# Report for the Final project of Programming and Scripting 2021 GMIT course

This is the final project for the 2021 Programming and Scripting course. The purpose of this project is to analyse the Fisher’s Iris data set using Python scripting.

The Fisher’s Iris data set comes from the 1936 R. A. Fisher paper “The use of multiple measurements in taxonomic problems”. In this paper, the author tackles the problem on how to distinguish between 3 different species of Iris flowers using the measurements of their 4 characteristics (Petal’s and Sepals length and width). The method used in this paper focuses on finding a linear function of these four measurements that would maximize the ratio of the difference between the means to the standard deviations within species. The larger this ratio gets the easier it is to distinguish between different species using the measurement data.

--------------------------------------------------------------------

Table 1: Simple descriptive statistics for the whole data set

====================================================================

0 sepal length sepal width petal length petal width

count 150.000000 150.000000 150.000000 150.000000

mean 5.843333 3.057333 3.758000 1.199333

std 0.828066 0.435866 1.765298 0.762238

min 4.300000 2.000000 1.000000 0.100000

25% 5.100000 2.800000 1.600000 0.300000

50% 5.800000 3.000000 4.350000 1.300000

75% 6.400000 3.300000 5.100000 1.800000

max 7.900000 4.400000 6.900000 2.500000

--------------------------------------------------------------------