologies and processes. Moreover, this fact supports that AirSage's industry leading insight is an accurate and cost-effective way to conduct mobility analyses.

AirSage's LBS data output has been compared to output obtained from the United States Federal Highway Administration's (FHWA's) National Household Travel Survey (NHTS).

For this review, AirSage provided a sample of long distance, inter-MSA and interstate travel. The expanded sample - representing all long-distance trips (>50-miles) conducted within a given time period - matched the magnitude of the reported trip rates from in-depth surveys conducted by FHWA.

Figure 1 (below) compares the trip distribution estimates from the NHTS and AirSage. The main difference was that FHWA's data cost millions of dollars and took years to collect, whereas AirSage's LBS data leveraged a larger sample size and was prepared at a fraction of the cost and time.

It is for these reasons that AirSage output continues to be trusted and procured by our repeat clients.

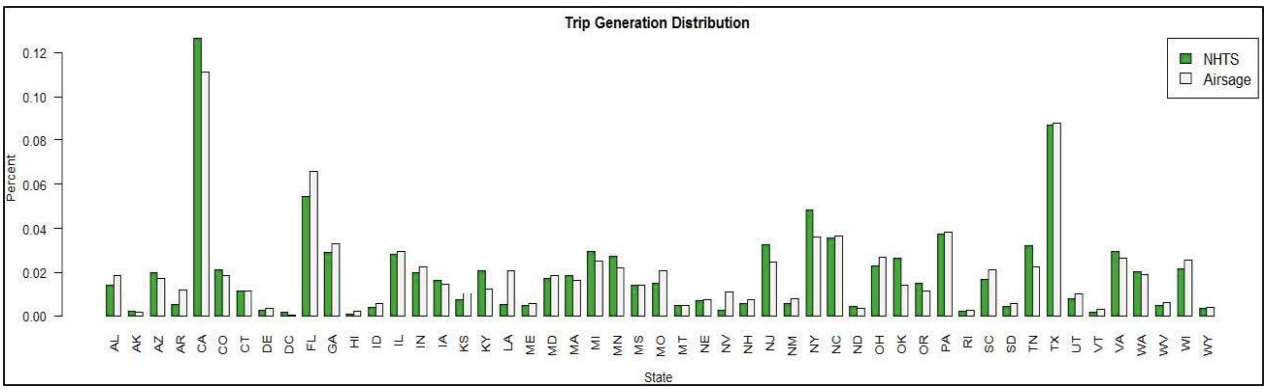



**Figure 1: Trip Generation Distribution Comparison: National Household Travel Survey (NHTS) from FHWA and AirSage Estimates**

## Data Integrity

AirSage prioritizes data integrity in an industry where data fraud has become a growing concern. As some data providers struggle with inaccuracies or even synthetic data, AirSage sets itself apart by continually monitoring for discrepancies. On an ongoing basis, AirSage tests its data against bridge closures, road closures and venue re-openings to ensure that AirSage LBS data perfectly matches the real world.

A recent study demonstrated how AirSage's data aligns with actual traffic patterns, proving its accuracy and reliability. In the study, three roadways were evaluated during a time where the roads were closed. When AirSage data was used to evaluate the road conditions, no data points were present, as anticipated.

Figure 1 below is a visualization of an I-95 segment evaluated in the study.

The image presents LBS data points (i.e., mobile device pings) in the vicinity of a catastrophic event wherein an entire segment of I-95's both northbound and southbound was destroyed in a crash involving a tractor trailer. The left image presents typical conditions and the right image presents conditions during the full roadway closure. As expected, the area of the roadway closures on I-95 during the roadway/bridge closure was clear of all LBS sightings.

By utilizing rigorous data cleansing and validation methods, AirSage ensures that its insights reflect real movement behaviors, making it a trusted source for transportation planning, mobility research, and market analysis.

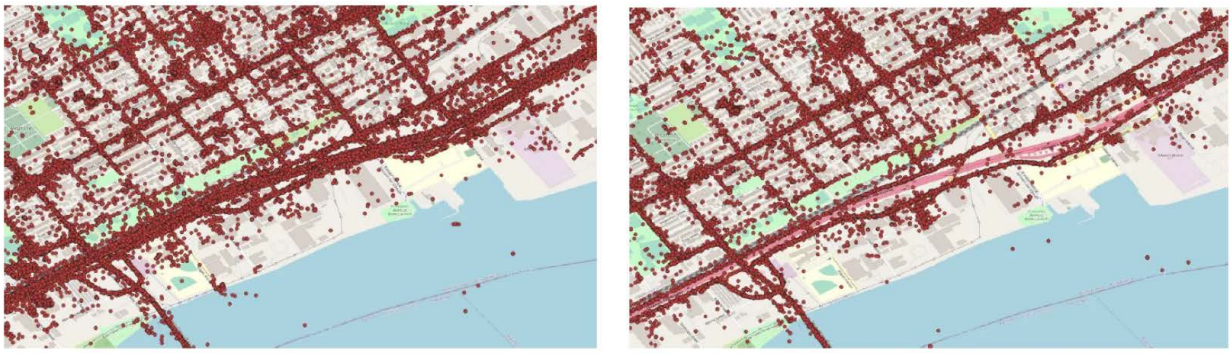


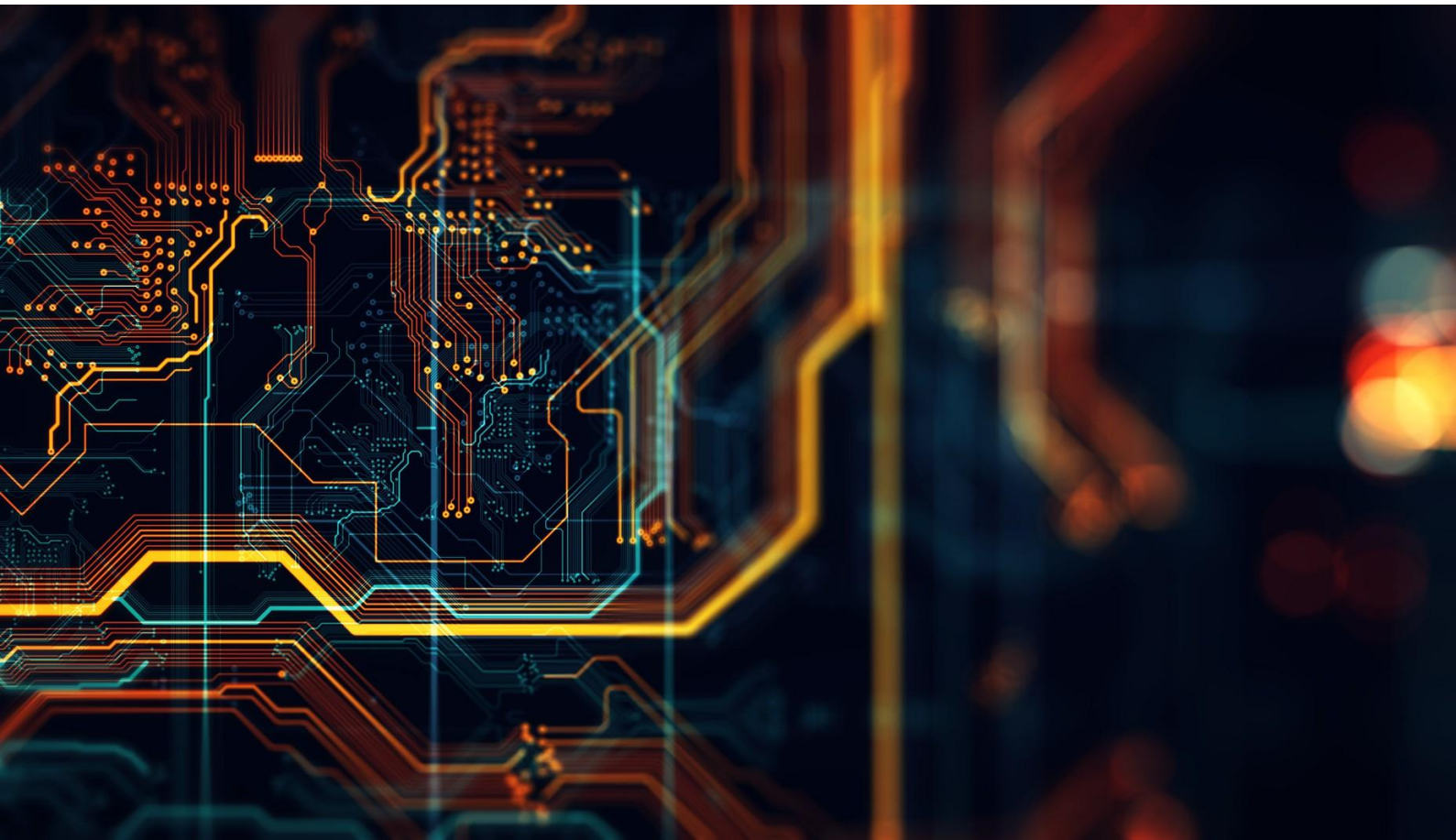
**Figure 1: I-95 Roadway Closure - Typical Conditions vs Post-Crash Event Conditions from AirSage LBS Data**

# DATA HISTORY & DELIVERY

As the pioneer in big data location intelligence, AirSage has extensive experience with collecting data.

Our high-quality GPS mobile device data goes back to 2018. With access to the raw data, we can go back in time and reprocess data to answer new questions even if they were not originally asked. The same goes for our ability to strive to improve our outputs by processing historical data with the newest methodologies as they are developed and become available.

The AirSage Team typically provides processed insights of mobile device data on a monthly basis, delivered via CSV file. Historical data can be made available quickly once the scope is defined. Custom datasets are tailored to the client's needs and, depending on size and complexity, may take a few weeks to ensure the highest quality results.



sales@airsage.com
(404) 809-2499
2146 Roswell Rd, Suite 108, PMB 862,
Marietta, GA 30062
https://www.airsage.com/