Fourth Industrial Revolution (4IR) Summer School
Module **2** – **D**ay **3**

Part 1 - Data Transformation and Grouping

Task 1: Data Transformation

Using the functionalities provided by the Pandas and NumPy libraries, write a Python program to read the given "stud_exams.csv" file, and then accomplish the following tasks:

A. Add a new column "Training" using mapping to indicate which students are eligible for training, such that students **with no test preparation course** are marked as eligible.
B. Encode the categorical variables in the dataset using different encoding methods.
C. Apply "equal-width binning" to generate three categories ['low', 'average', 'high'] for the scores in each subject column. Store the resulting columns in a new DataFrame.
D. Add a fourth column to the resulting DataFrame to indicate whether a student falls into the same category across all three subjects.

Task 2: Data Grouping

Write Pandas program to read the salaries file "salaries.csv"

"https://raw.githubusercontent.com/Apress/data-analysis-and-visualization-using-python/master/Ch07/Salaries.csv"

A. Group the data according to the "rank" column and show the generated groups.
B. Count the number of rows in each group.
C. Find out the number of unique values in each group.
D. Calculate the mean, median, minimum, and maximum salary, by groups, using the 'agg' method.
E. Rename all the generated columns so you capitalize each word.
F. Compare the number of males to the number of females in each rank.
G. Apply a function to increase the salary of employees in different ranks with different percentages.