

Project Outline

Kira Wolff

April 2021

Research Question

Data

The data with which I will investigate my research question is publically accessible via kaggle.com. The dataset is called “Stroke Prediction Dataset”, uploaded on the 26th of January 2021 by the user fedesoriano (<https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>). Unfortunately, no further sources for the data or information about the collection methodology or author are given. Thus, the validity of the data is uncertain and conclusions should be drawn with a grain of salt. Nevertheless, the data superficially seems fit to investigate the research question.

The dataset includes information about 5110 people. There is demographic data (age, gender), medical data (Hypertension, Heart Disease, Glucose Level, BMI, Stroke), and lifestyle data (Marriage, Work Type, Residence Type, Smoking Status).

Methods

Expectations