# **Take home exercise – Gender Pay Gap**

Suggested time for completion is ~4 hours

## Project Description

Large businesses are required to report the gap in pay by gender in many regions in the world (for UK see [1]). [2] shows how to calculate the Gender Pay Gap

The business believes there may be differences in pay between women and men and wants to understand whether the data they have supports this belief. They also want to understand what the drivers of these differences are and what changes they can make.

You are given the 3 csv files: Main\_dataset.csv, Function.csv and PartTime.csv. Use this information to investigate and convey your conclusions to the business.

## Project Completion

The completed project should take the form of a set of slides and visualisations but you are free to be innovative in the way you choose to present the analyses you have done. Your audience will be a business executive who does not have knowledge in Data Science. There will be a Director of Data Science supporting the business executive who will be interested about how the analyses were carried out.

Please send the presentation as well as any code/files you used to complete the project ahead of the meeting.

## Guidance

**Verification**: Verify whether the data supports the belief that there are differences in pay between genders and provide evidence for your conclusions

**Identifying Patterns**: Investigate whether there are any patterns in the data that may help explain any differences

**Solutions**: Suggest potential solutions or focus areas to remedy the problem

**Presentation**: Prepare a presentation to be presented to a business executive

## More about the Data

The data given here contains some patterns that may represent what is seen in some markets and companies (i.e. the US and UK). The data is purely constructed and does not pertain to an actual company – Please specify this in any presentations or material shared.

You can find the definitions of the fields and files below:

Main\_dataset.csv:

**Person\_ID**: The unique identifier of this person

**Gender**: The gender of the person. Either woman or man

**Age**: The age of the person in years

**LengthOfService**: The number of years this person has been in the organisation

**Admin Role**: Whether this person has a role that is Administrative in nature

**International Assignee**: Whether this person has taken a position abroad to fill a critical role

**Is Mobile**: Whether this person is considered to be mobile or not. Mobile means they have taken roles in different countries over their history at the organisation

**YearlyPay**: The yearly pay for this person in the same currency (£s). It is pro-rated meaning if someone on 50% part time is paid £10k, this translates to £20k

Function.csv:

**Person\_ID**: The unique identifier of this person

**Function**: The function this person currently works in

PartTime.csv:

**Person\_ID**: The unique identifier of this person

**PartTime**: Whether this person is currently working part time or not.