Personal Key Indicators of Heart Disease – EDA

By Darin Myers and Team

# Description of the Data Set

The dataset is a cleaned and paired down version of the 2020 Behavioral Risk Factor Surveillance System (BRFSS) data set that was collected by the CDC. The data set was paired down for use in Machine Learning models for prediction of heart disease.

# Overview of Data

Graphical user interface, application

Description automatically generated

There are 18 Variables in this data set with the first column, HeartDisease, acting as our dependent variable. There are no missing cells (NA) in this data set. While there are some duplicate rows, these are to be expected as the unique identifier for the rows has been removed from the data set and the number of columns has been paired down. With over 300K rows, there is likely to be some duplication. These are legitimate respondents, and it is not advised to remove these rows

# Data Frame Describe





Most of the Variables were Yes/No questions that resulted in a value of 0 or 1.

# Variables

## Dependent Variable – HeartDisease

Graphical user interface, application, Teams

Description automatically generated

The dependent variable for this data set is the first column HeartDisease. As shown in the figure above, the data is highly imbalanced, as most respondents did not have a diagnosis of heart disease. This is not unexpected when dealing with a large set of randomly selected individuals.

## Independent Variables

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Teams

Description automatically generated

Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes]

Graphical user interface, application, Teams

Description automatically generated

Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week

Graphical user interface, application, Teams

Description automatically generated

(Ever told) (you had) a stroke?

Graphical user interface, application

Description automatically generated

Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 was your physical health not good?Graphical user interface, application

Description automatically generated

Thinking about your mental health, for how many days during the past 30 days was your mental health not good?

Graphical user interface, application

Description automatically generated

Do you have serious difficulty walking or climbing stairs?

Graphical user interface, application

Description automatically generated

Are you male or female?Graphical user interface, application

Description automatically generated

Fourteen-level age category. We decided to bin this based on screening recommendations starting at 45 and highest risk from 65 and older.

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Teams

Description automatically generated

(Ever told) (you had) diabetes? Due to the extremely low categories of Prediabetes and Gestational Diabetes, we decided to roll those categories into No.

Graphical user interface, application

Description automatically generated

Adults who reported doing physical activity or exercise during the past 30 days other than their regular job

Graphical user interface, application

Description automatically generated

Would you say that in general your health is...

Graphical user interface, application

Description automatically generated

On average, how many hours of sleep do you get in a 24-hour period? We decided to bin this data into three groups, those who get less than the recommendation, those who meet the recommendation and those who are above the recommendation for hours of sleep. The recommendation is for 7-9 hours of sleep per day.

Graphical user interface, application, Teams

Description automatically generated

(Ever told) (you had) asthma?

Graphical user interface, application, Teams

Description automatically generated

Not including kidney stones, bladder infection or incontinence, were you ever told you had kidney disease?

Graphical user interface, application

Description automatically generated

(Ever told) (you had) skin cancer?