**Laboratory Work 1**

1. Upload the 3 libraries you will have to use for this exercise:

Numpy, pandas , maplotlib.pyplot,

1. Upload the dataset
2. Explore your dataset using the dataset.describe() function
3. List all the variable names with the dataset.columns.tolist() command

(0.5pt)

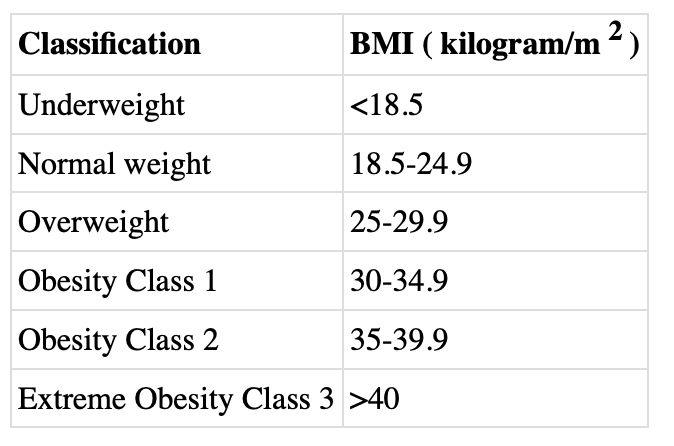
1. How many of the people in your dataset have diabetes and how many do not?The variable outcome takes values of 0 and 1. 1 means that the person has diabetes and 0 means that the person has no diabetes. (0.5pt)
2. Create a bar chart to represent the frequencies of those with diabetes versus those without diabetes, label the axes and name the chart.

(1pt)

1. Create a pie plot, change the colors, add the labels, add shade, and ‘explode’ the pie slice for the portion of people with no diabetes from the “Outcome”. Create the legend and put it into the right corner so that it does cover the pie plot.

(1pt)

1. Calculate and report the mean the standard deviation of the BMI variable. (0.5pt)
2. By using the Chebysheffs Theorem infer what is the range for the BMI variable for 75% of the people in your dataset. What is the range for 89% of the people in your dataset. (0.5pt)
3. Create a histogram for the BMI variable, plot the mean and meadian on the historgram to answer the following question:
4. Is the BMI distribution skewed based on the position of the mean versus the median?
5. Calculate the skewness of BMI
6. Based on the value of BMI median and on the chart below infer what percentage of people in your dataset are obese. (1.5pt)



1. Perform the same operations as you did in 10, but with the Blood Pressure variable. Conclude on the skewness. (1.5pt)

If you performed all of the steps above did you notice anything suspect in your histograms? Clue: it has to do with outliers on the left side. What is it? (0.5pt)

1. Remove all the observations where the BloodPressure' and the BMI variables take the value of zero. These observations are obviously measurement errors (no alive person has BMI or Blood pressure zero). (1pt)
2. Now perform once again all the steps in 8, 9, 10 and 11. Report the new results. Comment on how your measurements and thus your conclusions have changed. (2pt)