Methodology Documentation

AS DOC 001 - Version 1.0

May 17, 2018

www.AirSage.com ▪ 1350 Spring St NW Suite 475 ▪ Atlanta, GA 30309 ▪ 404.809.2499

Contents

Revision History

Date	Version	Description	By
05/17/2018	1.0	First Release	ksabba/vsivaraman

1.0 Introduction

AirSage is the leader in providing insights on population mobility from passive spatio-temporal data. AirSage delivers these insights through a suite of data products, which provides population mobility metrics across all of Continental US, Alaska, Hawaii, and Puerto Rico. Data products are developed from a nationwide sample of anonymized and encrypted GPS sightings (location traces) generated by communication devices (smartphones, tablets etc.).

2.0 Methodology Overview

AirSage using its proprietary algorithm analyzes GPS sightings generated by each sample communication device (from here on referred to as a device), across the nation over a period of one calendar month to estimate home and work locations for the sample device. On average, the algorithm estimates home and work census block for a sample more than 120 million devices per month.

Home census block is where a device registers the most number of sightings during the night hours (9:00 pm – 6:00 am) and work census block is where a device registers the most number of sightings during the day hours (10:00 am – 4:00 pm). The estimated sample attributed with home and work location is further used to produce a national home penetration table, which reports the number of resident sample devices and population across each US census tract. The device sample and population count for each census tract is subsequently used to compute device weights. The device weights (expanded to population) for each census tract is then is used to represent the mobility pattern of the population in such tract by applying these weights to mobility pattern exhibited by the representative sample devices.

Using the estimated home and work census block for each sample device, the sighting registered by a device within its home census block is classified as a "home point" or 'HP'. Similarly, any device sighting registered within its work census block is classified as a "work point" or 'WP'. Any device sighting registered within the same home and work census block is classified as a "home-work point" or "HWP".

Device sightings at all other locations are classified into either stationary points outside of home/work or moving points. All device sightings classified as stationary points outside of home/work are called "end points" or 'EP'. Similarly, all device sightings classified as moving points are called "transient points" or '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. All device sightings that do not satisfy the above rule set are classified as 'TP'. [1]Please note, duration of stay is no longer used as a criteria to classify stationary points outside home/work also known as 'EP'.

[1] For telecom based data products, stationary point 'EP' classification criteria is based on duration of stay greater than or equal to 5 minutes. For GPS based data products, since sightings are not aggregated to activity points based on a clustering algorithm, duration of stay cannot be computed using the aggregated activity point cluster start and end time.

The aforementioned parameters were arrived at after empirical analysis of GPS sightings across the US. Furthermore, the duration of stay is estimated for 'EP' point type only using a combination of time interval, distance, and empirical speed distribution between adjacent sightings.

The classified sightings along with home and work locations estimate for each sample device are used as input to develop AirSage data products. Figure 1 provides an overview of AirSage's data flow process, from ingesting raw GPS device sightings to classification, and subsequent application to produce the suite of AirSage data products. The next sections briefly discusses this process and each data product.

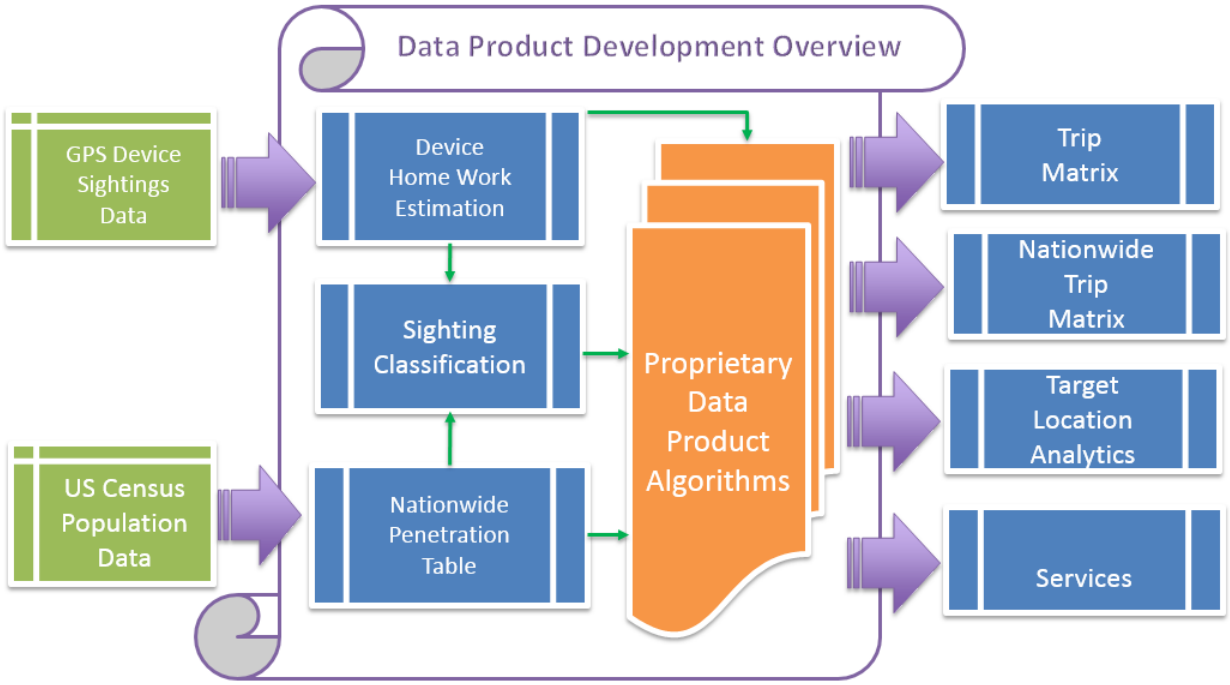

Figure 1 : Product Development Data Flow Diagram

3.0 Sample Weighting Overview

The home location estimated per device introduced in the last section, is used to compute weights to represent the entire US population. For this, weights are computed at a census tract geography for the entire US including Alaska, Hawaii, and Puerto Rico. These weights are calculated for a calendar month based on the number of devices seen in the calendar month at census tract level. Hence, the weights calculated at a census tract geography change every month, and are unique for each census tract. This weight is a simple ratio of number of resident census population in a census tract over number of AirSage devices with home location in a census tract. This weight is applied to AirSage sample devices to expand their movement across the nation to be representative of the entire US population. For instance, within census tract 1, assuming there are 1000 people living there as reported by census and

AirSage devices with home location is estimated to be 250 – each device gets a weight of 4 (1000/250). Now, if 10 of 250 devices were observed to make 20 sample trips between two analysis zones. These 20 sample trips would be expanded by weight of 4, resulting in a population estimate of 80 trips. Similarly within census tract 2, assuming there are 5000 people living there as reported by census and AirSage devices with home location is estimated to be 500 – each device gets a weight of 10 (5000/500). This weight (10) would be applied as presented in the example above to any sample trip within the US being reported by device that is a resident of census tract 2.

4.0 Trip Matrix Data Product

A trip matrix provides population travel metrics between study areas also called traffic analysis zones (TAZs) or census geographies such as block groups and tracts. It provides insights on population movement segmented by time of day, travel purpose (such as to work, home, other etc.), traveler type (resident or visitor) for an average weekday, weekend in a calendar month. The base trip matrix is configurable to add on features such as long distance filter, sub-matrix, and trip matrix with demographics. The next four sub-sections briefly discuss the features within a trip matrix. More detailed description such as inputs required and output attribute field data dictionaries can be found in Trip Matrix product documentation "ASDOC0013 – Trip Matrix Product Documentation". All the output data files are provided in comma-separated values file (csv) format.

Trip Matrix product can help answer the following questions:

- How many person trips are made between a high population area and employment area
- What time of day are most person trips made
- How many trips are made by residents of a study area vs visitors to that area
- What is the distribution shape for a study area
- How many long distance trips are made between two study area TAZs
- What is the origin and destination of people seen passing through a high use corridor
- How are people with different demographics making trips within a region such as income groups etc.

4.1 Regional Trip Matrix

In order to produce a regional trip matrix, the algorithm first draws sample data (classified sightings and nationwide penetration table) for the desired study month and geography using input (study area polygon TAZ shapefile, and parameters). The algorithm then constructs daily sample trip legs by chaining the stationary classified sightings produced by each sample device. A trip leg is a record reporting movement between two consecutive stationary sightings (HP, WP, HWP, or EP) in order of time generated by a device. Each constructed sample trip leg includes the following attributes: type of sighting, time of sighting, sighted study area TAZ, census block, device type (resident/visitor), and device weight. Trip legs constructed with the above attributes for each sample device for each day in the study month is then packaged into a trip matrix for the study region as per client requirement. Each

record of trip matrix deliverable file reports the extrapolated (weighted) number of person trips between two study area TAZs segmented by the features as per client requirements in its "Count" field. The trip legs predominately used to produce AirSage's standard transportation product, the regional trip matrix is also used to produce supplemental products such as the long distance trip matrix, sub-matrix, and socio-demographics discussed further below in this section.

4.2 Long Distance Trip Matrix

The long distance trip matrix, an offshoot of regional trip matrix, is developed from the sample trip legs used to produce the regional trip matrix. In that, the daily sample trip legs produced from classified stationary sightings generated by a given sample device is chained to form sample daily long distance trip legs, based on a predefined distance threshold (e.g.: 50 mile) by the client. Specifically, this procedure is implemented in two steps in the algorithm. First, the algorithm chains (or links) the daily sample trip legs sorted in order of time for a given device to produce daily sample trajectories. Each sample trip leg is chained to a sample trajectory being built, if and only if the distance to the trajectory origin, from the "to" end of the sample trip leg under consideration is further away from the "to" end of the previous sample trip leg chained to the trajectory. In essence, the algorithm in the first step, builds daily sample trajectories chaining the sequence of daily sample trip legs up until the sample trip leg that does not meet the above criteria. In the second step, the algorithm processes the sample trajectories to produce sample long distance trip legs using the attributes from the origin (first sighting) and destination (last sighting) of each sample trajectory. The daily sample long distance trip legs constructed as discussed above, are then filtered based on the pre-defined distance threshold (e.g.: 50 mile), eliminating chain of trips that are shorter than 50 miles in length. Further, each sample long distance trip leg is built with the same set of fields as a regular sample trip leg, and hence can be packaged to the same set of specification as a regular regional trip matrix discussed in the previous section.

4.3 Sub-Matrix

The sub-matrix, as its name suggests, is a subset of the trip matrix, in that it reports trips between TAZs in a study area for a sub-set of the population that are seen within a client defined sub-zone such as a roadway corridor, toll road etc. The sub-zone needs to be at least 2.5 miles in length for best desired results. The sub-matrix requires the client to provide information on the sub-zone such as a shapefile showing the corridor/toll road etc. in addition to the regular TAZ shapefile. The sub-matrix algorithm extracts classified sightings seen in the sub-zone and uses the select sightings in combination with trip legs to produce sub-matrix results. It primarily reports the daily travel pattern between study area TAZs for only the segment of population seen in a sub-zone. The sub-matrix product output, besides including all the attributes of a regular trip matrix, also includes an additional attribute called "SZCount". The "SZCount" reports the number of person trips seen passing through the sub-zone. The number of sub-zones are limited to less than three (3) per study area due to complexity involved in

producing the sub-matrix. The sub-matrix provides output data for each sub-zone separately. For instance, if a study area has two (2) sub-zones, each sub-zone selection is done separately and two (2) output files will be produced. The selection is not based on an intersection of device sightings seen in both the sub-zones.

4.4 Trip Matrix with Demographics

Trip matrix with demographics augments the regional trip matrix discussed in section 4.1 with demographic variables such as household income, age/gender, vehicle ownership, and ethnicity. The demographic variables are inferred based on a device home location as discussed in section 2.0. The demographic variables published by US Census ACS 5-year estimate is used in conjunction with sample device home locations at a census block group geography to tag sample devices with demographic variables. The demographic variables are added to the trip matrix broken down by census demographic bins. In essence, total count of person trips between TAZs within a study area are broken down by income groups, age/gender groups, vehicle ownership groups, and ethnicity groups.

5.0 Nationwide Trip Matrix Data Product

Nationwide trip matrix provides population travel insights across all of US including Alaska, Hawaii, and Puerto Rico. The output file format is not configurable and is produced on a monthly basis at a census block group geography. It provides extrapolated (weighted) person trips between origin census block group, and destination census block group segmented by home census block group, travel purpose, hour of the day, and day of the week. Nationwide trip matrix product produces one compressed csv file for every calendar month. More details on the data dictionary can be found in Trip Matrix product documentation "ASDOC0013 – Trip Matrix Product Documentation".

6.0 Target Location Analytics (TLA) Data Product

TLA product provides population insights for a point of interest (POI) or several points of interest (POIs). POI typically is a high population attractor such as a retail mall, fast food joint, event space, or a game stadium. Insights provided include trade area reach such as home location and work location distribution for population seen at a POI, hourly population distribution at a POI on a daily basis, frequency of visits to a POI, duration of stay at a POI, and population demographics such as income, age, and ethnicity. Device sightings are filtered to capture visibility at a POI based on input parameters. The device sightings are extrapolated using weights to represent population mobility. More details on the input parameters and the output file formats can be found in the TLA product documentation "ASDOC0012 - TLA Product Documentation". The output of TLA product is a series of compressed csv files.

TLA product can help answer the following questions:

- How many people are seen at a specific shopping mall in a month(s)
- How many times have they visited a specific shopping mall in a month(s)
- How much time did they spend at a specific shopping mall
- Where do they live and work
- What is the income distribution of the people seen at a specific shopping mall
- What is the age, gender, and ethnicity distribution of the people seen at a specific shopping mall
- How many people are seen a football or baseball game
- What is the demographics of the people seen at the game

7.0 Services

Few products currently in development are categorized as services. This is because these products are still in development, and do not have all the specifications required to complete the product. Further, AirSage is still in the process of customer discovery, an essential step to gain a comprehensive understanding of the marketplace requirements. In addition, the process to fulfill the services is also not yet streamlined, and requires significant amount of manual effort. Below is the description of two (2) of these services.

7.1 Activity Density

Activity density provides insight on population activity at a defined geography for a certain time period within a larger region. The geography to which population activity is aggregated is usually at a coordinate grid level such as 1000-meter grids, 100-meter grids, or 10-meter grids. Activity density is extrapolated to represent population activity using weights discussed in sections 2.0 and 3.0.

Activity density, for instance can provide insights at 100-meter grids in downtown Los Angeles, population activity by hour of the day for a specific day(s) or for all weekdays or weekend days. It can also provide population activity at 10-meter grids by hour of the day in the vicinity of a special event such as a marathon.

Input requirements and output file structure for activity density are discussed below.

The following input parameters are required to process activity density:

a) Bounding coordinate box with latitude, longitude. Activity density will be provided within this bounding box. The bounding box must contain 4 edge coordinates with latitude and longitude.
b) Study time period. The study time period can be a calendar month or specific day(s) within a month(s). If study period is calendar month – weekdays vs. weekend days vs. entire month. For calendar month, the activity provided will include totals across weekdays/weekend days/entire month.

c) Grid geography – 1000 meter grid aggregation vs. 100 meter grid aggregation vs. 10 meter grid aggregation.
d) Sighting filter – by default all sightings within the bounding box will be used. A filter can also be applied based on point type definitions mentioned in section 2.0. Filter can be applied to only include 'EP' point type or a combination of 'EP' and 'WP' point types etc. within the bounding box.

Based on the input parameters, the output is either one comma-separated values (csv) file or a series of csv files.

For instance, if the study time period selection is one full calendar month, then only one output csv file will be produced. Instead if it's weekdays vs. weekend days, then two csv files will be produced showing weekdays vs. weekend day totals. Similarly, for individual days, one csv file will be produced for each day. The output file attribute table is shown in the figure 2 below.

Figure 2: Activity Density Attribute Fields

Attribute_Field	Description
activity_lat	**based on input grid geography - one of the three will be in the output:** 1000 meter activity grid coordinate (latitude with 2 decimal places) 100 meter activity grid coordinate (latitude with 3 decimal places) 10 meter activity grid coordinate (latitude with 4 decimal places)
activity_long	**based on input grid geography - one of the three will be in the output:** 1000 meter activity grid coordinate (longitude with 2 decimal places) 100 meter activity grid coordinate (longitude with 3 decimal places) 10 meter activity grid coordinate (longitude with 4 decimal places)
hour	H00 to H23 (H00 - activity between midnight and 1am; H01 - activity between 1am and 2am etc.)
density	population seen in the hour at the coordinate grid during the study time period

7.2 Convention Visitors Bureau Studies (CVB)

CVB provides insights on visitor devices to points of interest within a region. The points of interest (POIs) are high attractor places such as adventure parks, small beach towns, boardwalks, beaches, museums, events etc. CVB provides home location report for visitor devices seen at the POIs weighted to population. It also provides insights on number of days visitor devices are seen in the region. Visitor devices to region and POIs within the region are defined and sampled based on the home location of a device. The region is typically defined by a buffer boundary. All devices seen at the POIs with home location outside the buffer boundary are tagged as visitor devices. It's important to have the POIs lie geographically within the buffer boundary. The buffer boundary is typically a 50-mile or 100-mile radius

from a study area centroid or boundary. The study region cannot be bigger than a census designated MSA (metro area).

Input required for CVB is two polygon shapefiles. One shapefile must contain all the points of interest (POIs). Each POI must represent a unique polygon within the shapefile. The second shapefile must contain one buffer polygon showing the 50-mile or 100-mile radius.

Study time period is typically a calendar month or few contiguous days for events. The visitor devices are further weighted to represent population. CVB doesn't have standard set of output files produced since it's still in development. However, a set of preliminary files are produced for testing as described below:

- For visitor population seen at each POI during the study time period, count of visitor population aggregated to home census block groups is provided in a csv file format.
- For visitor population seen at each POI during the study time period, count of visitor population aggregated to number of days seen in the region is provided in a csv file format.

CVB, for instance can provide the visitor population seen in Cape Canaveral, FL or Melbourne Beach, FL with visitor being someone with a home location outside 50 miles of Brevard County, FL. Similarly, it can provide visitor population seen in Central Park, NYC or Times Square, NYC with visitor being someone with a home location outside 100 miles of Manhattan. In the above examples, Cape Canaveral, Melbourne Beach, Central Park, and Times Square are POIs.

7.3 Custom Analysis

Anything outside of what is described in this document is categorized as custom work and will require prior approval from the technical team at AirSage. Custom work examples may include and is not limited to the following:

- Sub-matrix/long distance trip matrix with demographics
- Home census block group report of visitor population seen at the two or more POIs. Intersection of devices seen at two or more POIs