Mini Project III

Purpose

- Apply the **Data Science Process** applying skills you have learn in recent modules

Choose **one** of the following tasks:

1. Choose a **dataset** or select a **question** and identify and gather the data you need. Perform EDA on the data, identify features and a target variable, and apply predictive models including **decision trees** and **ensemble methods**. Evaluate these and present your findings.

Mini Project III

2. Choose a dataset for a text classification task – such data might be obtained via web scraping or may already exist in clean form. There needs to be a target variable containing the category of text. Perform the following steps.

   a) Perform tokenization and remove stop words.

   b) Apply stemming and/or lemmatization.

   c) Create features such as word counts, parts of speech, vectorisation (e.g. Word2Vec, TF-IDF).

   d) Select and apply two or more predictive models to predict the category of text.

   e) Evaluate the models and present your findings.

Present from **slides** in 8-12 minutes on Monday 17 May.