LISÄTEHTÄVÄT viikko 35

Jenni Ylisirniö

Tehtävä 1

Tallenna oheinen diabetes.csv ja lataa se pandasin dataframeen. Tulosta muutamia tunnuslukuja mm. count, mean, min, max, std.

Text

Description automatically generated

Screenshotissa näkyy vain pieni osa ohjelman ajosta, mutta ajo toimii täydellisesti.

Tehtävä 2

Piirrä histogrammi kuvaaja datasta**:**

Graphical user interface

Description automatically generated

Yksi histogrammi näyttää vähän siistimmältä:

Graphical user interface

Description automatically generated with low confidence

Tehtävä 3

Piirrä korrelaatioheatmap datasta:

import seaborn as sns

import pandas as pd

import matplotlib.pyplot as plt

healthData = pd.read\_csv('diabetes.csv')

plt.figure(figsize=(16, 6)) # Increase the size of the heatmap.

heatmap = sns.heatmap(healthData.corr(), vmin=-1, vmax=1, annot=True)

heatmap.set\_title('Correlation Heatmap - Diabetes', fontdict={'fontsize':12}, pad=12)

sns.heatmap(healthData.corr())

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

<https://medium.com/@szabo.bibor/how-to-create-a-seaborn-correlation-heatmap-in-python-834c0686b88e>

Tehtävä 4

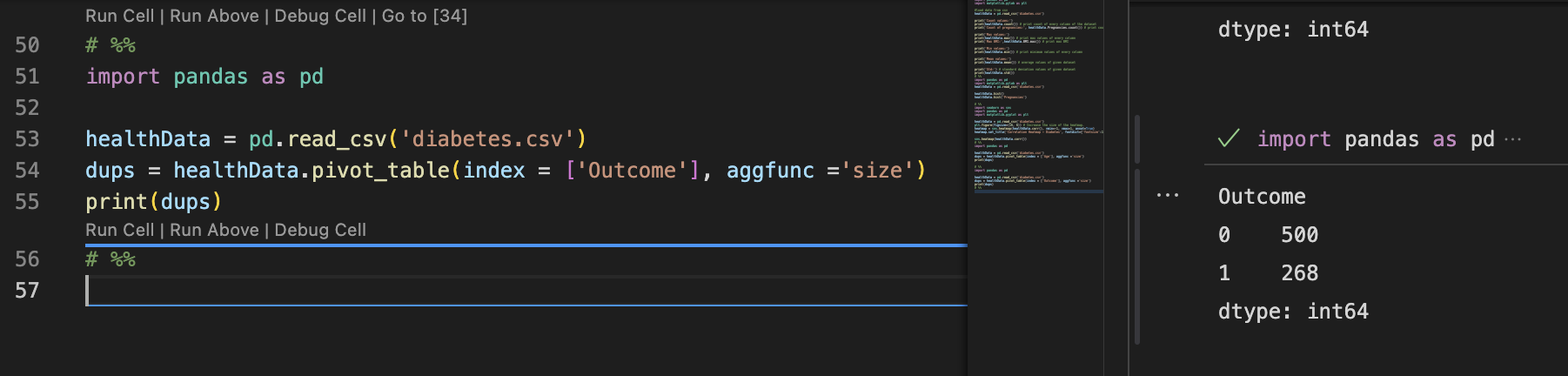
Laske potilaiden lukumäärä iän mukaan siten, että suurin lukumäärä on ensin.

Graphical user interface, text

Description automatically generated

Osasin laskea potilaiden lukumäärän iän mukaan, mutta järjestystä en osannut vaihtaa niin, että suurin lukumäärä olisi ensin. Yritin sort\_values , mutta ei jostain syystä onnistunut.

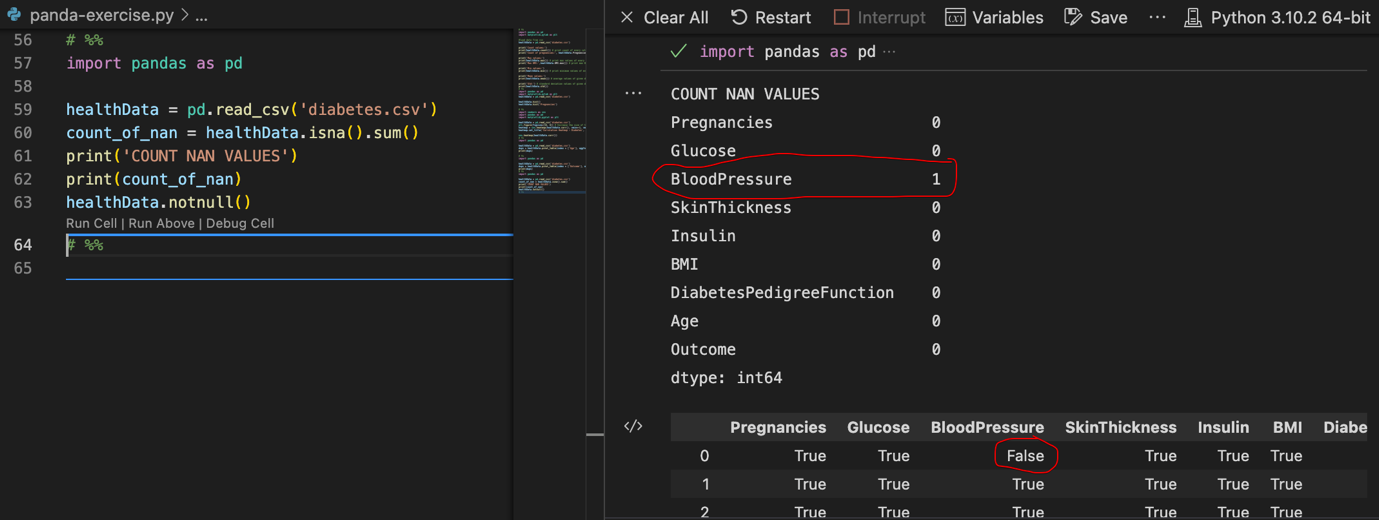
Tulosta myös montako diabetestapausta (1) ja ei diabtesta (0) on aineistossa**.**



Eli 500 ei ole (0) diabetestä ja 268 on (1) diabetes.

Tehtävä 5

Onko aineistossa nan arvoja?Jos on, missä sarakkeissa ja montako?



Aineistossa on yksi NaN (not a number) arvo. Se löytyy Blood Pressure -sarakkeesta, riviltä yksi.