June 23, 2022

The Honorable Elizabeth Warren
United States Senator
309 Hart Office Building
Washington, DC 20510

Dear Senator Warren:

SafeGraph appreciates the opportunity to respond to your May 17, 2022, letter (“Letter”).

We would like to start by explaining our business structure and clarifying a few key concepts misconstrued by the media. Our data products provide historical insights about places, not individual people. We strive to be the source of truth about the physical world. We do so by narrowly focusing on curating data sets with the best quality and coverage of commercial locations around the globe, making our data an open and trusted resource for all.

We are committed to enabling innovators in private industry, the public sector, and academia to solve society’s greatest challenges through the use of our high-veracity, privacy-protective data—such as responding effectively to the COVID-19 crisis—while preserving the privacy of individuals.

Unfortunately, many of the concerns articulated in your Letter reflect significant inaccuracies about SafeGraph and our products, likely originating from misstatements in certain publications that are simply wrong.

Through this letter, we will provide additional information on these and related points, including the following:

1. SafeGraph is not a “location data firm.” We are not a source of data on individual persons. Our products cannot be used for “tracking” or targeted advertising to a specific person. We simply compile and provide data about physical locations around the world. Even so, prior to receipt of your Letter, we took action in response to public concerns by removing even anonymized, aggregated data related to any type of family planning center.

2. SafeGraph does not sell data that identifies individuals, and SafeGraph’s data cannot be “de-anonymized” using any known method of re-identification. Here is why:
   a. SafeGraph does not sell any device-level or personally identifiable information, including individual mobility or geolocation data. The
information provided to our clients through our Patterns data set, which is the focus of your Letter, is the product of a statistical model that helps data scientists and researchers learn about the ways in which groups of people interact with a place without revealing any information about the individuals in those groups. In short, Patterns is designed to provide a useful form of measurement, while protecting individual privacy.

b. We utilize the industry-leading system, Differential Privacy, which allows users to understand group patterns while protecting the individual data underneath. This is the same system used by Apple, Google, Microsoft, and the U.S. Census to protect individual privacy.

c. We go even further to protect anonymity by reporting in our Patterns data set only aggregated statistics of visits to a location, which are infused with randomized noise, to further ensure that our data cannot connect back to individuals or individual devices.

3. Patterns is built from anonymized data provided by data aggregators. The data aggregators source anonymized data from opted-in devices, wherein the user has provided consent, representing a small fraction of mobile devices in the United States. SafeGraph does not collect this information directly from mobile applications. And SafeGraph does not have, and has never had, an SDK that collects location or user information.

4. SafeGraph’s data serves the public interest and fuels innovation. Thousands of academic researchers have relied on free access to SafeGraph’s data to produce hundreds of published research papers on an array of important subjects. As we discuss further below, researchers from many of the most prestigious institutions in the country, including Stanford, UC Berkeley, MIT, Harvard, Yale, the University of Chicago, Wake Forest University, Fordham University, the University of Washington, Seattle University, Columbia University and many, many more have used SafeGraph’s Patterns data to study the world and make impactful, often life-saving, discoveries.

Our Commitment to Privacy

SafeGraph is committed to privacy and remains at the forefront of privacy innovation in our industry. Not only is SafeGraph fully compliant with all applicable laws and regulations, our privacy practices and techniques are at the cutting edge of industry standards and norms. Our privacy safeguards include data aggregation, application of rigorous Differential
Privacy, and the exclusion of any device-level information or identifiers from its data products. SafeGraph was an early adopter of Differential Privacy\textsuperscript{1} tools and other rigorous privacy protections that ensure SafeGraph’s data cannot be “de-anonymized” using any known method of re-identification. These protections enable sophisticated data scientists to use statistical information to learn about the ways in which groups of people interact with a place without revealing any information about the individuals in those groups.

SafeGraph is committed to transparency. Detailed information about each of our products, including the precise contents of each product’s schema, can be found on our public website.

You have asked whether SafeGraph will commit to “a permanent ban on the sale of the location data of people who visited ‘Family Planning Centers’ and any other sensitive locations.” SafeGraph has never offered such location data tied to individuals, and has removed from our Patterns data set even aggregated, anonymized visit statistics for all businesses categorized by the NAICS Code for family planning centers. We have no intention of changing that course.

We appreciate the opportunity you have provided us to clarify the public record on these important matters.

I. **SafeGraph Produces High-Veracity Data About Places**

Your Letter indicates that your main interest is the SafeGraph product called Patterns, which is our only product connected in any way to mobility data. Our discussion thus begins with a brief summary of Patterns and SafeGraph’s flagship product, Places, which Patterns complements.

A. **SafeGraph Does Not Sell Individual Location Data**

Our Places product focuses exclusively on verifiable facts about commercial and public locations. The Places data set lists information about business locations, such as address, associated brands, and hours of operation, which is sourced from publicly available information—often provided on the website of the businesses being described.

The Places schema also includes North American Industry Classification System (“NAICS”) Codes to help categorize the primary business activity of a location. NAICS Codes are not

\textsuperscript{1} Differential Privacy is a method by which individual information is protected through the addition of randomized “noise” to data points.
created by SafeGraph; the Federal Statistical Agencies developed the Codes for “the collection, analysis, and publication of statistical data related to the US Economy.” Because there is no official assignment of NAICS Codes, and because SafeGraph primarily uses machine learning to assign the Codes to many of the locations in Places, this coding is not exact.

SafeGraph developed Patterns as a privacy-safe alternative to more sensitive, device-level location data. Our privacy goals are achieved by developing an aggregated and fully anonymized statistical model computed from opt-in device data, which is collected by third-party vendors and licensed to SafeGraph.

Aggregated: Patterns data is centered on points of interest (like a retail store), not on devices. Patterns users receive cumulative (or aggregated) counts of anonymous, opted-in devices to a specific place. No information is revealed concerning any individual user, and no device-level information is provided. In other words, Patterns offers an approximate composite of visits to a location rather than identifying specific devices or persons. This is done so that it is impossible to use Patterns to “track” a single device through various destinations.

Anonymized: SafeGraph’s privacy measures, including the insertion of statistical noise to device counts and other portions of the Patterns schema, filtering out rare events (such as low device counts to a particular location), and the application of Differential Privacy, help ensure that no Patterns data points can be connected to an individual. For these reasons, Patterns cannot be used for targeted advertising to a specific person.

Statistical: Patterns is not, and does not purport to be, a product that provides exact numbers of visits. Patterns is built off a sampling of visits to a given business location. The opt-in data on which Patterns is built represents a small fraction of the U.S. population. Patterns simply acts as a statistical model showing the relative popularity of one place compared to others.

B. SafeGraph Sources Only Anonymized Information to Build Patterns

SafeGraph does not partner with or directly collect information from any mobile applications to source Patterns. Similarly, SafeGraph does not, and has never, had an active software development kit (“SDK”), including an SDK designed to collect location data. Instead, SafeGraph sources data from a limited number of known vendors. These

---

vendors are subject to stringent contractual requirements regarding the sourcing of data and compliance with privacy laws, and source data only from applications that comply with native platform policies. SafeGraph has the ability to audit each vendor’s privacy practices. To date, SafeGraph has no cause or reason to believe any vendor is not fully compliant with required- and best-practices regarding privacy.

All data sourced by SafeGraph is anonymized before it reaches SafeGraph. SafeGraph never receives data that includes an individual’s name, address, gender, age, biographical or other personal information. The Device ID, assigned by the device manufacturer, is the only persistent identifier in the data that SafeGraph receives to build Patterns. The Device ID is not tied to any user profile; no information about the owner of the device is assigned to the Device ID. And SafeGraph does not attempt to identify the device owner.

SafeGraph removes the Device ID from the input data when building Patterns, so that the SafeGraph product available to customers does not include the Device ID. SafeGraph does not provide any device-level information in any of its products.

II. SafeGraph Employs Cutting-Edge Privacy Protection Tools to Protect Individual Privacy

The suggestion that our data can be used to identify individuals is false. SafeGraph employs industry-leading privacy techniques—including state-of-the-art Differential Privacy and data aggregation measures used by, among others, the United States Census

---

3 One of SafeGraph’s sources of data is its subsidiary, Veraset, LLC, of which SafeGraph is the majority owner. Veraset is a distinct legal entity, with independent contracting authority and a separate management team that controls its day-to-day operations. To the best of SafeGraph’s knowledge, the data provided to SafeGraph is representative of the data provided to other customers of Veraset. While certain Veraset data includes Device IDs, Veraset does not include any personal identification information in the data they provide. Furthermore, Veraset often anonymizes even the Device IDs provided in their data by giving hashed Device IDs that allow the data user to differentiate one anonymous device from another without any risk of ever being able to connect other, outside information to those Device IDs. Finally, SafeGraph neither attempts to de-anonymize data provided by Veraset nor makes data with Device IDs available in any product to anyone.
Bureau,4 Apple,5 and Google.6 SafeGraph is unaware of any method capable of re-identifying our data.

We believe reports alleging that is possible to re-identify SafeGraph data are inaccurate. Such reports, including those cited in the Letter, contend that it is possible to unveil individuals using “anonymized” data. But the analyses on which these reports rely are wholly inapplicable to SafeGraph’s data (just as they are to the similarly protected Census data) because each such study concerns data points at the individual-device level; that is, data that includes a Device ID or comparable persistent identifier. No SafeGraph product contains data points at the individual-device level, including Device IDs.

Patterns reports only aggregations of visits (the true numbers of which are further obfuscated by SafeGraph’s self-imposed privacy requirements, see Section II.B.), and applies Differential Privacy to data by infusing aggregated statistics with high amounts of interdependent, randomized Laplacian noise. These safeguards place SafeGraph’s data beyond any known means of re-identification.

A. Differential Privacy Prevents Data Re-Identification

Differential Privacy was first proposed as a privacy-preserving technique in a published 2006 joint paper authored by researchers from Microsoft, Ben-Gurion University, and the Weizmann Institute of Science. Its goal was to enable users of statistical databases “to learn properties of [a] population as a whole while protecting the privacy of the individual contributors.”7

Differential Privacy has since become the standard privacy-protection method for sensitive data of all types. It protects individual privacy through the addition of randomized “noise” to data points. In Patterns, this noise is introduced into various components of the schema,

---


including the device counts allocated to a physical location. This enhances privacy without degrading the statistical value of the data.

As noted, Differential Privacy is used by leading technology companies. It is also used by the United States Census Bureau to protect census information. In fact, the 2020 U.S. Census made the explicit and essential choice to employ Differential Privacy as the new disclosure avoidance system, “designed to withstand modern re-identification threats.”

Harvard University defines Differential Privacy as “a rigorous mathematical definition of privacy,” and explains that data “is said to be differentially private if by looking at the output, one cannot tell whether any individual’s data was included in the original dataset or not.”

Parameters titled “epsilon” and “delta” are used to quantify and compare the amount of privacy provided by Differential Privacy in any given database. Lower epsilon values mean higher levels of privacy protection. Epsilon values, sometimes called the “privacy budget,” in the data industry generally range from 0.4 up to 17.

For context, the epsilon values for some of the country’s largest companies using Differential Privacy are reported as follows:

- Apple: \( \epsilon = 2 - 8 \)
- Facebook: \( \epsilon = 0.45 - 2 \)
- Google: \( \epsilon = 1.68 - 2.64 \)
- LinkedIn: \( \epsilon = 0.3 - 28.8 \)
- Microsoft: \( \epsilon = 1.672 \)

---


9 Apple, supra note 5.


• 2020 U.S. Census:\textsuperscript{14} epsilon 2.47 – 17.14

SafeGraph sets its epsilon at 0.9, the low-end of the privacy budget range in the data industry, indicating a high amount of statistical noise and greater privacy in Patterns.

B. \textit{Aggregation Further Disassociates Data From People}

In addition to the use of Differential Privacy, SafeGraph aggregates data to build statistical Patterns models. This method further disassociates the data points from any individual person. This process is similar to how the U.S. Census Bureau produces its data in a privacy-sensitive manner.

Patterns aggregates cumulative counts of anonymous visitors to a physical location or to Census Block Groups. The former aggregation complements the Places dataset and provides a statistical sample illustrating the relative popularity of one location point versus another. The latter method serves the same purpose, but shows the relative popularity of a larger geography at different times. Neither version of aggregation can be used to “track” or “locate” individual people.

SafeGraph also controls for, and filters out of Patterns, rare aggregation events like statistics with low counts. If only one device from a Census Block Group appears in the source data, it is removed from the Patterns report. Similarly, SafeGraph rounds up statistics under a certain threshold.

III. \textit{Statistical Movement Data Serves the Public Interest}

No licensee has ever proposed, and SafeGraph has never approved, any commercial use of Patterns data relating to family planning centers. Commercial customers primarily use Patterns to better compete in their markets. For example, businesses use Patterns to compare the popularity of their locations to those of a competitor. They also use Patterns to assist with site selection, as the data shows relative traffic to different areas.

SafeGraph licenses data to federal public health agencies and local governments for use cases limited to supporting the ongoing COVID-19 response effort and related public-health functions, measuring demographic trends and the economic impact of policy initiatives, urban planning and development, and mass-transit strategies. At the onset of the pandemic, SafeGraph provided data for free to federal, state, and local governmental entities to assist

their responses to the COVID-19 crisis. Some of these entities became commercial licensees after seeing the data’s value to the efficient allocation of scarce government resources.

SafeGraph does not license data to any law enforcement agencies.

SafeGraph ensures that all of its licensing agreements contain strict requirements that the data remains anonymized, never linked to any personally identifiable information, never used in an attempt to reverse engineer, decompile, or otherwise re-identify the data by any method (even though Differential Privacy protection renders this impossible), shared with or disclosed to any third parties, or used in any other unauthorized manner. Moreover, SafeGraph demands the right to audit any licensee to ensure its compliance with the contractual terms regarding permissible uses of the data.

SafeGraph also makes data available for free to academic institutions and users with academic credentials for use solely in connection with noncommercial academic research and publication. More than 2,000 academics from America’s most prestigious institutions have relied on SafeGraph data in their published works. Some examples include:

- A team of researchers from Stanford and Northwestern University used SafeGraph data, including Patterns, to develop a model that can predict how COVID-19 spreads in cities. The study was published in the prestigious Nature magazine.

- Researchers at UCLA used Patterns data to contribute to the growing body of research around the disproportionate impacts of the pandemic on communities of color and to recommend data-driven policy solutions to inform assistance and recovery policies and programs.

- A researcher at Columbia University used Patterns to identify mobility trends linked to COVID-19 prone locations to assist policy makers in

---


determining when and how to safely reopen businesses and other gathering places.18

- Academics at Florida International University and the University of Florida used SafeGraph data to analyze foot traffic in the wake of Hurricane Irma and determine the impact such events have on a population’s visitation patterns.19

- Researchers at Purdue University used SafeGraph data to examine the foot traffic and economic recovery following Hurricane Maria in order to quantify the economic impact of disasters on businesses.20

- Economists at the National Bureau of Economic Research used SafeGraph data to measure possible racial disparities in the Paycheck Protection Program.21

- Academics from the University of Washington, examined how economic policy affects consumer spending and foot traffic.22

- Researchers at MIT, Harvard, and the University of Chicago used SafeGraph data to study the impact of COVID-19 on trips to urban amenities by specifically looking at travel behavior changes in Somerville, Massachusetts.23

---


Researchers at UC Berkeley, using SafeGraph’s data, were able to forecast the spread of COVID-19, which could allow for management of the disease at a local and global level by pinpointing policy measures to locations where they are most needed without “excessively stifling economic activity.”

Researchers at Yale and Columbia used SafeGraph data to examine the role of meteorological factors in the transmission of COVID-19.

A team of researchers from Yale, Harvard, Wake Forest, and Fordham Universities used SafeGraph data to explore the causal effects of chronic air pollution on the intensity of COVID-19.

Academics at Seattle University, used SafeGraph data to investigate college reopenings in the fall of 2020.

The list goes on with similar, important research conducted with SafeGraph data.

IV. SafeGraph Voluntarily Removed All Statistical-Movement Data Relating to Reproductive Health Locations

As of the writing of this letter, SafeGraph has removed all of the visit-related Patterns statistics aggregated to businesses categorized by the NAICS Code for family planning centers, 621410. In other words, it is now impossible to access any information about visits to family planning centers from our platform. In determining what data to remove, we erred on the side of over-exclusion because NAICS Code 621410 encompasses physical locations that offer services like family planning counseling services, fertility clinics, and birth control in addition to Planned Parenthood and abortion clinics. The data is no longer available to any of SafeGraph’s customers. SafeGraph voluntarily took this step, prior to


25 Yiqun Ma, Sen Pei, Jeffrey Shaman, et al., Role of meteorological factors in the transmission of SARS-CoV-2 in the United States, NAT COMMUN 12, 3602 (2021), https://doi.org/10.1038/s41467-021-23866-7.


receipt of your Letter, to ensure that all statistical visit information related to abortion providers is removed.

Prior to removing the data, it posed no privacy threat to individuals because it only showed representative numbers of devices that visited a physical location within the NAICS Code. It did not show anything about any individual. It did not reveal anything about the services a user procured at such a facility, if they even entered it. Nor could it be used to infer any medical information about any individual. SafeGraph’s data could not be used to send targeted advertising because there was no device-level information associated with the data.

V. SafeGraph Has No Knowledge Of Any Alleged Google Play Store Ban

SafeGraph has no knowledge of any purported “ban” imposed on SafeGraph by Google (or any other application platform). The origins of the rumored “ban” seem to trace to a single report from the Summer of 2021, which alleged that, in early June 2021, Google informed mobile application developers that they had seven days to remove SafeGraph’s SDK from their apps, or the apps would face removal from the Google Play Store. To the best of SafeGraph’s knowledge, none of the material “facts” in the report are true.

SafeGraph does not, and has never, used an SDK to collect information about mobile device users, location information or otherwise. This was true in June 2021, when the rumored ban was reportedly imposed.

SafeGraph can only speculate that the source of the confusion could be a limited exchange that occurred in 2018. Shortly after the company’s founding and before its decision to focus its business on data about places, SafeGraph briefly explored the concept of an open source SDK called OpenLocate. The OpenLocate SDK was never functional. Independent contractors assisting in development of the beta-test OpenLocate SDK created a private testing application (called “com.openlocate.example”) in the Google Play Store to allow a panel of testers to evaluate the beta-test SDK in the private, testing application. Because the testing application was designed solely to test the SDK, it was neither a functioning mobile application nor publicly available. Google removed the testing application from the Play Store under the functionality-provisions of its Policy Center. The removal concerned only the private testing application; it was not a ban on SafeGraph’s continued use of the Play Store.

Because SafeGraph had abandoned the idea of an SDK before it progressed beyond a nonfunctional, beta-test version and the testing application was solely related to the
abandoned SDK, SafeGraph did not appeal the testing application’s removal from the Play Store in 2018 or subsequently attempt to upload a functional version.

As noted, SafeGraph is unaware of any ban from the Google Play Store. Google has not informed SafeGraph directly of any “ban” applicable to the Company. No application developer has informed SafeGraph of any such outreach from Google. And SafeGraph still does not use SDKs to collect information about application users.

###

SafeGraph empowers problem solvers. SafeGraph’s data contributes to life-saving research, improves the strategic decisions of businesses and policy makers alike, and supports societal progress by helping some of the best innovators in private industry, the public sector, and academia unlock answers to the hardest questions. Access to data necessarily carries trade-offs; expanding access potentially allows a greater possibility of misuse, whereas limiting access demonstrably stifles innovation and research. As explained in this response, SafeGraph’s stringent privacy protections—applied to data that focuses on places, not people—support numerous beneficial uses while minimizing or eliminating the potential for its data to be used in privacy-invasive ways. SafeGraph is, and always will be, intently focused on privacy considerations as it pursues its mission to provide safe, anonymous data to companies and researchers as we partner with them to improve the world for us all.

cc: The Honorable Tammy Baldwin
    The Honorable Patty Murray
    The Honorable Tina Smith
    The Honorable Bernard Sanders
    The Honorable Edward J. Markey
    The Honorable Richard Blumenthal
    The Honorable Cory A. Booker
The Honorable Amy Klobuchar
The Honorable Christopher S. Murphy
The Honorable Ron Wyden
The Honorable Tammy Duckworth
The Honorable Alex Padilla
The Honorable Ben Ray Lujan