Framework for Evaluating the Accuracy of Location Data Technologies

Featuring proprietary research and data from Euclid

January 2018





Executive Summary

Location data is becoming an increasingly important data set for retailers seeking to better understand their customers and engage with them on a personalized level. As younger generations of customers rise through the ranks of purchasing power, there is a growing expectation on the part of consumers that retailers leverage location data to deliver a better overall brand experience. In fact, location-based advertising has been proven to be 20 times more effective than generic ads¹.

As the value of location data builds both in the eyes of brands and consumers, it is important to understand the landscape of technologies available and - perhaps more importantly - the benefits and limitations of each technology. In this paper, we will introduce key parameters for understanding these data and evaluate each of the major technologies on each parameter.

The Framework

To help evaluate location data technologies, it is helpful to have a framework to consistently measure against. There are three main attributes that are particularly important to keep in mind for this data:



Precision

How precise is the location the dataset provides? Will you be able to know the individual's exact location? Or know they are in one building or store versus the one next door?



Frequency

How consistently is this dataset captured? How frequent are events captured from this technology? Is there a high enough probability of capturing a store visit?



Scale

Is this dataset scalable? Can it reliably identify a large enough number of store visits to deliver a return on the investment?

We will evaluate three main categories of data using this framework:

GPS Technology

Extracted from mobile applications that collect latitude/longitude via GPS, we take a look at SDK and Bid-stream datasets as the two primary GPS-based location data sources.

Beacon Technology

Beacons use Bluetooth technology to detect the proximity of a mobile device with the corresponding, beacon-powered mobile application installed.

Wi-Fi Technology

Wi-Fi technologies collect device ID's pinging the network in a location and tie them to identity when the visitor logs in or with transaction time stamp matching.

1. Future of Location Based Marketing, Entrepreneur, 2016



GPS Technology

Precision

SDK datasets have variable precision. Typically, 50% or less of data from the leading SDK datasets have precision that is within 20 meters. This makes it extremely difficult to understand which store the shopper is in when in dense shopping environments, such as shopping malls or urban centers.

	<10m	<20m	<50m	<100m
SDK A		55.9%		93.6%
SDK B		50%		75%
SDK C	21%		36%	64%

Cumulative Distribution of Dataset Precision (Advertised)

Bid-stream datasets have low precision. Estimates are that **more than 90% of the bid-stream have inaccurate location data**². These are also highly dependent on publisher and user enablement of GPS, which each have a high effect on the precision capabilities.

Developers also used different strategies and direct the OS to use more coarse or granular methods. Therefore, even when "Location Services" is enabled and a lat/long fix is retrieved, it may not represent an accurate or recent location. iOS 11 has also reduced the persistence of these data sets because consumers must actively choose location permissions and may choose to only send location data when the app is 'In Use'.

Frequency

Below is the advertised distribution of events/device/day across a number of SDK location datasets, as well as Euclid's independent analysis:

Ping Frequency Findings - SDK Location Dataset

	<10 events/day	11-100 events/day	>100 events/day
SDK A	~40%	~42%	18%
SDK B	28%	31%	41%
SDK C	35%	62%	3%

Distribution of Events/Device/Day (Advertised)

2. Less than 10% of bid-stream location data is high-quality - and we know how to find it, Medium, 2017



Number of Days in Month That Device is been (Actual)				
	<5	5-10	11-20	21-30
SDK B	50%	15%	14%	21%
SDK C	50%	18%	16%	17%

Number of Days in Month That Device Is Seen (Actual)

Distribution of Events / Device / Day (Actual)

	1 event/day	<10 events/day	11-100 events/day	>100 events/day
SDK B	12%	34%	52%	2%
SDK C	25%	51%	22%	2%

There is a meaningful difference between what SDK datasets advertise as their ping frequency, and what our data science team found. **Based on our findings, a significant portion (46%-76%) of SDK datasets capture <10 events/device/day.** This leads to a low probability that an individual visit is captured at a particular location.

Scale

There is a meaningful delta between advertised scale and relevant scale. What matters is the scale of MAIDs that are of high precision, and that are seen with a high level of frequency.

	Advertised Scale	Relevant Scale		
Bid-stream				
Bid-stream Location Data	200 MM devices/20 billion daily bid requests	90%+ of data contains incomplete, inaccurate, or illegitimate data1		
SDK				
SDK A	45 MM uniques/13 MM DAUs	~7.3 MM DAUs w/ precision <20 m, and ~2.3 MM DAUs w/ >100 events/device/day*		
SDK B	15 MM uniques/6 MM DAUs	~3 MM DAUs w/ precision <20 m, and ~0.1 MM DAUs w/ >100 events/device/day*		
SDK C	11 MM uniques/4 MM DAUs	~0.8 MM-1.4 MM DAUs w/ precision <10-50 m, and ~0.1 MM DAUs w/ >100 events/device/day*		

* SDK A data based on advertised ping frequency. SDK B/SDK C data is based on verified ping frequency.



Beacon Technology

Precision

Beacon technologies have the highest amount of precision among all location data sets today because the data is based on the Bluetooth signal strength from a device nearby a beacon. Most beacon technologies tout a location accuracy between 1 and 4 meters, but the precision ultimately depends on how close the device is to the beacon³.



Frequency

Beacon technology frequency depends on the mobile device manufacturing specs, which set how frequently it scans for Bluetooth. Most devices scan for Bluetooth every 2-5 seconds thus providing a highly consistent reading on whether a device is or is not within range of the beacon⁴.



Scale

Scale is the difficult factor for beacon technologies to achieve, because it is entirely dependent on the customer having a specific app installed on their mobile device. Therefore mobile app penetration for the brand is the limiting factor. Although retail app penetration is increasing over time, it is often a low percentage to rely on for location data.



Number of Mobile Retail Apps that US Smartphone Owners Currently Have on Their Phone⁵

3. How many apps do Americans keep on their phones on average? Not that many, really, PhoneArena, 2017

5. Shoppers Are Downloading More Mobile Retail Apps, eMarketer, 2016



^{4.} Beacon FAQ, Locatify, 2015

Wi-Fi Technology

Precision

Wi-Fi-based technologies' precision is almost entirely dependent on the Wi-Fi network that it is built on. Wi-Fi signal pings typically provide location granularity within a range of 5-10 meters. This granularity enables a brand to understand if someone is or is not in their physical location, but not necessarily information on where they are within a location.



Frequency

How frequently a device can be detected by Wi-Fi is determined by the device specifications and how often it does a probe request to look for available networks. Because these are very frequent, this dataset has a high probability of identifying a device within the range of the Wi-Fi network.

	ction Every 60 - 90 Seconds
Events / Device / Day ~500	e / Day ~500

*Aggregated from data across Euclid's customer base

Scale

Scalability for Wi-Fi technology is dependent on the size of the overall network. The larger the network, the more likely a device can be identified by either matching against transactions or through a passive visit.

Number of Unique Devices (Total)	>800M
Number of Unique Devices (per Month)	>120M

*Devices within Euclid's network

6. Survey: Consumers willing to share personal data for deals, better customer service, MarketingLand, 2017



Summary

As we've illustrated, there are many factors to evaluate when exploring location data technology providers. But regardless of the technology, it is crucial to keep the customer at the center of your decision process. 43% of customers are more inclined to share personal data to receive personalized promotions or deals⁶, and this comes as no surprise because they are directly benefiting from this shared information. These location datasets are best used in mutual benefit for all parties involved.

Relevancy is an additional consideration for brands evaluating location data sets because the locations where this data is captured have an impact on consumer transparency and trust. While some technologies only collect this data in the commercial locations that actually matter to a brand, others can collect it from anywhere that an app is opened.

Ultimately, brands should be aware of the accuracy of the location data they are using to deliver ads to customers. Each has varying levels of precision, frequency, and scale - but can empower brands to engage with visitors on a more personalized level and ultimately influence the path to purchase.

Glossary

SDK - Software Development Kit. Used by third-party app developers to collect location data through an ad placed within an app on a mobile device.

Bid-stream - Provides location data through mobile device native location-based services7.

Events - We define an "event" as an identified ping containing location data from an SDK.

MAID - Mobile ad ID. An identifier provided by the mobile device's operating system which allows developers and marketers to track activity for advertising purposes. They can also be used to target and serve ads to groups of users.

DAU - Daily Active Users

PII - Personally identifying information.

7. How Location Data Providers Can Filter Out Bidstream 'Noise' For Quality Attribution, GeoMarketing, 2016

