

Hypothesis-1: Highly populated states are at higher risk of getting asthma.

This metric is created in two steps:

1) In the initial step, we have considered Population, Population density and Asthma.

After checking the skewness of the data, we have used three quantile ranges for our analysis:

- Less than 50 quantile- **row_value < quantile_50**
- Greater than or equal to 75 quantile- **row value >= quantile_75**
- Greater than or equal to 50 quantile and less than 75 quantile- **quantile_50 <= row_value < quantile_75**

We have added 26 permutations and combinations in the form of if-else conditions.

While deciding on the metric, the priority of the quantile ranges were as follows:

1. Population- Total number of residents
2. Asthma- Number of persons who reported being told by a health professional that they currently have asthma
3. Population Density- The population density of the state

Population	Population Density	Asthma	Metric
x>=75	y>=75	z>=75	3.25
50<=x<75	50<=y<75	50<=z<75	3.25
x<50	y<50	z<50	3.25
50<=x<75	y>=75	50<=z<75	3
x<50	50<=y<75	z<50	3
50<=x<75	y>=75	z>=75	2.75
x<50	50<=y<75	50<=z<75	2.75
x>=75	y>=75	50<=z<75	2.5
50<=x<75	50<=y<75	z<50	2.5
x>=75	50<=y<75	z>=75	2.25
50<=x<75	y<50	50<=z<75	2.25
50<=x<75	y<50	z<50	2
x>=75	50<=y<75	50<=z<75	2
50<=x<75	50<=y<75	z>=75	1.75
x<50	y<50	50<=z<75	1.75
x>=75	y<50	z>=75	1.5
x>=75	y<50	50<=z<75	1.25
x>=75	y>=75	z<50	1
x>=75	50<=y<75	z<50	0.75
50<=x<75	y<50	z>=75	0.5
x>=75	y<50	z<50	0.25
x<50	y>=75	z>75	0
x<50	y<50	z>=75	-0.1
x<50	y>=75	50<=z<75	-0.25
x<50	y>=75	z<50	-0.5
x<50	50<=y<75	z>=75	-0.75

The above table mentions all the combinations that we have used along with the metric.

In the above metric, there are a few combinations that are grouped together and assigned a metric value for it. We adapted this method because all the data that we have is w.r.t the population and same goes with the hypothesis-1.

One limitation that this metric can have is that we have assigned same metric for two proportionate combinations.

For example:

Population	Population Density	Asthma	Metric
$50 \leq x < 75$	$y \geq 75$	$50 \leq z < 75$	3
$x < 50$	$50 \leq y < 75$	$z < 50$	3

Once we get the metric using these three columns, we further investigate our hypothesis with Smoking and Air Pollution in a similar way.

1. Smoking-The number of persons who reported smoking.
2. Air Pollution- Average exposure of the general public to particulate matter of 2.5 microns or less measured in micrograms per cubic meter

Here, we have performed a few additions on the previous metric based on the metric score.

As, our hypothesis focuses on highly populated states, we have used the top 4 metric values that are: 3.25, 3, 2.75 and 2.5. Also, we have grouped two metric values as the metric value can change over the years.

Previous_metric	Air Pollution	Smoking	Metric
$x = 3.25$ or $x = 3$	$y \geq 75$	$z \geq 75$	previous_metric+1.75
$x = 3.25$ or $x = 3$	$50 \leq y < 75$	$50 \leq z < 75$	previous_metric+1.75
$x = 3.25$ or $x = 3$	$y < 50$	$z < 50$	previous_metric+1.75
$x = 3.25$ or $x = 3$	$50 \leq y < 75$	$z \geq 75$	previous_metric+1.5
$x = 3.25$ or $x = 3$	$y > 75$	$50 \leq z < 75$	previous_metric+1.5
$x = 3.25$ or $x = 3$	$50 \leq y < 75$	$z < 50$	previous_metric+1.25
$x = 3.25$ or $x = 3$	$y < 50$	$50 \leq z < 75$	previous_metric+1.25
$x = 3.25$ or $x = 3$	$y \geq 75$	$z < 50$	previous_metric+1.0
$x = 3.25$ or $x = 3$	$z < 50$	$y \geq 75$	previous_metric+1.0
$x = 2.75$ or $x = 2.5$	$y \geq 75$	$z \geq 75$	previous_metric+1.75
$x = 2.75$ or $x = 2.5$	$50 \leq y < 75$	$50 \leq z < 75$	previous_metric+1.75
$x = 2.75$ or $x = 2.5$	$y < 50$	$z < 50$	previous_metric+1.75
$x = 2.75$ or $x = 2.5$	$50 \leq y < 75$	$z \geq 75$	previous_metric+1.5
$x = 2.75$ or $x = 2.5$	$y > 75$	$50 \leq z < 75$	previous_metric+1.5
$x = 2.75$ or $x = 2.5$	$50 \leq y < 75$	$z < 50$	previous_metric+1.25
$x = 2.75$ or $x = 2.5$	$y < 50$	$50 \leq z < 75$	previous_metric+1.25
$x = 2.75$ or $x = 2.5$	$y \geq 75$	$z < 50$	previous_metric+1.0
$x = 2.75$ or $x = 2.5$	$y < 50$	$z \geq 75$	previous_metric+1.0