Deloitte.

PeoplePrism™

Dashboard Methodology

Child Care Deserts

Deloitte.

PeoplePrism™ reveals unique and compelling insights on your population.

About

PeoplePrism provides an innovative, data-driven predictive analysis tool that provides access to timely and rich data and includes interactive, geospatial mapping capabilities that can illuminate capability and service deserts, or sectors where demand exceeds supply or capabilities, by a user-defined area. An important component of the datasets used for the predictive analysis is its compliance with data security and privacy laws, including HIPAA, data sharing laws, data safeguarding, and ethical AI modeling.

PeoplePrism is one of the largest social drivers datasets with data on 335M people in the US, leveraging 1,700+ social drivers data variables on all adults and 8+ years of data history to identify populations of interest, supporting point-in-time analyses that can inform optimal outreach and deployment of resources.

Our Difference

PeoplePrism mobilizes one of the largest multi-faceted social drivers datasets in the country to help our clients deploy strategies designed to improve outcomes and access to social services and resources in vulnerable populations. Through this data, PeoplePrism identifies opportunities to achieve optimal outcomes that can be experienced by populations that have been disadvantaged by their social or economic status, geographic location, and environment.

Predictive Modeling Methodology

The PeoplePrism data solution comprises a collection of numerous individual- and household-level characteristics spread across socioeconomic, climate, environment, demographic, behavioral, and lifestyle attributes. These data points are used to train and build machine learning models that predict the probability of risk for a variety of vulnerabilities including chronic health conditions such as type 2 diabetes; behavioral health conditions such as opioid misuse; and economic vulnerabilities such as unemployment or food insecurity, etc.

Where traditional risk modeling techniques rely primarily on an individual's clinical or electronic health records and demographic data to model health risk, and financial or credit activity data to model economic risk, we take a comprehensive view of an individual across a variety of lenses including access to health care, language spoken, health literacy, food insecurity, transportation insecurity etc. Such an approach allows us to build models that capture the complex inter-relationship between the myriad factors correlated with adverse outcomes, especially for marginalized populations.

We leverage the power of semi-supervised machine learning (ML) methods to train explainable predictive models. Demographic, behavioral, economic, geographic, and other variables corresponding to each individual were used to build models to predict the probability of an individual to be at risk of a specific condition such as opioid misuse or uninsurance etc. Modeling datasets are generated that represent the US population with respect to protected groups, including age, race, and gender. We extend the bias checks further to include representation across income groups and the urban-rural continuum as defined by USDA[1]. Each feature input into the model is examined for its distribution with respect to age, race, gender, income, and the urban-rural continuum to ensure that they do not act as a proxy for any of those variables.

Dashboard Indicators

Filtered Adult Population

The PeoplePrism Child Care Deserts dashboard displays the total count of households for Washington, DC, which satisfies the selected risk indicators and social drivers. The dashboard visualizes child care insecurity as the rate of filtered adult population per 100,000 households within each geographic boundary. When the rate falls under 100, the data on the dashboard is obfuscated to protect the identity of certain groups.

Access to Child Care Indicator

The Access to Child Care indicator helps reveal Child Care Deserts across Washington, DC. To determine if a household has insufficient access to child care, a distanced-based approach was used to measure the available supply per child care provider and demand per household. This distance-based approach assumes families are interested in nearby providers despite their location.

PeoplePrism collected publicly available data on licensed and registered child care providers in DC and geographic information was captured for all providers. If the provider information did not include a latitude and longitude, the data was geocoded using Esri's Streets Premium network data to determine these coordinates. Service areas were defined as the accessible network of places within a 20-minute drive of a household. Within this service area, the supply is defined as the sum of the available child care slots from child care facilities; the demand is the sum of the total number of child care aged children (0-4 years of age).

Finally, a “capacity-to-population ratio” was generated for every household, which is the total supply (number of available, nearby child care slots) divided by the total demand (total count of child care aged children). This supply to demand ratio represents the pressure an individual household is facing to find available child care. All values below 1 represent a household with more children than available slots in the area, and as the values approach 0, the higher the insecurity pressure. This metric does not consider additional barriers to access, including financial comfort or transportation access.

The Center for American Progress’s 2017 report “Mapping America’s Child Care Deserts” establishes that true child care insecurity is faced when there are 3x as many children competing for the same available child care slots[2]. In line with CAP’s definition, given a household-level supply-to-demand ratio, a fixed cutoff for facing “High Insecurity” pressure was set at 0.3. Households with a ratio between 0 and 0.3 were assigned high risk of child care insecurity, while households at 0.3+ were assigned as low risk.

Child Care Provider Data

Child care provider data was scraped from publicly available sources to calculate the access to child care indicator and to display as an additional map data layer as points of interest to visualize child care provider locations. For each provider, data was collected on their facility name, address, and capacity. The following facility types were classified as a child care provider:

- Licensed camp
- Public school
- Licensed group
- Licensed family
- Certified family
- Out of state program

To be classified as a child care provider, each facility needed to be operational, have capacity for children 0-4 years of age, and be licensed.

Parent Status

The parent status filter in the dashboard provides two options for selection: single parent with child and two parent with child. Multiple qualifying household types are contained within these two groups and are defined below for reference.

Single parent with child:

- One person (female householder) with children present
- One person (male householder) with children present

Two parents with child:

- Married (husband and wife present) with children present
- Married (husband and wife present) with no children present
- Two persons, one male and one female (marital status unknown) with children present
- Male householder with one or more other persons of any gender with children present
- Female householder with one or more other persons of any gender with children present

Dashboard Models

The PeoplePrism Child Care Deserts dashboard employs multiple advanced models to predict the risk of various health, behavioral, and economic conditions across DC. These models are utilized in the dashboards to filter and analyze data to better understand child care deserts across different population segments. This section briefly describes the PeoplePrism models available within the dashboard.

Race and Ethnicity Model

PeoplePrism contains self-reported race data for nearly 70 million adults in the US. For the remaining individuals whose race is unknown, a multi-class classification model that predicts a person's race is trained using the self-reported race and ethnicity data. The model is trained using first and last names, geographical information, household characteristics, and socioeconomic fields to predict a person's race.

Gross Income as % of 100% Federal Poverty Level

Federal Poverty Level (FPL) is a measure of income used to determine eligibility for certain programs and benefits (e.g. Medicaid)[3]. PeoplePrism contains estimates of household level income in the form of income bins which is converted into continuous income using a predictive model. The household income is combined with household composition as specified by Federal Poverty guidelines[4] to calculate FPL for a given household.

Contact

For more information, contact our team at PeoplePrismContact@deloitte.com.

References

[1] https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/
[2] https://www.americanprogress.org/article/americas-child-care-deserts-2018/
[3] https://www.healthcare.gov/glossary/federal-poverty-level-fpl/
[4] https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines