# Seminar 5 (Week 5)

Q1. Please perform the following activities on the penguin dataset:

I. Please clean the dataset. (Remove null values) [We have done this in W2]

II. Please transform the non-numerical values into numerical values. Please focus on the 'island' and 'sex' feature and consider to transform them using get_dummies() method.

III. Please apply Decision Tree Classifier to measure the accuracy for both testing and training set.

IV. Please create a Decision Tree plot for the algorithm.

V. Please apply GridSearchCV for to find the best parameters for learning. And create a new model with the best parameters.

VI. Please generate the classification report and the decision tree for the new model.

Q2. Please perform the following activities on the Global Cybersecurity Threats Analysis dataset

I. Import the dataset into pandas dataframe and clean & process the dataset if required.

II. Please plot the following information using the dataset
   a. Attacks on different countries per year.
   b. Type of attacks per year.
   c. Top 10 target industries by attack count
   d. Top four attack sources.
   e. Top four defense mechanism used
   f. Top four security vulnerabilities.