BAD GEO
IS THE NEW BOT

Understanding Location Accuracy
for Data-driven Marketing

FOURSQUARE

Location-based targeting is in its infancy, with mostly adequate solutions for reaching people based on online behaviors and regional proximity. Cookies allowed advertisers to understand online browsing behavior, and to use behavior insights to better target consumers. If someone spends time on fashion sites, advertisers can target them with specific retailer ads as they browse the rest of the internet. While Big Tech leaders have allowed marketers to understand how consumers move through the internet to better personalize messaging and smooth the path to purchase for e-commerce, the same has never been true of the offline world.

If you know an individual has visited luxury car dealerships, big box electronics stores, or upscale supermarkets, you can target them with relevant ads. Nearly ninety-three percent of commerce happens offline, but these real-world visits and behaviors have not helped inform ad targeting. The places people visit are stronger indicators of who they are than the websites they visit, and should be a huge targeting opportunity for marketers. Building behavioral location segments is the cookie of the real world.

Two factors have made this level of targeting possible. First, there is the ubiquity of smartphones and sensors. The iPhone launch just eight years ago, and subsequent rise of iOS and Android mobile applications, have embedded mobile into nearly all facets of everyday life. Second, the advent of location-based services facilitates a mapping of those locations to specific businesses. These two developments, layered together, allow for an unprecedented view into consumers' real world mobility patterns. It provides a new window into what people love, hate; what individuals hope for, and what communities value.

7 billion check-ins

250,000,000 photos

70,000,000 tips

90,000,000 "tastes"

From Foursquare's view of the world, your 7 billion check-ins have taught us about more than 65,000,000 places that exist in the world. The 250,000,000 photos and 70,000,000 tips you've left have shown us what's great inside those places. The 90,000,000 "tastes" you've added to your profiles (since the launch of Foursquare 8.0) have taught us about what matters to you and what you want us to help you find. Beyond Foursquare and Swarm exists an ecosystem of app partners making hundreds of millions of daily Foursquare API calls to contribute to the freshness and accuracy of our dataset.

For years, we've been asked to bring our location technology outside of the Foursquare domain. And, after spending the the last couple years honing the technology, we can now allow high-precision matching to behavioral location segments for targeting anywhere.

We're excited to introduce Pinpoint, the world's most accurate location advertising intelligence available.

However, the focus of this paper is not Pinpoint, nor is it solely about Foursquare. While developing this product, we discovered a number of systemic discrepancies in the advertising ecosystem which we will detail at a high level in the following pages.

IN THIS WHITE PAPER, YOU'LL LEARN ABOUT...

- **The challenges and misperceptions of current location technology**

- **Our internal approach for assessing location data**

- **Which questions to ask when evaluating data partners**

WHY LOCAL IS HARD

Indoor location, offline analytics and proximity or place-based ad targeting are semi-independent developments all connected or enabled by the adoption of smartphones. They are not part of a single, coherent market. Nonetheless, the convergence of these trends will have a direct or indirect impact on hardware spending and software licensing, merchant operations, "shopper marketing," loyalty programs, email marketing and digital advertising as a whole.

- Mapping the Indoor Marketing Opportunity,
  Opus Research Inc.

As location becomes an increasingly important part of marketing strategies, dozens of ad tech companies, new content platforms, and established media players are looking for ways to incorporate location signals into ad targeting, creative, performance, and delivery. This enthusiastic adoption of emerging technology has created a crowded environment of "me too" solutions, rather than strategic innovation leveraging proprietary data sets. Although there are a number of players with partial pictures and potentially interesting ways of analyzing the world, most do not offer a comprehensive solution of local information and actionable conclusions for marketers.

To contemporize John Wanamaker's famous dilemma, "I know half of my location advertising budget is wasted...I just don't know which half." For the most part, the media community knows that not all location data sources are created equal; but in the nascent days of geo-based solutions, it is hard to separate the wheat from the chaff.

To an extent, the dilemma that advertisers face with location data comes down to two core factors:

1 **There are fundamental limitations to hardware's location accuracy.**

2 **The mid to long tail of businesses are diverse, densely populated, and subject to high rates of turnover.**

> I know half of my *location* advertising budget is wasted... I just don't know which half."

Close Only Counts in
Horseshoes & Hand Grenades

It is difficult to get an accurate location fix on a mobile device. Why? Because most modern devices use a combination of GPS, WiFi, and cell tower positioning; and thus, under different circumstances, can provide location fixes of varying degrees of accuracy. Most people think about a "location fix" as a single point in space, but a location fix is better described as a spatial region of varying probability — an estimation of latitude and longitude.

Often times, this region has a radius greater than 50 meters, which equates to about 85,000 square feet or roughly 1½ NFL football fields. Imagine how many different stores can fit into a football field. Now, imagine the number of different businesses and establishments that could occupy that same amount of space in an urban environment, and you will realize how imprecise these location fixes can be in practice. It's easy to find a big box retail store on a map, but the challenge in unlocking the potential of location is in understanding the more granular data. The important qualitative data associated with a place, beyond latitude and longitude, is increasingly difficult to discern when you're in a densely populated city.
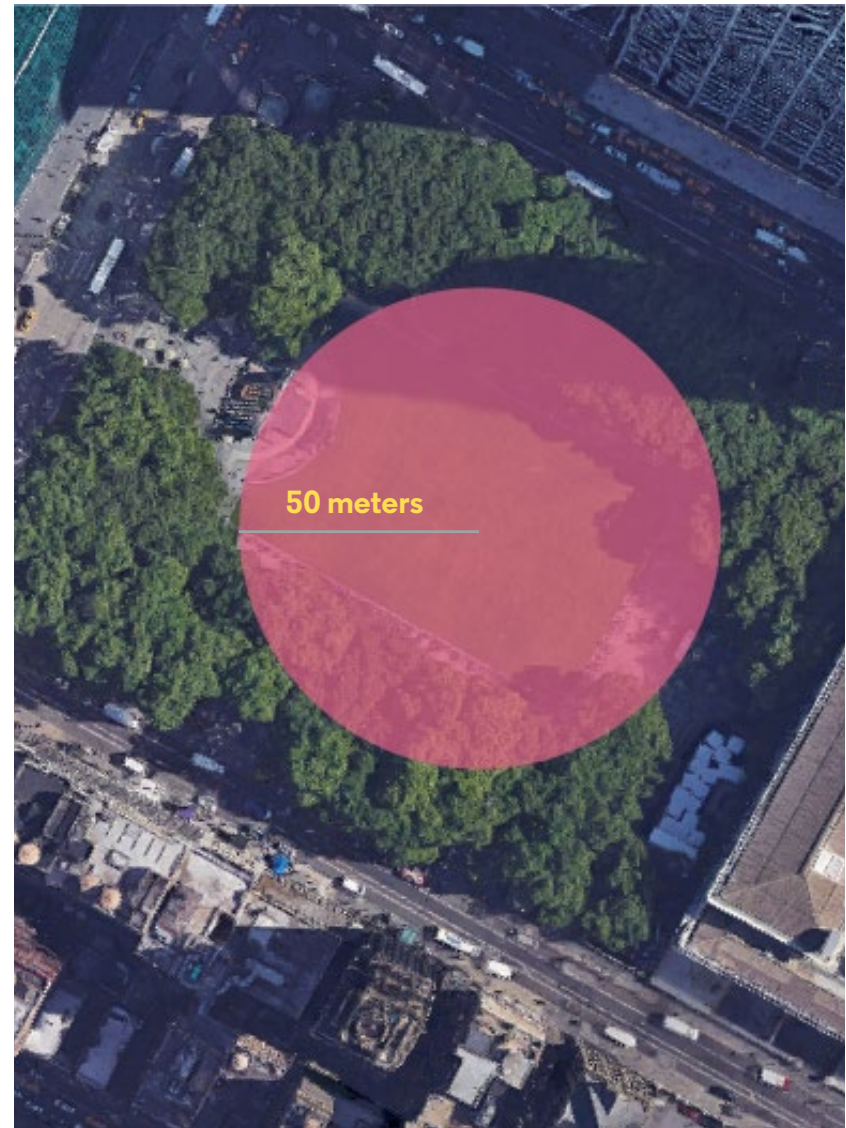
Despite your device being firmly in one store, you may get a location fix suggesting that you are across the street or at the business next door — particularly while visiting smaller businesses. For this reason, it's important to carefully model the location fixes, accuracy of these fixes, as well as the raw sensor readings like WiFi strength.

The world isn't flat. Being bound by lat/long and two-dimensional aerial mapping results in an incomplete representation of physical spaces. Most think of location as a set of X and Y coordinates — latitude and longitude. A few might even attempt to understand the height of physical structures as a Z-axis. Foursquare's unique advantage goes far beyond, incorporating dozens of dimensions, including time, popularity, category, wifi, signal strength, accuracy, and more. This allows us to learn how mobile phones view the world. These comprehensive spatial models provide a significant improvement in accuracy when compared to traditional techniques, which include measuring the distance to the street address associated with the venue, polygon representations of venues, or grid-cell approaches.

Moving at the Speed of Local

As businesses change, the composition of the world changes, too. Turnover proposes a challenge in getting an accurate read of where local businesses are and what their location looks like. The challenge to date has resided mainly in the fragmented data sources which inherently have a lag time baked into their infrastructure. With millions of businesses evolving, opening, and closing every year, relying on data inputs from various sources caused decreased vendor accountability and an increased risk of inaccuracy.

Foursquare's consumer foundation puts us in a unique position to overcome the hurdles in location data. We have billions of user-verified signals, as well as trillions of passive location signals, where time and again a person or their mobile device told us where they were. With massive amounts of historical data to match against, you can begin to smooth out the statistical variation seen in GPS technology. Secondly, our user base has crowdsourced 65 million unique places. Each place has its own digital thumbprint that is continuously updated and verified from our user base and the millions of Foursquare API calls from partners like Twitter, Pinterest, Waze, Microsoft, and 85,000 other apps and companies.

Unique ID

Latitude - Longitude

Time Stamp

IP Address

Publisher

Sent from thousands of publisher partners and mobile exchanges

A SYSTEM OF CHECK-INS & BALANCES

To build a machine model, you need tons of data and that data needs to be of good quality. Foursquare has it; and because we can extract meaningful signals from the data, we are able to tell actual locations vs. approximate locations.

– David Sinsky, Product Manager, Foursquare Labs

The development of Foursquare's Pinpoint ad technology requires a large-scale engineering effort to filter and make sense of roughly a billion geo-tagged ad impressions per day. The goal was to create scalable audience profiles based on our core belief that the places you go are the best indicator of who you are. Although it was not our intention at the outset, we discovered systemic data discrepancies during this process which are summarized in the following.

Our Approach

In order to prioritize the quality of the ad impression sources, our engineering team evaluated and categorized publisher sources into three groups: Whitelist, Greylist and Blacklist. As the names imply, these categories are a reflection of the quality of the location data we are receiving from the inventory sources.

To determine these classifications, we first needed to make sense of the geo-coordinates associated with an ad call. We had a unique advantage during this process because our two consumer-facing apps and network API partners provide a wealth of first-party location data that is directly linked to the sensor readings from the phone. Location is core to our consumer value proposition, so a focus on GPS accuracy, refinement, and maintenance sit at the heart of our team's daily responsibilities.

By running a verification analysis using Foursquare and Swarm data as a "panel," we were able to determine the veracity of the location data received from publisher sources. Although we called this process the "user panel" internally, it is more accurately characterized by the term First-Party Location Quality Assurance (QA). This unique method of QA helped prioritize sources of data for the next phase of Pinpoint audience mapping, namely logging device IDs against commercial places.

This process is not how we build user segments, but rather how we determine which sources have the highest likelihood of quality data, to which we can map against our 65 million user-generated venues.
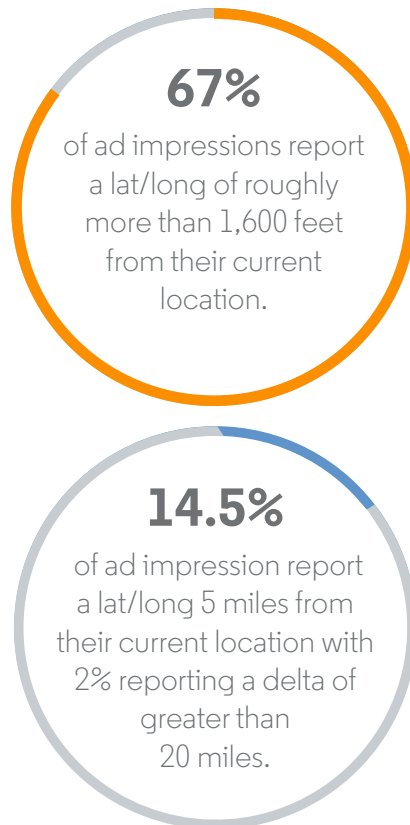
Our methodology is a simple one, and does not require complex equations or any inference. We matched Foursquare and Swarm device IDs with the exchange device IDs associated with a geo-tagged ad impression. When these two data points overlapped, we were able to compare our first-party location data during the time the ad impression fired and the reported location from the exchange impression.

Looking for these overlapping sets of data is certainly the most transparent methodology in our view; but it did seem like it would limit the sample size pretty dramatically. However, we started with a very large pool of data; so at the time this study went to print, our sample to run this Location QA was already 49,881,676 instances where we could do an apples-to-apples lat/long comparison. For our purposes, let's just say that N = 50 million. Using this as our baseline sample for evaluation, we began to sort publishers based on the difference in distance between what we record and what is passed as the user's location.
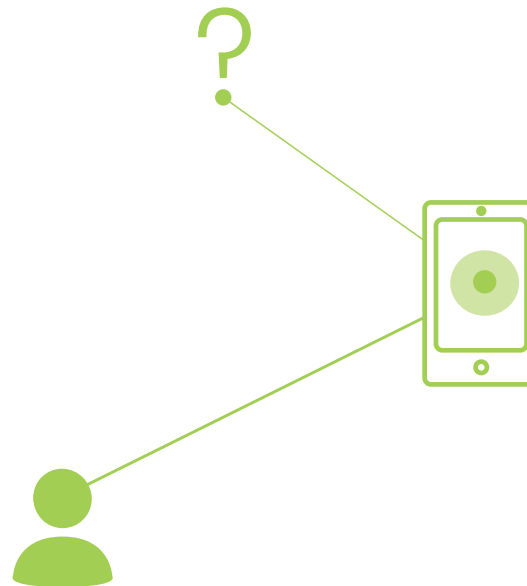
LAT/LONG DATA

Top Level Findings

BLACKLIST PUBLISHERS

67%

of ad impressions report a lat/long of roughly more than 1,600 feet from their current location.

14.5%

of ad impression report a lat/long 5 miles from their current location with 2% reporting a delta of greater than 20 miles.
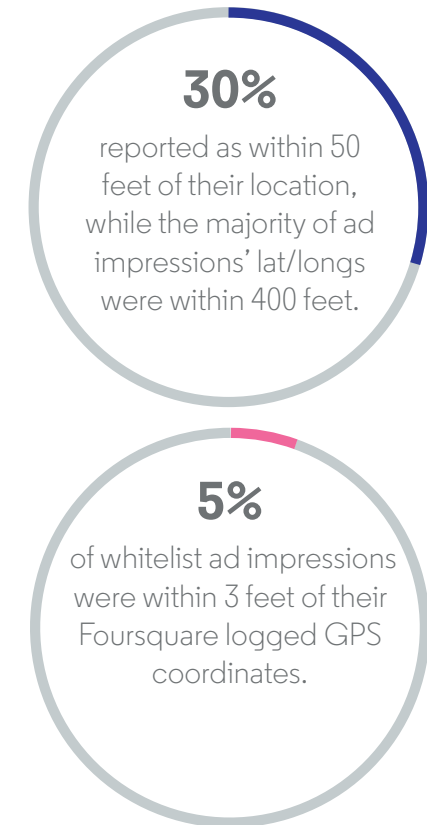
A small (but still significant) number of impressions from the sample (roughly 50,000) reported a lat/long pairing more than 100 miles from their actual location.

WHITELIST PUBLISHERS

30%

reported as within 50 feet of their location, while the majority of ad impressions' lat/longs were within 400 feet.

5%

of whitelist ad impressions were within 3 feet of their Foursquare logged GPS coordinates.

This level of accuracy only occurred for two-thirds of one percent (.66%) for the publishers on the blacklist. Note that this is a discrepancy of 750%.

Marketers are becoming increasingly familiar with mobile location technologies; however a firm understanding of how to unlock the value of mobile remains elusive for many.

Location targeting shouldn't be reduced to simply targeting a zip-code or DMA or implementing a new, trendy strategy or technology.

The best mobile location ad campaigns combine raw location data with strong insights into how mobile is used and its relevance in the lives of consumers to deliver compelling and engaging experiences.

- Mobile Location Use Cases and Case Studies, IAB
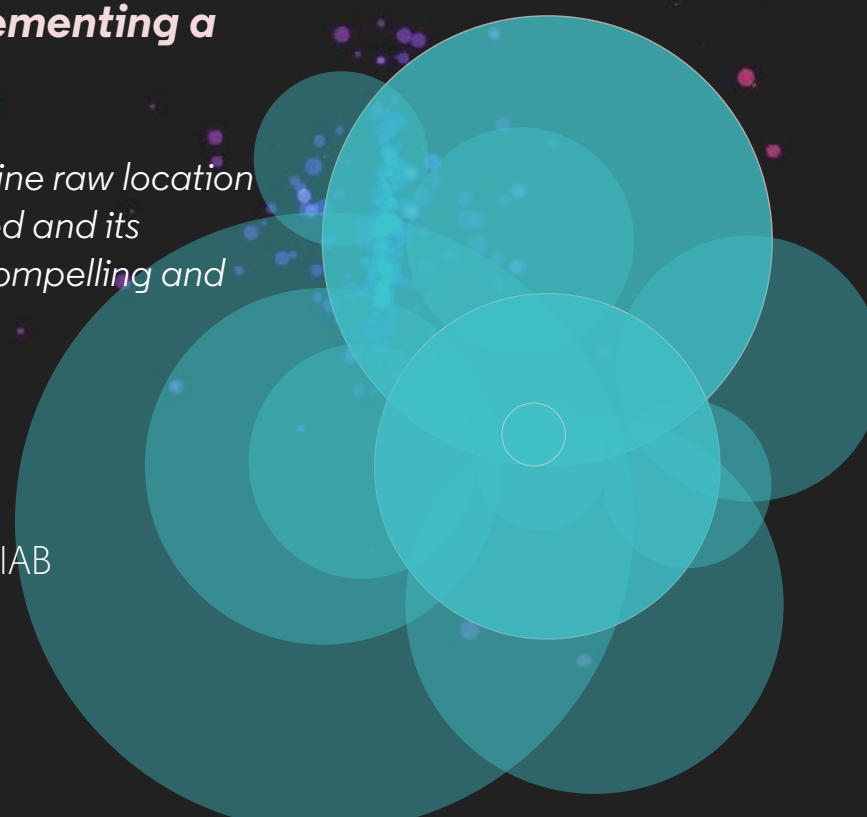
WHITELIST % Closeness to Lat/Long Actual	BLACKLIST % Closeness to Lat/Long Actual	Meters	Miles	Feet	% Discrepancy
3.06%	0.00%	0	0.000	0	..
5.61%	0.66%	1	0.001	3	750%
10.93%	1.29%	3	0.002	10	747%
19.77%	2.14%	7	0.004	23	824%
30.73%	4.32%	15	0.009	49	611%
40.24%	7.36%	31	0.019	102	447%
45.96%	11.82%	63	0.039	207	289%
50.07%	14.83%	127	0.079	417	238%
55.13%	20.17%	255	0.158	837	173%
61.50%	32.93%	511	0.318	1,677	87%
69.28%	52.36%	1,000	0.621	3,281	32%
77.70%	66.29%	2,000	1.243	6,562	17%
85.55%	76.60%	4,000	2.485	13,123	12%
91.47%	85.54%	8,000	4.971	26,247	7%
95.64%	92.54%	16,000	9.942	52,493	3%
98.04%	97.03%	33,000	20.505	108,268	1%

Bad Geo is the New Bot

Location inaccuracy will become the next frontier for the industry to address with the same vigor that it's addressing click fraud and viewability. If you're relying on location data for an accurate understanding of who people are and how to message to them, the findings above should elicit the same reaction the industry is currently having about wasted impressions and artificial inflation of campaign numbers.

There will always be a margin of error in datasets this large, but the current state of affairs requires industry attention and action to correct rampant discrepancies and, in some cases, outright deception.

LOCAL KNOWLEDGE

Connecting these dots between offline and online, physical and digital, isn't simply about Big Data. It's about smart data.

– Josh Engroff, MediaPost

The MMA recently determined that, "for location data to be meaningful, it needs to be tied to specific places…The quality of a location-based audience profile is also dependent on the quality of the underlying place's data." Given the commoditization of data sources in the marketplace and data discrepancies from these vendors, how can advertisers ensure that they are leveraging quality data to reach the right people?

Recognizing that advertisers must become "fluent in the language of location," the IAB's Mobile Marketing Center of Excellence released *12 Questions Every Buyer Should Ask About Location Data.* These questions help marketers to ask the right questions — but what answers should advertisers expect?

Based on the attributes we look for in our data partners, and standards to which we hold ourselves accountable, we've answered these twelve questions with the responses that advertisers should expect from vendors who are capable of delivering the best-in-class media targeting options.

> " for location data to be meaningful, it needs to be tied to specific places... **The quality of a location-based audience profile is also dependent on the quality of the underlying place's data.** "

What is the source of your "place" information (i.e., data about what businesses, points of interest or addresses are found at specific lat/longs in the physical world)?

The Foursquare Places database, which includes 65 million verified venues, was created over the last 6 years from 7 billion explicit visits to locations. It's maintained and sharpened by our millions of users and 85,000 API partner apps and developers.

What is your overall share of first versus third-party place data? Do you have a proprietary mapping system/address data or utilize a third-party database?

Foursquare has its own proprietary place data, and uses Mapbox which is built on top of OpenStreetMaps for mapping services.

What is your approach to organizing places/place data (e.g., polygons, geo-fence radii, etc.)?

Foursquare uses its own proprietary "shapes" for each of the 65 million places in our database, created by observing the

mobile devices that visit these locations. We use many factors, including the "WiFi fingerprint" to locate places by the strength of the observed access points.

How precise is your place information (e.g., are you able to discern the location of a specific store in a mall versus the parking lot)?

Foursquare is able to determine the X and Y accuracy of places on maps, as well as the Z accuracy of places in a building using WiFi signals. This data can be used granularly to define visits to different stores in a mall or on a city block.

How comprehensive is your place information (e.g., What percentage of business, addresses, or points of interest do you have place data for)?

Foursquare has over 65 million commercially viable places mapped around the world. Only Foursquare is able to identify venues within malls, airports, and dense city blocks, and more.

How do you qualify and/or verify your place information? For example, how do you address the scenarios below:

A. How do you compensate for bad addresses?

Foursquare has a comprehensive system in place for accurately determining correct addresses, as well as fixing information that needs to be updated. Through our consumer apps, users notify us if information needs to be corrected. Through partnerships with other companies we can continue to train our system to be accurate. We also look at the shapefiles of places to determine actual information of a location versus a postal address.

B. Do you have a way to account for recent opening and closing of locations?

Foursquare has a system for consumers to "flag" a place to identify if it is open or closed. We also have hours of operation and other data on each venue to help consumers find the right information.

What are your sources of device location data (i.e., data used to locate a device) and how do you receive that data? E.g., first-party (O&O properties/servers), direct third-party deal/relationship, impression/exchange data, etc.

Foursquare sources data from our owned and operated properties on the web, our published apps on multiple operating systems, and from partnerships with other 85,000+ developers (Twitter, Pinterest, Waze, etc.), applications and mobile exchanges.

Foursquare also has third-party relationships with leading app developers, audience data companies, and ad tech platforms to facilitate its advertising offerings.

How do you identify and filter out the types of targetable location data that are not appropriately accurate for my campaign's needs?

Foursquare uses its own proprietary methodology to ensure the right audience at the right time at the right place. We use social signals from friend activity at a place, time of day, proximity to a venue, your own activity at that location, the place's hour of operation, and the overall proclivity to match the place to the needs and objectives of the campaign at hand.

What types of device location data do you use? E.g., device GPS, cell tower/triangulation, user-reported (check-in), user-reported (registration), WiFi, IPS, beacons, low power Bluetooth, zip - local content, centroids, NFC, etc.

Foursquare uses a multitude of mobile signals, including but not limited to:

1. GPS
2. Carrier tower information
3. Latitude
4. Longitude
5. IP address
6. Horizontal Accuracy
7. Vertical Accuracy
8. WiFi
9. User reported profile data
10. Third-party reported profile data

How long is your location data stored/considered relevant?

A. Is your device data time stamped?

All data processed by Foursquare has a timestamp. From this, we can identify whether someone is in a bar or a pharmacy in the same building by looking at the timestamp and identifying hours of operation of the businesses as well as typical visitation behavior.

B. If you offer dwell times, how are these calculated?

Foursquare is able to look at dwell times of its user base using passive location tracking on the device's GPS signals.

How do you verify/substantiate that the device location data you are using is accurate?

Foursquare's 7 billion explicit check-ins are both a large-scale source of first-party location data but also a proprietary "golden training set" to train our best-in-class background (ie. passive) location detection algorithms.

What is the scale of your device location data?

Over 800 billion WiFi scans of venues collected worldwide. We have over 65 million locations mapped across the world and 20 million US devices in the Pinpoint audience. We see over two billion location trails every day.

CONCLUSION

Gone are the days of mass-marketing. To effectively connect with mobile users, we have to understand the places, moments and circumstances that trigger the desire or need for an advertiser's products or services.

- Mobile Location Use Cases and Case Studies, IAB
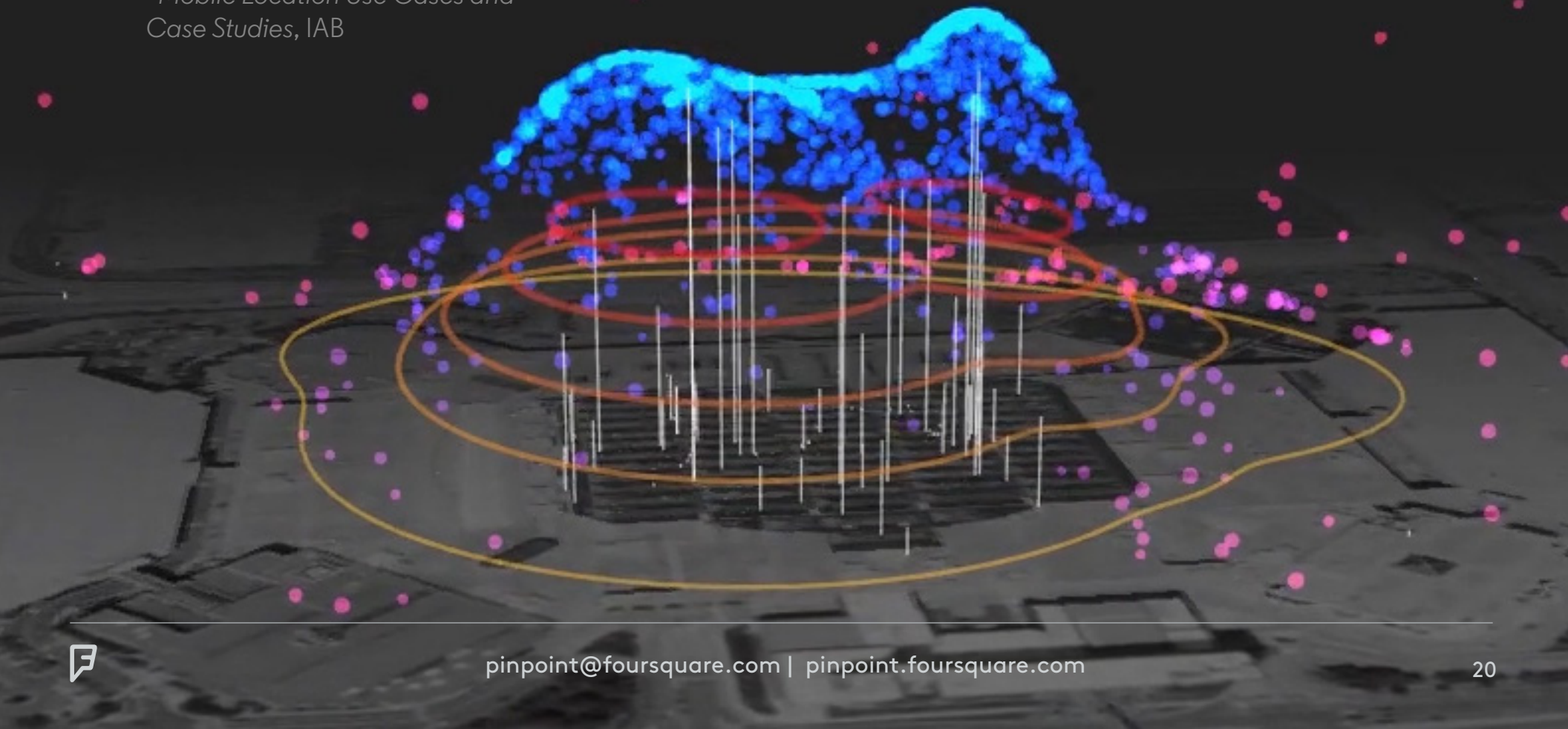
There's a reason that we're one of the only companies doing proactive and predictive local search and firing off contextual notifications — it's hard. We believe the places you go are the best indicator of who you are, and we've spent six years developing the most powerful tools to understand location data and the places consumers go. In order to provide you with the most accurate, relevant data available, we work with 85,000 publishers and app developers. It takes a village — this ecosystem of app developers, publishers, consumers, data scientists — to make the world a more familiar, personalized place.

The ability to target consumers based on their habits and movements throughout the physical world, rather than their browsing history, brings a new level of relevancy and effectiveness to advertising. In 2015, brands must take a closer look at how they use location data to deliver the right messaging to the right people. BIA/Kelsey projects by 2018, more than half (52%) of all mobile ads will be location-targeted, creating an ad opportunity of more than $15.7 billion. The bottom line? Brands that connect the dots between location, behavior, and intent will win because this type of experience is quickly becoming the UX standard for a great consumer experience. The future of mobile and local is predictive, proactive and personalized. This is the first inning in a very big industry.

To learn more about location-driven
behavioral targeting or Pinpoint, contact us:
pinpoint@foursquare.com

"Quarterly Retail E-Commerce Sales, 4th Quarter 2014."
U.S. Census Bureau News. U.S. Department of Commerce, 17 Feb. 2015.
Web. <http://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf>.

"Using Location for Audience Targeting."
MMA Documents. MMA NA Location Committee Audience Working Group, 1 Feb. 2015.
Web. <http://mmaglobal.com/documents/leveraging-location-audience-data>.

"Using Location for Audience Targeting."
12 Questions Every Buyer Should Ask About Location Data. IAB Mobile Marketing
Center of Excellence, 1 July 2014.
Web. <http://www.iab.net/locationdataguide>.

"Think Global, Target Local: Multi-Location Advertising in the Mobile Age."
Research & Analysis: Insight Paper. BIA/Kelsey, Jan. 2015.
Web. <http://www.biakelsey.com/Research-and-Analysis/Coverage-Areas/Mobile/
insight-report-multilocation-advertising-mobileagethankyou.asp>.