**Public Democracy, Inc.**

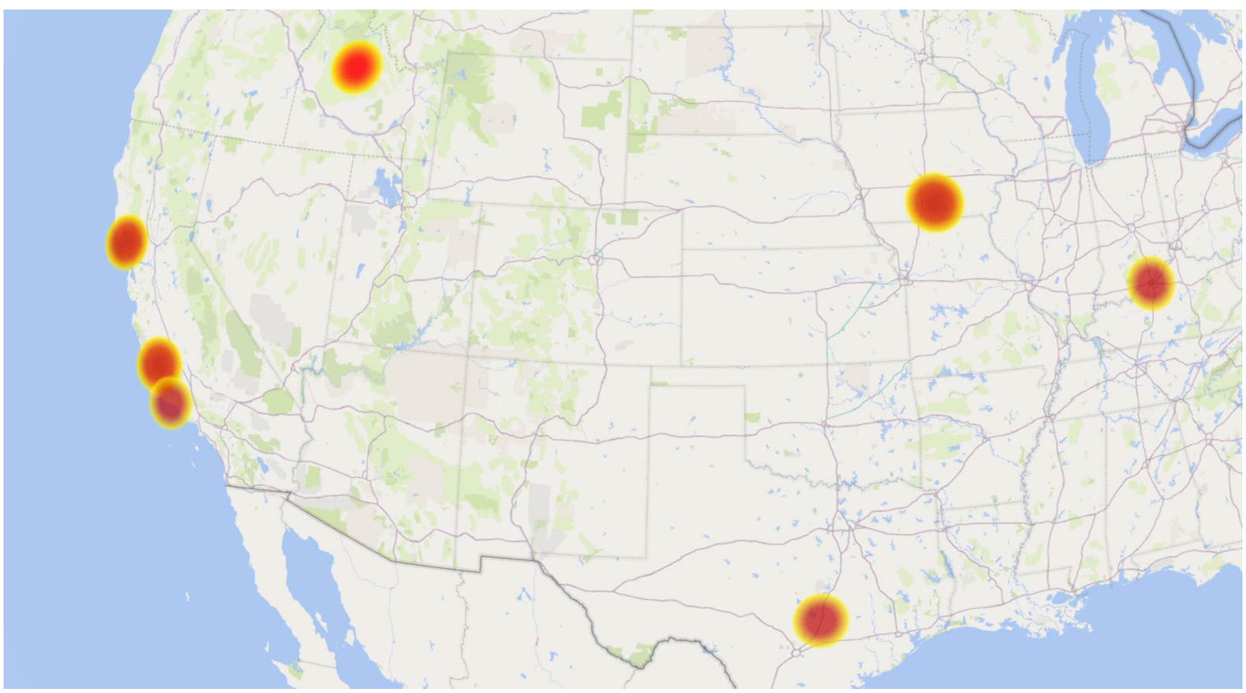
1901 N Ft Myer Dr, Suite 900, Arlington, VA 22209

publicdemocracy.io

DUNS #: 081303416
CAGE code: 84R65

Updated: April 4, 2020

# *Digital Tools to Map COVID-19 Spread and Deliver Interventions*



**Beginning March 13, Public Democracy's behavioral models identified the COVID-19 outbreaks shown above up to a day before the first confirmed case was reported in each region.**

*Digital Tools to Map COVID-19 Spread and Deliver Interventions*

## SECTION 1: Product Summary, Technical Description, and Capabilities

### 1.1 Overview

Over the past decade, Public Democracy, Inc. (a certified Public Benefit Corporation) has developed a proprietary psychometric database of over 100 million Americans with insights into the core values that motivate people to engage in the real world and join together in common purpose. In recent years, we have applied this data (and our decade-plus experience as a top communications firm) to build best-in-class behavioral models and empathy-driven machine learning systems. Our tools operate in a PII/PHI-free environment – instead of tracking identifiable individuals, our system uses human and machine learning to adaptively analyze patterns of behavior.

Those patterns then allow us to identify individuals and groups by their behavior and intent. We can filter within those behavioral segments using the hundreds of thousands of additional identifiers available through the commercial data market.

We have refined a novel, iterative process that leverages Big Data (commercial, public, and proprietary), enterprise-level marketing systems and adware, predictive analytics, and an iterative *human-companion-in-the-loop* process to identify populations groups in space and time, model behavioral patterns, identify and predict trends, and assess risk.

Over the weekend of March 13, **our early COVID-19 models successfully identified localized outbreaks prior to any media or public health reports** about a first case in numerous locations (Austin, TX; Blaine County, ID; Lucas County, OH; Santa Barbara, CA; Sonoma County, CA; and San Luis Obispo, CA). In another location with only a few confirmed cases that our model flagged as a hot spot (Kentucky), 10 new cases were reported in the following 48 hours, with over 100 cases over the following week.

These breakthroughs led to being invited to join the White House COVID-19 Data and Technology/Research Task Force, and we cohosted a session with the Atlantic Council at the Skoll World Forum on applying "Data For Good to Combat COVID-19." We also began to form new partnerships with tech and communications partners, including a collaboration with the Ad Council to develop moment-in-time messaging and an offer from Splunk to include our data as a layer in the Johns Hopkins COVID-19 map once we develop the system to unify our data into a format and stream that can be more easily integrated into an API for visualization. Our tools have the capability to:

- Leverage social media scraping, natural language processing and keyword associations, and apply proven machine learning systems to create **real-time reporting** on emergent COVID-19 clusters, prevalence of COVID-19, risk of infections spikes, and contributing factors in a given population;
- Refine reports in physical space, allowing **for analysis down to the city block or military base level**,
- Build **epidemiological data** on COVID-19 spread to enable better historical analysis and supplement under-reported rates based on individual testing alone;
- Utilize existing online/mobile applications for **self-reporting of symptoms** to improve need-specific information flow to users and better chain of command notifications and population-level data analytics;
- Backtrack confirmed cases and incorporate contact tracing, geofencing, and the ever-growing suite of public tools to alert those most likely to be exposed; and
- Most significantly, allow for customizable, moment-based and **need-based direct digital communication** with different population groups, ensuring that they receive the best information and, in a format, and moment when they are most able and willing to follow it.

## 1.2 Technical Summary

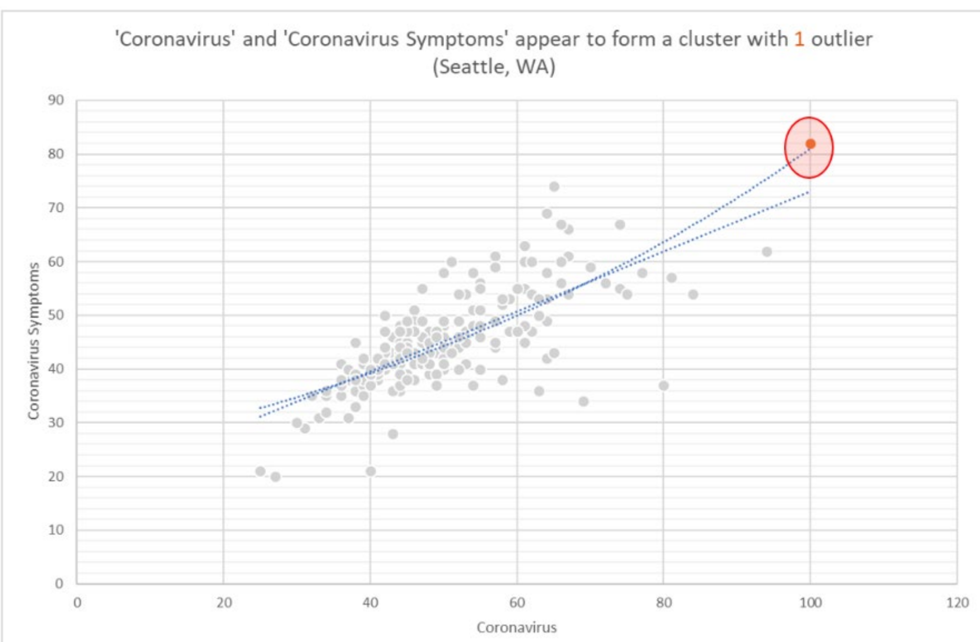
### *Behavior Segment Identification and Mapping*

Our innovative technology iteratively builds on psychometric and off-line data, layering on real-time online/mobile commercial, behavioral, and geospatial data through a machine learning process that also analyzes and tests online behavior to predict outcomes, refine understanding, and deliver information. The models we create from this process allow for both population-level insights and direct education/intervention where and when it is needed most.

Our early models have already allowed us to predict and measure outbreaks by identifying patterns that indicate who has developed COVID-19. Because of the interrelated nature of our methods, significant technical investments are needed to: 1) establish thresholds and confidence intervals for our behavioral models, 2) test direct engagement campaigns, and 3) develop a dashboard/API for others to access our reporting in real time. These investments will allow us to improve our ability to identify patterns that indicate a person has COVID-19 or is in another high risk/value category, where they are in their personal learning/interest curve, and when the best moment is to engage them.

The first step in our proprietary process for building our behavioral models (which power our population segment mapping and engagement tools) is the identification, acquisition, and creation of offline and online data. We have begun that process for COVID-19 but are currently seeking additional resources to scale and speed its completion. This is an iterative process that includes research into available public datasets, creation and delineation of increasingly refined behavioral and interest segments, keyword mapping, and general topic and population-level research to improve the *human-companion-in-the-loop* element of our machine learning process.

The chart below shows a simple two-variable analysis that served as an early validation and starting point in our process. During the 30 days prior to March 10, residents of Seattle, WA, distinguished themselves from all other major metro areas in that they were disproportionately more likely to view content that included both “coronavirus” and “coronavirus symptoms.” This is one powerful example of the type of preliminary data we use to direct our machine-learning to discover other correlative and differentiating behaviors that distinguish target population segments.




By using 30-day look-backs in the data to test and validate, and by applying machine learning and our *human-companion-in-the-loop* processes to these insights, we can identify early patterns in existing data that enable us to further refine our machine learning targets to improve our signal-to-noise detection ratios in population data. This process also identifies important gaps in our data we should target to improve. At that point, we seek out other sources of data, or devise new ways to generate the data necessary, for the machine learning to continue.

For example, the Topical and Keyword report in the image below shows an important early analysis of Americans searching for “Coronavirus Symptoms” vs. those searching for “Symptoms of Coronavirus.” The measurements for each group reflect topical content type and then keywords in content, based on what users in each of those two segments are reading and their anonymized digital/social media communication. This simple two-variable example initially differentiates population segments based only on Search (which, as a data input, makes up only about 5% of the mobile/online behavioral data we use to develop these behavioral models). Yet this example is quite informative.

It would be natural to assume that people searching for “Coronavirus Symptoms” and those searching for “Symptoms of Coronavirus” are merely expressing the same intent/need/behavior using slightly different words. But by following the path from that initial Search input, it becomes clear that these are (or, at least were, during the week of March 10) very distinct groups.

**The “Coronavirus Symptoms” group is interested in general information on COVID-19 and the “Symptoms of Coronavirus” group is seeking information for more personal and medically specific reasons.** The top keywords in content being read and communicated by these two groups show a clear distinction in intent and interest, with the Trending measurements reflecting the newest and most significant increases in interest.

| "Coronavirus Symptoms" | | |
|---|---|---|
| **Related Topic** | **Relevancy Score** | **RISING** |
| Coronavirus | 99 | 350% |
| Virus | 7 | 200% |
| Influenza | 4 | |
| Common cold | 3 | |
| Fever | 1 | |
| Mortality rate | 1 | |
| Outbreak | 1 | |
| Sore throat | 1 | |
| | | |
| **Content keyword Correlations** | | |
| usa coronavirus | | |
| what is coronavirus | | |
| what is coronavirus symptoms | | |
| coronavirus china symptoms | | |
| coronavirus china | | |
| coronavirus update | | Trending |
| coronavirus us | | Trending |
| what is the coronavirus | | Trending |

| "Symptoms OF Coronavirus" | | |
|---|---|---|
| **Related Topic** | **Relevancy Score** | **RISING** |
| Coronavirus | 100 | 500% |
| Symptom | 98 | 750% |
| Virus | 8 | 300% |
| Medical sign | 3 | |
| Breathing | 1 | 160% |
| Cough | 1 | |
| Fever | 1 | |
| Shortness of breath | 1 | 300% |
| | | |
| **Content keyword Correlations** | | |
| what is symptoms of coronavirus | | |
| symptoms of the coronavirus virus | | |
| what is the coronavirus | | |
| what is the symptoms of coronavirus | | |
| what is the symptoms of the coronavirus | | |
| signs of coronavirus | | Trending |
| signs symptoms of coronavirus | | Trending |
| symptoms of the flu | | Trending |

It is noteworthy that the “Symptoms of” group had already identified and was reading about the three symptoms that differentiated COVID-19 from the seasonal flu and the common cold, indicating that the “Symptoms of” group was much farther along on their knowledge-discovery curve, as well as far more personally-focused. Answering how and why subpopulations within the “Symptoms of” segment group reached this advanced point before the broader population is key to distinguishing between understanding individual need, intent, and lived experience.

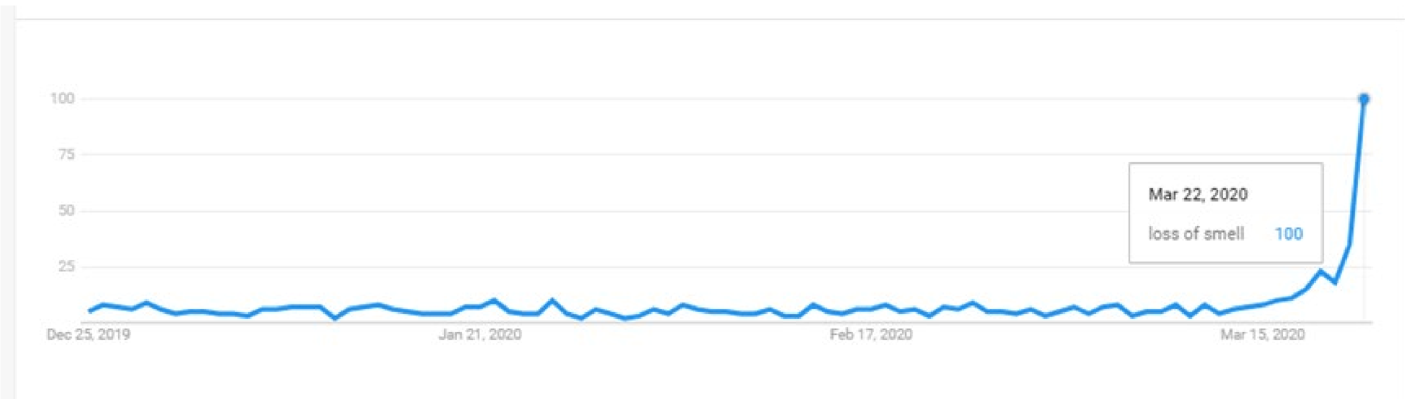
Specifically, our subsequent success identifying the subgroup that entered this more refined (but still quite broad) segment because of a personal/off-line experience with these symptoms was the key to our identifying regional outbreaks before the first person in that geographic area tested positive for COVID-19.

Furthermore, the process of “filtering out” other subgroups not personally experiencing symptoms creates insights necessary for the **audience segmentation that is vital for effective moment-based and need-based differential communication strategies**. We believe this will be the most valuable and necessary application of these tools, which we discuss later in this paper.

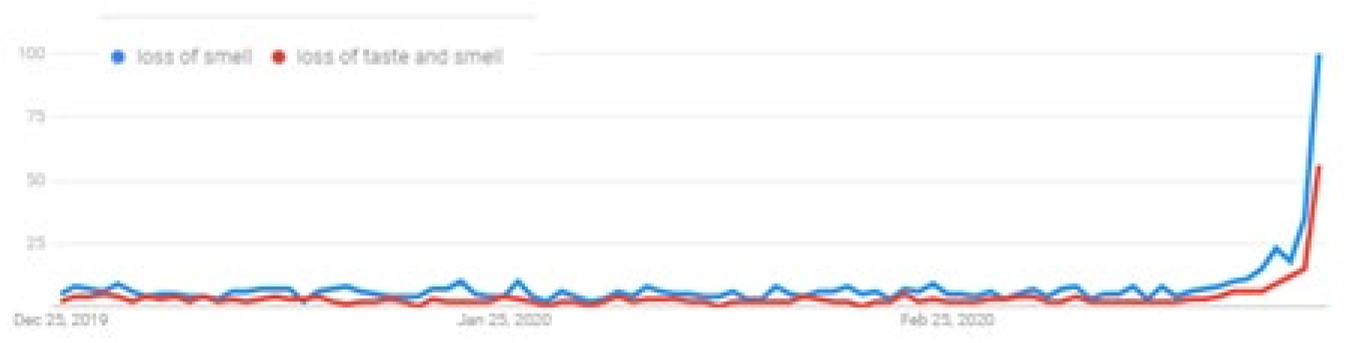
*Our identification of these kinds of bright lines – in intent and personal experience between seemingly synonymous terminology – have allowed us to successfully identify people at various stages of opioid Substance Use Disorder, resulted in award-winning efforts to develop online trust and engagement funnels for veterans with severe PTSD and many other successful applications of our tools (see: http://publicdemocracy.io/overview).*

Machine learning helps us focus in specifically on the targeted behavior and intent of our different population segments, which highlight new trends and outliers that inform our continued iteration and refinement. For example, a week before USA Today headline story on how loss of smell was the "new" symptom doctors were associating with COVID-19, we had included Anosmia in our targeting models and had flagged it in conversations with several federal agencies.

As the size of our predicted COVID-19 population both grew and became more refined, we began to notice that "loss of smell" was an odd (if not statistically significant) associated term in different data reports. A quick search for possible reasons identified an anecdotal observation by a Seattle-based intensivist who said one-third of his COVID-19 patients had anosmia. That discovery then led us to explore trends around national interest in "loss of smell" (see chart below showing Search for "loss of smell" mapped over time). The connection was starting to become clear, directing the next round of our iterative process.




Loss of taste does not appear to be as common a symptom as loss of smell. But once the combination of machine learning and off-line data had helped us identify loss of smell as a COVID-19 identifier, the *human-companion-in-the-loop process* allowed us to explore other medically-related factors that were not be quite as significant on their own in the data but might fit when brought into context (see the next chart where the red line tracks "loss of taste"). As we bring more and more of these points together through human intuition and apply machine learning, we are able to cut through more of the "data noise" and generate additional high quality signal data necessary for continued refinement of the machine-learning process.




As these examples illustrate, even simplified two-variable analyses can create very meaningful and useful insights. But machine learning and multi-factor analyses are necessary for us to identify the more refined behavioral patterns, which then allow for a more granular segmentation of these populations. Our iterative *human-companion-in-the-loop* process and lessons from past applications of these tools also continue to play a key role in that process.

Even though our machine-learning tools are able to access the massive array of commercially available behavioral datasets, with their billions of anonymized measurements on users, a human element is necessary to identify and prioritize off-line and disconnected datasets that can improve and filter the machine learning process. This *human-companion-in-the-loop* process also enables the use of empathy and social insights to intuit possible new directions to explore through directed machine learning and data creation. Our team's diverse expertise in national security, pastoral care, crisis communications, community organizing, economic development, and political/mobilization campaigns has proven especially effective in the development and application of those empathy-based insights.

## *Moment-and-Need-Dependent Digital Engagement, Education, and Intervention*

One of our most powerful tools for generating data to fill holes and bridge the machine learning gaps that we identify in existing datasets is the use of quick, targeted, and insightful digital engagement campaigns. We can access much thicker behavioral data (and make better use of the enterprise-level marketing platforms that contribute to our learning and mapping processes) when we are able to run digital ads to test new models and refine our population and risk segments. These quick engagement campaigns also provide real-world and third-party-validated outcomes and proof of concept, which speed adoption and will attract follow-on funding.

Our ability to use digital campaigns to intercept individuals during specific points in their information acquisition curve can efficiently direct or redirect the searches of specific populations as needed. This allows us to triage the flow of individuals seeking information and to identify those most in need of assistance (medical or informational). *The ability to communicate directly based on need and interest will be especially important as the COVID-19 crisis continues.*

The algorithms that power the internet and Google Search cannot keep up with the pace of developments around COVID-19. For example, Governor Newsom's announcement of the statewide shelter-in-place order for all of California did not appear on Google News for over 45 minutes after it was announced. That is because the shelter-in-place orders for San Francisco had too much algorithmic dominance for keywords related to California, lock-down, Governor Newsom, etc. As public health best practices, rules, and resources change on a daily (and even hourly) basis, the only option for effective public health communication will be promoted, personalized communication.

We need to develop systems immediately that directly connect people with the most up-to-date public health information they need in that moment. **If we do not start building these systems now, we will not have them in place when health officials and government leaders need to start triaging and redirecting people to available beds, providing the latest testing options, and countering misinformation.**

Hospital networks are already working with schools, office real estate companies, and other similar groups to identify the overflow space that will be needed to triage and treat patients in the hope of avoiding the need to resort to tents in parking lots or public parks. None of those new locations, or the hospital closures that necessitate them, will show up in Search or mapping sites in a timely manner. Greater capacity for real-time communication with people in the greatest need with the latest information is crucial. Furthermore, traditional Public Service Announcements, such as population-wide TV or radio (or even broad digital advertising) to inform the population about new testing centers and services will likely prove ineffective, if not counterproductive. Taking approaches that are not targeted towards people in the most need will, unfortunately, result in these triage locations being overrun by worried citizens who show up to get tested but do not need immediate care.

Lastly, but crucially, we want to emphasize just how important a nuanced approach will be over the days to come. Different communities have different levels of trust in government sources and different abilities to process and act on information. Useful, believable, and actionable information in our current times and in this particular situation will require curated and personalized communication. Public Democracy has demonstrated our ability to achieve these outcomes in numerous past campaigns, but we must start now and scale quickly to ensure we have developed the data refinement and messaging insights necessary to execute these more personalized public health communications as they become increasingly vital.

## *Need and Opportunity to Act*

The data “noise” around COVID-19 is increasing every day and more Americans are moving farther down their own learning curves, making the data creation from direct engagement campaigns all the more necessary to maintain our models’ accuracy and granularity. Fortunately, we have seen in past efforts that fully implementing campaigns like this one can “teach the internet” about these more refined population groups and their needs, speeding and improving organic referrals as well.

Our tools can actively insert accurate, authoritative, and comprehensive information that public health agencies wish to deliver at the moments when each specific audience is most receptive to this information and using language and values that will resonate best. This will be vital in mitigating impact and slowing COVID-19 spread, in meeting the informational and emotional needs of our communities, and in combating the spread of misinformation.

Rather than relying on people to sign up for alerts from accurate sources of information (e.g., the CDC, WHO, and local governments), or passively waiting for their browsing activity to bring them to the right website at the right time, our outreach strategies work far more effectively. We can tailor each message to each audience and actively deliver them seamlessly across multiple platforms, in the moments when people are most receptive to them.

Mitigation requires meeting people where they are—emotionally, on their learning curve, and in their physical space. In numerous past case studies (see below), our models have proven Public Democracy’s ability to identify these moments, and our demographically-and-socioemotionally-targeted messaging efforts have effectively helped combat misinformation, empower the public, and assist in the nation’s responses to public health concerns. **We know COVID-19 will be the greatest challenge we have yet encountered, and we have shifted our full resources towards this effort.**

The need to begin direct and differentiated communication efforts is not just a matter of healthcare for those with COVID-19, or even for public health more generally. Getting this right will improve America’s public safety, economic recovery, and national security.

Investments made early in this process create outsized returns and impacts. Data learns on a curve, and so speeding our ability to onboard new data, to create data to fill holes, and to expand our capacity to dive deeper into findings at each iterative step will have a dramatic impact on our success.

## *A Unique Solution*

Our process utilizes a great deal of public and commercially-available data, and we also use enterprise-level marketing and adware tools that are non-proprietary. But **our iterative process—comprising human insight, our proprietary Values Data™, and our empathy-driven machine learning techniques—is cutting edge in the marketplace.** The Office of National Drug Control Policy determined that our unique ability to map risk and engage individuals with SUD justified their seeking legal authority for sole source contracts in support of our technology. Google also sought us out as a partner, providing free advertising and unlocking all the blocked keywords associated with SUD to improve our behavioral models. And as stated above, our ability to track COVID-19 spread and identify contributing factors and related symptoms was recognized by the White House, the Atlantic Council, the Ad Council, and the creators of the Johns Hopkins COVID-19 map.

**In short, there are numerous data generation and analysis tools in the market, but our proprietary process, tradecraft, and ability to incorporate our best-in-class Values Data™ make our offering and capabilities uniquely different from others.**

## 1.3 Past Successes

Outside of our COVID-19 applications, we have already proven the efficacy of our tools, which combine off-line data and online behavior modeling, to develop behavioral mapping and digital engagement campaigns that:

- predict risk of substance use disorder (SUD) and map populations in the different stages of SUD in space and time,
- identify and deliver services to veterans with severe PTSD,
- find and recruit patients who developed cancer from specific environmental factors,
- map human trafficking patterns and effectively off-ramp individuals who drive demand, and
- develop market-actionable insights into the behavior and priorities of underserved communities.

Our work with veterans was named *Best in Show for Machine Learning and Data Analytics* in 2018 by the National Association of Political Consultants. But most importantly, it saved lives and built an incredible community. Monsanto is nearing a $10 billion settlement with the initial claimants in the case over RoundUp whom we recruited. The Office of National Drug Control Policy determined that our unique ability to map risk and engage individuals with SUD justified their seeking legal authority for sole source contracts in support of our technology. Google also sought us out as a partner, providing free advertising and unlocking all the blocked keywords associated with SUD to improve our behavioral models.

# SECTION 2: Testing, Development Schedule, Compliance

## 2.1 Testing and Validation

Our proprietary and innovative process uses a combination of machine learning, anonymized data to crowdsource solutions, and human intuition to create PII/PHI-free behavioral models, which allows us to achieve three key objectives:

1) Identify patterns indicative of COVID-19 infection and/or increased relative risk;
2) Understand and refine population segments with distinct temporal, geo-spacial, socioemotional, and informational needs; and
3) Allow the delivery of vetted ultra-targeted, authoritative public health messaging.

These three objectives are interrelated and, through a combination of machine learning, human refinement, crowdsourcing, and message testing, each interconnected step adds to the totality of the efficacy of our efforts.

*Objective 1: Pattern identification*. We will measure success in our models by validating our ability to predict initial outbreaks and subsequent spikes in new cases using data released by public health officials after cases have been confirmed. Depending on how quickly test infrastructure improves, we may also use results from other models developed by researchers to establish likely total cases based on confirmed ones as validations.

*Objective 2: Audience Segmentation*. Success in Objective 1 is in part dependent upon and also helps improve Objective 2. One of the ways that we refine models is to identify and filter out audience segments not exhibiting the behavior we are targeting. To remove a group engaging with COVID-19 content for reasons other than their or a loved one's infection, we need to understand what motivates that other segment to engage in this space. We will validate different segment types through keyword associations and by tracing their different paths in self-learning. We may also validate segments by demonstrating their interest in engaging with promoted content and digital ads.

*Objective 3: Information Dissemination*. We will stage testing for this objective, first by demonstrating our ability to generate engagement with our promoted content, which can be measured through response metrics. We also believe we will be able to measure success through metrics that track shifts in behavior of those who engaged our content compared to others who were not exposed to our digital interventions. This first-stage evidence will also support and improve outcomes for Objective 1 and Objective 2. Second-stage testing and rollout will involve larger scale curated and need-based communication campaigns directed by the government that will also include message/image/keyword testing to improve communication outcomes. From there, we will scale communication as directed and needed by government and health partners.

## 2.2 Development Schedule

We began collecting early data sets on COVID-19 on January 23, 2020, with the expectation that there would be a need to repurpose past models for this crisis. We began developing the COVID-19 models in earnest the weekend of March 7 and were working on models to predict outbreaks by March 13. Over the past two weeks, our team has sought input from over one hundred different potential partners, in order to develop the approach we are proposing here. We have also engaged with dozens of public health experts, dozens more leaders in local government, and a wide array of tech and data leaders – and last week, we officially joined the White House COVID-19 Data & Research Tech Task Force – all of which will support our efforts, and all of which will be improved by what we learn through this work.

With full funding, we believe that within a month we can have more robust and refined models, which will be able to cut through the increasing background noise, as all of America engages at deeper levels of interest and learning. Integrated data reporting systems and basic dashboards can be built concurrently and be available within 30 to 60 days. Soon thereafter, APIs can then be quickly developed to meet partner specifications. We expect that more robust visualization dashboarding tools can be developed within 60 to 90 days of full funding.

Through our prior work around economic development and with the US Census, we already have relationships with over 50 city leaders, economic developers, and groups like the National League of Cities and U.S. Conference of Mayors, which will speed adoption and utilization of these tools once they are more fully developed.

## 2.3 Regulatory Strategy

We expect this project to generate anonymized geospatial and behavioral data using commercially available tools primarily designed for digital advertising. We will also generate related but more specific PII-free user response data from direct engagement education/intervention campaigns. In addition, we will use Public Democracy's proprietary Values Data™, which is PII-based direct response data.

All of our data have already been vetted through public and commercially available channels and processes, and we will abide by all appropriate HIPPA, privacy, and other legal guidance for the use, storage, and handling of data.

We take privacy and ethical usage of data <u>very seriously</u>. As a Public Benefit Corporation, it is core to our governance and a key purpose for our existence, which we have advocated for vocally and explain clearly on our website. We expect any data created from this project, beyond Public Democracy's own privately held data, to be anonymized population-level data or anonymized user response data. The PII-free nature of most of this data reduces privacy risks, but we have an internal ethics review of all data usage and are very aware of the vulnerable nature of the populations we are dealing with.

One benefit of Public Democracy's diverse team is that we bring significant academic and personal experience in the area of ethics and balancing multi-factor moral needs.  Beyond our internal review and compliance with our own data usage standards, we will collaborate with government, business, and civic partners to weigh different usage and publication policies to protect both individual rights and to promote the common good.

We will restrict the use of all data generated through this project to internal purposes unless otherwise specified in a contract with a partner or deemed be necessary for the public good.  We expect federal and local governments, business, and non-profits to be interested in our data to help mitigate impacts of COVID-19.

In general, we believe visualizations and conclusions from aggregated data and trends will be more valuable than the raw data to partners and the public, which will add another layer of data protection.

Contact

Eric Sapp
President, Public Democracy
esapp@publicdemocracy.io