Buyer's Guide

Foursquare Visits

Understanding where people are going can inform where your business' next steps forward should be when making investment decisions or deciding where to set up your next location.

This guide focuses on the value of buying our first party visit data, used for trade area analysis, site selection, demand forecasting, quantitative investing, and more.

It details the background on the product, the schema, and our data normalization process which serves as a key differentiator.

Table of Contents

Powering analysis for site selection, demand forecasting, quantitative investing & more

Foursquare Visits provides clients with a granular daily feed of all visits to individual venues using first party only data. This feed allows clients to see location data and visitation trends from the past three years across the United States and can be delivered through a S3 bucket.

The value of Visits is made possible due to a number of benefits including:

Venue-level granularity. Our data provides you with the ability to understand venue-level visits at scale based on the movement of the millions of people in our first party panel to up to six million locations (depending on the year).

Normalized data. Visits are weighted based on users' representativeness in our panel relative to their demographic's true representation and size according to the United States Census, mitigating demographic panel bias and fluctuations. This enables you to understand total visits to venues for the entire population.

Relevant analysis. We will showcase foot traffic to the chains, categories, and/or markets that matter most to you.

User demographics. Visit data is accompanied with age, gender, and home zip allowing you to learn more about users visiting places and any changes in demographics and catchment area over time.

People plus Places. Places, our robust mapping file, is included to enable trend analyses of both people and places.

Privacy first. Scale and accuracy doesn't come at the cost of the user data privacy. We will always protect our users' data and work within our clients' compliance needs by filtering out home/work visits. We also provide hourly granularity while protecting consumers' privacy by obfuscating the exact time of the visit.

Analyzing Trends

We recommend indexing normalized visits in order to compare trends across different point-of-interest (POI) aggregations (venues, chains, categories). This allows for comparison of normalized visits trending on the same scale, so that time series can be easily compared. This is particularly useful when comparing growth or decline across categories, competitive sets, etc. and understanding the relative impact of external forces on a given set of POI.

Here are a few examples of indexing:

- **Monthly indices.** At a given level of POI ontology, index summed monthly norm visits to a single baseline month. For instance, if January is selected as a baseline month, the January index would begin as 100 (January normalized visits / January normalized visits * 100). A subsequent February index would then be February normalized visits / January normalized visits * 100, and so one for future months. Once done for each category or chain being compared, the indexed trending can be illustrated on a single line graph.
- **Rolling weekly indexing.** To derive smoothed weekly indices, rolling 7-day sums can be derived. So for a given date in the feed, sum norm visits for that day and all six preceding days. Each of these rolling weeks can then be indexed to a single baseline week as determined by the user, following the logic above. For example, since COVID-19, Foursquare has been indexing different categories and chains to a pre-COVID week in mid-February 2020.
- **YoY index differences.** To compare relative growth year-over-year, differences in indices can be useful. For instance, following the monthly indices guidance above, one could derive two sets of indices for 2020 and 2019 respectively and compare the point difference of those indices to see whether relative growth has changed significantly YoY. Say, for instance, a February 2020 index for a given category is 110 (Feb 2020 normalized visits / Jan 2020 normalized visits * 100), and the February 2019 index is 100 (Feb 2019 normalized visits / Jan 2019 normalized visits * 100), the relative growth is 10, or 10%.

Data Delivery

- Daily, weekly or monthly delivery with a 24 hour lag
- Historical data available beginning January 2018
- Option to receive all visitation data or only for select DMAs, from chains or from categories
- Available in the United States only
- Sent via Foursquare-owned S3 bucket or via Amazon Data Exchange

Over the last three years, the average file size of the daily file was 215MB/day (average ranges from 60MB to 600MB). Historical One-time 3 years = 750GB (240GB compressed)

Stats for visits and visitors 2020

Total number of visits: 1.5B

Total number of visitors: 11M

Total number of unique venues: 5M

Stats for visits and visitors 2019

Total number of visits: 4.9B

Total number of visitors: 21.2M

Total number of unique venues: 6.2M

Stats for visits and visitors 2018

Total number of visits: 3.1B

Total number of visitors: 16.1M

Total number of unique venues: 5.4M

[Visits by top level category: November 1 2019 to November 1, 2020]

Category	Total Visits
Shops & Services	1,688,058,590
Food	1,044,679,060
Professional & Other Places	680,974,312
Outdoor & Recreation	518,305,520
Travel & Transport	408,560,451
Arts & Entertainment	168,271,850
Nightlife Spots	163,531,465
Residences	102,268,993
Colleges & Universities	101,751,623
Events	1,304,606

Total Visits	**4,877,706,470**

Schema

[FSQ Visits]

Column	Type	Description	Example
venueID	string	Foursquare's unique identifier for a particular venue/location; displayed as a 24-character string	5c6b234e037be1002ce54dda
utc_date	date	Date of the visit in UTC time	2019-07-24
utc_hour	number	Hour of the visit in UTC time	10
local_date	number	Date of the visit in local time	2019-07-24
local_hour	number	Hour of the visit in local hour	11
gender	string	Modeled gender of the visitor	Male or female
age	string	Modeled age cohort of the visitor	20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84
Full_Panel_Reweighted_sag_score	number	Normalization score assigned to visit, based on the user's state, age & gender (e.g. this 1 visit should be weighted as 87.6813336 visits, to derive insights representative of the overall US population)	87.6813336
dwell_time	number	Number of seconds a phone is in a venue (difference between start and end time)	193

Home zip	number	Home zip code of user, for zips with >1k in population *This data point can be optional and available in early 2021*	10016
home_CBG*	string	Indicates the census block group in which the visitor lives	482012231001
work_CBG*	string	Indicates the census block group in which the visitor works	482012231001

Foursquare may be able to provide a unique anonymous id tied to each visit depending on the client's use case and with approval from Foursquare's privacy team.

*Note: Census Block Groups will be available on a go-forward basis beginning January 2022

[Places Mapping File]

Column	Type	Description	Example
venueID	string	Foursquare's unique identifier for a particular venue/location; displayed as a 24-character string	5c6b234e037be1002ce54dda
venue_name	string	Name of the venue	Old Navy
address	number	Address of the venue	1217 Rockville Pike
city	string	City the venue is located in	Rockville
dma	number	DMA the venue is located in	Washington, DC
zip	string	Zip code the venue is located in	20852
state	string	State the venue is located in	MD
country	number	Country the venue is located in	US
geo_lat	number	Latitude coordinates of the venue	39.054581
geo_long	number	Longitude coordinates of the venue	-77.11638
level_1_cat	string	Primary category of the venue	Shops & services

level_2_cat	string	Secondary category of the venue	Clothing stores
level_3_cat	string	Tertiary category of the venue	Women's stores
chain_id	string	Foursquare's unique identifier for a particular chain; displayed as a 24-character string	556d1d36aceaff43eb0a9e60
chain_name	string	Name of the chain	Old Navy
parent_venue_id	number	The ID of the parent venue if this venue is located in an airport, shopping centre, etc.	4eb0307f6c250ddebea09895

Normalization for first party data & SAG Scores

How normalization works and why it's important

Foursquare projects 'normalized visits' based on observed visits from the millions of consumers in our always-on foot traffic panel.

We apply a weighting to each observed visit. That weighting is based on state, age, and gender and thus referred to as a 'SAG' Score. By accounting for any age, gender, and regional skews within the panel, we are able to estimate real world trends.

We recommend using normalized visits rather than 'raw' observed visits. Weightings are adjusted to account for fluctuations in the size of our panel, as well as other technical factors (such as changes to how we calculate a visit in our SDK due to an OS update) that occur periodically. Normalization will inherently de-bias our data and control for fluctuations in panel size (since the raw panel visitation in our feeds will show demographic skew and is susceptible to step changes in the scale of our panel).

Normalized visits should always be used in compiling visits trend analyses to mitigate the effects of changes to our panel user base (both positive and negative). Raw visits will show significant changes in the volume of data over time, due to a variety of factors including:

- Changes to our check-in methodology and the attributes used in our Pilgrim SDK
- Monthly active user (MAU) changes in our partners' apps due to new version releases, new features, user acquisition efforts such as press and marketing, etc.
- OS updates such as iOS 13
- Changes in our panel (e.g. new apps contributing to our panel)
- Changes in consumer behaviors related to world events like COVID-19

Example:

On a given day, we see three panel visits to a Starbucks in New York by users who are female, age 20-24 living in New York State (Cohort A), and four visits from users who are male, 35-39, also living in New York State (Cohort B).

We have 500 active panelists in Cohort A, and the United States Census tells us there are 50,000 people in that demographic. All panelists in that demographic will carry a SAG score of 100 (50,000 / 500). 200 active panelists are in Cohort B, and the Census tells us there are 100,000 in that demographic. Panelists in Cohort B will each carry a SAG score of 500.

Each of Cohort A's visits will constitute 100 population visits, and each of Cohort B's will constitute 500.

So, total normalized visit volume on that day to that Starbucks in New York would be

(100 * 3) + (500 * 4) = 2,300

Because of how SAG scores are derived, raw visits and normalized visits will not necessarily always follow the same trend at a given point in time. For instance, it is possible for normalized visits to increase or remain stable, while raw visits decrease. SAG scores not only debias our data; they scale and stabilize total normalized visit volume. For instance, with any backend panel changes, we may see an increase/decrease in raw active user volume. This is usually met with a corresponding increases/decreases in raw visit volume. Accordingly, remaining users' SAG scores will increase/decrease in order to stabilize our projected normalized visit volume and counteract the panel-specific changes in visit volume.

Best practices for monthly address-level analysis

Thresholds and aggregation

Normalized visits may show some volatility due to the nature of our normalization process. Foursquare recommends **analyzing visits by venue and month**, rather than daily or weekly metrics, for reasonably stable patterns. For example, we may see a visit from a user with a large weighting (SAG score) because that user is a part of a highly underrepresented demographic in our panel, and this can lead to occasional anomalous spikes in visits, sometimes even for venues with a large number of raw visits. This volatility is particularly evident at granular reporting levels (e.g. venue-level, daily) where visit volume may be sparser.

Minimum raw visit thresholds should be used to determine feasibility. At low raw visit volumes, normalized visit trends can become noisier and less reliable. By assessing variation in venue level patterns against overall chain and category level patterns, we can further expand the number of feasible venues with a lower raw minimum, while excluding venues with potentially volatile, anomalous monthly trending over time.

The following thresholds are derived by assessing deviations of foot traffic indices at the venue-level against broader chain-level indices, flagging venues that have unreasonably high average deviations from chain patterns over the last 12 months.

Monthly Visit Feeds*	Minimum Raw Visits	Maximum Variation
High Confidence	>75 average raw visits/ month	Raw Visits: Monthly Standard deviation < 100% of mean SAG Score: Standard Deviation <350

Medium Confidence	>30 and <=75 average raw visits/month	Raw Visits: Monthly Standard Deviation <70% of Mean SAG Score: Mean <300 & Std <300
Low Confidence	Venues that do not meet above criteria	

These thresholds are provided as a general guideline for a broad set of large chains and categories. When analyzing a specific set of venues, we may recommend applying different thresholds, depending on the vertical and geographic skews of the specific venues. As a final filter, Foursquare also recommends omitting any individual SAG scores that exceed 1,000 prior to aggregating to a venue, month level.

FAQ

Are employees' visits included?

No, both home and work visits are filtered out of our first party visits. To calculate a user's work, Foursquare looks at at least 56 days worth of visit data, focusing on the visits of long duration (at least two hours). Foursquare uses these visits to build a set of places where each user has spent a large majority of time. We then assign these location clusters as home/work, and tag visits to these locations accordingly. Foursquare differentiates home from work based on the proportion of the time of day the user has spent at said location. Typically if the user visits the location during business hours, then we tag that location as work. If the user spends long periods of time in a place outside of business hours, then we tag that location as home.

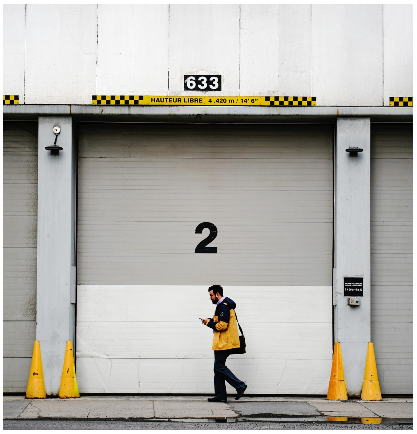
Why am I seeing visits even when we know the location is not open?

In rare cases, Foursquare data may show a visit to a store location even when the store is not open. For example, me may capture some visits to Chick-fil-A on Sundays due to individuals who are standing in close proximity of the store for several minutes, resulting in a visit being recorded (inaccurate snaps). We're consistently improving our clustering methodology to increase our visit precision and reduce false positives.

Why would a venue have dwell time but no normalized visits?

The visit may be missing a normalization weight (SAG score) because we have not been able to infer the user's demographics (state, age, and gender), which are required inputs for the SAG score. Once we can infer the user's demographics, we attach a normalization weight to their go-forward and historical visits. We are exploring ways to potentially remove visits without normalization weights. In the meantime, please disregard these visits in the dataset.

About Foursquare



Foursquare's technology is built atop the **most robust set of places and visit data in the industry.**

We're the only player in the mobile location space with a scaled, first-party user base and over 13 billion user-confirmed visits from our owned and operated apps, as well as from our partner apps, thus able to deliver a unique "phone's eye view" of millions of places around the globe.

Our proprietary multi-sensor, stop detection technology accurately detects how people move in the real world beyond the simple signals or polygons used by competitors. Context clues like time-of-day and venue popularity provide an additional layer of insight to optimize our model predictions. Our real-time feedback system confirms & continuously improves our model accuracy. Nobody else has this. **Only Foursquare.**

Data Sources

We evaluate nine billion visits to places globally each month via data from our first-party users and from our partners. Millions of users in the United States have given Foursquare consent to persistently measure their location via our owned and operated apps or the apps of partners who have integrated our SDK, Pilgrim. Additionally, knowing the visit history of a user allows us to distinguish a frequent store visitor from an infrequent one. This is the industry's largest always-on dataset. **No other company can do this at the scale that we can.**

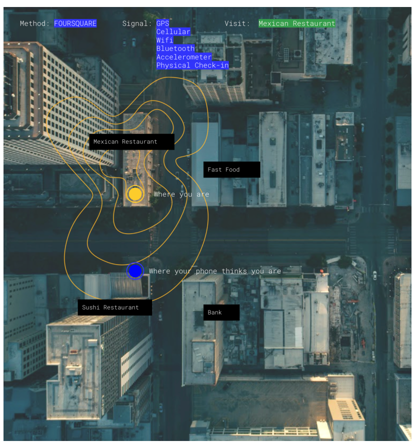
Foursquare is committed to protecting consumer privacy, ensuring that 100% user consent is obtained for all first party location data and the ways in which that data is used is always clear to the user.

Defining a visit

Our ability to measure a visit in the physical world with precision is made possible by accurate stop detection technology, actively refreshed venues data, and quality first-party data sources. Other location data companies are not able to provide this. Only Foursquare combines our crowdsourced venues database of 105 million points-of-interest (POI) and superior stop detection methodology to generate robust visit data.

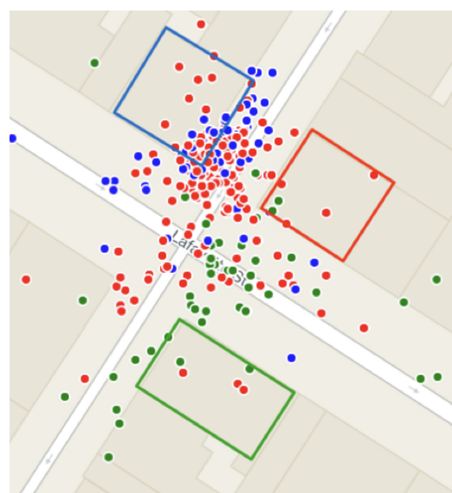
When a user enters a venue and dwells for at least two minutes, our Pilgrim technology records all of the signals available on the phone. It then matches that person to confirmed signals from our panel of 13 billion in order to register a visit.

Utilizing stop detection technology and dwell time is crucial for reporting visits because we are capturing true visits as opposed to someone driving by or sitting in traffic nearby. We utilize stop detection and entrance/exit time (dwell time), in addition to a number of other signals, in order to passively log visits at the highest degree of accuracy. Our other signals include WiFi triangulation, GPS distortion measurement, lat/long, accelerometer, barometer, compass, bluetooth, timestamp, cell tower signals and time of day popularity ratios.



Signal interference is common in dense urban locations and busy indoor environments. We also utilize place shapes to detect visits within such dense locations. We take where a person is and where their device is (based on the signals listed above), and "snap" that person to a venue based on their likelihood being there. We know the likelihood of that person being at that location because of our always-on proprietary first party panel.

To visualize this, think of place shapes as amorphous 3D shapes. It's not simply a drawn shape or polygon. Certain places have a higher probability of snapping you to that venue and thus the boundaries of the venue extend past the way you would see them on a map.



This is why we can snap the red users in the blue shape to the red shape because we know they're more likely to be in the red shape. We don't rely on GPS coordinates and we don't see a map in the way that a person sees a map.

Many of our competitors use polygons to form their POI and detect a visit. Polygons are only taking GPS and lat/long coordinates, which does not accurately represent where someone is. In an internal study we ran from February to April 2018, Foursquare determined that our snap-to-place method accurately attributed the right venue details 3x better than both polygons and venue radii (30m).

Being able to track real visits by collecting a multitude of signals differentiates our SDK from competitors' who rely on limited or singular indications.

Only Foursquare uses duration and rate of movement to distinguish stops from passers by with accuracy. Pilgrim's stop detection and multi-sensor data precisely ensure that a user is snapped to the right location.