



UBERMEDIA

Overview Of UberMedia Vista Data Sets

A review of the contextualized data sets available through the Vista Data API. Included in this document is a discussion of the contents and application for each type of data set.

Overview

UberMedia Vista interfaces with a location data store which contains over 13 trillion mobile location observations since January 2014. These mobile location data points are collected from location SDKs and from advertising SDKs, and then screened to remove any non-high quality data.

The Vista API and the Vista Explorer both allow access to formatted data sets which can provide a variety of insights concerning a location's visitors. These data sets are broadly categorized into five distinct types:

- **Foot Traffic Reports:** Breaks down counts of visitors to a set of locations.
- **Point-in-Polygon Reports:** Provides specific information about visits by specific devices to a set of locations.
- **Path-to-Purchase:** Provides data points related to a device's movements on the way to and from a set of locations, or in a set window of time.
- **Location:** Derived data sets based on activity associated with a location.
- **Device ID Reports:** Provides information about visiting devices.

This document outlines each report in these categories and what business insights can be derived from each report. For information on how to access the API or Vista Explorer, contact your UberMedia representative.

Glossary

Polygon: Generic shorthand for any shape upon which location analysis is done. In practice, the Vista Data API accepts ESRI™ Shapefiles, geoJSON or KML files as standards for defining polygons or shapes to use for reports.

Gold Standard Polygons: Some reports outlined in this document leverage UberMedia's Gold Standard Polygon (GSP) set. GSPs are hand-drawn and double-verified shape files which exactly outline the boundaries of high-value retail properties. The amount of GSPs grow over time, adding more value to the reports leveraging these shapes.

Study Locations: The (set of) location(s) used in the specific analysis.

Baselined Mobile Data: Due to growth of mobile data over time, for certain reports, UberMedia will provide a baselined measurement which is an attempt to measure organic growth or decline in visitation apart from the natural growth of observability over time.

Monthly Active Uniques (MAUs): The average number of unique devices seen in a month.

Friendly Dates: A conversion from Unix timestamp (number of seconds elapsed since January 1, 1970 UTC) into a comprehensible date in the form of date-time-day of week-timezone.

Common Evening/Daytime Locations: For devices with sufficient data, UberMedia derives a Common Evening (CEL) location which is an expression of the location where the device is most frequently observed in the evenings and on weekends. The Common Daytime Location (CDL) is where the device spends its weekday days. More specific information on the derivation of these values is available from your UberMedia representative.

Foot Traffic Reports

Foot Traffic reports are designed to return information about counts of visitors to the study locations. It is most useful for answering business questions related to volumes of visitors and baselined visitors.

Visitor Count

Returns total count of visitors and visits for the timeframe for each study location. The visit count is the count of days the device was observed in the study geography, not the total number of observations within the geography. The format of the report is:

Polygon	Unique Visitor Count	Visit Count
---------	----------------------	-------------

Visitor Count by Day

Returns total count of visitors to each study location, broken out by day. The format of the report is:

Date	Polygon	Observed Devices
------	---------	------------------

Extrapolated Visitor Report

The extrapolated visitor report uses the visitation behavior of a set of “always-on” panelists to extrapolate an estimate of real-world visit and visitor numbers.

Polygon Name	Panelist Visitor Count	Panelists Visits Count	Extrapolated Visitor Estimate	Extrapolated Visit Estimate	Average Visits per Visitor
--------------	------------------------	------------------------	-------------------------------	-----------------------------	----------------------------

Retail Heartbeat

The Retail Heartbeat report is a baselined estimate of visitors to the study locations developed to offset the rising observability of mobile devices. Since 2014, the UberMedia Vista location data store has grown from 80MN MAUs to 400MN MAUs in the US. Thus, this report provides a visitation index which outlines visitation trends to locations, had the volume and observability of

data been consistent through time. This is baselined against the time period of maximum reach and observability—which is the most recent time periods. This report includes a .png of a plot of the report data and the raw data in the following format:

Location	Date	Observed Devices	Visitation Index	Uncertainty
----------	------	------------------	------------------	-------------

Due to its nature of calculating baselines to an index, this report can only be run from January 2016 forward through the end of the past month. These time restrictions are a result of a large offline processing procedure which must be done to create the baseline and extrapolation model. A separate document that does a deep dive into the methodology behind this report is available upon request from your UberMedia representative.

Point In Polygon Reports

Point In Polygon reports can be used to track visitation patterns to locations for cross shopping behavior, or day of week/time of day reporting and frequency of visit analysis.

Zero Point

A zero point report provides a list of unique devices and the timestamp the device was observed in the study locations. The format is:

Polygon ID	Hashed Device ID	Unix Timestamp of Visit
------------	------------------	-------------------------

Pin Report

A pin report provides the same information as the zero point report, but with the addition of the exact lat-long the of device when while seen in the polygon. This becomes useful for analyzing traffic patterns of larger areas like shopping malls, high streets, or neighborhoods. The format of this report is:

Polygon ID	Hashed Device ID	Latitude of Visit	Longitude of Visit	Unix Timestamp of Visit
------------	------------------	-------------------	--------------------	-------------------------

Like all our reports that include timestamps, an option exists to convert the output of the report to contain the timestamp of point-in-polygon visits to be “friendly” dates, as well as a choice to convert to local time.

Path To Purchase Reports

Path to Purchase reports provide detailed information about the travels of individual devices, either contextually in time to a visit to the study locations, or alternatively, following the travel over a specific interval or independent period of time.

Pre-Post Report/Pre-Post Sampled

Provides, for all study locations, all the points the device was seen before and after visiting a specific polygon. The timeframe for the before-and-after window is set at the time the report is requested. The report is formatted as follows:

Polygon ID	Hashed Device ID	Unix Timestamp of Visit	Pre/Post-Visit Timestamp	Pre/post-visit Latitude	Pre/Post-Visit Longitude
------------	------------------	-------------------------	--------------------------	-------------------------	--------------------------

This report can be tricky to work with when first using it, as data for both pre-visit observations and post-visit observations are mixed together in one column. The appendix of this document gives a specific example of the interpretation of this report.

The Pre-Post Sampled report is delivered in the same format, but only a sampling of pre-and-post visits is used. This sampled report is good for use cases where the amount of data returned in the pre-post regular report is overwhelming and unwieldy, making analysis by non-big data engineers a challenge. A random sampling of pathing behavior is often sufficient to pick up broad trends for pathing.

Pathing X Report

The Pathing X Report, like the Pre-Post report, also contains information about the travels of a device before and after being seen in a study location. The data, however, is presented in a different format:

Polygon ID	Hashed Device ID	Latitude of Observation Point	Longitude of Observation Point	Time Before Appearance In Polygon	Unix Timestamp of Observation
------------	------------------	-------------------------------	--------------------------------	-----------------------------------	-------------------------------

In this case, the timing of the observation in relation to the visit is expressed in a +/- format. So a value of 60 in the “Time Before” column means the observation happened 60 seconds before arrival at a study location. A value of -120 means the observation happened 2 minutes *after* the appearance in a study location.

Note that there will be some lines in this file that have a Time Before value of zero—these are the so-called “zero points” and represent observations inside the study location.

Pathing X with Context/Context-Only

The Pathing X Report contains all the points a device was observed. The Pathing X with Context enhances this report by adding 3 additional columns to the report:

Polygon ID	Polygon Name	Polygon Category
------------	--------------	------------------

This report leverages UberMedia’s Gold Standard Polygons. If an observation point appears inside a GSP location, then the information about that location is appended to the observation.

This report also has an option to receive *only* those pathing reports that are within GSPs, instead of all points enhanced with the context.

Feeder Report

The Pathing X Rollup report, or Feeder report provides a summarized output of where visitors to a location visit before/after visiting the study locations. The results are delivered in a series of four files that are at the polygon, polygon category, sub-sector and sector level.

The sector analysis file contains some special values in it, in addition to the sector information. The report has a “CEL” value which is the percentage of study location visitors who were seen at their CEL within the report’s pre/post time window. The “CDL” value is the percentage of visitors who were seen at their CDL in the pre/post window. And then finally, “Transient” indicates that the visitor was only seen in points more than 20 miles away from the study location that was not near their home or work area (thus being likely transient through the area).

The sector file is thus formatted as follows:

Destination Category	Sector Name	Number of Devices Analyzed	Percent Appearing in Sector in Pre/Post Window
----------------------	-------------	----------------------------	--

Subsector:

Destination Category	Sector Name	SubSector Name	Number of Devices Analyzed	Percent Appearing in Subsector in Pre/Post Window
----------------------	-------------	----------------	----------------------------	---

Category:

Destination Category	Category ID	Category Name	Number of Devices Analyzed	Percent Appearing in Category in Pre/Post Window
----------------------	-------------	---------------	----------------------------	--

Polygon:

Destination Category	Polygon ID	Polygon Name	Number of Devices Analyzed	Percent Appearing in Polygon in Pre/Post Window
----------------------	------------	--------------	----------------------------	---

Custom Context

Instead of using the UberMedia Gold Standard POI polygon set, Vista users can pass in their own custom POI to use for context. This is done by means of a standard geoJSON input which can contain categories and polygon names.

Device History Report

The Device History report delivers a list of observation points for a set of devices during a specific period of time. This is different than the Pathing reports described above because the data is not associated with a specific visit to a study location. Thus, the input to this report is a list of device IDs, generally pulled from an external data source. This report is best used to get a full picture of the movements of a device through a given time period. Its output is:

Hashed Device ID	Observation Latitude	Observation Longitude	Timestamp of Observation
------------------	----------------------	-----------------------	--------------------------

Like its pathing brethren, this report can also be run through the Gold Standard Polygon set to produce a **Device History with Context** report. This appends columns to the report: Polygon ID, Name and Polygon Category.

Polygon ID	Polygon Name	Polygon Category
------------	--------------	------------------

Cluster Report

The cluster report is a specialized pathing report which attempts to algorithmically identify popular stopping points for devices on their way to and from study polygons. Unlike the “with context” family of reports, this report can identify areas that are outside of identified POI.

The output of the report is a geographic file (geoJSON or geoDB) which can be imported into mapping software. Each cluster or stopping point is identified by a contour shape and a number associated with the contour shape which outlines how many unique devices were identified as stopping within that contour shape.

Numerous parameters can help adjust the ruleset for when to create a contour, based upon the underlying mobile data.

Common Evening Location (CEL) and Common Daytime Location (CDL) Reports

The CEL-CDL Expanded Report

Using a proprietary methodology, UberMedia derives CEL and CDL values for devices on a monthly basis. Not all devices in a study will have one of these values. UM has one main CEL-CDL report (the “Expanded” report), but offers a few variations of CEL-CDL reports so that each partner can select the variant that is optimized to their use cases.

The *standard* CEL-CDL Expanded report is delivered as a gzip file three different reports: a CEL report, a CDL report and a summary report. The CEL and CDL have the same 17 columns as follows:

Polygon ID	Hashed Device ID	CEL Latitude	CEL Longitude	Timestamp	CEL Country	CEL Census	CEL Micro	CEL Municipality
CEL Admin	CEL Province	CEL Postal 1	CEL Postal 2	CEL Custom 1	CEL Custom 2	CEL Distance	Visit Count	

If “Friendly Dates” is identified to append to the report, the following 4 columns will also appear:

Local Date	Local Time of Day	Local Day of Week	Local Timezone
------------	-------------------	-------------------	----------------

The CDL file will have the SAME columns but the column headers will read “CDL” and the data contained within will all refer to the device Common Daytime Location instead.

Some Important Notes about Timestamps and Dates:

- The default timestamp is in Unix Time.
- The timestamp represents the **first** visit to the study geography/polygon ID from the CEL/CDL on the line in question.
- If a device has visited the study geography multiple times over the course of the study period *and* moved during the study period, then the device will be listed multiple times. Each entry in the report will represent a pair of study geography-CEL location with the timestamp representing the first time they visited the study geography from each CEL location.
- If a device has not moved during the study time period there will be only one entry in the file for each study geography with the timestamp representing the first observed visit within the study time period.
- The times, whether Unix/epoch or “friendly” represent the time and timezone of the visit to the polygon, not the local time at the CEL point. A visit to the Eiffel Tower by a New York City resident will have its time expressed in Central European timezone, not in Eastern Standard Timezone.

An Important Note about Visit Counts and Distance:

- Visit count within the context of this report is not to be confused with device observations or pings available in other reports.
- Similar to the Visit Count report discussed above, the visit count is a count of the number of distinct days that the device appeared in the study geography.
- “Distance” is a value in “as-the-crow-flies” miles from the CEL or CDL to the centroid of the study geography where the visit occurred.

Variants of the CEL-CDL Report

In order to help cover more use cases, Vista offers some variants on the report which may be better suited for a specific project.

The Detailed Version

The detailed version of the report has a critical difference in the each visit-day to the study geography is recorded in the report. For instance, if a device visited 10 times to the study geography, all appearances are noted in the report. While the columns remain the same as the standard output, the amount of rows in the report will be larger as each visit is captured as its own row.

The Legacy Version

UberMedia has been producing CEL - CDL reports since we first entered the location data business 7 years ago. During that time, we have been constantly innovating to try to cover the most use cases for our clients by improving the quality and outputs of the reports we offer. The legacy version of the CEL-CDL report was known (and is still available through the UI and the API) as the CEL-CDL Distance Report. This report is formatted as follows:

Polygon ID	Device ID	CEL Latitude	CEL Longitude	CEL Distance	CDL Latitude	CDL Longitude	CDL Distance
------------	-----------	--------------	---------------	--------------	--------------	---------------	--------------

There are a few critical differences with this report from the Expanded report:

1. CELs and CDLs are all in one file, leading to larger file sizes.
2. The available boundaries (“full geo”) are limited to a smaller set of boundaries from a smaller amount of countries.
3. The “full geo” options are not included by default in the report.

The Summary File

By default, a summary file is included with the CEL-CDL reports. This summary provides an aggregation of the detailed data contained in any of the CEL-CDL report variants. The summary file includes:

- Total count of devices/visitors¹ with a CEL (repeated for CDL)
- Total count of visits² with a CEL (repeated for CDL)
- Median Distance traveled from CEL to the centroid of the study geo (repeated for CDL)
- A pseudo-histogram with buckets for various distances traveled to get to the study geography (repeated for CDL)
- Top 20 lists of geographic drivers of visits to the study geography by:
 - Postal 2
 - Postal1
 - Census
 - Micro
 - Admin
 - Municipality
 - Province
 - Country

Notes:

¹ Visitors in the summary are any device + cel/cdl combination. Any one device ID can be considered as multiple visitors if they have moved CELs or CDLs during the study period. Keep this in mind if using the standard report and attempting to reproduce the values in the summary file.

² Visits are defined as 1 visit per study geography per device per day. Repeat visits within the same day are not counted as more than 1 visit per day.

The Visitor Home Report

Combines information from the point-in-polygon report and the CEL reports into one convenient format:

Device ID	Polygon ID	Timestamp seen in Polygon	CEL Latitude	CEL Longitude
-----------	------------	---------------------------	--------------	---------------

Generator Report

The Generator report uses less history of a device's movements to determine where the device spends its time in the previous 7 days leading up to a location visit. The Generator might be most useful when analyzing behaviors in, say, a tourist town, where something like a hotel would be a device's "home" for a preceding 7 days, as opposed to their actual CEL. This report contains:

Resident-Worker Report

Device ID	Generator Evening Latitude	Generator Evening Longitude	Generator Daytime Latitude	Generator Daytime Longitude
-----------	----------------------------	-----------------------------	----------------------------	-----------------------------

This report generates a list of all people who have either a CEL *OR* a CDL inside the (set of) study locations:

Polygon ID	DeviceID	Resident	Worker	CEL Latitude	CEL Longitude	CDL Latitude	CDL Longitude	Commute Distance
------------	----------	----------	--------	--------------	---------------	--------------	---------------	------------------

The "resident" column is set to TRUE if the CEL is within the study polygon boundaries, FALSE if the CEL exists but is outside the study polygon boundaries, or UNKNOWN if no CEL exists for the device. The "worker" column works the same way, but applies to CDL values. If a device has neither a CEL nor a CDL within the study polygon boundary, it is excluded from this report. It is also possible that a device lives and works within the same study polygon, and in this case, both of these columns are TRUE.

Additional Device-Level Reports

Device ID Reports contain information that UberMedia has derived about the device, based on all the information associated with a specific device—location history, audience/segment membership, demographics, etc.

Demographics Report

The demographic report is based on the latest-available census data from governmental agencies in the supported countries. For example, the US data is based on US Census Bureau's ACS dataset. The demographics are paired through the census ID associated with a device's CEL. There are 3 files delivered with this report. The summary file gives a holistic picture of the household level demographics of visitors to all locations using a weighted average methodology. The input file gives a breakdown by the input passed to Vista—in most cases this input is a set of polygons, thus the input file shows a breakdown of visitor demographics per polygon. Finally the census_ID file gives a listing of each census ID found to contain a visitor's CEL and the associated demographics of that census ID. The formats of each of these files is as follows:

Summary

Category	Sub_Category	Value	Devices Count	Sub_Cat Sort Order	Unit
----------	--------------	-------	---------------	--------------------	------

Input

Input	Category	Sub_Category	Value	Devices Count	Sub_Cat Sort Order	Unit
-------	----------	--------------	-------	---------------	--------------------	------

Census_ID

Census ID	Category	Sub_Category	Value	Devices Count	Sub_Cat Sort Order	Unit
-----------	----------	--------------	-------	---------------	--------------------	------

Notes:

1. as of the ACS 2016 demographics, the US Census has made changes to the numerical format representing census values which come from too small a sample to report on. They have now changed to a large negative number for things like home value to represent "not enough data." This will have the side effect of potentially generating negative numbers in some census block report cells.
2. Supported countries as of this writing of this document: USA, AUS. However, UberMedia has partnered with MB-International to enable the licensing of additional socio-demographic data from over 100 countries. Contact your UberMedia representative for more information.

Audience Affinity Report

The Audience Affinity report leverages UberMedia’s Gold Standard Audiences. These audience segments are built monthly by UberMedia leveraging various mobile data cues: locations visited, apps used, demographics, etc. The report is formatted as follows:

Audience Code	Audience Name	Count of Visitors in Audience	Percentage of Visitors in Audience
---------------	---------------	-------------------------------	------------------------------------

*Note: As of the writing of this document, this report is not baselined. Meaning that large scale audiences may artificially appear to have a high percentage of devices in them. Thus, it is **highly recommended** that a baseline audience affinity be run (a random selection of devices from the same city or DMA for example), and reports presented as indexed against a baseline.*

Audience Membership Report

The Audience Membership report gives a visitor-by-visitor breakdown of audience membership. The report is formatted as such:

Location Affinity Report

Hashed Device ID	Audience Name
------------------	---------------

The Location Affinity report assesses visitors to set of study locations by examining which other Gold Standard Point-of Interest Polygon locations these visitors were observed. It then baselines these visits against an equally-sized random set of devices through a statistical sampling method called bootstrapping. This baseline can be either regional or national level—say baselining visitors to a study location to the average American, or the average Bostonian, or the average Californian.

The location affinity report returns results at a sub-sector level (Big Box, Casual Dining, Fitness/Wellness) and at a category level (Best Buy, Home Depot, Baja Fresh, Gold’s Gym, etc). Each of these levels contain two files, an expanded report and a summary report.

The summary report contains just the statistically significant results, while the expanded report contains all results. A full discussion of the methodology and how to interpret results can be found in the Location Affinity white paper which can be requested from your UberMedia representative.

GeoSocial Affinity Report

UberMedia has partnered with [Spatial.ai](#) to provide a GeoSocial affinity report which provides a profile of the social media segments of your actual visitors. These profiles can be impactful for marketing and merchandising use cases. A full white paper which delves into the detailed methodology and interpretation of the report outputs is available from your UberMedia representative. The report offers the following information:

Polygon Name	GeoSocial Segment Name	Market Baseline Social Score	Market Baseline Score Rank	Weighted Raw Social Score	WRSS Rank	Social Score Differential	GeoSocial Affinity Score	GSA Rank
--------------	------------------------	------------------------------	----------------------------	---------------------------	-----------	---------------------------	--------------------------	----------

Location Reports

Location Reports are intended to provide more information concerning the study locations. They are derived from analyzing behavioral patterns of a location's visitors.

Optimal GeoSpace Report

The Optimal GeoSpace (OGS) report using both pathing data and CEL/CDL data to create a single-variable draw area shape.

OGS shapes have a number of applications:

- **Trade Areas:** These OGS shapes are developed using real-world mobile phone location data instead of traditional drive-time or in-store survey methods often used to determine trade areas. These single-variable shapes can be used as an input with other data to refine trade area analyses.
- **Cannibalization Study:** They can be used to evaluate cannibalization between stores or competitors.
- **Ad Targeting:** drive maximum efficiency of ad spend by targeting the true areas where a location derives its visitors/customers from.

The OGS report is delivered with the following files:

- Full and simplified geoJSON files of the OGS shapes.
- Stats file which shows results by input polygon with information like the status for each polygon, the areas in square feet and square miles of the resultant shapes.
- Summary file of results with a count of source polygons by OGS status. These status could include: could not resolve shape, success, too few devices.
- A Google Earth/.kmz file with the source polygon, full and simplified OGS shape and a heat map of the data for each study polygon.

Dwell Time Report

The dwell time report gives a location-by-location breakdown of the time that visitors spend in a location or set of locations.

Study Location	Total Visitors	Measured Inner Visits	Median Inner Dwell Time (seconds)	Measured Outer Visits	Measured Outer Dwell Time (seconds)	Average Dwell Time (minutes)
----------------	----------------	-----------------------	-----------------------------------	-----------------------	-------------------------------------	------------------------------

By default, employees are screened out of the report, yet options exist to remove this screening and look at the visit dwell times of all available visitors.

Each location in the study has a line in the report. Each line gives the total count of visitors who passed the various screening to be evaluated in the report (dwell minimums, employee screening, etc), as well as the counts of visits which pass the requirements for pathing data (must have at least one point before/after arrival in the study location within a defined interval of time.

For each location, a median inner dwell time and a median outer dwell time is presented. As well as a single metric to rule them all—the average dwell time per location.

The final row in the sheet gives the overall dwell times for averaged across all the study locations.

It is also possible to get well time on a visitor-by-visitor basis as well.

Note: some values in this report are in seconds, some in minutes.

For a complete discussion of dwell, please refer to the Dwell Time information sheet.



Appendix

Pre-Post Report Interpretation

As mentioned in the main body of this paper, the pre-post and pre-post sampled report can be tricky to interpret without guidance. The format is as follows:

Polygon ID	Hashed Device	Unix Timestamp of Visit	Pre/post-visit timestamp	Pre/post-visit latitude	Pre/post-visit longitude
------------	---------------	-------------------------	--------------------------	-------------------------	--------------------------

An example of a report could look like this:

Polygon ID	Hashed Device	Unix Timestamp of Visit	Pre/post-visit timestamp	Pre/post-visit latitude	Pre/post-visit longitude
Polygon A	Device 1	150050	150020	39.55440	98.656031
Polygon A	Device 1	150050	150000	39.55443	98.656029
Polygon A	Device 1	150050	150175	39.55601	98.656031
Polygon A	Device 1	153950	153900	39.55478	98.656004
Polygon A	Device 2	150154	150006	39.875541	98.487265
Polygon B	Device 3	150695	150716	39.546819	98.586512

In this example, during the timeframe of the report, Polygon A was visited by 2 different devices, Device 1 and Device 2. Device 1 visited Polygon A twice (timestamps 150050 and 153950). The first three rows relate to the first visit to Polygon A by Device 1. On the way to Polygon A, Device 1 was observed twice (rows 1 and 2 show timestamps lower/before the timestamp of the visit). Row 3 shows a point where Device 1 was seen once after the visit to Polygon A (pre-post timestamp higher/after the visit timestamp).

Device 2 visited Polygon A and was seen only once in the before-after window, in this example case, just one time at 150006 before the visit at 150154.