# PORTFOLIO & GROUP ASSIGNMENT

MSc Data Science, Coventry University UK (2024 Batch)

# Portfolio

**Question 1:**

You are tasked with completing a data analysis project. This project should include data collection, exploratory data analysis (EDA), and proposing a solution to a real-world problem based on the insights.

**Requirements:**

1. **Data Collection:**
   - Identify a dataset or collect data relevant to your chosen problem.
   - Clearly explain the data collection method (e.g., web scraping, surveys, APIs, existing datasets).

Add cloud touch AWS/Azure

2. **Exploratory Data Analysis (EDA):**
   - Perform a detailed EDA to understand the dataset. This includes:
     - Data cleaning and preprocessing.
     - Descriptive statistics and visualizations to uncover patterns or anomalies.
     - Highlight key insights from the data.
3. **Problem Solving:**
   - Identify a specific problem related to the dataset.
   - Propose a feasible solution based on the insights obtained from your EDA.
4. **Collaboration and Documentation:**
   - Use a Git repository to manage your work.
   - Ensure you are making meaningful contributions (tracked via Git commits).
   - Document your progress, decisions, and code clearly in the repository.
   - Include a final report summarizing the findings and proposed solution.

**Deliverables**:

- A well-organized Git repository for your individual project.
- Analysis notebooks/scripts demonstrating the ETL process and other steps.
- A summary report including:
  - Insights from your dataset (e.g., key patterns, trends).
  - Clustering analysis results and their implications.
  - Model accuracy and its potential business impact.
  - Recommendations for operational improvements based on your findings.

**Question 2:**

Pick **one** of the following topics and compile a report to answer the question. You may analyze existing data and provide necessary plots/tables to justify your answer from a business statistics viewpoint. Your answer should not exceed three A4 sheets (Times New Roman, 12pt, 1.15 line spacing). *Note: The use of AI tools for report composition is strictly prohibited and will be given zero marks if detected*.

a) Yahoo's decline from a dominant internet player to its current state?

b) Mihin Lanka fail in Sri Lanka

c) One of the largest companies in the world, General Motors also resulted in one of history's largest bankruptcies.

## Group Assignment

You are tasked with analyzing and improving operations for given company data. The company wants to better understand its performance, customer satisfaction, and demand trends to optimize their operations.

Your task is to:

1. **Perform an ETL process** to clean, transform, and integrate the data for analysis. Ensure proper documentation of your pipeline.
2. Use **clustering techniques** to identify patterns or groupings within the data (e.g., customer segmentation, service performance, demand forecasting, etc.).
3. Develop a **forecasting model** to predict future trends based on your dataset (e.g., sales performance, customer satisfaction, order volumes).
4. Document and version-control your project using **Git**, ensuring clear commit messages and a structured workflow.
5. Present your findings and actionable recommendations based on your analysis.

**Work Division:**

- Divide the following tasks among team members:
    1. Data Preprocessing.
    2. Exploratory Data Analysis.
    3. Problem Identification and Solution Proposal.
    4. Git Repository Management and Documentation.
- Ensure equal participation from all members.

**Submission:**
Submit the Git repository link and a brief presentation summarizing your project, including:

- The problem you addressed.
- Key findings from your EDA.
- Your proposed solution.

Groups:

Group 1: 2024.1 batch Index 1 to 5

Group 2: 2024.1 batch Index 6 to 10

Group 3 : 2024.1 batch Index 11 to 15

Group 4 : 2024.2 batch Index 1 to 5

Group 5 : 2024.2 batch Index 6 to 10

Group 6 : 2024.1 batch Index >15 and 2024.2 batch Index >10