v1.3

Monthly Patterns

Patterns is a dataset of visitor and demographic aggregations available in the US only. For Patterns data in Canada, see Weekly Patterns.

Monthly patterns are available starting from January 1st, 2019.

Table of Contents

Monthly Patterns Schema

File Names: [patterns_*.csv.gz]

Column Name	Description	Type	Example
`placekey`	Unique ID tied to this POI. This ID IS NOT GUARANTEED to be persistent.	String	`222-222@222-222-222`
`parent_placekey`	If place is encompassed by a larger place (e.g. mall, airport), this lists the placekey of the parent place; otherwise `null`.	String	`223-223@222-222-222`
`location_name`	The name of the place of interest.	String	`Salinas Valley Ford Lincoln`
`street_address`	Street address of the place of interest.	String	`1100 Auto Center Circle`
`city`	The city of the point of interest.	String	`Irvine`
`region`	The state, province or county of the place of interest.	String	`CA`
`postal_code`	The postal code of the place of interest.	String	`92602`
`iso_country_code`	The 2 letter ISO 3166-1 alpha-2 country code.	String	`US`
`brand_ids`	Unique and consistent ID that represents this specific brand.	List	`BRAND_59dcabd7cd2395a2, BRAND_8310c2e3461b8b5a`
`brands`	If this POI is an instance of a larger brand that we have explicitly identified, this column will contain that brand name.	List	`Ford, Lincoln`
`date_range_start`	Start time for measurement period in ISO 8601 format of YYYY-MM-DDTHH:mm:SS±hh:mm (local time with offset from GMT). The start time will be 12 a.m. Monday in local time.	String	`2020-03-02T00:00:00-06:00`

date_range_end	End time for measurement period in ISO 8601 format of YYYY-MM-DDTHH:mm:SS±hh:mm (local time with offset from GMT). The end time will be the following Monday at 12 a.m. local time.	String	2020-03-09T00:00:00-06:00
raw_visit_counts	Number of visits in our panel to this POI during the date range.	Integer	1542
raw_visitor_counts	Number of unique visitors from our panel to this POI during the date range.	Integer	1221
visits_by_day	The number of visits to the POI each day (local time) over the covered time period.	JSON [Integer]	[33, 22, 33, 22, 33, 22, 22, 21, 23, 33, 22, 11, 44, 22, 22, 44, 11, 33, 44, 44, 44, 33, 34, 44, 22, 33, 44, 44, 34, 43, 43]
poi_cbg	The census block group the POI is located within.	String	560610112022
visitor_home_cbgs	The number of visitors to the POI from each census block group or dissemination area based on the visitor's home location.	JSON {String: Integer}	{"360610112021": 603, "460610112021": 243, "560610112021": 106, "660610112021": 87, "660610112021": 51}
visitor_home_aggregation	The number of visitors to the POI from each census tract or aggregate dissemination area based on the visitor's home location.	JSON {String: Integer}	{"17031440300": 1005, "18089021500": 522, "17197883516": 233, "17031826402": 5, "17031826301": 4, "04013115802": 4}
visitor_daytime_cbgs	The number of visitors to the POI from each census block group or dissemination area based on primary daytime location between 9 am - 5 pm.	JSON {String: Integer}	{"360610112030": 9872, "880610112021": 8441, "569610112020": 5671, "160610112041": 2296, "980610112021": 1985}
visitor_country_of_origin	The number of visitors to the POI from each country based on visitor's home country code.	JSON {String: Integer}	{"US": 98,"CA": 12}

distance_from_home	Median distance from home travelled by visitors (of visitors whose home we have identified) in meters.	Integer	1211
median_dwell	Median minimum dwell time in minutes.	Double	5
bucketed_dwell_times	The distribution of visit dwell times based on pre-specified buckets. Key is the range of dwell time in minutes and value is number of visits that were within that range.	JSON {String: Integer}	{ "<5": 40, "5-20": 22, "21-60": 45, "61-240": 3,">240": 5}
related_same_day_brand	Other brands that the visitors to this POI visited on the same day as the visit to this POI. Limited to top 20.	JSON {String: Integer}	{"mcdonalds": 7,"amc": 5,"target": 3}
related_same_momth_brand	Other brands that the visitors to this POI visited in the same month as the visit to this POI. Limited to top 20.	JSON {String: Integer}	{"mcdonalds": 7,"amc": 5,"target": 3}
popularity_by_hour	The number of visits in each hour over the course of the date range, in local time. First element in the array corresponds to the hour of midnight to 1 am, second is 1am to 2am, etc.	JSON [Integer]	[0, 0, 0, 0, 0, 0, 0, 222, 546, 444, 333, 232, 432, 564, 456, 345, 678, 434, 545, 222, 0, 0, 0, 0]
popularity_by_day	The number of visits in total on each day of the week (in local time) over the course of the date range.	JSON {String: Integer}	{"Monday": 3300,"Tuesday": 1200,"Wednesday": 898,"Thursday": 7002,"Friday": 5001,"Saturday": 5987,"Sunday": 0}
device_type	The number of visitors to the POI that are using Android vs. iOS.	JSON {String: Integer}	{"android": 6, "ios": 8}
[1] carrier_name	The number of visitors to the POI based on the wireless carrier of the device.	JSON {String: Integer}	{"Verizon": 342, "T-Mobile": 288, "AT&T": 265}
normalized_visits_by_state_scaling	raw_visit_counts scaled using Advan's best current	Float	715.08396...

	methodology for estimating actual visits.		
`normalized_visits_by_region_naics_visits`	`raw_visit_counts` divided by the `sum(raw_visit_counts)` to the `naics_code` in the same state or province during the same time period. This measures changes in the category-specific popularity of the POI over time.	Float	`0.00411...`
`normalized_visits_by_region_naics_visitors`	`raw_visit_counts` divided by the `sum(raw_visitor_counts)` to the `naics_code` in the same state or province during the same time period. This measures changes in the visits per devices that visited the same category in Advan's panel to the POI over time.	Float	`0.0127...`
`normalized_visits_by_total_visits`	`raw_visit_counts` divided by the `total_visits` in the same state or province during the same time period. This measures changes in the relative popularity of POI over time.	Float	`0.0000567...`
`normalized_visits_by_total_visitors`	`raw_visit_counts` divided by the `total_devices_seen` in the same state or province during the same time period. This measures changes in the visits per device in Advan's panel to the POI over time.	Float	`0.0000913...`

We do not report data is less than 2 visitors are observed from that group. If there are between 2 and 4 visitors this is reported as 4.

[1] carrier_name is a premium column. Please *Contact Us* for more details.

Panel Overview Data

Along with the Monthly Patterns file, we also deliver Panel Overview Data (see tables below) to help you better understand the context of the data appearing in Monthly Patterns.

Home Location Distributions by State/Census Block Group

File Names: [home_panel_summary.csv]

Column Name	Description	Type	Example
`year`	Calendar Year	Integer	`2018`
`month`	Calendar month starting from 1 as January	Integer	`1`
`state`	Lowercase abbreviation of U.S. state or territory	String	`ca`
`census_block_group`	US FIPS code for this Census block group	String	`530330080012`
`number_devices_residing`	Number of distinct devices observed with a primary nighttime location in the specified census block group.	Integer	`54481`
`number_devices_primary_daytime`	Number of distinct devices observed with a primary daytime location in the specified census block group.	Integer	`54482`

Number of Visits/Visitors by State

File Names: [visit_panel_summary.csv]

Note: Includes three rows designated ALL_US, ALL_CA, and ALL to provide total visitors seen during the week in the US, in Canada, and in both countries (might be less than sum of visitors by state due to same visitors having visits in multiple states).

Column Name	Description	Type	Example
`year`	Calendar year that the first day of the Monthly data is in	Integer	`2018`

month	Calendar month that the first day of the Monthly data is in	Integer	1
state	Lowercase abbreviation of U.S. state or territory	String	ny
num_visits	Number of point-of-interest visits observed in the specified state	Int	8900
num_unique_visitors	Number of unique visitors observed with at least 1 point-of-interest visit in the specified state	Integer	966

Normalization Stats

File Names: [normalization_stats.csv]

We also provide a file containing information on our panel for each day of the week to help with normalization. This file will have rows for each region for each day of the week.

Note: This includes the ALL_US, ALL_CA, and ALL rows similar to `visit_panel_summary.csv`

Column Name	Description	Type	Example
year	Calendar year	Integer	2019
month	Calendar month starting from 1 as January	Integer	1
day	Calendar day	Integer	1
region	When iso_country_code == US, then this is the USA state or territory. When iso_country_code == CA, then this is the Canadian Province or territory.	String	CA
total_visits	All visits we saw on the given day in local time (includes visits to POI and visits to homes)	Integer	200
total_devices_seen	Total devices in our panel which we saw on the given day with	Integer	50

	any visit in local time (POI or home visit)		
`total_home_visits`	Visits we saw on the given day in local time to the device's home geohash-7	Integer	`120`
`total_home_visitors`	Total devices we saw on the given day with at least 1 visit to the device's home geohash-7	Integer	`35`

Lastly, each delivery of Monthly Patterns contains a metadata file about the release itself.

Key Concepts

- **Visit Attribution:** we compute the visits/visitors and other metrics inside a POI using the POI's geometry. We do not apply any dwell time or any concept of "stops"; we rely on the polygon for accuracy. We have tested our data on 1,500 publicly traded tickers versus (a) top line revenue as reported from the companies and (b) credit card transaction counts on physical locations, and we have determined consistently that in the vast majority of cases filtering for dwell time reduces the signal and makes the correlation/forecasting worse.
- **Determining Home Location**: we compute a device's home/work (night/day) location by computing the time a device spent in each building in the country; then taking the most frequented building.
- **Backfills**: Backfill is when we take our most recent locations (i.e., addresses + geofences) and run our visit attribution algorithm backward in time to generate a new history of "backfilled" Patterns. Backfills are typically generated every time new Advan POIs are added (typically monthly, with the exception of August and December) on Advan POIs only. This means historical Patterns will only be present for all Advan POIs, including over 20,000 **Industrial POIs**, or any other POIs that were released on or before December 2022.

The best way to think about the two products is that the underlying visits and algorithms are the same, but they are aggregated at different timescales and delivered at difference frequencies.

Below are some differences between Monthly Patterns and Patterns:

1. Each delivery of Monthly Patterns covers one week starting Monday and ending end of day on Sunday. The data will be available three days later on Wednesday of each week, providing more frequent actionable data.
   (Note: A very early, now deprecated, version of Monthly Patterns (v1) went from Sunday to end of day Saturday and was delivered on Tuesdays)

2. We have Monthly Patterns starting from January 1, 2019. For historical beyond that, the regular Patterns file should be used.
3. In Monthly Patterns, we include a `visits_by_each_hour` column to enable you to get a more detailed view of the week.
4. Monthly Patterns does not include `popularity_by_hour` (covered by `visits_by_each_hour`) and `popularity_by_day` (covered by `visits_by_each_hour`).
5. We update our Places file monthly and start using the new file for our visits generation on the first of each calendar month. This means that if we introduce a new place on the 1st of a month and the Monthly Patterns file straddles two months, you will only see visits for the new place in those days of the week that are in the new month. This edge case only affects a tiny fraction of places each month.
6. Monthly Patterns is also the only product where Canada foot traffic data is available. This means there are also some subtle changes to how columns behave and names of columns that make it distinct from the Monthly Patterns product (e.g., the `state` column in Supplemental Files is named `region` in Monthly Patterns).

Column Name Detailed Descriptions

`street_address`

- We implement a number of steps to clean, validate and standardize `street_address`.
- You should expect `street_address` to be title-cased, consistent, and friendly for human reading. Please send us your feedback if you see otherwise.
- If you care about street addresses as much as we do, we also have more specific address columns to split out address components. These are optional and available upon request for future deliveries.
    - `primary_number`
    - `street_predirection`
    - `street_name`
    - `street_postdirection`
    - `street_suffix`

`city`

- In the US, all centroids (latitudes/longitudes) are referenced against a geospatial file of city boundaries as defined by the US Census Bureau (browse the boundaries here). In edge cases, the preferred city name in the address line reflects a pre-annexed city name, and we try our best to preserve those city names where possible.
- In Canada, city names are the output of normalized address strings from POI sources.
- In Great Britain, city names are the output of normalized address strings from POI sources, but in edge cases, we allow POIs to have a null city name as long as `region` is populated. The `region` column in Great Britain refers to county boundaries, and counties are a decent alternative to cities for geographic filtering.

region

- When `iso_country_code == US`, then this is the US state or territory.
- When `iso_country_code == CA`, then this is the Canadian Province or territory.
- When `iso_country_code == GB`, then this is the United Kingdom county.

postal_code

- When `iso_country_code == US`, then this is the US 5 digit zip code.
- When `iso_country_code == CA`, then this is the Canadian postal code in the form of a 3 digit Forward Sortation Area (FSA), a space, and the 3 digit Local Delivery Unit (LDU).
- When `iso_country_code == GB`, then this is the British postal code. Learn more about Great Britain postal code precision here.

raw_visit_counts

These are the aggregated raw counts that we see visit the POI from our panel of mobile devices. The duration of the visit must last at least 4 minutes to count as a visit to a given POI.

visits_by_day

- This is an array of visits on each day in the week, Monday through Sunday.
- We are breaking up days based on local time.

visits_by_each_hour

- This is an array of visits for each hour that exists in the week.
- Only the start of a visit is used for this column so multi-hour visits will only be counted once unless the visit crosses a UTC day boundary. Note that this is a subtle difference compared to how `popularity_by_hour` is calculated in Monthly Patterns
- We are breaking up days based on local time.

visitor_home_cbgs

- These are the home census block groups (U.S.) or dissemination areas (Canada) of the visitors to the POI.
- For each census block group, we show the number of associated *visitors* (as opposed to the number of *visits*). If visits by home cbg is desired, we recommend taking the *visitors* from each CBG and multiplying by the average visits/visitor (i.e., `raw_visit_counts` / `raw_visitor_counts`) as an approximation.
- We do not have a home census block group for each visitor and not each visitor originates from the US. The number of US visitors listed in the `visitor_country_of_origin` column represents the total number of visitors which we have determined originate from the US versus Canada.

`visitor_home_aggregation`

- This is similar to `visitor_home_cbgs` except they represent the home *census tracts* (U.S.) or *aggregate dissemination areas* (Canada) of the visitors to the POI.
- We recommend using this column when you do not need to know visitor homes areas at such a fine level (CBGs represent 600-3000 people while DAs represent 400-700), but can aggregate to the next-level-up geographic unit (CTs represent 2,500 to 8,000 people while ADAs represent 5,000 to 15,000).

`visitor_daytime_cbgs`

- These are the daytime census block groups of the visitors to the POI.
- For each census block group, we show the number of associated *visitors* (as opposed to the number of *visits*).

`visitor_country_of_origin`

- These are the countries of origin of the visitors to the POI.

`distance_from_home`

- This is the median distance from home to the POI in meters for the visitors we have identified a home location.
- This is calculated by taking the haversine distance between the visitor's home geohash-7 and the location of the POI for each visit. We then take the median of all of the home-POI distance pairs.
- If we have fewer than 5 visitors to a POI, the value will be null.
- We do not adjust for visits - each visitor is counted equally.

`median_dwell`

- This is the median of the minimum dwell times we have calculated for each of the visits to the POI.
- We determine the minimum dwell time by looking at the first and last ping we see from a device during a visit. This is a minimum dwell because it is possible the device was at the POI longer than the time of the last ping.
- It is possible to have a minimum dwell of 0 if we only saw 1 ping and determined the visit based on factors such as wifi.

`bucketed_dwell_times`

- This is a dictionary of different time spans and the number of visits that were of each duration.

- The time spans are in minutes.
- Data contains the following bins: { "<5", "5-10", "11-20", "21-60", "61-120", "121-240", ">240"}

`related_same_day_brand`

- These are the brands that the visitors to this POI also visit, on the same day that they visit the POI. The number mapped to each brand is an indicator of how highly correlated a POI is to a certain brand. The value is a simple percent of POI visitors that visited the other brand on the same day.
- Only the first 20 brands are returned.

`related_same_week_brand`

These are the brands that the visitors to this POI visit over the course of the week. Interpreted and calculated in the same way as `related_same_day_brand`.

`carrier_name`

- This is a premium column that maps wireless carrier names to the number of visitors to the POI whose device uses that wireless carrier.