De dataset ‘cardio\_train.csv’ bevat 70000 rijen en 13 kolommen. De inhoud van de kolommen wordt hieronder uitgelegd. Daarnaast worden er een aantal statistieken meegegeven.

Id: in de kolom ‘id’ staan de id’s van elke rij. Genummerd vanaf 0 tot en met 69999.

Age: in de kolom ‘age’ staat de leeftijd van de patiënt in dagen. De minimale leeftijd is 10798 dagen wat omgerekend naar jaren 29 is. De hoogste leeftijd is 23713 dagen, dit is omgerekend 64 jaar. De gemiddelde leeftijd is 53 jaar.

Gender: De kolom ‘gender’ bevat het geslacht van de patiënt, een ‘1’ staat voor een vrouw en een ‘2’ staat voor een man. 65% van de patiënten zijn vrouw. Uit de data blijkt dat 32,31% van de vrouwen een cardiovasculaire ziekte heeft en 17,66% van de mannen.

Height: in de kolom ‘height’ staat de lengte van de patiënt in centimeters. De kleinste persoon is volgens de data 55 centimeter en de langste persoon is volgens de data 250 centimeter. De gemiddelde lengte is 164,36 centimeter.

Weight: De kolom ‘weight’ bevat het gewicht van de patiënten in kilogrammen. De lichtste persoon in volgens de data 10 kilogram en de zwaarste persoon is 200 kilogram. Het gemiddelde gewicht is 74,2 kilo.

Ap\_hi: De kolom ‘ap\_hi’ bevat de bovendruk van de patiënten in mmHg. De hoogste waarde is 16020 en de laagste waarde is -150. Het gemiddelde ligt op 128.

Ap\_lo: De kolom ‘ap\_lo’ bevat de onderdruk van de patiënten in mmHg. De hoogste waarde is 11000 en de laagste waarde is -70. Het gemiddelde ligt op 96.

Colesterol: de kolom ‘colesterol’ bevat het cholesterol gehalte van de patiënt. De kolom bevat drie waarden ‘1’ normaal, ‘2’ boven normaal en ‘3’ ruim boven normaal. De meesten vallen in categorie 1.

Gluc: de kolom ‘gluc’ bevat het glucose gehalte van de patiënt. De kolom bevat drie waarden ‘1’ normaal, ‘2’ boven normaal en ‘3’ ruim boven normaal. De meesten vallen in categorie 1.

Smoke: de kolom ‘smoke’ geeft aan of de patiënt rookt. De kolom bevat twee waarden, ‘0’ de patiënt rookt niet en ‘1’ de patiënt rookt. Uit de data blijkt dat 4,18% van de patiënten die roken een cardiovasculaire ziekte heeft en 45,79% van de patiënten die roken een cardiovasculaire ziekte heeft.

Alco: de kolom ‘alco’ geeft aan of de patiënt alcohol drinkt. De kolom bevat twee waarden, ‘0’ de patiënt drinkt geen alcohol en ‘1’ de patiënt drinkt alcohol. Uit de data blijkt dat 2,60% van de patiënten die drinken een cardiovasculaire ziekte heeft en 47,37% van de patiënten die niet drinken een cardiovasculaire ziekte heeft.

Active: de kolom ‘active’ geeft aan of de patiënt actief is. De kolom bevat twee waarden, ‘0’ de patiënt is niet actief en ‘1’ de patiënt is actief. Uit de data blijkt dat 39,45% van de patiënten die actief zijn een cardiovasculaire ziekte heeft en 10,51% van de patiënten die niet actief zijn een cardiovasculaire ziekte heeft.

Cardio: de kolom ‘cardio’ geeft aan of de patiënt een cardiovasculaire ziekte heeft. De kolom bevat twee waarden, ‘0’ de patiënt heeft geen cardiovasculaire ziekte en ‘1’ de patiënt heeft een cardiovasculaire ziekte. In de data heeft 49,97% van de patiënten een cardiovasculaire ziekte.