Task 1:

a) First, I removed all the majors and kept only the top 10 majors, so that the graph looks better. Then, I replaced every ENGINEERING word with ENG. To reduce the amount of the white space.

b) First, I removed all the categories majors and kept only the top 5.now the graph looks clearer. Then, I removed the outliers from the median column for each category that way we can focus on the distribution and the graph look a lot nicer. Last, I added the row samples so that we can see at each median how many samples we have. Also, I rotated the X axis labels by 90 degrees to make the labels more readable.

c) I plotted the Violin plot for the top three major categories that shares women in their work.
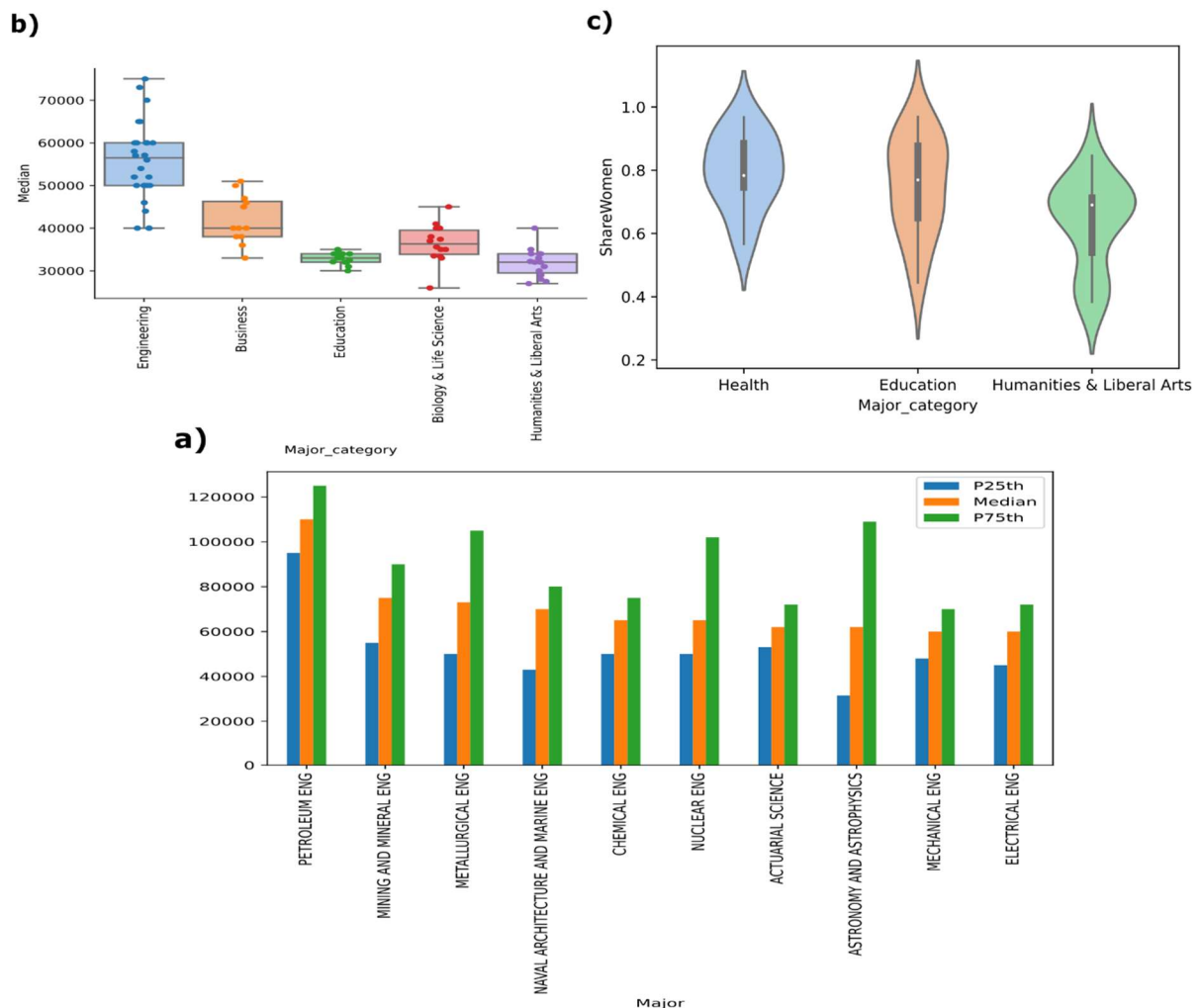

Figure 1: Characteristics of the College Majors dataset. a) The top 10 majors depending on the median earnings of full-time, year-round workers. b) The boxplots of median earnings for the top 5 categories (as measured by median earnings. c) The distribution of the major categories that shares women in their work. Only the top 3 were considered.

Name: Amr Aly
student (matriculation) number: 23225160
Code: https://github.com/amr1235/DSS/blob/main/exercise%203/Homework_4.ipynb