DATA-ANALYYSIN JA TEKOÄLYN PERUSTEET

Jenni Ylisirniö

Viikko 36

Tehtävä 1

Tsekataan, mitä dataa csv:t sisältää:

Graphical user interface, text

Description automatically generated

Describe-funktiolla molemmat datasetit:

Graphical user interface, text

Description automatically generated

Huomataan esimerkiksi, että max palkka on 10 000(€?) ja minimi 2 000(€?). Mutta esimerkiksi puhelinnumeroiden kohdalla describe ei anna mitään järkeviä arvoja (puhelinnumeron keskiarvo???).

Datasettien info-funktio:

Text

Description automatically generated

Info kertoo, kuinka monta saraketta datassa on, datatyypit ja muistin käytön.

Tsekataan null-arvot:

Text

Description automatically generated

Employees datassa on 10 NaN -arvoa sarakkeessa phone2.

header=0 ja header=None ero

Graphical user interface, text, application

Description automatically generated

Merge:

Graphical user interface, text

Description automatically generated

Poistetaan image sarake (del employeesData[’image’]):

A screenshot of a computer

Description automatically generated with medium confidence

Image poistui phone2 ja gender sarakkeiden välistä.

Tehtävä 2