

# Week 1 - Assessed Exercises

## Data Programming with Python

In this week's lecture you learnt how to import a data set into Python, perform some basic data analysis and plot the data. For the assessed exercises, you will perform similar tasks for the *Heart Disease UCI* data set.

The *Heart Disease UCI* data set contains different heart related measurement for a number of patients along with an indicator of whether or not heart disease was detected. The attributes (data columns) are as follows:

- age - in years
- sex - 1 male, 0 female
- cp - chest pain type
- trestbps - resting blood pressure
- chol - serum cholestoral in mg/dl
- fbs - fasting blood sugar 120 mg/dl
- restecg - resting electrocardiographic results (values 0,1,2)
- thalach - maximum heart rate achieved
- exang - exercise induced angina
- oldpeak - ST depression induced by exercise relative to rest
- slope - the slope of the peak exercise ST segment
- ca - number of major vessels (0-3) colored by flourosopy
- thal - 3 = normal; 6 = fixed defect; 7 = reversable defect
- target - 1 if heart disease detected, 0 if not

Import the data into Python and answer the following questions.

1. (a) How many rows and columns there are in the *Heart Disease UCI* data set?  
(b) What is the age of the 3rd person in the data set, i.e. on the third row?

Compute the table of different chest pain types.

2. How many people have type 3 chest pain?
3. (a) What is the age of the youngest person included in this study?  
(b) What is the age of the oldest person included in this study?

Look up what the cut function (*pd.cut*) does and use it to create a new variable which is age grouped into 20-30, 30-40, 40-50, 50-60, 60-70, 70-80.

4. How many people are in the group (50,60)?

Answers, along with all of your code, should be written into the *.py* template. Save your filled *.py* file with the following name structure *SurnameFirstname\_Week1.py* (where *Surname* and *Firstname* should be replaced with your name) and upload it to Brightspace.