BIOMEDICAL DATA SCIENCE HOMEWORK #1
DUE DATE 9/12/2019

- Please turn your homework code as a link to a Google Colab Notebook.

- No late homework will be accepted (you will not be able to submit the repository link to Canvas after the due date). Do not change the repository after the due date; your homework will not be graded.

- <u>20 </u>points total

Use the accompanied dataset for this homework. Read the dataset description carefully here and make sure you understand the dataset features and values.

a) Load the dataset in a Google Colab notebook (4 points).

b) Seaborn is a Python data visualization library that provides many different types of plots. Explore the dataset using a bar plot to show the relationship between each one of the categorical features and "comfort", with respect to the class label. Interpret the results in a markdown cell (8 points).

c) Perform any necessary preprocessing steps, as if you were using this in a machine learning model. Explain why each step is needed. Use the markdown cells of your notebook for explanations (8 points).