

# Rice Datathon 2023

## Visualizing Gaps in U.S. Mammography Coverage

(Beginner Track)

In our analysis, we rank each state with a score that is based upon 5 key metrics. They are as follows:

$p$  = total population  
 $w$  = number of women over the age of 50  
 $m$  = number of certified mammogram facilities  
 $d$  = death rate due to female breast cancer  
 $incr$  = projected percent increase of  $w_{state}$  from 2019 to 2024

We define the set,  $A$ , to contain all 51 U.S. states, and  $a \in A$ , which each contain the 5 key metrics above as attributes. These attributes are used to create weighted priority scores for each state through the discrete weight function which is defined by:

$$\sum_{a \in A} f(a)w(a)$$

We have defined  $f(a)$  and  $w(a)$  as follows:

$$f(a) = \left\{ \begin{array}{l} (1), \quad \frac{w}{m} \text{ (Normalized)} \\ (2), \quad \frac{d}{m} \text{ (Normalized)} \\ (3), \quad \frac{w}{p} \text{ (Normalized)} \end{array} \right\}$$

Where (1) is the number of women over the age of 50 per certified mammogram facility, (2) is the death rate due to breast cancer per facility, and (3) is the number of women over 50 in relation to the total population

$$w(a) = \left\{ \begin{array}{l} (1) * 1.5 \\ (2) * 2 \\ (3) * 1 \end{array} \right\}$$

(Note that we normalized each of the 3 elements in  $A$ . such that they are represented by a value between 0 and 1, for easier comparisons and so that we can factor in the relative importance of each of the metrics.)