UBERMEDIA

Derivation of Common Evening and Common Daytime Locations

Overview

The UberMedia user database is developed by associating a user device ID (Apple IDFA or Android AAID) with a Common Evening Location (CEL) and a Common Daytime Location (CDL), then associating each of those locations with various geographic districts. District assignments always includes the country. In several countries, various additional districts are available, including the province, postal code and census region.

Data Source Discussion

UberMedia sources data from location SDKs and also extracts location data in the course of mobile advertising. Ideally, these coordinates are returned from the device "location services" which is a device-manufacturer-proprietary combination of satellite and terrestrial signals that yields the most accurate possible location of the mobile device itself.

In order to ensure that the data signals are derived from interaction with the GPS on devices and not by some other means, UberMedia's Mobile Intelligence Engine scores each ingested data point, and discards those data points deemed low quality—points that were determined at the time of the device registration, or that have been derived via an IP lookup, or that were cached while the owner had location services turned off.

All ad requests are time-stamped with UTC time. For efficiency, local time is approximated by adjusting the recorded UTC time by the reported longitude; detailed timezone information is ignored.

Analysis Timeframes

The **Common Evening Location (CEL)** for a device is estimated by determining where a device most frequently appears during the "non-work" hours (evening through morning and weekends). The overnight hours are defined as after 6pm and before 8am, weekends include Saturday and Sunday.

The **Common Daytime Location (CDL)** for a device is estimated by determining where a device most frequently appears during the "work" hours (daytime on weekdays). The hours are defined as after 8am and before 6pm from Monday through Friday.

Estimating Locations

For each device with sufficient data, estimates of CEL and CDL are made monthly, combining data from the most recent month with results from the two previous months to form updated location estimates. CEL and CDL are estimated independently for each device, each using the same process described below.

Location data for a device over the course of the month may contain many distinct points that represent the presence in the same building or land parcel. However, we are not concerned in small variations in position (e.g., if a device is at one end of a house or the other.) In order to collect statistics on position data which represent the same meaningful location, we group all the points which fall close to each other (within approximately 100 feet) into “groupings” and count the number of points in that grouping. If a device has at least two groupings, or one grouping of two requests, during the month then the estimation will proceed. A device’s groupings are ranked by the number of points contained in each grouping. The top five groupings per device, per month are retained indefinitely.

For each monthly update, the groupings from the current and two previous months are considered. Groupings are scored by the number of points in each, subject to a fading memory weighting. A grouping of 10 points from this month will have a higher score than a different grouping of 10 points seen one month ago. Similarly, a third grouping seen with 10 points in each of two distinct months (20 total points) will have a higher score than either of the previous two examples, but a value less than 20. The current month’s CEL or CDL update will use the highest scoring grouping.

To combine the points in the highest scoring grouping efficiently, the simple average latitude and average longitude of all points in the grouping over all months are formed. These become the updated CEL or CDL coordinates for the month. Recall that these points are no more than about 100 feet apart, so average-distorting outlier points will not be present; the averaging simply acts as a minor geographic refinement in position.

From January 2018-forward, each monthly update is stored in order to maintain a historic record of movements of the devices CELs and CDLs over time. These movements are then reflected in the various CEL-CDL reports output.

District Assignments

Once the CEL or CDL locations are derived for a given months, the data is loaded into Vista. When a CEL or CDL data point is set to be returned within a data set, any geographic districts associated with the CEL or CDL is applied.

A set of geographic boundary files have been collected which define country boundaries, postal districts, municipality boundaries, administrative and/or demographic reporting regions and similar (e.g., US Census Block Groups, Australian SA1 boundaries, Romanian Comunes, or Obces in Czechia). These are used to assign district codes to any CEL or CDL points which fall within one of the defined shapes. The list of these available and automatically assigned boundaries is available for download here.

Some specific caveats on the geographic district assignments. For all countries, at a minimum, devices seen in the country and meeting viewability thresholds in the CEL-CDL database are assigned

a home country. This is generally taken from the country corresponding to the CEL location, but if only a CDL is available, the home country will be assigned as the country of the CDL. The country boundaries are defined by a country's high resolution district files when available, or by a coarser public-domain global country boundary shapefile. Note that because the district boundaries come from various sources, the spatial resolution is variable and borders are not necessarily consistent at all scales. Some inconsistency may occur in assignment when a CEL or CDL is very near a geographic border defined by these shapefiles.

Boundary Licensing
UberMedia offers the ability to license any of the geographic shapes through a partnership (MBI) which specializes in gathering these boundaries and generating structured outputs. Contact your UberMedia representative to discuss further.

Beyond standard government-created geographic boundaries, there may be an interest in applying a custom context shape to the output of any given CEL or CDL report. UberMedia has developed a system whereby our partners can license or create shape files for their own use and turn those over to Vista. When a report is requested by that company (and that company alone), these special district assignments will appear within the report. Segmentation tiles, geohex, custom grid shapes, congressional districts, etc are all examples of districts supported on a company-by-company basis. other companies do not have rights to shape files sourced from another partner.

Notes on Privacy

In addition to hashing, all CEL and CDL locations output by Vista are jittered 50m in a random direction in order to help maintain the de-identification of device-level data.

In the EU, no lat-long points, including jittered points, are present in the output of the CEL and CDL reports. The latitude and longitude columns are empty.

Expectations of CEL-CDL Placement

Note: while the following analysis references the United States, the general logic applied is also applicable to other countries as well.

About half of the US devices in the CEL/CDL database have very little spatial separation (less than 100ft) between the CEL and CDL. This is expected and consistent with the US labor force participation rate and other labor statistics described in more detail below. When interpreting CDL reports, it is important to keep in mind that it is an estimate where the device most often spends its weekday daytimes. Although the CDL time window (8am-6pm, M-F) corresponds to traditional "business hours", not all devices are owned by people employed in traditional businesses. In fact, a

large fraction of the US adult population is not in the labor force at all. For example, in 2017 only 60% of adults in the US were employed [1].

Overall, multiple factors contribute large number of devices having an estimated daytime location in about the same place as their common evening location:

- The sizable population not in the labor force (retirees, homemakers, people on permanent disability, etc.), as well as the unemployed. This is about 40% of the adult population.
- Teenage device owners, who may be home in the afternoon and over the summer. Teenagers make up about 12% of the population [2].
- Shift workers (e.g. hourly retail, restaurants, etc.) whose work hours are variable and may be home on some weekdays.
- Work-from-home workers, and a growing fraction of employees who do so part of their time. As many as 43% of employed people spent some time working remote [3].
- "Gig economy" workers who are potentially not in one place consistently enough to establish a single daytime location (e.g. Uber/Lyft drivers).
- Some tablet devices may stay at home permanently.
- Device owners may preferentially use some apps at home rather than at work.
- Some workplaces have restrictive policies which do not allow phone use.

Together these factors should be considered when evaluating CDL reports. The CDL estimates do show where the device most often spends its weekday daytimes, even if it is not a traditional workplace. A large fraction of the population is either not employed, does not have consistent daytime working hours or does not have an out-of-home workplace. UberMedia's location reports can help reveal these diverse patterns of life across the population.

References:

[1] Bureau of Labor Statistics, *The Employment Situation - December 2017*, https://www.bls.gov/news.release/archives/empsit_01052018.pdf

[2] US Census, *American Community Survey 2012-2016*, https://factfinder.census.gov/bkmk/table/1.0/en/ACS/16_5YR/DP05

[3] New York Times, 2017-02-15, https://www.nytimes.com/2017/02/15/us/remote-workers-work-from-home.html