# Introduction to Data Science
# Coursework 1
# Deadline: 12.00, Friday, 25th February, 2022
# Submission: ELE

This assessment is worth 40% of the overall mark. This is an individual assessment. You are reminded of the University's regulations on plagiarism.

**Brief**

As data science students you know all about the important and growing role that big data, machine learning and data science plays in society. As such, it is important for all practitioners of data science to not only have technical knowledge, but also to maintain a keen awareness of the consequences of their work. If an engineering firm designs a faulty bridge, they can and should be liable for any accidents. If a data engineering firm designs an incorrect, biased or otherwise harmful algorithm, what should their responsibility be?

In this coursework you are asked to critically reflect on a situation where data has been misused (e.g. The Facebook-Cambridge Analytica data scandal), leaked (e.g. the Equifax data breach) or analysis of data has had unintended consequences (e.g. predictive policing; racial discrimination in facial recognition). You should research your topic thoroughly and do plenty of reading. The deliverable is a video presentation.

**Presentation:**
- Length: **10 minutes!** Presentations over this will be penalised by 10%.
- You should include references on your slides, this means when you use a figure or cite a source you should tag it with a number [1] or multiple numbers [2,3], and you should give that reference in a standard format on the slide itself or at the end of the presentation on a 'Bibliography' slide. Use any standard academic referencing style e.g. Harvard.
- Include graphs/pictures/tables as appropriate.
- You do not have to appear on camera (though you can if you wish). I am expecting most people to submit a "voice over slides" presentation. The style is up to you.

**Guidance**

**Choosing your Topic**
You are free to choose your own topic but some examples are
- Facebook-Cambridge Analytica scandal
- Various Data Breaches e.g. LinkedIn, eBay, Facebook...
- "Fake News" groups on WhatsApp during the 2018 Brazilian General Election
- Predictive Policing
- Algorithmic Sentencing
- Man is to Computer Programmer as Woman is to Homemaker…
- Facebook Emotional Contagion Study
- Github Co-pilot
- Racial bias in image analysis

Choose a topic from this list or find another case of insecure, biased or misused data and data analysis. Some good books giving a general overview of data science problems are

**Perez CC.** *Invisible women: Exposing data bias in a world designed for men.* Random House; 2019 Mar 7.

**O'Neil C.** *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown; 2016 Sep 6.

**Noble SU.** *Algorithms of oppression.* New York University Press; 2018 Feb 20.

These are useful but not required reading, there are plenty of articles and other resources freely available online.

## Marking

The marking scheme below is a guide to how the report should be structured.

| Section | What is expected for a good mark? | Marks |
|---|---|---|
| **Background** | A clear and concise account of the actors involved and a discussion of the specific problem or incident. | **10%** |
| **Causes** | What were/are the causes of the issue e.g. careless analysis, existing biases, weak security, fraud... | **10%** |
| **Consequences** | What were the consequences for the organisation/users or for society at large? If any changes have been made in response to the issue, discuss them. | **10%** |
| **Critical Reflection** | Reflect on the wider issues that your example raises on the role of data in society. Discuss how data scientists should respond to avoid future problems. For example, if you discuss a specific case of algorithmic bias in this section you should discuss the use of algorithms in society as well as algorithmic bias in general, its causes and what should be done about it. | **20%** |
| **Presentation** | Marks for each section will incorporate marks for good presentation: meaning correct referencing, clear and relevant graphs/figures, good structure, attractive slides etc. | **Overall** |

## Peer Marking

You will notice that the above marks only add up to 50%. The remaining marks will be decided by your fellow students. This means

- You will each be given 3 videos to watch.
- You will mark the videos (using an online form to be provided) against each criteria.
- Your own video will be marked by 3 of your peers - the scores will be averaged and added to the university assessor's score.
- **Until you have completed the 3 video marking tasks, your own video will not be marked!**

**Making a video**

- Use any software you like, or a phone camera
- Output should be shareable, preferably MP4 format. If you use a screen recorder that generates a unique format - convert it!
- A very good, free, screen recording software is OBS Studio: https://obsproject.com/
- You can do the whole thing in one take, but if you need to do some simple video editing there is lots of free software available e.g. https://kdenlive.org/en/
- Try to keep the video/audio quality reasonably high, but likewise please don't submit an enormous file.
- The maximum upload limit on ELE is 256MB, if your file is higher than that please submit a link e.g. to a Dropbox/Sharepoint/Drive folder. The video needs to be downloadable!
- The simplest format is to add a voiceover to some slides, this is perfectly fine. But you could also use animation, video clips, puppets, interpretive dance… The choice is yours!