Lab3: In the following two labs, we will apply the CRISP-DM methodology for data mining in a simple form. For this assignment, we will start with the data understanding, followed by data preparation. The overall goal of this and next lab is to build and train a Neural Network that can recognize handwritten digits from 0 to 9. You can use the following code to download the MINIST data of handwritten digits.

```
import tensorflow as tf
import numpy as np
from tensorflow import keras
import matplotlib as mpl
import matplotlib.pyplot as plt

(X_train_full, y_train_full), (X_test, y_test) =
keras.datasets.mnist.load_data()
```

Data understanding

a) Investigate the shape of the data and explain what you see.
b) Check if the labels are skewed by plotting their count distribution per class.
c) Select a random image from the training set (for example with the random.randint() function to select a random index of X_train) and print the resulting array. Calculate the maximum value and explain the meaning of that value.
d) Plot the image using the matplotlib function imshow(). Compare the image to the label and make sure it is correct

Data preperation

a) Separate the training data into training and validation set. Explain why this additional data split is needed.
b) Scale the data. In this simple case you can simply divide by the max value for scaling
c) Name three different performance measures and explain their meaning and advantages/disadvantages.